TRADE ADJUSTMENT COSTS IN DEVELOPING COUNTRIES: IMPACTS, DETERMINANTS AND POLICY RESPONSES

Trade Adjustment Costs in Developing Countries: Impacts, Determinants and Policy Responses

1818 H Street, NW, Washington, DC 20433, USA

ISBN: 978-1-907142-08-6

Centre for Economic Policy Research

The Centre for Economic Policy Research is a network of over 700 Research Fellows and Affiliates, based primarily in European universities. The Centre coordinates the research activities of its Fellows and Affiliates and communicates the results to the public and private sectors. CEPR is an entrepreneur, developing research initiatives with the producers, consumers and sponsors of research. Established in 1983, CEPR is a European economics research organization with uniquely wide-ranging scope and activities.

The Centre is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. CEPR research may include views on policy, but the Executive Committee of the Centre does not give prior review to its publications, and the Centre takes no institutional policy positions. The opinions expressed in this report are those of the authors and not those of the Centre for Economic Policy Research.

CEPR is a registered charity (No. 287287) and a company limited by guarantee and registered in England (No. 1727026).

The World Bank

The World Bank Group is a major source of financial and technical assistance to developing countries around the world, providing low-interest loans, interest-free credits and grants for investments and projects in areas such as education, health, public administration, infrastructure, trade, financial and private sector development, agriculture, and environmental and natural resource management. Established in 1944 and headquartered in Washington, D.C., the Group has over 100 offices worldwide. The World Bank's mission is to fight poverty with passion and professionalism for lasting results and to help people help themselves and their environment by providing resources, sharing knowledge, building capacity and forging partnerships in the public and private sectors.

Trade Adjustment Costs in Developing Countries: Impacts, Determinants and Policy Responses

Edited by:

GUIDO PORTO

AND BERNARD M. HOEKMAN

Contents

PART D. ADJUSTMENT PROGRAMS

List of Tables

List of Figures

Preface

The process of globalization that has been ongoing in recent decades has lifted millions out of poverty and greatly increased average incomes in a large number of countries. Large developing countries such as China, India and Brazil have become major players in the world economy. The relative economic weight of Europe, Japan and the United States is declining. While very beneficial from a development perspective, these global trends present multifaceted challenges to policymakers, including management of the associated increase in risk and volatility – as exemplified by the crisis that erupted in 2008; inclusion of those most marginalised; adapting to the shift that the emerging economic powers represent; and addressing the environmental repercussions of a rapidly developing world.

Making the results of globalization more beneficial to the poorest households and poorest countries is critical for the sustainability of the gains that have been achieved by the world as a whole in recent decades. Identifying measures that can help achieve this objective is an objective of the Global Trade and Financial Architecture (GTFA) project. This project, which is supported by the UK Department for International Development (DFID was originally set up to support follow-up activities that build on the report of the UN Millennium Taskforce on Trade (*http://www.ycsg.yale.edu/core/forms/Trade_for_Development.pdf*). It is piloted by a Steering Committee of researchers and policymakers and co-chaired by Ernesto Zedillo, Yale Center for the Study of Globalization, and Patrick Messerlin, Groupe d'Economie Mondiale de SciencesPo. The GTFA's objectives are to identify and promote concrete policy options for reinvigorating and strengthening the multilateral economic system that has supported the process of globalization and making it more sustainable and inclusive.

One of the activities funded by this project comprise the contributions to this volume: an effort to bring together leading researchers to summarize the state of knowledge on the magnitude and determinants of the costs of adjustment to greater trade openness, including the factors that affect the distribution of the potential gains from trade. The basic idea was to commission a series of short papers that synthesize existing knowledge and to use this as the basis for identifying where further research would be most fruitful.

The support provided by the GTFA project is gratefully acknowledged. Thanks are also due to all the contributing authors for their willingness to pull together their knowledge and thoughts on the subject of adjustment to trade, and to Olivier Cattaneo, Michelle Chester, Rebecca Martin, Cecilia Peluffo and Anil Shamdasani for their assistance with the logistics of putting together the volume.

Otaviano Canuto
Vive-President
Poverty Reduction and Economic Management
The World Bank

Stephen Yeo
Chief Executive Officer
CEPR

1

Trade Adjustment Costs in Developing Countries: Impacts, Determinants and Policy Responses

BERNARD HOEKMAN AND GUIDO PORTO

Integration into the global economy offers an enormous opportunity for reducing poverty, hunger, and economic injustice. Increasing competition on the domestic market, thus lowering prices and increasing choices for consumers and improving access to new knowledge, products and technologies, generates potential sources of aggregate efficiency gains. The gains from trade openness are not guaranteed, however. They are conditional on various factors, such as a sound investment climate (Freund and Bolaky, 2008), a sufficiently large "absorptive capacity" (e.g., human capital) to capture the spillover benefits from trade (Keller, 1996; Borensztein, De Gregorio and Lee, 1998), and an absence of major domestic distortions (Bhagwati, 1971; Krishna, this volume) such as credit constraints (Chesnokova, 2007; Manova, this volume).

Globalization also generates costs. Trade liberalization will result in a re-allocation of factors of production within and between firms and sectors. This is the source of the efficiency improvements that underpin the gains from trade, but it also brings with it adjustment costs. There are winners and losers. This distributional conflict is dynamic – the transition to a more open trade situation can be hazardous to both losers and winners. How agents adjust to take full advantage of the new trading opportunities, what they do to ease the burden of adjusting to reforms, and the factors that drive the adjustment are questions to which only limited attention has been devoted by trade economists. Attenuating the negative effects of integration for disadvantaged groups is an important task for governments. In principle, the gains from trade generate the resources that can be used by governments to do so. The design of public policies to facilitate the transition and smooth the adjustment process needs to be informed by an understanding of the impacts of trade reforms and the responses by firms, workers and households.

The goal of this book is to summarize the state of knowledge in the economic literature on trade and development regarding the costs of adjustment to trade openness and how adjustment takes place in developing countries. The book com-

prises 23 contributions by experts who focus on different dimensions of adjustment processes in developing countries. Each contributor was selected on the basis of having made a significant contribution to the relevant literature on adjustment to trade in developing countries. The authors were asked to summarize their own research and the state of play in their area of expertise. The objective was to pull together in one place what is known and to use this as the basis for identifying where additional work would have a high rate of return.

In this overview chapter we start with a brief discussion of the "stylized facts" on the impacts of increasing trade openness. We then summarize the main findings of the contributions to this volume. A final section turns to the implications for further research and for the design of policies to address adjustment costs.

1. SOME STYLIZED FACTS

Some of the stylized facts that have emerged from research on the impact of trade openness on firms and workers are well-known. Thus, there has been a significant increase in the wage premium for skilled labour around the world, a rise in the ratio of skilled-to-unskilled employment in *all* sectors, and rising relative inequality between the skilled and unskilled. At the same time, there has not been a large decline in the relative price of goods that use low-skilled labour relatively intensively.[1] The inference that has been drawn is that greater trade and trade reforms can only explain a small fraction of the general increase in wage inequality observed in both developed and developing countries. While typically skilled-biased technical change is seen as the main driver, recent research by Galiani and Porto (2010) shows that barriers to factor mobility and unions can also play a role.

Thus, globalization rewards the relatively more skilled disproportionately. This does not imply that unskilled workers are necessarily worse off. To the contrary: in the post-1970 period there has been a substantial decline in the number of people in absolute poverty, reflecting sizeable increases in the incomes of the poorest households in many developing countries, especially those with large populations such as Brazil, China, India and Indonesia (Ravallion and Chen, 2004; Ferreira and Ravallion, 2008). However, this trend has not been universal — poverty increased in Sub-Saharan Africa.

Basic trade theory predicts that trade reforms will result in labour reallocation across sectors. Surprisingly, much of the empirical research on this question does not find strong evidence for this using available data for developing countries. For example, Wacziarg and Wallack (2004) conclude that liberalization episodes are followed by a *reduction* in the extent of inter-sectoral labour shifts at the economy-wide 1-digit level of disaggregation. There is a weak (insignificant) positive effect at the 3-digit level and no evidence of trade-induced structural change

[1] From a trade theory perspective the goods price channel is the one through which greater trade should affect labour outcomes: those that are most dependent on production of import competing goods should be more affected.

at the more disaggregated 4-digit industry level. This somewhat counter-intuitive result – also found by many other studies (see Hoekman and Winters, 2007 for a survey) – in part reflects the relatively short time frame of many empirical analyses, and the fact that many tend to focus on the formal manufacturing sector, on which there is generally (much) better data. In a longer-run perspective, of course, by definition economic development entails significant structural change, with large numbers of people leaving agriculture and finding employment in manufacturing and services industries.

There is more evidence for the reallocation and adjustment processes *within* industries: the more productive domestic firms in an industry expand by drawing resources from less productive firms that shrink or go out of business. Recent theoretical developments and empirical analysis have emphasized the importance of recognizing that there is much heterogeneity of firm performance and efficiency/productivity within industries, and that this is a significant source of the welfare gains from trade liberalization (Melitz, 2003). Recognition of the heterogeneity of firms within and across industries helps to understand the empirical observation that there is much churning within sectors following trade reforms. It also helps to understand why trade liberalization is important for economic growth over time. As the more efficient firms expand and the less efficient ones contract, the overall productivity of the economy increases. If there are scale economies and imperfect competition, liberalization will allow more efficient firms to further reduce unit costs as their market expands.

Whether the impacts of more open trade operate more or less through wages as opposed to employment depends significantly on labour market institutions, the efficiency of capital markets and social policies. In developing countries, wage responses seem to be greater than impacts on employment. There is substantial evidence that trade liberalization decreased industry wage premiums in those sectors that experienced the largest tariff reductions. The recent global crisis has generated additional evidence that wages bear the brunt of adjustment to external shocks. Based on a sample of 41 middle-income developing countries, Khanna, Newhouse, and Paci (2010) conclude that the impact of the economic downturn during 2008–2009 fell disproportionately on the quality of employment rather than on the number of jobs. Slower growth in earnings accounts for nearly three quarters of the total adjustment for the average country, driven by a reduction in working hours, as well as a shift away from the better-paid industrial sector and toward informal or rural employment.

One implication of the finding that large-scale reallocation of workers *across* sectors is not the norm following a trade liberalization episode is that the direct effects of trade reform on *aggregate* employment tend to be limited. Policymakers are often very concerned about the effects of trade on overall employment. It is important to recognize that in principle trade opening or trade shocks should not have an effect on overall employment levels in the long run—this will be determined by macroeconomic variables and labour market institutions. Trade may affect the quality of jobs – through increased demand for workers with higher skills or by providing workers with greater access to productivity-enhancing

equipment and tools – but generally not the quantity of employment. This observation may not hold, however, for developing countries where there is significant under-employment prior to trade opening – trade opportunities may translate into investment in tradable sectors and increase formal employment.

In the short run, unemployment may rise as a result of reforms or a trade shock. Some studies have concluded that transitional unemployment is not very large relative to total employment.[2] However, there is little evidence on the nature and extent of transitional unemployment in developing countries, at least in part owing to the difficulties of measurement. Economically meaningful work cannot be equated with formal employment in low-income developing countries as most employment is informal (Maloney, 2004). A further unknown is whether those who lose employment as a result of a trade shock are disproportionately poor.[3] Whatever the overall size of the short-run labour market effects, the impacts will often be significant for those who lose their jobs. Indeed, the contributions to this volume focus on such short(er) run effects of trade reforms and openness – and the factors that affect the magnitude of the adjustments induced by changing trade opportunities and incentives.

2. OVERVIEW OF THE CONTRIBUTIONS

This book is divided into four parts. The first comprises analyses of the magnitude of trade adjustment costs in the presence of frictions in factor markets. The second discusses the impacts of trade shocks and greater trade openness. A wide range of topics are explored including the overall adjustment of the manufacturing sector, the nature of labour reallocation, the consequences of offshoring and migration on labour markets, the role of labour income risk, adjustment in child labour and schooling, the consequences of increased FDI, and general patterns of adjustment to changes in trade policies. The third part deals with some of the factors that affect the way trade, especially exports, adjust, including various types of transaction costs (e.g., transportation, search, market penetration and learning costs), access to credit and finance, or the need to comply with stringent product standards in export markets. Finally, the forth section provides a brief overview of trade adjustment assistance programs in the U.S. and compensation schemes for farmers in the EU.

[2] A multi-country study of trade liberalization before 1985 (Papageorgiou, Michaely and Choksi, 1991) argued that experiences varied from case to case, but that, on the whole, transitional unemployment was quite small. In a survey of more than fifty studies of the adjustment costs of trade liberalization in the manufacturing sector, mostly in industrialized economies, Matusz and Tarr (1999) also find that unemployment duration is generally quite short.

[3] Some evidence is available on the relationship between public sector job loss and poverty. Although such job losses are not due to trade shocks, they do provide information on transitional unemployment resulting from a reform. In Ecuador, employees dismissed from the central bank earned, on average, only 55 per cent of their previous salary 15 months later (Rama and MacIsaac,1999). In Ghana, Younger (1996) found that most laid off civil servants were able to find new work, albeit at substantially lower income levels, but that income levels and poverty incidence after job loss were not substantially different from the average for the country.

The Magnitude of Adjustment Costs

The cost of adjustment to increased trade openness depends importantly on the presence and magnitude of frictions in the operation of factor markets. Carl Davidson and Steven Matusz describe a model with frictions in labour adjustment. There are two sectors in the economy, one of which is initially protected by tariffs. Labour adjusts freely in the protected sector, but to enter the unprotected sector workers need to acquire skills (via training) and must search for jobs. When trade is liberalized, the protected sector shrinks and the unprotected sector grows, but does so only slowly because of the need for job training and search processes. The authors utilize this setting to illustrate that adjustment costs to greater trade can be quite large. Using numerical simulation exercises they conclude that adjustment costs can range from one third to 80 percent of the gross benefits from trade reforms. In addition, the process of adjustment takes time: output dips immediately after the trade reform and it takes between one and two and a half years to return to pre-liberalization output levels.

Similar results are reported by Artuc and McLaren, who develop a structural model of trade shocks and labour adjustment. The model can be estimated with limited data, a feature that facilitates replication in many developing countries. Using data for Turkey, Artuc and McLaren simulate the effects of a hypothetical trade liberalization on the Turkish labour market. They find that the costs for workers of switching between industries are high. This makes adjustment slow: the post-liberalization steady state would only be reached after a decade or so. In the presence of labour mobility costs that impede costless labour reallocation following trade liberalization, wages in the de-protected sector can decline by as much as 20 percent. The authors argue that despite the high switching costs, adjustment during the transition allows for a substantial wage recovery after a few years. This result reinforces the notion that the ease of adjustment via job reallocations can mitigate any initial losses from liberalization.

Carlos Casacuberta and Nestor Gandelman use Uruguayan manufacturing data to measure the costs of reallocation of capital, blue collar and white collar workers between 1982 and 1995, a period of significant tariff reforms. They provide two sets of findings, one on the extent of factor reallocation and one on the nature of factor adjustment. The Uruguayan rates of creation and destruction of capital and blue and white collar jobs were high and pervasive during 1982-1995. Greater exposure to international trade was associated with much higher job and capital destruction but only with slightly higher job creation. Moreover, trade liberalization was associated with higher reallocation rates for all factors of production. However, the authors report that the overall job reallocation rate in Uruguay was 14 percent, lower than in the rest of Latin America. To explore whether this is due to high costs of labour adjustment, Casacuberta and Gandelman estimate factor adjustment functions, which reveal the percentage of the factor (employment, capital) gap that is actually closed when factors adjust. The key result is that Uruguayan firms tend to find it easier to create employment than to destroy it. For instance, if a firm desires to cut factor employment by half, it

would in fact only reduce white collar workers by 5 percent, blue collar workers by 10 percent, and capital by 5 percent. In contrast, if a firm wants to double factor usage, it would only increase white collar workers by 12.5 percent, blue collar workers by 15 percent and capital by 5 percent. This suggests high adjustment costs in both hiring and firing. The authors also find that the creation and destruction of factor employment is interdependent in an asymmetric way. When firms want to hire more factors, they adjust one factor at a time. However, when they want to employ fewer factors (labour or capital) they tend to reduce the use of all factors together. Finally, the pattern of adjustment depends on the intensity of tariff protection: firms that experienced higher tariff cuts destroyed more capital and jobs than firms that faced lower tariff cuts.

The chapters by Jaime de Melo and by Olivier Cadot, Laure Dutoit and Marcelo Olarreaga take a different approach. The focus of much of the literature on the magnitude of adjustment costs tends to be on adjustment in labour markets. But in low-income developing countries, trade involves primary products that are, in many cases, produced by small farmers. What is the nature of the process of factor use adjustment in such countries? De Melo focuses on the cashew sector in Mozambique and the vanilla sector in Madagascar following market liberalization reforms in the 1990s. Analyzing the scarce available data, he concludes that while the reforms did increase producer prices, there were negligible supply responses at the farm level. In both cases, the simulations run by de Melo show very small gains, and thus limited impacts on either the distribution of income or on poverty. He attributes the low supply response following these reforms to two major factors. The first is the initially low participation in cash cropping. The second is high sunk costs in agriculture. For both vanilla and cashew production, there are significant sunk costs associated with planting new trees, an investment that requires a credible pricing policy. This credibility is, in turn, linked to the nature of prevailing institutions. The implication is that the impacts of agricultural (trade) policy reforms depend primarily on other factors.

Cadot, Dutoit and Olarreaga provide concrete examples of the role of transaction-related costs as determinants of adjustment following reforms (or shocks) in Africa. They note that while market agriculture dominates subsistence production, many African farmers remain in virtual autarky. Using data for Madagascar, the authors calculate that subsistence farmers could increase household incomes by over 40 percent by switching to selling for the market. This finding can be largely explained by high barriers to exit from subsistence, including risk, missing markets, and various transactions costs. In consequence, to understand why markets fail and farmers are not responsive to price incentives it is necessary to identify which transaction costs are prohibitive, for whom, and why. Three such costs stand out. First, variable transaction and transportation costs create a wedge between food farm-gate prices and local market prices (from neighbours or local dealers if there is a village market). Cadot et al. report estimates of such variable transaction costs ranging from 15 to 30 percent. Second, there are fixed transaction costs. These include searching for partners, enforcing contracts with distant buyers, and establishing quality. Here, the estimates range from a low of 15

percent to as high as 77 percent. Finally, there are sunk costs in shifting to market agriculture. Using farm data for Madagascar, the authors use estimates of earnings differentials in market vis-à-vis subsistence agriculture to estimate the sunk costs of leaving autarky. These range from 124 to153 percent of the value of annual output at market prices. This is interpreted as a once-and-for-all sunk cost since it is calculated from a comparison of lifetime earnings. Given the size of the various costs many farmers are effectively isolated from the price incentives associated with trade liberalization. Importantly, these various costs prevent not only the realization of the potential gains from trade, but also any other gains that can arise from adjustment to trade.

Adjustment Impacts

The contributions in this section of the book span research on the impacts of trade policy and trade shocks. Focusing on the case of Brazil, Marc Muendler provides a comprehensive description of labour reallocation following episodes of trade liberalization during the early 1990s. Muendler begins with a labour demand decomposition. Within the traded-goods sector, there is a reduction in demand for low-skill workers in favour of better educated workers. Between sectors, there is a labour demand shift towards both the least skilled (used intensively by traded-goods industries) and the most skilled (used intensively by nontraded-output industries). Using a very rich linked employer-employee dataset, Muendler shows that workforce changeovers mostly occur because as trade-exposed industries shrink they fire low-skill workers more frequently than high-skill workers. The displaced workers tend to shift to nontraded-output industries or out of recorded employment (informality). In particular, the Brazilian trade liberalization experience reveals that job losses occurred in protected industries that are not absorbed by either comparative-advantage industries or exporters. To explore why, Muendler goes on to combine the linked employer-employee dataset with information from a Brazilian manufacturing survey. His analysis reveals that labour flows away from comparative-advantage sectors and from exporters because their labour productivity increases faster than their production. This happens because for these firms and industries the larger market potential resulting from trade liberalization offers stronger incentives to improve efficiency. However, if productivity increases faster than production, then output shifts to more productive firms but labour does not.

Gordon Hanson discusses the evidence on overall adjustment in the manufacturing sector with an emphasis on outsourcing. Based on current developments in trade theory and empirics, Hanson argues that traditional factor abundance models cannot fully describe the most relevant type of adjustments in the manufacturing sector. This is a recurring theme in the volume. Hanson enumerates various channels that are bound to be important in the adjustment process. The evidence shows that falling trade barriers are associated with the exit of less productive firms, rising average industry productivity, greater fragmentation of production, greater volatility of employment, and possibly more informality. To

assess wage impacts, Hanson focuses on how the global fragmentation of production affects the wage structure in developing countries and concludes that outsourcing has played a major role (both in increasing wage inequality and in raising the volatility of employment), especially in countries like Mexico or China. However, these adjustments are not present in other developing countries like Argentina, Brazil, Chile and India, where quality upgrading, or skilled biased technological change accompanying liberalization are more likely to play a bigger role.

Margaret McMillan takes a deep look at production off-shoring, the reallocation of physical manufacturing processes outside of a country's border, and the impact on labour markets (employment and wages) both in developed and developing countries. Her review of the literature for developing countries corroborates the conclusions reached by Gordon Hanson above. Indeed, most researchers find that foreign firms pay higher wages and conclude that FDI has beneficial effects on host country labour markets. However, the magnitude of these effects varies substantially. The employment effects of FDI in developing countries are less well understood but they are likely to be important. It is noteworthy that outsourcing to Mexico, for instance, has caused an increase in employment volatility. For developed countries (drawing heavily on studies about the U.S. labour markets), the evidence on employment is mixed, with several studies finding complementarities between the operations of U.S. multinationals abroad and domestic activity and others reporting instead evidence that "jobs abroad do replace jobs at home." These effects are, however, small.[4] The impact on U.S. wages is often small too. McMillan ends with an interesting observation: in recent years, we have started to see off-shoring from developing countries like China and India.

The following three chapters address additional themes related to labour market adjustments in developing countries. Pravin Krishna and Mine Senses look at increased labour income risk following liberalization. This could happen if openness exposes import-competing sectors to a variable international economic environment. If trade induces reallocations of capital and labour across firms within and between sectors and if similar workers experience different outcomes during this reallocation process, openness will raise individual labour income risk. Another link between trade and income volatility emerges when increased foreign competition increases the elasticity of demand for goods and thus the elasticity of derived labour demand. This, in turn implies that shocks to labour demand may result in larger variations in wages and employment, and hence increase volatility in the labour market. Using U.S. data, Krishna uncovers two observations: first, those workers who switched industries experienced higher income

[4] McMillan also addresses the source of this discrepancy and concludes that for U.S. parents primarily involved in horizontal activities, affiliate activity abroad substitutes for domestic employment but for vertically-integrated parents, home and foreign employment are complementary. She furthermore claims that offshoring is not the primary driver of declining domestic employment of U.S. manufacturing multinationals between 1977 and 1999 (the culprit instead being falling prices of investment goods, falling prices of consumption goods, and increasing import competition).

risk than those who stayed; second, those who switched to the non-manufacturing sector experienced higher income risk than those who switched within manufacturing. A key finding is that within-industry changes in income risk are strongly related to changes in import penetration: quantitatively, an increase in import penetration by ten percent is associated with an increase in the standard deviation of persistent income shocks of about 20 to 25 percent. This increase in persistent income risk is (certainty) equivalent to a reduction by roughly 4 to 11 percent of lifetime consumption.

Eric Edmonds reviews the adjustment impacts on child labour and schooling. This is important for developing countries because it gives rise to intertemporal repercussions of trade liberalization via human capital accumulation of children. The main finding of research on this issue is that the dominant channel through which trade influences child time allocation and schooling is through impacts on adult labour and family asset income. Despite many possible channels, there is very little evidence supporting any connection between trade and child time allocation other than through the impact of trade on the living standards of the very poor. This reflects two fundamental reasons. First, among the poor, the level of income (i.e., the standard of living) is one of the most important determinants of child time allocation. Second, children are rarely engaged in work that will be easily connected to international trade. In poor countries, children work mostly in agriculture (much of which is for home consumption). Outside of agriculture, children are not intensively involved in traded sectors in general, and within manufacturing, firms involved in trade tend to be relatively more skill intensive.

Gordon Hanson tackles the issue of adjustment to international migration both in sending and in receiving countries, and of migrants themselves. While international migration is limited (considering the large income differences across countries), the economic impacts, especially for developing countries can be sizeable. Looking at migrants, there is ample evidence of large gross income gains to migrants (the net gain, however, is unknown, given little evidence on migration costs). Also, through remittances, migrants share a portion of their income with family members at home, who thus benefit as well from the process. On the other hand, no study suggests the existence of large negative consequences from global migration. In the U.S., the net impact of immigration is negligible (though some studies show negative wage effects). Hanson concludes that unless there are large unmeasured negative externalities from migration or migration exacerbates existing distortions in ways that have not yet been detected, it is difficult to justify restrictive barriers to global labour flows.

The chapter by James Harrigan switches the attention to trade preferences and describes the adjustment of exporters from developing countries to the elimination of the "Multi Fibre Arrangement" (MFA). The end of the MFA in 2004 led to big changes in the pattern of U.S. imports of textiles and apparel from the developing world and offers a unique opportunity to learn how adjustment to a specific (global) change in trade policy takes place. Harrigan emphasizes four major findings: (i) China, which had been severely constrained by the MFA, registered a steep decline in export prices and soaring export volumes after the MFA ended, reflecting the

country's comparative advantage in low-wage products; (ii) Other low-wage MFA-constrained exporters also saw big increases in their exports to the U.S.; (iii) The "East Asian Miracle" exporters experienced steep declines in export values, despite having large shares of filled quotas in 2004 – the consequence of higher real wages and the end of preferential access to the U.S.; and (iv) Mexico suffered losses from the end of the MFA, as it lost its previously-privileged access to the U.S. market, but these losses were somewhat cushioned by its proximity to the U.S., a factor that insulated Mexico from competition from lower wage Asian exporters.

Beata Javorcik discusses the adjustment of indigenous producers to FDI inflows. Multinational corporations (MNCs) are characterized by large endowments of intangible assets which translate into superior performance. This has three implications. First, foreign affiliates contribute directly to increasing the level of productivity of the host country. Second, superior performance means much stronger competition for indigenous producers. Third, MNCs are also a potential source of knowledge spillovers. Javorcik provides different pieces of evidence to support these implications. Empirical evidence from Indonesia shows that new foreign entrants taking the form of greenfield projects exhibit higher productivity than domestic entrants or mature domestic producers. Furthermore, evidence on foreign acquisitions of Indonesian plants suggests that such acquisitions lead to large and rapid productivity improvements taking place through deep restructuring of the acquisition targets. Javorcik also reviews evidence from enterprise surveys and econometric firm-level studies. She concludes that FDI inflows increase competitive pressures in their industry of operation and lead to knowledge spillovers within and across industries. Finally, she reviews the implications of inflows of FDI into service sectors and argues that the presence of foreign services providers may increase the quality, range and availability of services, thus benefiting downstream users in manufacturing industries and boosting their performance.

In the final chapter of this section Chad Bown discusses how domestic economies adjust when other countries change their trade policy. Bown makes the point that most countries that belong to the WTO face limited scope for large-scale tariff liberalization. Instead, they can resort to "exceptions" (safeguards and anti-dumping) that have impacts both on specific countries and specific industries. These exceptions, which are often applied in a discriminatory manner, imply that a given policy can have impacts not only on those countries and industries directly affected but also on third-partly players. Bown argues that this provides a very fruitful setting to explore in details the type of microeconomic adjustments to trade policies that are relevant for the agenda of this volume. Furthermore, and importantly, these experiments have nice properties facilitating the econometric identification of those patterns of adjustments. The author provides three examples of adjustments identified via this channel: (i) the sizeable "trade deflection" and "trade depression" of U.S. antidumping and safeguard restrictions on Japanese exports; (ii) the adjustment of farms in Vietnam to the U.S. antidumping on catfish; and (iii) the behaviour of Indian steel exporters in the face of U.S. safeguards imposed on imports from major world exporters (not including India).

Factors that Affect Export Responses

The next set of chapters investigate the effects of factors that help determine the way an economy adjusts to trade shocks. David Hummels discusses the interaction between transportation costs and trade shocks. His main point is that transportation costs are not an exogenous friction (as in the traditional iceberg-cost specification) but are endogenous to how production is organized and what is traded. Hummels lays out a simple framework to think about non-iceberg transports in which shipping charges are an increasing function of the product price (because higher value goods require more careful handling and higher insurance premiums). This framework, which better fits the stylized facts on transportation costs, helps to uncover a number of interesting issues. First, transportation costs depend on the composition of what is shipped. As a result, trade shocks matter for transport costs. For instance, in periods of rapidly rising demand, shipping capacity becomes scarce, ports become congested, and spot shipping prices rise quickly. Over longer periods however, rising demand for shipping may actually lower shipping prices, especially in smaller countries with initially low trade volumes. In these cases, tariff liberalization may cause reductions in shipping costs. Second, the ratio of weight to value of a product, which varies widely across goods, matters for transportation costs. Differences across countries in the product composition (weight/value ratio) of trade largely explain why developing countries pay nearly twice as much as developed countries for transporting goods internationally. Third, the existence of per unit transport charges raises the relative demand for high quality goods – the Alchian-Allen effect. Increases in input prices used in transportation (like fuel) would generate similar effects of shifting demand towards high value goods. Fourth, the non-iceberg nature of transport costs act as a kind of shock absorber, dampening the transmission of product price shocks to delivered prices. Hummels ends with a discussion of "transportation costs" broadly defined to include non-tariff costs of trade like information about foreign markets, marketing and distribution costs, product adaptation to local tastes and regulatory requirements, and timeliness. Some of these costs are discussed in greater depth in the next three chapters.

Tibor Besedes and Thomas Prusa study the duration of trade relationships in the United States. The authors apply survival econometric analysis to U.S. import data and find that international trade relationships are far more fragile than previously thought, with the median duration of exporting to the U.S. ranging from two to four years. However, the value of trade embodied in the small number of long-lived relationships is much larger than in the short-lived ones. They also find that if a country is able to survive in the exporting market for the first few years it will face a very small probability of failure and will likely export the product for a long period of time. Finally, they find that product differentiation significantly affects trade duration: the hazard rate is at least 23 percent higher for homogeneous goods than for differentiated products. Besedes and Prusa conclude that these findings are consistent with a matching model of trade formation whereby buyers need to find reliable sellers to secure a consistent supply of exports.

James Tybout and Costas Arkolakis and Olga Timoshenko take a broader look at export costs. Tybout argues that assumptions on trade costs are crucial to our understanding of adjustment because these costs govern the reallocation of resources after trade reforms. He questions two assumptions of the literature on heterogeneous firms: that the costs of breaking into foreign markets and of staying in are exogenous and common across firms; and, in line with Hummels' work, that the variable costs of exporting are proportional to the physical volume of goods exported. The first assumption is rooted in early evidence on the sunk and fixed costs of exporting. These one-time costs, which can be significant, capture the fact that, in order to begin exporting, firms must learn bureaucratic procedures, establish distribution channels, and repackage or even re-design their products for foreign consumers. While large sunk costs can explain why only few firms enter export markets and why there is persistence in exporting, they are inconsistent with a new set of stylized facts that are emerging from transactions-level data on international shipments: on the one hand, one-third to one-half of all the commercial exporters observed in customs data did not export in the previous year; on the other, most of these firms ship only small amounts for one year and then revert back to the domestic market in the following year. Tybout proposes a model with "market penetration costs" and "customer signalling" to take account of these observations.

Market penetration costs are discussed in detail in Arkolakis and Timoshenko. In their framework, the search process shifts from buyers to sellers. Exporters are assumed to easily find the first few customers in a destination market but incur increasingly higher costs in reaching more hard-to-find buyers. In consequence, exports costs rise more than proportionately with export sales. In this setting, initial non-exporters who experience favourable productivity shocks can easily enter export markets but, at the same time, can face larger costs down the road and exit if the search for additional customers fails. This model can explain the large volume of short-lived, small-scale exporting episodes, the surge in shipments among a small set of successful new exporters, and the growth slowdown as firms' exporting relationships mature.

Tybout discusses an extension of this theory in which sellers search for buyers and learn from this search. Search costs are increasing in sales, but each successful sale gives the exporter a noisy signal about the appeal of the product to consumers in the destination market. A large order from a new buyer signals that the product is likely to be popular with others, while a small order signals the opposite. Each time a match is made and a signal is conveyed, the exporter updates his priors concerning the product's appeal and adjusts his search intensity. Early signals are the most informative, so they result in the largest adjustments in search intensity. These types of models can potentially fit the data well. They are also helpful tools to illustrate what type of export costs is more realistic and how they affect the expected adjustments to trade reforms.

Kala Krishna turns the attention to how the existence of distortions affects how domestic economies adjust to – and are affected by – trade liberalization. Based on the theory of the second best, she argues that the interaction of distortions

(broadly defined) with trade liberalization helps to explain why poor countries sometimes fail to realize the gains from trade. Thus, it is important to identify these distortions *ex ante* and alleviate them concurrently with any trade liberalization effort. Krishna offers an important conclusion: more often than not, product market distortions are less likely to be made worse by trade than are factor market distortions. While the result is not a general one, the examples given by Krishna are compelling and provide support for the contention that policy-makers should look first at factor market distortions. Krishna argues that the same logic can be used to think about other features of developing countries that act as "distortions" (broadly defined) in the context of a disguised second best theorem. She illustrates the argument with examples of deficient infrastructure, corruption, inefficient and weak legal systems, etc. These "distortions" and their consequences, can be made worse by liberalization – the example of hold-up problems in agriculture provides an illustration.

Kalina Manova analyzes the role of credit constraints. Standard trade theory assumes that firms can expand and contract at no cost. In particular, the models abstract from market frictions that may arise from agency problems. In practice, however, the various costs of exporting (some of which were described above) may need to be covered up-front and this requires external financing. Imperfections in credit and financial markets then translate into barriers to trade and impediments to adjustment. Credit constraints, for instance, interact with firm heterogeneity and reinforce the selection of only the most productive firms into exporting. This means that credit constraints affect both the extensive margin (the number of firms exporting; the number of export destinations) and the intensive margin (firm-level exports) of trade. Manova reviews the literature and concludes that: (i) There is robust empirical evidence that credit constraints are an important determinant of global trade flows; and (ii) credit constraints reduce countries' total exports by affecting all margins of trade. More specifically, she argues that a third of the effect of financial development on trade values is attributable to firm selection into exporting, while two-thirds is due to firm-level exports. This suggests that firms face binding credit constraints in the financing of both fixed and variable export costs. Manova also studies the role of foreign financial flows. She explores the effect of equity market liberalizations, and finds that they increase countries' exports disproportionately more in sectors intensive to external finance and intangible assets. These results suggest that pre-liberalization, trade was restricted by financial constraints, which foreign portfolio investments relax to a certain degree.

Jo Swinnen and Miet Maertens explore the role of product standards as a determinant of exports of food and agricultural products, and whether they can obscure the benefits from trade liberalization. They investigate the role of standards as barriers or catalysts to trade and as barriers or catalysts to development. On the question of trade, they raise concerns that standards can act as non-tariff barriers for countries which face constraints in their ability to document compliance with stringent standards. However, Swinnen and Maertens show that there is only limited evidence that standards act as barriers. Instead, standards

may facilitate trade between countries with diverging norms. On the question of standards as catalysts to development, the authors raise concerns that standards can facilitate the exclusion of poor farmers from the supply chain and that they can facilitate the exploitation of smallholders, who would lose bargaining power (vis-à-vis large food exporters and multinational food companies). The evidence on the exclusion of smallholders because of high compliance costs and increasing levels of vertical coordination is mixed: there are cases of complete vertical integration with hardly any smallholder involvement (tomatoes in Senegal or fruits and vegetables in Zambia), and there are cases in which export production remains dominated by smallholders (vegetables in Madagascar and Ghana). In contrast, the evidence against exploitation of smallholder producers is more compelling. Swinnen and Maertens enumerate some of the benefits of high-standards contract production, including productivity gains, increased household income, reduced income volatility, technology spillovers, employment (downstream), and poverty reduction.

Adjustment Assistance Programs

The volume concludes with reviews of the trade assistance program in the United States and the agricultural support program of the E.U. Most developing countries do not have trade-specific adjustment assistance programs, raising the question what can be learned from the experience of high-income countries in this regard. The first paper, by David Richardson, analyzes the Trade Adjustment Assistance (TAA) program of the U.S. After a brief historical account of the American TAA, its mandate and coverage, the author assesses the role of a revised TAA program that could successfully deal with the modern process of global integration. Richardson argues that there are two main features of the current "integrated integration" process: large gains for the best-fitted agents (most productive firms, more able or motivated workers) and an increasingly unbalanced distribution of those gains against the less fit. In this context, Richardson claims that a successful TAA, one that would actually help improve and economy's overall performance and welfare, should increase its scale, its scope (and be transformed into a sort of structural adjustment assistance), and its constituency to harbour both "natural" American institutions (labour unions, community colleges) and new American institutions (like not-for-profit social services or insurance companies).

The last chapter by Jo Swinnen describes the history of the E.U. Common Agricultural Policy (CAP). The main objective of agricultural policies in the E.U. was to support agriculture in order to protect farm incomes and employment from more general market liberalization. Initially, this was done by setting high import tariffs and export subsidies, and by fixing prices. Following the Uruguay Round, multilateral disciplines were negotiated for agricultural support policies in the WTO, and tariffs, export and production subsidies were partially replaced by "compensation" (direct) payments to farmers in the 1990s and, later, by so-called decoupled payments to farmers (not linked to output) in reforms in 2003 and 2008. These reforms can be regarded in some sense as adjustment policies – the

aim being to compensate farmers for the changes while still achieving the objective of the CAP of ensuring a "fair standard of living" for farmers and "stabilizing markets." Swinnen argues that while farm incomes are directly affected by the CAP, the observed catch-up to incomes in other sectors has been the result of non-CAP payments (for example, the integration of rural areas in factor markets and in the rest of the economy). Furthermore, there is no evidence of any impact on long-run employment levels in EU agriculture. Second, he argues that the evidence on stabilization is more nuanced. The old CAP system of price interventions reduced price variability but did not necessarily provide a good safety net. The current direct payment system has less or no impact on price variability, but does reduce income variability and reduces risk in farming households by providing a guaranteed source of income. Looking forward, Swinnen argues that the periodic reforms of the CAP have been successful in reducing market distortions but he doubts the Single Farm Payment system will address key challenges of the future, such as climate change and ensuring food quality.

3. IMPLICATIONS FOR POLICY AND RESEARCH

The benefits of trade and trade reforms are conditional on many factors. The contributions to this volume make clear that a variety of domestic distortions and transactions costs can be major impediments to adjustment. Government policies (or the absence of policies) therefore can play an important role in the adjustment process. Indeed, a key responsibility of governments is not only to ensure that economic units confront the "right" incentives to induce investment in activities in which a country has a comparative advantage, but also to assist in facilitating adjustment to technological changes and policy shocks. Such assistance will generally reflect a mix of economic and social motivations, i.e., ranging from a focus on overcoming market failures and other distortions that constrain adjustment to the realization of equity (distributional) objectives.

The conventional wisdom in the trade literature is that the aggregate gains from trade will generally exceed aggregate losses, in principle allowing the winners to compensate the losers while still remaining better off. In practice, of course, losers often are not compensated, in part because compensation is difficult to implement—governments may not have the instruments needed.[5] The chapters in this volume have generally not focused on the policy implications of research. However, one conclusion that can be drawn is that the focus of policy should not be limited to assisting the losers or attenuating negative impacts—issues that have tended to attract much attention in the policy literature (see, e.g., the chapter by Richardson) – but should span efforts to remove or reduce the transactions and other costs that limit desirable adjustment and therefore reduce the aggregate gains from trade. Ad-

[5] Verdier (2005) notes that trade integration can be expected to affect the redistributive capacity of governments in several ways. Trade opening may change the structural parameters of the economy, making redistribution more or less difficult. From a political perspective, it may affect the pattern of political power and coalitions, preventing or promoting compensation through the redistribution of resources inside the economy.

justment costs may be so high that the potential benefits of trade are only partially realized if at all or are distributed in a very asymmetric way, e.g., accruing primarily to higher income, urban households.

The appropriate measures to encourage and support adjustment to greater trade openness are not the primary focus of this volume. One general conclusion that can be drawn from the development literature is that the business environment or investment climate broadly defined should be at the centre of policy attention. Without a stable macro-economy and realistic exchange rate, adequate infrastructure, human capital and functioning input markets, the benefits of openness are reduced. The research focusing on adjustment to trade helps to identify other, more specific areas for government policy or action. These centre on the functioning of factor markets, rural product markets and other input markets – reducing the various fixed costs discussed above and enhancing the productivity of domestic firms and farmers.

Much of the more policy-oriented literature on adjustment costs centres on the need for, and design and effectiveness of programs to assist poor or vulnerable households to cope with shocks. This is not a focus of most of the chapters in this book for the simple reason that such programs should not be targeted towards or limited to the social adjustment costs of changes in trade. There is a long history of direct assistance for restructuring of firms or industries in developed countries. This spans subsidies and bailouts and government involvement in downsizing industry or managing supply through "crisis cartels" and forced consolidation through mergers. Such policies are often very costly, in part be prolonging the adjustment period and distorting competition (Noland and Pack, 2003). Policies are better directed at facilitating adjustment through pro-active labour market policies, retraining programs and financing for skills enhancement (Richardson, this volume). Khanna, Newhouse, and Paci (2010) suggest a policy package that combines (1) income maintenance programs – that is, cash transfers to low-paid poor workers; (2) interventions that facilitate flexible-hours arrangements; and (3) policies that compensate workers for temporary reductions in standard working hours – for example, by granting partial compensation from the unemployment benefit system or by providing paid training opportunities.

Taking advantage of the opportunities created by trade and investment liberalization often requires substantial effort and investment in upgrading the production process. At a general level, neither theory nor experience provides unambiguous guidance regarding the design of policies to support such investments. As discussed in the chapter by Javorcik, much depends on whether there are spillovers, whether these are international or intra-national, and on the capacities of firms and workers to absorb and adapt new technologies.[6]

Exit by low performers and entry of new firms with incomplete information on their "capacity" is a major channel for the efficiency gains from trade reform.

[6] Hoekman, Maskus and Saggi (2005) argue that appropriate policies in this area follow a technology ladder, with the priority in poor countries with weak institutions and limited R&D capacity being to improve the business environment and encourage imports of technology embodied in goods.

An implication is that governments should promote entry by new firms and remove barriers to exit, which may include restrictive labour market regulation – see for example Besley and Burgess (2004). Given that managers of firms confront incentives to improve performance as trade openness increases, measures to promote innovation (R&D) and to assist on upgrading of existing firms can make sense (Motohashi, 2002; Pakes and Ericson, 1998). This can include policies that facilitate learning about managerial performance and measures that encourage the use of new technology to improve performance (Hoekman and Javorcik, 2004). More generally, many of the chapters in this volume suggest that taking action to reduce transactions costs, improve access to credit, and improve access to information through trade support services can be very beneficial from a poverty reduction and trade expansion perspective. Much of this agenda is increasingly being supported through various aid for trade programs that are provided by bilateral donors and multilateral development agencies.

Turning to factor market "distortions", a key feature of the theoretical models that are used to assess the magnitude of adjustment costs is that factors of production cannot move freely from sectors that shrink due to liberalization to those that expand. The trade literature on adjustment costs has emphasized frictions in labour markets so that the costs of adjustment to trade occur during labour reallocation. Displaced workers cannot easily find jobs in expanding sectors because, for instance, the new jobs require specific skills that need to be acquired, a process that takes time.

The quantification of the welfare costs of these barriers to labour mobility in the literature tend to use of structural models and involve calibration exercises. A potential next step in this research is to make more use of micro (survey) data in the estimation of costs. Another potential extension of research in this area is to explore additional sources of labour immobility such as adjustment costs in the capital stock and in investment (as in the business cycle literature). In the presence of complementarities between investment (i.e., desired capital stocks) and labour, the cost of adjusting the capital stock can affect employment and wages. If capital cannot easily adjust to reforms, is this a sizeable source of welfare losses for workers?

The work done on adjustment costs in Africa provides some answers to this question. The chapters by de Melo and by Cadot, Dutoit and Olarreaga (and the literature they review) argue that the welfare costs of reforms are generated by feeble supply responses in agriculture and by barriers to exit from subsistence into market agriculture. Given that the resource allocation involves farm labour (among other factors), the problem can be framed in terms of a labour mobility cost theory. A major reason identified by these authors to explain the limited adjustment in rural Africa is insufficient capital investment. The sunk costs argument that is presented in these chapters is, to a large extent, a manifestation of the impossibility to adjust capital and non-labour inputs like trees (vanilla, coffee, cashews), seeds, machinery, pesticides, etc. The list of possible barriers to exit from subsistence is large. It would be very helpful for policy analysis and design to have a broad menu of the possible barriers and a tentative ranking of their rel-

ative importance for smallholders. Moreover, while this problem is surely very general and widespread in rural economies, from an adjustment to trade perspective research should focus on tradable crops such as cashews in Mozambique, coffee in Uganda, cotton in Zambia, or vanilla in Madagascar.[7]

From a policy perspective the intra-household impacts of trade reforms are also of interest. When a worker is unable to move into an expanding sector, or if a farmer can in fact escape subsistence farming due to improved opportunities for exports, how are other family members affected? Some intra-household effects, such as child labour and schooling, are already the focus of research that is reviewed in this volume (e.g., the chapter by Edmonds). The exploration of other impacts and mechanisms of adjustment within the household or community could generate interesting additional research results.

More generally, much more can be done to deepen our understanding of the impacts of adjustment. The research on labour market developments in Brazil following the liberalization of the 1990s, the response of wages and employment to trade (including outsourcing), the impacts on income risk, the effects of migration, the response of exporters in developing countries to abrupt market access changes, the role of FDI, the impacts of corporate governance, are just a few examples of the type of adjustment impacts that can arise. There are many more – such as impact of trade on the quality of inputs (including services), on the quality of outputs, or on innovation. More research on the spillover effects of global shocks or specific trade policy measures – such as the abolition of the MFA or the imposition of discriminatory trade restrictions by one or more countries – is another example.

The research agenda on factors that affect export responses is also open. Much work is being done to better understand the nature and effects of trade costs faced by exporters. The standard notion of sunk and fixed costs of exporting is now being extended to assess market penetration costs, learning effects, search costs, coordination between buyers and sellers, etc. In part this effort reflects the increasing availability of data on trade transactions. As more and richer datasets become available, the range of questions related to trade costs that can be subjected to empirical research will widen.

The contributions in this volume cover only a subset of the factors that impede or affect adjustment, including factor market distortions, credit constraints and transactions costs. This reflects the focus of existing research, but does not imply that these are the only relevant factors. Identification of general market failures that affect adjustment is an important avenue for research. Similarly, more research on the factors that affect the ability of firms to enter markets and expand market share – along the lines of the work on product standards – will surely be informative from both a research and policy perspective.

[7] This does not imply that the constraints to food production are less relevant than those for exportable crops, but the latter are more relevant from the perspective of trade adjustment.

There are several themes that are not covered by the chapters but that are potentially relevant for the adjustment costs/impact agenda. Two areas stand out. One is the nature of adjustment in factor markets in low-income developing countries. The research on labour reallocation typically focuses on more advanced, middle-income economies, like Brazil or Uruguay; as well as developed countries. There is much less knowledge on adjustment in low-income countries, especially in Africa, where the share of manufacturing is generally very small and issues of agricultural adjustment at the smallholder level are much more relevant. The papers on the dichotomy between cash crops and food crops in Africa are good examples of the type of work that is needed, but needs to be expanded to include additional issues. Data limitations are of course a major constraint. The linked employer-employee data available in Brazil, for instance, does not exist in most poor countries with a large rural sector. Survey data on households, firms, and workers are, however, increasingly becoming available and will hopefully be used more to study trade-adjustment issues.

Another priority area for research is the role of domestic institutions, building on the types of insights that are developed in the chapter on distortions by Kala Krishna. While the term ``institutions'' is often too broad to generate specific policy insights or guidance, it is clear that a deeper understanding of how a given economy adjusts to trade requires deep knowledge of the setting that characterizes the functioning of the economy. This depends heavily on the institutional context that governs incentives and therefore drives the adjustment process.

The synthesis of research on adjustment to trade in developing countries collected in this volume demonstrates that our knowledge is still very imperfect. However, it does suggest that for developing countries – especially low-income economies with large informal and agricultural sectors – the key issues go beyond adjustment by (wage) workers to a new equilibrium and the transitional costs of unemployment, job search and so forth. These are issues that have attracted much of the attention in the trade and labour literature, in part driven by political economy considerations (e.g., Sapir, 2000; Verdier, 2005). While certainly also relevant for developing countries, identifying and addressing the constraints that impede the ability of households to leave subsistence and the informal sector, and that limit the scope for firms, farmers and communities to benefit from greater trade opportunities should be a research priority.

Bernard M. Hoekman is Director of the International Trade Department of The World Bank and a CEPR Research Fellow.

Guido Porto is Professor of Economics at Universidad Nacional de La Plata, Argentina.

REFERENCES

Besley, T., and R. Burgess (2004). "Can Labour Market Regulation Hinder Economic Performance? Evidence from India," *Quarterly Journal of Economics* 119(1): 91–134.

Bhagwati, J. (1971). "The generalized theory of distortions and welfare," in: J. Bhagwati, R. Jones, R. Mundell and J. Vanek (eds.), *Trade, Balance of Payments and Growth: Papers in International Economics in Honor of Charles P. Kindleberger*, North-Holland, Amsterdam.

Borensztein, E., J. De Gregorio and J-W. Lee (1998). "How Does Foreign Direct Investment Affect Economic Growth?" *Journal of International Economics* 45: 115–135.

Chen, S. and Martin Ravallion (2004). "How Have the World's Poorest Fared since the Early 1980s?," *World Bank Research Observer* 19(2): 141–69.

Chesnokova, T. (2007). "Immiserizing deindustrialization: A dynamic trade model with credit constraints," *Journal of International Economics* 73(2): 407–20.

Freund, C. and B. Bolaky (2008). "Trade, regulations, and income," *Journal of Development Economics*, 87(2): 309–21.

Galiani, S. And G. Porto (2010). "Trends in Tariff Reforms and in the Structure of Wages," *The Review of Economics and Statistics*, 92(3).

Hoekman, B. and B. Javorcik (2004). "Policies to Encourage Firm Adjustment to Globalization," *Oxford Review of Economic Policy*, 20(3): 457–73.

Hoekman, B. and L. A. Winters (2007). "Trade and Employment: Stylized Facts and Research Findings," in Antonio Ocampo, K.S. Jomo and Sarbuland Khan (Eds.), *Policy Matters: Economic and Social Policies to Sustain Equitable Development*, New York, Zed Books, London, Opus, New Delhi.

Hoekman, B., K. Maskus, and K. Saggi (2005). "Transfer of Technology to Developing Countries: Unilateral and Multilateral Policy Options," *World Development*, 33(10): 1587–602.

Keller, W. (1996). "Absorptive Capacity: On the Creation and Acquisition of Technology in Development," *Journal of Development Economics* 49: 199–227.

Khanna, G., D. Newhouse, and P. Paci (2010). "Fewer Jobs or Smaller Paychecks? Labour Market Impacts of the Recent Crisis in Middle-Income Countries," *Economic Premise* No. 11. Washington DC: World Bank.

Maloney, W.F. (2004). Informality Revisited. *World Development* 32 (7): 1159–1178.

Matusz, S.J., and D. Tarr (1999). "Adjusting to Trade Policy Reform," Policy Research Working Paper No. 2142, World Bank, Washington, DC.

Melitz, M. (2003). "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica,* 71(6): 1695–725.

Motohashi, K. (2002). "Use of Plant-Level Micro-Data for the Evaluation of SME Innovation Policy in Japan," OECD Science, Technology and Industry Working Papers 2002/12, OECD, Directorate for Science, Technology and Industry.

Noland, M. and H. Pack (2003), *Industrial Policy in an Era of Globalization: Lessons from Asia.* Washington DC: Institute of International Economics.

Papageorgiou, D., M. Michaely, A. Choksi (eds) (1991). *Liberalizing Foreign Trade.* 7 volumes. Basil Blackwell, Oxford for the World Bank.

Pakes, A. and R. Ericson (1998), "Empirical Implications of Alternative Models of Firm Dynamics", *Journal of Economic Theory* 79, 1–45.

Rama, M. and D. MacIsaac (1999). "Earnings and Welfare after Downsizing: Central Bank Employees in Ecuador," *World Bank Economic Review* 13: 89–116.

Sapir, A. (2000). "Who Is Afraid of Globalization? The Challenge of Domestic Adjustment in Europe and America," CEPR Discussion Paper 2595.

Verdier, T. (2005). "Socially Responsible Trade Integration: A Political Economy Perspective," in F. Bourguignon, B. Pleskovic, and A. Sapir (eds.), *Are We on Track to Achieve the Millennium Development Goals?* Washington DC: World Bank.

Wacziarg, R. and J. Seddon Wallack (2004). "Trade Liberalization and Intersectoral Labour Movements," *Journal of International Economics* 64: 411–439.

Younger, S.(1996). "Labour Market Consequences of Retrenchment for Civil Servants in Ghana,"in David E. Sahn (ed.). *Economic Reform and the Poor in Africa*. Oxford University Press, New York.

PART A
ADJUSTMENT COSTS

2

Modeling, Measuring, and Compensating the Adjustment Costs Associated with Trade Reforms

CARL DAVIDSON AND STEVEN MATUSZ

1. INTRODUCTION

One of the most deeply rooted tenets in neoclassical economics is that movement toward freer international trade increases aggregate economic welfare. While it is generally understood that there are costs associated with the reallocation of resources induced by trade reforms, there has been surprisingly little research focused on modeling, characterizing, measuring, and analyzing these costs in the aggregate.[1] Matusz and Tarr (2000) provide a reasonably comprehensive survey of the state of the literature as it existed in the late 1990s. Even a cursory reading of that paper reveals that there was only a small handful of studies aimed explicitly at furthering our understanding of these costs. In the absence of more concrete research, the authors of that study pieced together bits and pieces of several dozen studies that were tangentially related to adjustment in order to argue that adjustment costs are likely small relative to the overall gains from liberalization.

The work by Matusz and Tarr (ibid.) piqued our interest in the topic of adjustment costs, and much of our research since 2000 has been aimed at this topic.[2] This note is intended as a brief summary of our methodology and findings. Our approach is first to sketch a slightly simplified version of the model that we have found useful in our study of adjustment costs, and then follow this with a discussion of a variety of results that we have been able to tease from this model.

[1] There are several prominent studies that have measured the personal costs of worker dislocation. For example, Jacobsen, et al. (1993); Kletzer (2001); and Hijzen et al. (2010).

[2] We have not been alone in noting this gap in the literature, as evidenced by the other research notes commissioned for this project. Two notable pieces that were published in recent years and that complement our work are Treffler (2004) and Artuç et al. (2008).

2. A BASIC MODEL[3]

2.1 Labor dynamics

We assume that there are two sectors and labor is the only factor of production. Define L as the total (time-invariant) endowment of labor with $L_s(t)$ being the mass of workers associated with sector s at time t, so that $L_1(t)+L_2(t) = L$. We eventually explain how $L_s(t)$ is exogenously determined. For now, however, we take the distribution of workers across sectors as a given.

We assume that jobs in sector 1 are always available instantly to anyone who wishes to work in this sector, and that jobs in this sector are never subject to involuntary separation. In contrast, workers who wish to obtain jobs in sector 2 must first invest time and resources in training, then search for those jobs, with the search process taking further time. Moreover, jobs in sector 2 are subject to involuntary separation. Some workers who lose their jobs may be fortunate and retain their skills, thereby being able to reenter the search process immediately. However, others are less fortunate and must again start at the bottom. Clearly, time is an essential ingredient in this (or any) model of adjustment costs. The arithmetic used in analyzing continuous-time models tends to be neater and cleaner than the arithmetic used for discrete-time analysis. As such, we set our model in continuous time.

We think of transitions between states as following a Poisson process. Figure 2.1 illustrates the dynamics in sector 2 and sets out the notation. The rates of transition out of training, unemployment, and employment are represented by the exogenous parameters τ, e, and b, all of which are positive.[4] Given the Poisson

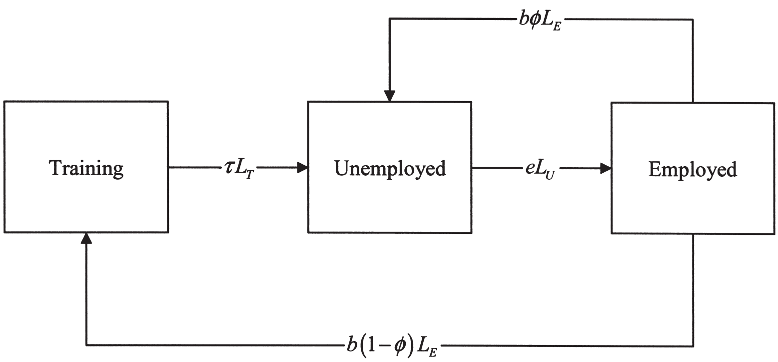

Figure 2.1: *Worker Flows*

[3] This model is a slightly simplified version of the model that we develop in Davidson and Matusz (2000; 2004b; 2004c; and 2006). The main simplification is to strip one sector of all job turnover and assume that the marginal product of labor in that sector is independent of worker ability.

[4] In other work, we have investigated congestion externalities and their implications for gradual adjustment to trade shocks. In order to do so, we endogenize the transition out of unemployment, making it a function of the number of workers searching for jobs. See for example Davidson and Matusz (2004a).

process, the expected duration of training and unemployment, as well as the expected duration of a job are given by $1/\tau$, $1/e$ and $1/b$.

Define ϕ as the share of sector 2 workers who retain their skills after losing their job, and let $L_T(t)$, $L_U(t)$, and $L_E(t)$ be the mass of sector 2 workers who are training, unemployed, and employed at time t.

The labor market dynamics of sector 2 are represented by the following set of differential equations, where a dot over a variable signifies a derivative with respect to time:

$$(1)\quad \dot{L}_E(t) = eL_U(t) - bL_E(t)$$

$$(2)\quad \dot{L}_U(t) = \phi bL_E(t) + \tau L_T(t) - eL_U(t)$$

$$(3)\quad \dot{L}_T = (1-\phi)bL_E(t) - \tau L_T(t)$$

Each of the above equations indicates that the change in the mass of workers in a given state equals the difference between the flow into that state and the flow out of that state. For example, $eL_U(t)$ is the flow from unemployment to employment, while $bL_E(t)$ is the flow out of employment.

Let $L_E(\infty)$ represent the steady state mass of employed workers, with analogous notation defining other steady state variables. For now, take $L_2(\infty)$ as given. We can then solve (1) to (3) to obtain

$$(4)\quad L_E(\infty) = \left(\frac{e\tau}{(1-\phi)eb + (e+b)\tau}\right)L_2(\infty)$$

$$(5)\quad L_U(\infty) = \left(\frac{b\tau}{(1-\phi)eb + (e+b)\tau}\right)L_2(\infty)$$

$$(6)\quad L_T(\infty) = \left(\frac{(1-\phi)eb}{(1-\phi)eb + (e+b)\tau}\right)L_2(\infty)$$

As we show below, the steady state distribution of labor across sectors will depend, in part, upon trade policy. Imagine, for example, that trade liberalization results in workers moving from sector 1 (where they were fully employed) to sector 2 (where they must begin at the bottom, by training for new work). The immediate impact is that total output falls because of the reduction in sector 1 output that is not instantly compensated by greater sector 2 production. Moreover, there may be real resource costs involved in the training of sector 2 workers, as well as in the search process. As time goes on, those workers who moved to sector 2 eventually finish training and move through the search process to obtain employment. Adjustment costs in this context are measured by the lost output and resources spent in training and search along the adjustment path to the

new steady state. Because of the simplicity of (1) to (3), closed-form solutions for the adjustment path are easily derivable.[5]

2.2 General equilibrium

To close out the model, we need to endogenize $L_2(t)$. To do so, we assume that workers are heterogeneous in ability, indexed by $a \in [0,1]$. We assume that ability only matters for the production of good 2. In particular, each worker in sector 1 can produce q_1 units of output, whereas a worker in sector 2 produces q_2a units of output. We assume that output markets are perfectly competitive. Combined with the assumption that labor is the only input, all revenue reverts to the worker so that wages are $w_1=q_1$ and $w_2=pq_2a$ where p is the price of good 2 and good 1 is numeraire.

The basic decision that each worker faces is whether to take a job in sector 1 or start the training process in sector 2. Once this decision is made, the worker is carried along by the dynamics of the model (either remaining employed in sector 1 or moving through the training-search-employment process in sector 2), unless some exogenous change causes the worker to reevaluate his choice of activity.

In order to decide the appropriate course of action, each worker has to calculate the present discounted value of the expected utility that would result from each activity. Let V_1 represent this value for workers employed in sector 1, while V_T, V_U, and V_E represent the same terms for workers training, seeking employment, or employed in sector 2. Letting $v(w_i, p)$ represent indirect utility and ρ represent the discount rate, the present discounted values for each type of worker can be found by solving the following Bellman equations:

$$(7) \quad \rho V_1 = v(w_1, p) + \dot{V}_1$$

$$(8) \quad \rho V_E(a) = v(w_2(a), p) - b\left[V_E(a) - \left(\phi V_U(a) + (1-\phi)V_T(a)\right)\right] + \dot{V}_E(a)$$

$$(9) \quad \rho V_U(a) = 0 + e\left[V_E(a) - V_U(a)\right] + \dot{V}_U(a)$$

$$(10) \quad \rho V_T(a) = -v(c, p) + \tau\left[V_U(a) - V_T(a)\right] + \dot{V}_T(a)$$

In (9), we assume that unemployed workers earn no income. The real resource cost of training is represented by c in (10).

Because transition rates and wages are both time-invariant (except for the possibility of discrete jumps in response to changes in policy), these discounted values are also time-invariant so the time derivatives in (7) to (10) are zero.

Substituting for wages, equations (7) to (10) can be solved for discounted incomes in terms of parameter values.[6] Doing so results in an expression for V_T that

[5] See Davidson and Matusz (2002; 2004c) for explicit closed-form solutions to this system of equations.
[6] For solutions, see Davidson and Matusz (2004c).

is a (linearly) increasing function of ability, whereas V_1 is independent of ability. For a wide range of relative prices, there exists a critical cut-off ability $\bar{a}$ such that $V_T(\bar{a})=V_1$. Workers indexed by $a<\bar{a}$ maximize expected lifetime utility by working in sector 1; those with $a>\bar{a}$ do so by first training in sector 2, then moving up the ladder to employment. Of course, the marginal worker with ability index $\bar{a}$ is just indifferent between sectors. The determination of $\bar{a}$ is illustrated in Figure 2.2.

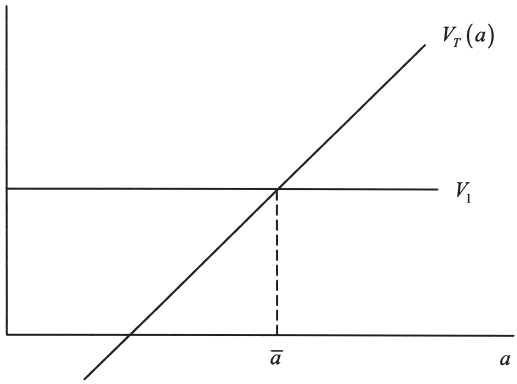

Figure 2.2: *Finding the Equilibrium Allocation of Workers*

Let $F(a)$ be the distribution function for ability. Then we have:

$$(11) \quad L_1(t) = F(\bar{a})L$$

$$(12) \quad L_2(t) = (1 - F(\bar{a}))L$$

2.3 Welfare

We measure social welfare by adding up expected lifetime utilities for all individuals in the economy. Welfare at time t is then:

$$(13) \quad W(t) = L_1(t)V_1 + L_E(t)V_E + L_U(t)V_U + L_T(t)V_T$$

This is equivalent to the present discounted value of utility net of training costs if we were to freeze the allocation of labor as it appears at time t. By replacing the time t allocations of labor with their steady state values, we obtain the steady state value of welfare.

Let $W_{FT}(\infty)$ represent the steady state level of welfare with free trade and $W_{TD}(\infty)$ the steady state level of welfare in an initially trade-distorted equilibrium. If we were able to jump from one steady state to another, the present discounted value of the gain from trade would be $W_{FT} - W_{TD}$. We can refer to this difference as the potential gain from trade reform. However, the adjustment to the new steady state is time consuming. Let W_A represent the present discounted value of utility along the economy's adjustment path moving toward free trade. That is:

$$(14) \quad W_A = \int_t^{\infty} e^{-\rho t} v\big(Y(t) - C(t), p\big)$$

where $Y(t)$ is the aggregate value of gross output at time t and $C(t)$ is aggregate training cost.

Equation (14) takes a different form to (13), because we are allowing the distribution of employment to change along the adjustment path in (14), but not in (13). Note that the relative simplicity of the model permits us to trace out the entire adjustment path and calculate (14).

The actual gain from trade reform is $W_A - W_{TD}(\infty)$. Adjustment costs (AC) represent the gap between the potential gain and the actual gain:

$$(15) \quad AC = W_{FT}(\infty) - W_A.$$

2.4 Numeric implementation

We need a few key parameters to use this model to evaluate numerically the adjustment costs of reform. One key parameter is b, the job breakup rate. The theoretical interpretation of this parameter matches up well with the measures of job destruction originally conceived and calculated by Davis et al. (1996) for the United States. Moreover, since their seminal work, many researchers have followed suit and calculated job destruction rates for many countries, including the United Kingdom (Konings 1995); Poland (Konings et al. 1996); Norway, (Klette and Mathiassen 1996); Canada (Baldwin et al. 1998); Bulgaria, Hungary, and Romania (Bilsen and Konings, 1998); Slovenia (Bojnec and Konings 1999); Russia (Brown and Earle 2002); Estonia (Haltiwanger and Vodopivec 2002); and Ukraine (Konings et al. 2003).

The other key parameter, the job acquisition rate (e) is more difficult to pin down. This is not closely related to the job creation measures of Davis et al. (1996). The basic difference is that their measure is a rate based on firm-level employment, whereas the theory calls for a rate based on sector-specific unemployment. One way around this is to note that the inverse of this measure is the average duration of a spell of unemployment, which is clearly observable at the level of the economy. However, we need this measure to be sector-specific, and that might be problematic.

Other parameters include the time and resource costs of training, as well as the probability that a worker needs to retrain after losing their job. There exists at least some research touching on these issues. The final set of parameters includes preference parameters, the discount rate, and the parameters of the technology (determining the equilibrium wage rates). All of these can be set so that the resulting equilibrium matches up well with what a real economy might look like.

Using parameter estimates from the literature (where available), and choosing other values that seemed sensible when they were not available from the literature, we undertook a thought experiment where a 'low-tech' sector where workers could train very quickly without any resource costs and move straight from training into a job was initially protected by a 5 per cent import tariff. We then

removed the tariff, causing some workers to shift into a 'high-tech sector' where training was both time-consuming used real resources. Moreover, job search subsequent to training was non-trivial. But jobs in the high-tech sector were more durable and paid better wages than those in the low-tech sector.[7] We found the following:[8]

- Adjustment costs accounted for at least one-third of the gross benefits from trade reform under the most optimistic scenario where training costs were low, the protected sector was large, and the discrepancy in job durability between the two sectors was minimized.
- Adjustment costs ate up as much as 80 per cent of the gross benefits from trade when training costs averaged 15 months of the average high-tech worker's wage and when the protected sector was relatively small.
- Ignoring the resource costs of training reduced the adjustment costs considerably, though these costs could still be as high as 25 per cent of the gross gains from trade. On the low end, however, adjustment costs were much smaller, accounting for only 5 per cent of the gross gains from trade.
- As expected, net output dips right after liberalization, as workers shift into a sector that requires significant training and search. Net output did not return to its pre-liberalization level for more than a year, and in some scenarios it took as long as two and a half years before net output reached its pre-liberalization level. These figures were roughly halved when we looked only at gross output, not adjusting for the real resource costs of training.[9]

[7] See Davidson and Matusz (2000; 2002; 2004c). In this exercise, we had job turnover in both sectors, but the breakup rate for jobs was higher in sector 1 (the low-tech sector) than in sector 2 (the high-tech sector). In particular, we assumed that the breakup rate in sector 2 varied from b_2=0.1 (implying that the average duration of a job is 10 years) to b_2=0.167 (so that the average duration of a job in this sector is approximately six years. For sector 1, we assumed that either b_1=1.0 or b_1=0.5 (the average duration of a job varied between one and two years). We maintained the assumption that training was only required in sector 2 (justifying our reference to this as the 'high-tech' sector) and calibrated the model so that τ=3 (the average duration of training is four months) and we varied the real resource cost of training between a minimum of a month of the average sector-2 worker's wage to a maximum of 15 months of the average sector-2 worker's wage. We assumed that ϕ=0.3, though our sensitivity analysis showed that the results were almost completely insensitive to the value of this parameter. We set $e=4$ so that the average duration of unemployment was 13 weeks, consistent with US experience over the long run. We assumed Cobb–Douglas preferences with consumers evenly splitting their income between the two goods. Again, sensitivity analysis showed that the preference parameter had virtually no impact on the magnitude of the results. Finally, we chose productivity and price parameters so that the initial equilibrium was characterized by a certain fraction of the labor force in the high-tech sector, with that fraction ranging from 0.20 to 0.66.

[8] These results are summarized in Table 2.1, which is adapted from Table 1 in Davidson and Matusz (2004c).

[9] Using a different model of adjustment (where workers are always fully employed, but incur non-pecuniary idiosyncratic moving costs in switching sectors), Artuç, et al. (2008) find that approximately eight years are required before adjustment is 95 per cent complete. This metric is somewhat different than the one we used, but a casual glance at figures 4 and 5 in Davidson and Matusz (2000) would suggest that in our framework adjustment would be 95 per cent complete somewhere in the range of five years.

Table 2.1: *Simulated Adjustment Costs*

One third of labor force in the initially-protected sector	
Training Cost	Adjustment cost as a share of gross benefit of trade reform
1 month	0.39
15 months	0.80
Two thirds of labor force in the initially-protected sector	
Training Cost	Adjustment cost as a share of gross benefit of trade reform
1 month	0.34
15 months	0.72
Training Cost	Years before output returns to its pre-liberalization level
1 month	1.3
15 months	1.5

Notes: Training cost is measured as the number of months' worth of wages for the average high-tech worker. The number of years before output returns to its pre-liberalization level is insensitive to the initial share of the workforce in the protected sector. These examples were calculated assuming that the average duration of a job in the low-tech sector is 2 years, while the average duration of a job in the high-tech sector is 10 years.

3. OTHER ISSUES

We have studied various features of this model in our exploration of adjustment costs. We note some of our findings in this section.

Despite the fact that labor is the only input, there are winners and losers from trade liberalization. This follows from the fact that labor is heterogeneous in ability. Workers who are in the export-oriented sector prior to liberalization (the 'incumbents') gain unambiguously. Those who remain in the import-competing sector subsequent to liberalization (the 'stayers') are unambiguously harmed. Those who switch sectors (the 'movers') bear the entire adjustment cost. The lower-ability workers in this segment are harmed by liberalization, while the higher-ability workers benefit. This is consistent with the empirical evidence on the experience of displaced workers.[10]

Any labor market policy (short of lump-sum transfers) aimed at compensating those harmed by trade reform effectively results in the replacement of one distortion with another. Our research in this area shows that the least distorting policies are those that are tied to the *ex post* wage (for example, wage subsidies) if the target group involves the movers, whereas the least distorting policies are independent of the wage (for example, employment subsidies) if the target group centers on those trapped in the import-competing sector. In the first instance, the subsidy encourages excessive movement, but a wage subsidy minimizes this effect because the marginal mover has a relatively low wage compared with the average mover, so the subsidy that fully compensates the group of movers is

[10] Kletzer (2001, Table 3.1) finds that more than one-third of displaced workers find reemployment at wages equal to or higher than their pre-displacement wage.

relatively small for the marginal mover. In contrast, policies softening the blow to the stayers discourage movement. Since the marginal stayer in this case has higher ability than the average stayer, policies that are independent of the wage (ability) will minimize the resulting distortion.

We work out the full analysis of compensation in Davidson and Matusz (2006b). In that paper, we apply numeric techniques similar to those we used in getting a handle on the magnitude of adjustment costs to get some sense of the costs of compensation (measured as the size of the resulting distortion relative to the gain from trade). We find that using the correct policy creates very little deadweight loss, ranging from well under 1 per cent of the gains from trade to approximately 30 per cent of the gains from trade, depending upon parameter assumptions and the particular scenario studied. However, using the wrong policy causes these costs to balloon, in some cases more than doubling for the same set of parameter values and the same scenario.

One might think that even the free-trade equilibrium of our model is distorted since wages differ by sector. Indeed, the marginal worker is indifferent between the low-wage and the high-wage sector. Since the wage reflects productivity, it would seem sensible to induce at least some workers near the margin to move to the high-wage sector. But the *laissez faire* equilibrium in this model is not distorted. Workers make the right choices by maximizing the discounted value of expected lifetime utility. What the above story misses is that it is not possible to move a worker employed in the low-wage sector directly into employment in the high-wage sector. While the value of output may ultimately increase by moving in this direction, there are initial losses that more than offset any future gain. We analyze this issue in detail in Davidson and Matusz (2006a).

One of the key assumptions that leads to efficiency of the *laissez faire* equilibrium is the assumption that the job acquisition rate is exogenous. We pursue a different route in Davidson and Matusz (2006b) where the acquisition rate is subject to congestion externalities. In that framework, escape-clause policies that are explicitly temporary trade limitations in response to unexpected global shocks can increase social welfare by slowing the adjustment process and partially offsetting the externality.

4. STRENGTHS AND WEAKNESSES

We clearly think that our approach to modeling adjustment has much to offer. The analysis is framed within an internally consistent general equilibrium model that is squarely in the tradition of the simple general equilibrium models often used to explore trade and policy issues. The arithmetic for undertaking comparative steady state analysis is clear (while somewhat messy), and it is possible to derive closed form solutions explicitly for the adjustment path. The key parameters are small in number and (with perhaps one key exception to be noted below) empirically observable. The model allows for heterogeneous outcomes among workers and permits calculation of an explicit, well-defined measure of adjustment costs as well as an explicit, well-defined measure of the gains from liberalization. The numeric results are plausible.

Some of the benefits of our approach can be seen by a comparison with Trefler's (2004) excellent empirical study of the short-run costs and long-run benefits of the Canada–United States free trade agreement (CUSTA). As in our methodology, Trefler (ibid.) is able to examine the effects of the creation of trade preferences on a variety of outcomes (including employment and productivity) within the context of a coherent framework. Despite his examination of 213 industries, however, his is not a true general equilibrium analysis. Moreover, he is able to conclude that CUSTA reduced Canadian manufacturing employment by 12 per cent (an adjustment cost), while creating large increases in industry-wide productivity. However, he is only able to suggest that the employment effect was transitory (amounting to an adjustment cost), while the productivity effect was permanent (translating into the benefit of liberalization). In any event, the magnitudes of the employment effect and the productivity effect cannot be meaningfully aggregated to draw any conclusions about the magnitude of the adjustment cost relative to the benefits of trade.

On the flip side, Trefler's (ibid) analysis also has strengths that shine a light on some of the weaknesses of our model. Trefler's work is, after all, an econometric undertaking. As such, he is able to talk about issues such as statistical significance and so on, whereas our model is not able to draw those sorts of conclusions. In addition, by ignoring the general equilibrium effects, Trefler is able to examine many sectors simultaneously. While it is possible to add sectors to our model, the results become muddied. For example, imagine just three sectors such that workers with the lowest ability self-select into sector 1, those with intermediate ability self-select into sector 2, and those with the highest ability go to sector 3. Suppose now that liberalization results in a simple change in the price of good 2. For example, suppose that the price of good 2 falls relative to the prices of goods 1 and 3. This will cause exits from sector 2, with lower-ability workers in this sector moving to sector 1, and higher ability workers in this sector moving to sector 3. This is only one possible scenario, but already it is possible to see how the complications can mushroom.[11] Another weakness of our model is that the numeric results depend upon the magnitude of the job acquisition rate (e in equation 1). As we noted earlier, empirical measures of job destruction are conceptually close to our job breakup rate (b in equation 1), but the same cannot be said for empirical measures of job creation. Finally, our model cannot get at the within-sector changes in productivity identified by Melitz (2003) as resulting from the expansion of high-productivity firms as low-productivity firms are eliminated by trade reform.[12]

While not an inherent weakness of our model, we should note that our frame of reference is an industrialized economy that is generally free from distortions other than those pertaining to trade. Application to developing countries might

[11] This does not even take into account the standard issues regarding the difficulty of determining trade patterns and resource allocation when the number of sectors exceeds two.

[12] We touch on this issue in Davidson et al. (2008), though we only explore comparative steady states in that model.

require suitable modifications to account for an informal sector, substantial state-owned enterprises, and other characteristics of such economies.

Carl Davidson is Professor of Economics and Department Chairperson, Michigan State University.

Steven J. Matusz is Professor of Economics, Michigan State University.

BIBLIOGRAPHY

Artuç, Erhan, Chaudhuri, Shubham and John McLaren (2008). Delay and Dynamics in Labor Market Adjustment: Simulation results. *Journal of International Economics*, 73: 1–13

Baldwin, John, Dunne, Timothy and John Haltiwanger (1998). A Comparison of Job Creation and Job Destruction in Canada and the United States. *Review of Economics and Statistics*, 80: 347–56

Bilsen, Valentijn and Jozef Konings (1998). Job Creation, Job Destruction and Growth of Newly Established, Privatized and State-Owned Enterprises in Transition Economies: Survey Evidence from Bulgaria, Hungary, and Romania. *Journal of Comparative Economics*, 26: 429–45

Bojnec, Štefan and Jozef Konings (1999). Job Creation, Job Destruction an d Labour Demand in Slovenia. *Comparative Economic Studies*, 41: 135–49

Brown, J David and John Earle (2002). Gross Job Flows in Russian Industry Before and After Reforms: Has Destruction Become More Creative? *Journal of Comparative Economics*, 30: 96–133

Davidson, Carl and Steven J Matusz (2000). Globalization and Labour-Market Adjustment: How Fast and At What Cost? *Oxford Review of Economic Policy*, 16: 42–56

—(2002). Globalisatoin, Employment and Income: Analysing the Adjustment Process, in *Trade, Investment, Migration and Labour Market Adjustment*, David Greenaway, Richard Upward, and Katharine Wakelin, (Eds.), Houndmills, Basingstoke, Hampshire and New York, N.Y.: Palgrave, Macmillan.

—(2004a). An Overlapping-generations Model of Escape Clause Protection, *Review of International Economics*, 12: 749–68

—(2004b). International Trade and Labor Markets: Theory, evidence, and policy implications, Kalamazoo, MI: W E Upjohn Institute for Employment Research

—(2004c). Should Policy Makers Be Concerned About Adjustment Costs? in Devashish Mitra and Arvind Panagariya, (Eds.), *The Political Economy of Trade, Aid and Foreign Investment Policies*, Amsterdam, The Netherlands: Elsevier

—(2006a). Long-run Lunacy, Short-run Sanity: A simple model of trade with labor market turnover, *Review of International Economics*, 14: 261–76

—(2006b). Trade Liberalization and Compensation, *International Economic Review*, 47: 723–47

Davidson, Carl, Matusz, Steven J and Andrei Shevchenko (2008). Globalization and Firm Level Adjustment with Imperfect Labor Markets. *Journal of International Economics*, 75: 295–309

Davis, Steven, Haltiwanger, John and Scott Schuh (1996). *Job Creation and Destruction*. Cambridge, MA: MIT Press

Haltiwanger, John and Milan Vodopivec (2002). Gross Worker and Job Flows in a Transition Economy: An analysis of Estonia, *Labour Economics*, 9: 601–30

Hijzen, Alex, Upward, Richard and Peter Wright (2010), The Income Losses of Displaced Workers, *Journal of Human Resources*, 45(1): 243–269

Jacobson, Louis, LaLonde, Robert and Daniel Sullivan (1993), Earnings Losses of Displaced Workers, *American Economic Review*, 83: 685–709

Klette, Tor and Astrid Mathiassen (1996), Job Creation, Job Destruction, and Plant Turnover in Norwegian Manufacturing, *Annales D'Economie et de Statistique*, 41/42: 97–124

Kletzer, Lori (2001). *Job Loss from Imports: Measuring the Costs*. Washington, D.C.: Institute for International Economics

Konings, Jozef (1995), Job Creation and Job Destruction in the UK Manufacturing Sector. *Oxford Bulletin of Economics and Statistics*, 57: 5–24

Konings, Jozef, Lehmann, Hartmut and Mark Schaffer (1996). Job Creation and Job Destruction in a Transition Economy: Ownership, firm size, and gross job flows in Polish manufacturing 1988–1991, *Labour Economics*, 3: 299–317

Konings, Jozef, Kupets, Olga and Hartmut Lehmann (2003). Gross Job Flows in Ukraine, *Economics of Transition*, 11: 321–56

Melitz, Marc (2003). The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity, *Econometrica*, 71: 1695–1725

Trefler, Daniel (2004). The Long and Short of the Canada–US Free-trade Agreement, *American Economic Review*, 94: 870–95

3

A Structural Empirical Approach to Trade Shocks and Labor Adjustment: An Application to Turkey[1]

ERHAN ARTUÇ AND JOHN MCLAREN

INTRODUCTION

In trade liberalization, in trade shocks, in dealing with large-scale public-sector downsizing, a major issue facing policy analysts is how to assess the costs faced by workers in moving from the afflicted sector of the economy to another sector. If it is prohibitively costly to switch industries, for example, a trade liberalization that decimates an import-competing sector may raise the present discounted value of real GDP but cause serious harm to the population of workers who had grown dependent on the sector, and this harm needs to be assessed and taken into account.

In this paper, we summarize a method for estimating these costs based on a dynamic rational-expectations model of labor adjustment, and for using these estimates in policy simulations to try to assess exactly these distributional impacts of policy. The approach has been developed in a number of papers by Cameron et al. (2007), Chaudhuri and McLaren (2007) and Artuç et al. (2008; 2010). The method can be summarized as follows. First, specify a model of the labor market for the whole economy in which each period each worker has the opportunity to switch sectors, but at a cost, which varies for each worker over time according to a distribution whose parameters are to be estimated. The time-varying idiosyncratic costs allow for gradual reallocation of workers to a shock, and they also allow for anticipatory reallocation to an expected future shock, both of which are important features of real-world labor adjustment. Second, derive from this model an equilibrium condition analogous to an Euler equation, which can then be brought to the data to estimate the underlying parameters of the distribution

[1] We would like to thank Guido Porto and Kamil Yilmaz for their comments. We gratefully acknowledge the financial support of the Scientific and Technological Research Council of Turkey (TUBITAK) and Turkish Academy of Sciences (TUBA).

of idiosyncratic moving-cost shocks. Third, estimate those parameters by fitting this equilibrium condition to data on gross flows of workers and wages, across sectors of the economy, and across time. Fourth, use these estimated parameters to simulate policy experiments.

Artuç et al. (2010) apply this method to the Current Population Surveys of the US Census, but they are applicable to a wide array of other countries.[2] We will demonstrate the techniques here on a data set from Turkey. In particular, we use a very limited data set—a worker survey with modest sample sizes and only three years of data. Nonetheless, structural parameters of the labor adjustment process can be easily estimated and a rich variety of questions can then be explored using convenient simulation methods. We thus show that a well-grounded analysis of the dynamic response to trade shocks can be accomplished quite easily, without much computer power and with very modest data.

1. A SUMMARY OF THE MODEL

The model is developed in detail in Cameron et al. (2007) and Chaudhuri and McLaren (2007). Essentially, the basic model is a Ricardo–Viner trade model with the addition of costly inter-industry labor mobility.[3] The essential idea can be summarized as follows. Workers can always change their sector of employment, but must incur costs to do so. At the same time, each individual worker faces time-varying idiosyncratic shocks that either make it either costly for that worker to change sectors, or, at times, costly not to change sectors. As a result, a certain fraction of workers are always changing sectors—the labor market exhibits gross flows. When a trade shock hits a sector adversely, the workers whose idiosyncratic moving costs are currently low leave the sector while those currently with high moving costs wait. This induces gradual adjustment to a trade shock. It also implies that option value is important in workers' utilities, as each worker is aware that no matter what sector they are in at present, there is some probability that they will choose to move to another sector in the future.

1.1 Basic setup

Consider an n-good economy, in which all agents have preferences summarized by the indirect utility function $v(p,I) \equiv I/\phi(p)$, where p is an n-dimensional price vector, I denotes income, and ϕ is a linear-homogeneous consumer price index. Assume that in each industry i there are a large number of competitive employers, and that their aggregate output in any period t is given by $x_t^i = X^i(L_t^i, K^i, s_t)$, where L_t^i denotes the labor used in industry i in period t, K^i is a stock of sector-

[2] Another example is Artuç (2009), which applies a different estimation method but the same simulation method to National Longitudinal Survey of Youth data.

[3] In principle, the model can accommodate geographic as well as inter-industry mobility. Instead of n industries, we could have n industry–region cells, for example; all of the logic below would carry through without amendment. In practice, we have limited the discussion to inter-industry mobility because we have not found enough inter-regional mobility in the data to identify the parameters of interest.

specific capital,[4] and s_t is a state variable that could capture the effects of policy (such as trade protection, which might raise the price of the output), technology shocks, and the like. Assume that X^i is strictly increasing, continuously differentiable and concave in its first two arguments. Its first derivative with respect to labor is then a continuous, decreasing function of labor, holding K^i and s_t constant. Assume that s follows a stationary process on some state space S.[5]

The economy's workers form a continuum of measure $\bar{L}$. All workers are homogeneous, and each of them at any moment is located in one of the n industries. Denote the number of workers in industry i at the beginning of period t by L_t^i. If a worker, say, $l \in [0, \overline{L}]$, is in industry i at the beginning of t, theworker will produce in that industry, collect the market wage for that industry, and then may move to any other industry. In order for the labor market to clear, the real wage w_t^i paid in industry i at date t must satisfy $w_t^i = \left(p_t^i(s_t)/\phi(p_t(s_t))\right)\left(\partial X^i(L_t^i, K^i, s_t)/\partial L_t^i\right)$ at all times, where the $p_t^i(s_t)$ are the domestic prices of the different industries' outputs and may depend on s_t as, for example, in the case in which s_t includes a tariff.

If worker l moves from industry i to industry j, the worker incurs a cost $C^{ij} \geq 0$, which is the same for all workers and all periods, and is publicly known. In addition, if the worker is in industry i at the end of period t, the worker collects an idiosyncratic benefit $\varepsilon_{l,t}^i$ from being in that industry. These benefits are independently and identically distributed across individuals, industries, and dates, with density function $f: \Re \mapsto \Re^+$, $f(\varepsilon) > 0 \forall \varepsilon$, cumulative distribution function $F: \Re \mapsto [0,1]$. Without loss of generality, assume that $\int \varepsilon f(\varepsilon) d\varepsilon \equiv 0$. Thus, the full cost for worker l of moving from i to j can be thought of as $\varepsilon_{l,t}^i - \varepsilon_{l,t}^j + C^{ij}$. The worker knows the values of the $\varepsilon_{l,t}^i$ for all i before making the period-t moving decision.[6] We adopt the convention that $C^{ii}=0$ for all i.

Note that the mean cost of moving from i to j is given by C^{ij}, but its variance and other moments are determined by f. It should be emphasized that these higher moments are important both for estimation and for policy analysis, as will be discussed below.

All agents have rational expectations and a common constant discount factor $\beta<1$, and are risk neutral. An equilibrium then takes the form of a decision rule by which, in each period, each worker will decide whether to stay in their industry or move to another, based on the current allocation vector L_t of labor across industries, the current aggregate state s_t, and that worker's own vector $\varepsilon_{l,t}$ of

[4] Adjustment of capital over time is obviously important, but in this study we set it aside to focus on labor.

[5] We need to allow for shocks to sectoral labor demand to estimate the model, because otherwise the model would predict that all aggregates would converge non-stochastically to a steady state over time. Obviously, the data do not behave in that way, because of ongoing aggregate shocks. However, these exogenous shocks to labor demand are a distraction from our questions of interest and would generate enormous computational difficulties in simulations, so we drop them in our simulation exercises.

[6] It is useful to think of the timeline as follows: The worker observes s_t at the beginning of the period, produces output and receives the wage, then learns the vector $\varepsilon_{l,t}$ and decides whether or not to move. At the end of the period, the worker enjoys $\varepsilon_{l,t}^j$ in whichever sector j the worker has landed.

shocks. In the aggregate, this decision rule will generate a law of motion for the evolution of the labor allocation vector, and hence (by the labor market clearing condition just mentioned) for the wage in each industry. Each worker understands this behaviour for wages, and thus how L_t and the wages will evolve in the future in response to shocks; and given this behaviour for wages, the decision rule must be optimal for each worker, in the sense of maximizing their expected present discounted value of wages plus idiosyncratic benefits, net of moving costs.

To close the model, we need to determine the prices p_t^i. We do this in two ways in two different versions of the model. In the first version, all industries produce tradeable output, whose world prices are determined by world supply and demand and are exogenous to this model; the domestic prices p_t^i are then equal to the world price plus a tariff. In the second version of the model, a subset of the industries produce non-tradeable output, whose prices are determined endogenously. At each moment, the allocation of labor L_t determines the quantity of each industry's output, and hence the supply of each non-tradeable good; this, combined with the prices of the tradeable goods, allows us to compute the price of each non-tradeable good that equates domestic demand with that supply. Note that we do not need to concern ourselves with any of these price-determination issues for the estimation of the model, but we will need them later for the general equilibrium simulation of the model.

1.2 The key equilibrium condition.

Suppose that we have somehow computed the maximized value to each worker of being in industry i when the labor allocation is L and the state is s. Let $U^i(L,s,\varepsilon)$ denote this value, which, of course, depends on the worker's realized idiosyncratic shocks. Denote by $V^i(L,s)$ the average of $U^i(L,s,\varepsilon)$ across all workers, or in other words, the expectation of $U^i(L,s,\varepsilon)$ with respect to the vector ε. Thus, $V^i(L,s)$ can also be interpreted as the expected value of being in industry i, conditional on L and s, but before the worker learns their value of ε.

Assuming optimizing behavior, that is, that a worker in industry i will choose to remain at or move to the industry j that offers them the greatest expected benefits, net of moving costs, we can write:[7]

$$U^i(L_t,s_t,\varepsilon_t) = w_t^i + \max_j \{\varepsilon_t^j - C^{ij} + \beta E_t[V^j(L_{t+1},s_{t+1})]\} \tag{1}$$

$$= w_t^i + \beta E_t[V^i(L_{t+1},s_{t+1})] + \max_j \{\varepsilon_t^j + \overline{\varepsilon}_t^{ij}\}$$

where:

$$\overline{\varepsilon}_t^{ij} \equiv \beta E_t[V^j(L_{t+1},s_{t+1}) - V^i(L_{t+1},s_{t+1})] - C^{ij}. \tag{2}$$

[7] From here on, we drop the worker-specific subscript l.

Note that L_{t+1} is the next-period allocation of labor, derived from L_t and the decision rule, and s_{t+1} is the next-period value of the state, which is a random variable whose distribution is determined by s_t. The expectations in (1) and (2) are taken with respect to s_{t+1}, conditional on all information available at time t.

Taking the expectation of (1) with respect to the ε vector then yields:

$$V^i(L_t, s_t) = w_t^i + \beta E_t[V^i(L_{t+1}, s_{t+1})] + \Omega(\bar{\varepsilon}_t^i), \tag{3}$$

where $\bar{\varepsilon}_t^i = (\bar{\varepsilon}_t^{i1}, \ldots, \bar{\varepsilon}_t^{iN})$ and:

$$\Omega(\bar{\varepsilon}_t^i) = \sum_{j=1}^{N} \int_{-\infty}^{\infty} (\varepsilon^j + \bar{\varepsilon}_t^{ij}) f(\varepsilon^j) \prod_{k \neq j} F(\varepsilon^j + \bar{\varepsilon}_t^{ij} - \bar{\varepsilon}_t^{ik}) d\varepsilon^j. \tag{4}$$

The average value to being in industry i can therefore be decomposed into three terms: (1) the wage, w_t^i, that an industry-i worker receives; (2) the base value of staying on in industry i, that is, $\beta E_t[V^i(L_{t+1}, s_{t+1})]$; and (3) the additional value, $\Omega(\bar{\varepsilon}_t^i)$, derived from having the option to move to another industry should prospects there look better (and which is simply equal to the expectation of $\max_j\{\varepsilon^j + \bar{\varepsilon}_t^{ij}\}$ with respect to the ε vector). We will call this the option value associated with being in that industry at that time. Note that, since $\bar{\varepsilon}_t^{ii} \equiv 0$, this is always positive.

Using (3), we can rewrite (2) as:

$$C^{ij} + \bar{\varepsilon}_t^{ij} = \beta E_t[V^j(L_{t+1}, s_{t+1}) - V^i(L_{t+1}, s_{t+1})]$$

$$= \beta E_t[w_{t+1}^j - w_{t+1}^i + \beta E_{t+1}[V^j(L_{t+2}, s_{t+2}) - V^i(L_{t+2}, s_{t+2})]$$

$$+ \Omega(\bar{\varepsilon}_{t+1}^j) - \Omega(\bar{\varepsilon}_{t+1}^i)], \text{or}$$

$$C^{ij} + \bar{\varepsilon}_t^{ij} = \beta E_t[w_{t+1}^j - w_{t+1}^i + C^{ij} + \bar{\varepsilon}_{t+1}^{ij} + \Omega(\bar{\varepsilon}_{t+1}^j) - \Omega(\bar{\varepsilon}_{t+1}^i)]. \tag{5}$$

Note that $\bar{\varepsilon}_t^{ij}$ is the value of $\varepsilon^i - \varepsilon^j$ at which a worker in industry i is indifferent between moving to industry j and staying in i. Condition (5) thus has the simple, common sense interpretation that for the *marginal* mover from i to j, the cost (including the idiosyncratic component) of moving is equal to the expected future benefit of being in j instead of i at time t+1. This expected future benefit has three components. The first is the wage differential. The second is the revealed expected value to being in industry j instead of i at time t+2, as revealed by the cost borne by the marginal mover from i to j at time t+1, or $C^{ij} + \bar{\varepsilon}_{t+1}^{ij}$. The last component is the difference in option values associated with being in each industry. Thus, if I contemplate being in j instead of i next period, I take into account the expected difference in wages; then the difference in the expected values of continuing in each industry afterward; and finally, the differences in the values of the option to leave each industry if conditions call for it.

Put differently, condition (5) is an Euler equation. Given appropriate choice of functional forms, this can be implemented to estimate the moving-cost parameters. We turn to that task next.

1.3 The estimating equation

Let m_t^{ij} be the fraction of the labor force in industry i at time t that chooses to move to industry j, i.e., the *gross flow* from i to j. With the assumption of a continuum of workers and i.i.d idiosyncratic components to moving costs, this gross flow is simply the probability that industry j is the best for a randomly selected i-worker. Now, make the following functional form assumption. Assume that the idiosyncratic shocks follow an extreme-value distribution with parameters $(-\gamma\nu,\nu)$:

$$f(\varepsilon) = \frac{e^{-\varepsilon/\nu-\gamma}}{\nu}\exp\left\{-e^{-\varepsilon/\nu-\gamma}\right\}$$

$$F(\varepsilon) = \exp\left\{-e^{-\varepsilon/\nu-\gamma}\right\},$$

implying:

$$E(\varepsilon) = 0, \text{and}$$

$$Var(\varepsilon) = \frac{\pi^2\nu^2}{6}.$$

Note that while we make the natural assumption that the ε's be mean-zero, we do not impose any restrictions on the variance. The variance is proportional to the square of ν, which is a free parameter to be estimated, and crucial for all of the policy and welfare analysis.

By assuming that the ε_t^i are generated from an extreme-value distribution we are able to obtain a particularly simple expression for the conditional moment restriction, which we then plan to estimate using aggregate data. Specifically, it is shown in the (web-only) Appendix to Artuç et al. (2010) and in the Appendix to the 2007 working paper that, with this assumption:

$$\bar{\varepsilon}_t^{ij} \equiv \beta E_t[V_{t+1}^j - V_{t+1}^i] - C^{ij} = \nu[\ln m_t^{ij} - \ln m_t^{ii}] \tag{6}$$

and:

$$\Omega(\bar{\varepsilon}_t^i) = -\nu \ln m_t^{ii} \tag{7}$$

Both these expressions make intuitive sense. The first says that the greater the expected net (of moving costs) benefits of moving to j, the larger should be the observed ratio of movers (from i to j) to stayers. Moreover, holding constant the (average) expected net benefits of moving, the higher the variance of the idiosyncratic cost shocks, the lower the compensating migratory flows.

The second expression says that the greater the probability of remaining in industry i, the lower the value of having the option to move from industry i.[8]

[8] Note that $0 < m_t^{ii} < 1$, so $\Omega(\bar{\varepsilon}_t^i) = -\nu \ln m_t^{ii} > 0$

Moreover, as the variance of the idiosyncratic component of moving costs increases, so too does the value of having the option to move. This also makes good sense.

Substituting from (6) and (7) into (5) and rearranging, we get the following conditional moment condition:

$$E_t\left[\frac{\beta}{\nu}(w_{t+1}^j - w_{t+1}^i) + \beta(\ln m_{t+1}^{ij} - \ln m_{t+1}^{jj}) - \frac{(1-\beta)}{\nu}C^{ij} - (\ln m_t^{ij} - \ln m_t^{ii})\right] = 0. \quad (8)$$

This condition can be interpreted as a linear regression:

$$(\ln m_t^{ij} - \ln m_t^{ii}) = -\frac{(1-\beta)}{\nu}C^{ij} + \frac{\beta}{\nu}(w_{t+1}^j - w_{t+1}^i) + \beta(\ln m_{t+1}^{ij} - \ln m_{t+1}^{jj}) + \mu_{t+1}, \quad (9)$$

where μ_{t+1} is news revealed at time $t+1$, so that $E_t\mu_{t+1} \equiv 0$. In other words, the parameters of interest, C^{ij}, β and ν, can then be estimated by regressing current flows (as measured by $(1\mathrm{n}m_t^{ij} - 1\mathrm{n}m_t^{ii})$) on future flows (as measured by $(1\mathrm{n}m_{t+1}^{ij} - 1\mathrm{n}m_{t+1}^{jj})$) and the future wage differential with an intercept.

The basic idea of the estimating equation (9) can be summarized as follows. We regress current flows of workers from i to j on next-period flows in the same direction and on next-period j-sector wages minus i-sector wages. If there are a lot of flows in all directions, that implies a high value for the intercept of this equation, which in turn implies a high variance for the idiosyncratic shocks ν relative to average moving costs C^{ij}. On the other hand, for a given overall level of flows, if those flows are very responsive to the expected next-period wage differential, that implies a large slope coefficient in the regression equation, which implies a low variance ν of the idiosyncratic shocks. That is how this simple regression can identify the mean and variance parameters of moving costs. In practice, for this exercise, we will constrain all average moving costs to be the same, or $C^{ij} = C \forall i, j$.

2. DATA

Our estimation strategy hinges on observing aggregate gross flows across industries. We construct gross flow measures from retrospective questions in the Hane Halki Isgücü Anketi (HHIA), or Household Employment Survey, of the Turkish Statistical Institute (TUIK), 2004–6. The survey asks, among other questions, what industry the worker is in at present and what industry the worker was in last year. This enables us to construct rates of flow, m_{t-1}^{ij}, for each date t. We also obtain industry wages w_t^i as the average wage reported in the HHIA samples for industry i at date t. These are deflated by the CPI, and normalized so that over the whole sample the average annualized wage is equal to unity. We restrict the sample to males aged 21 to 64 currently working full time or seasonal for wages (who do not own a business). Although agriculture is 30 per cent of the economy, the number of workers who list themselves as full-time or seasonal agricultural workers is very small (3 per cent of the sample). We include all full-time and seasonal agricultural workers but exclude those who own a farm, on the

assumption that they are better treated as owners of a fixed, industry-specific factor than as a mobile worker. This selection process yields us 47,064 observations for 2004, 47,723 for 2005, and 49,394 for 2006.

If we have n industries, then there are n^2 rates of gross flow to keep track of each period (or $n(n-1)$ if one excludes the fraction of workers in each industry who do not move). Thus, the number of directions for gross flows proliferates rapidly as the number of industries increases, leading in finite samples to zero observations and observations with very small numbers of individuals. As a result, we need to aggregate industries, and we aggregate to the following four:

1. Agriculture, forestry, fisheries, hunting.
2. Manufacturing, mining, utilities, construction.
3. Trade, hotels and restaurants.
4. All other services including: education, public (administration and social security, military, and so on.), health, finance, real estate, transportation, and communication.

As a result of our aggregation, the sample size for each regression is 24, since we have 3 years minus 1 to allow for lags, and 4 times 3 directions of flows.

Table 3.1: *Descriptive Statistics: Gross Flows, 2004–6*

	Agric/Min	Manuf/Const	Trade/Hotels	Service
Agric/Min	0.8616	0.0669	0.0338	0.0376
Manuf/Const	0.0015	0.9770	0.0120	0.0095
Trade/Hotels	0.0011	0.0313	0.9482	0.0194
Service	0.0007	0.0085	0.0076	0.9832

Notes: Origin sector is listed by row, destination sector by column. Each cell of table contains mean flow rate, averaged over the three years.

Descriptive statistics for the resulting data are shown in Tables 3.1 and 3.2. Table 3.1 summarizes gross flows. Each cell of the table shows the fraction of workers, averaged across years, in the row sector who moved to the column sector in any given period; for example, on average, 3.38 per cent of Agriculture/Mining workers in any year moved to Trade/Hotels. The rates of gross flow range from below 0.1 per cent for the move from Service to Agriculture/Mining to 6.7 per cent for the move from Agriculture/Mining to Manufacturing/Construction. A tendency for workers to exit Agriculture/Mining in favor of Manufacturing/ Construction and Services is evident in the figures.

Table 3.2 shows descriptive statistics for wages. Wages vary a great deal across sectors. Normalized wages (that is, normalized to have a unit mean) averaged across time range from 0.5642 for Agriculture/Mining to 1.1882 for Services, suggesting that the shifts of workers observed in Table 3.1 are driven by a rise in the demand for labor in Services and Manufacturing/Construction relative to Agriculture/Mining.

Table 3.2: *Descriptive Statistics: Wages, 2004–6 (normalized)*

Agric/Min	0.5648
Manuf/Const	0.9253
Trade/Hotels	0.8046
Service	1.1882

3. RESULTS

Before showing estimations, we should point out that we do not attempt to estimate β. This model is not designed to estimate rates of time preference, and although it could be done in principle, in practice it turns out that that one parameter is very poorly identified. It turns out that estimating and simulating the model with different values of β produces nearly identical time paths for key observable variables, so it is not surprising that it is hard to identify econometrically. We simply impose a value of β in all that follows; to check for sensitivity to the choice of β, we report estimations with both $\beta = 0.9$ and $\beta = 0.97$.

Table 3.3 shows the results from the basic regression. As mentioned above, we impose $C^{ij} \equiv C \forall i \neq j$, so that the mean moving cost for any transition from one industry to any other is the same. Throughout the table, the t-statistics are reported in parentheses.

Table 3.3: *Regression Results for the Basic Model*

$\beta = 0.97$		$\beta = 0.9$	
ν	C	ν	C
2.56	22.89	1.62	9.50
(3.50***)	(3.23***)	(5.36***)	(5.36***)

Notes: T-statistics are in parentheses; one-tailed significance: 1-percent***; 5-percent**; 10-percent*.

The first two columns report results for $\beta = 0.97$, and the last two report results for $\beta = 0.9$. Notice that the estimated moving costs are enormous. Given our convention on normalizing wages, the value of C is estimated at 9.50 and 22.88 times average annual income, for the two cases respectively. The two values for ν translate to a standard deviation for the idiosyncratic shocks of just over one-year's average annual earnings. This reflects both the modest levels of gross flows observed in Table 3.1 and the fact that the flows respond only modestly to wage differentials across sectors. This suggests that the labor market will likely be sluggish in reallocating workers following a trade shock, as will be confirmed in the simulations.

Strictly speaking, OLS is likely to be biased in this case, because the disturbance term, μ_{t+1}, contains any new information at date $t+1$ and so will in general be correlated with date $t+1$ regressors. This implies a need for instrumental variables. The theory implies that past values of the flows and wages will be valid instruments, and the optimal weighting scheme can be derived as in the Generalized Method of Moments (GMM) (Hansen 1982). This strategy is employed

in Artuç et al. (2010). However, in that study, the data set was 26 years long; in the present case, with only three years, that strategy is not available to us. We are hoping that after the Turkish Statistical Institute releases two more years of household survey data, we will be able to use instruments, which is left over here for future research. As an imperfect fix, we note that in the earlier paper with US data, estimating with OLS instead of with instruments increased the value of C by 67 per cent and ν by 54 per cent when $\beta = 0.97$, and C by 34 per cent and ν by 31 per cent when $\beta = 0.9$, so we divide our parameter estimates by 1.67, 1.54, 1.34 and 1.31 for an alternative set of simulations. After simulating the model with original estimates (for both $\beta = 0.9$ and $\beta = 0.97$), we repeat the simulation exercises with the corrected parameters, and show that the bias in the OLS estimates probably does not affect the adjustment path of workers and values significantly.

Table 3.4: *Regression Results for Sector-specific Entry Costs*

	Beta=0.97		Beta=0.90	
	Estimate	t-statistic	Estimate	t-statistic
ν	1.60	(2.98***)	1.17	(3.95***)
C^1 (Agriculture)	0.00	(0.00)	3.10	(1.53*)
C^2 (Manufacture)	17.67	(2.41**)	7.17	(3.54***)
C^3 (Trade)	9.90	(1.31)	5.55	(2.61***)
C^4 (Service)	20.81	(5.52***)	8.65	(6.03***)

T-statistics are in parentheses. One-tailed significance: 1-percent***, 5-percent**, 10-percent*.

Another possible source of bias comes from the fact that we have imposed uniform moving costs for all sectors, so that $C^{ij} = C \forall i, j$. Degrees-of-freedom concerns prevent us from estimating the full set of C^{ij} parameters without restriction, but we have also estimated the model with a slightly richer specification allowing for sector-specific entry costs. In this approach, $C^{ij}=C^{j}$ for $i=1,\ldots,4$. Table 3.4 shows the results of this regression.

Compared with the Basic Model regression from Table 3.3, we find that all sectors exhibit lower entry costs. However, we were not able to identify entry cost of Agriculture when $\beta = 0.97$, therefore we are not using results from Table 3.4 in simulations. This identification problem probably arises because of our very small sample size. We find that entry to the Service sector (which includes government and professional sectors) is the most costly one, while entry to Agriculture (which includes only workers who do not own a farm) is the least costly.

4. SIMULATION: A SUDDEN TRADE LIBERALIZATION

Now, we use the estimates to study the effect of a hypothetical trade shock through simulations. Note that for the estimations, the only functional form assumption we needed was the density for the idiosyncratic shocks, but to simulate the model we need to choose functional forms (and parameter values) for production and utility functions as well. We assume that each of the four

sectors has a Cobb–Douglas production function, with labor and unmodelled sector-specific capital as inputs. Thus, for our purposes, the production function for sector is given by:

$$y_t^i = A^i (L_t^i)^{\alpha^i} (K^i)^{1-\alpha^i}, \tag{10}$$

where y_t^i is the output for sector i in period t, K^i is sector-i's capital stock, and $A^i>0$ and $\alpha^i>0$ are parameters. Given the number of free parameters and our treatment of capital as fixed,[9] we can without loss of generality set $K^i = 1 \forall i$. This implies that the wages are given by:

$$w_t^i = p_t^i A^i \alpha^i (L_t^i)^{\alpha^i - 1}, \tag{11}$$

where p_t^i is the domestic price of the output of sector i.

For simulations, we need to choose values of production function parameters to provide a plausible illustrative numerical example, broadly consistent with quantitative features of the data. To do this, we set the values A^i and α^i to minimize a loss function given our assumptions on prices (see below). Specifically, for any set of parameter values, we can compute the predicted wage for each sector and that sector's predicted share of GDP using (11) and (10) together with empirical employment levels for each sector and our assumptions about prices as described below. The loss function is then the sum across sectors and across years of the square of each sector's predicted wage minus mean wage in the data, plus the square of labor's predicted share of revenue minus the actual share, plus the square of the sector's predicted minus its actual share of GDP. The sectoral GDP and labor's share of revenue figures are from the Turkish Statistical Institute (www.turkstat.gov.tr) input–output table for 2002, but the remaining figures are from our sample. In addition, we assume that all workers have identical Cobb–Douglas preferences, using consumption shares from 2002 input–output table of the Turkish Statistical Institute for the consumption weights. The parameter values that result from this procedure are summarized in Table 3.5.

Table 3.5: *Parameters for Simulation*

	A^i	α^i	Domestic price	World price
Agric/Min	0.2131	0.1330	1	1
Manuf/Const	1.3497	0.3708	1	0.7
Trade/Hotels	0.9379	0.2237	1	1*
Service	1.9494	0.3407	1	1

Then, to provide a simple trade shock, we assume the following: (i) Units are chosen so that the domestic price of each good at date t=−1 is unity. (Given our

[9] We assume that capital is fixed in order to focus on the workers' problem and to keep the model manageable. Of course, that is an important limitation, since capital should also be expected to adjust to trade liberalization, and that should also be expected to affect wages.

available free parameters, this is without loss of generality.) (ii) There are no tariffs on any sector aside from manufacturing, at any date. (iii) The world price of manufacturing output is 0.7 at each date. The world price of all other tradeable goods is equal to unity at each date. (iv) There is initially a specific tariff on manufactures at the level 0.3 per unit, so that the domestic price of manufactures is equal to unity. (v) Initially, this tariff is expected to be permanent, and the economy is in the steady state with that expectation. (vi) At date $t=-1$, however, after that period's moving decisions have been made, the government announces that the tariff will be removed beginning period $t=0$ (so that the domestic price of manufactures will fall from unity to 0.7 at that date), and that this liberalization will be permanent.

Thus, we simulate a sudden liberalization of the manufacturing sector. We compute the perfect-foresight path of adjustment following the liberalization announcement, until the economy has effectively reached the new steady state. This requires that each worker, taking the time path of wages in all sectors as given, optimally decides at each date whether or not to switch sectors, taking into account that worker's own idiosyncratic shocks. This induces a time path for the allocation of workers, and therefore the time path of wages, since the wage in each sector at each date is determined by market clearing from (11), given the number of workers currently in the sector. Of course, the time path of wages so generated must be the same as the time path each worker expects. It is shown in Cameron et al. (2007) that the equilibrium exists and is unique. The computation method is described at length in Artuç et al. (2008), and programs for executing the simulations are contained in the web-only appendix for Artuç et al. (2010). Simulations converge quickly and require modest computing power.

As a simplification, all goods are assumed to be traded, so all output prices are exogenous, and we use the case with $\beta = 0.97$.

The simulation output is plotted in the figures. Figure A3.1 shows the time path of the allocation of workers. The Manufacturing/Construction sector has 32.8 per cent of the workers in the tariff steady state, but under free trade has only 28.5 per cent. All of the other sectors gain workers due to the liberalization. The adjustment is fairly slow, taking approximately a decade. Figure A3.2 shows the time path of real wages in each sector. With the abrupt drop in the price of Manufacturers/Construction output, real wages in all of the other sectors jump up at the liberalization date, but then gradually decline as workers stream into those sectors from Manufacturing/Construction and push the equilibrium down the labor demand curve. Still, the real wage in each of those sectors is above the tariff steady-state level in the short run and the long run of the liberalization simulation. The story for Manufacturing/Construction wages is the reverse: An abrupt drop of about 20 per cent on the date of the liberalization followed by a gradual improvement, as workers leave the sector and push the equilibrium up the labor demand curve in that sector. Still, the real wage in Manufacturing/Construction is at each date below what it was in the tariff steady state.

Figure A3.3 shows the effect of the liberalization on lifetime utility of the various workers. Each plot shows the expected present-discounted value of utility

for a typical worker in each of the sectors at each date.[10] The main thing to notice is what happens at the sudden liberalization date. For all workers except those in Manufacturing/Construction the effect is positive: Lifetime utility jumps up, as should be expected, since those workers expect a rise in their real wages. Importantly, these lifetime utilities take into account the probability for each worker that they will move to another sector down the road due to idiosyncratic shocks. The only way a services sector worker, for example, would be hurt by the liberalization is if they moved into Manufacturing/Construction, which is a positive probability event but not very likely.

On the other hand, the lifetime utility of a Manufacturing/Construction worker falls at the liberalization date by about 14 per cent. The drop in utility is much smaller than the drop in wages for two reasons: The later increases in real wages in the sector, as noted above (see Figure A3.2), and the fact that workers in Manufacturing/Construction have the option to move to other sectors; this option value is increased by the liberalization, because it increases the real wage in the other sectors.

We repeat the exercise under the assumption that $\beta = 0.9$ and keep all other assumptions the same. Figures A3.5 to A3.8 show results of the case where $\beta = 0.9$. We find that unlike Artuç et al. (2010) (which uses US data), changing the discount factor does not affect how much Manufacturing workers are made worse off. Figure A3.7 shows that, similar to the previous case, manufacturing workers are worse off by about 13 per cent. This may be due to the fact that Turkish workers are less mobile compared to US workers, as reflected in the parameter estimates of Tables 3.3 and 3.4.

In order to get a better understanding of the effects of OLS bias on the adjustment path of labor allocation, wages and values of workers, we repeat the simulations using adjusted estimates as we described in the previous section. The comparison of simulations with adjusted parameters and original OLS parameters are shown in Figures A3.9 to A3.14. The OLS bias seems to change the simulation results only slightly and qualitative results are robust. For example, for the $\beta = 0.97$ case, lifetime utility of Manufacturing workers falls by 10 per cent rather than 14 per cent after the trade shock if we use corrected parameters (shown in Figure A3.11), while for the $\beta = 0.97$ case it falls by 12 per cent rather than 13 per cent with the corrected parameters (shown in Figure A3.14).

Therefore, this analysis suggests that since Turkish labor market data are consistent with very high intersectoral moving costs, workers in the liberalizing sector are likely to suffer a significant welfare loss that will be only partially alleviated over time. Workers in other sectors all enjoy a significant welfare benefit, as does the economy as a whole. Thus, the estimates suggest a significant political conflict over trade liberalization, with a strong motive for the government to find cost-effective ways to compensate the workers in the import-

[10] Note that this is different from the present discounted value of each sector's real wage, because it must take into account each worker's option value. The value is computed using equations (3) and (7) over the simulation; once again, see Artuç et al. (2008) for details.

competing sector. Note that these simulations find much more scope for distributional conflict than the US simulations in Artuç et al. (2010) because of the high estimated moving costs for Turkish workers. Note also that the nature of conflict is sharply different than the type highlighted by Heckscher–Ohlin-type models, in which the conflict would be between blue-collar and white-collar workers. In this paper, we find strong evidence of the potential for inter-sectoral conflict over trade policy.

In addition, we have identified a pitfall of reduced-form wage analysis: A regression of sectoral wages on sectoral tariffs using data at date $t=-1$ and $t=1$ would significantly overestimate the long-run effect of the liberalization on the Manufacturing/Construction wage (and standard approaches would not even allow for an effect on the real wages in the other sectors). In addition, the wefare loss for those workers is considerably more modest than the reduction in wages, once the dynamics of the equilibrium adjustment and the effects of option value are taken into account.

Finally, it should be emphasized once again that this analysis has imposed very modest data requirements; the estimation requires only a simple regression; and the simulation requires minimal computing power. As a result, this type of exercise is very portable across data sets and countries.

Erhan Artuc is Assistant Professor of Economics at Koç University, Istanbul

John McLaren is Professor of Economics at the University of Virginia.

BIBLIOGRAPHY

Artuç, Erhan 2009. *Intergenerational Effects of Trade Liberalization* http://www.econart.com

Artuç, Erhan, Chaudhuri, Shubham and John McLaren 2008. Delay and Dynamics in Labor Market Adjustment: Simulation Results. *Journal of International Economics,* 75(1): 1–13

Artuç, Erhan, Chaudhuri, Shubham and John McLaren 2010.Trade Shocks and Labor Adjustment: A structural empirical approach. *American Economic Review*, 100(3).

Cameron, Stephen, Chaudhuri, Shubham and John McLaren 2007. Trade Shocks and Labor Adjustment: Theory. NBER Working Paper 13463

Chaudhuri, Shubham and John McLaren 2007. Some Simple Analytics of Trade and Labor Mobility. NBER Working Paper 13464

Hansen, Lars P 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica,* 50: 1029–1054

APPENDIX

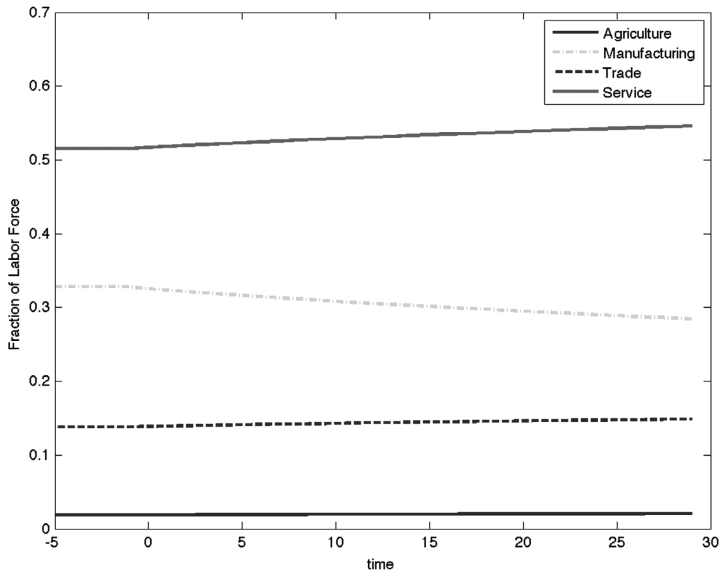

Figure A3.1: *Simulated Trade Liberalization I – Labor Allocation*

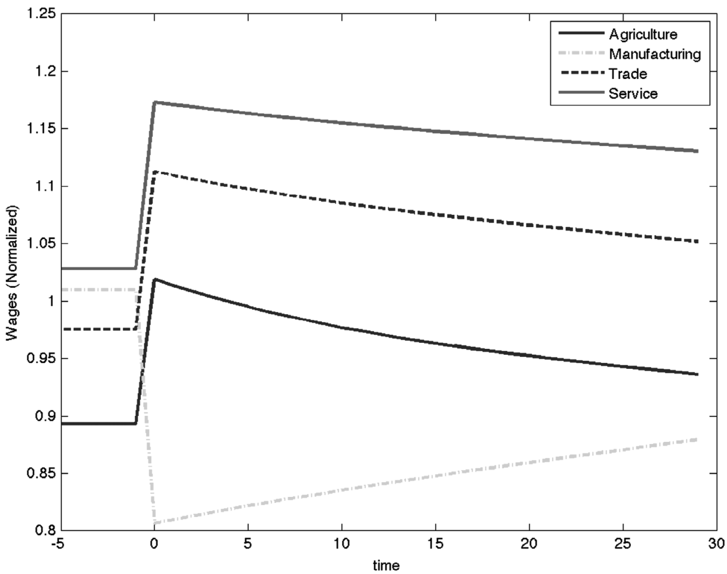

Figure A3.2: *Simulated Trade Liberalization I – Wages*

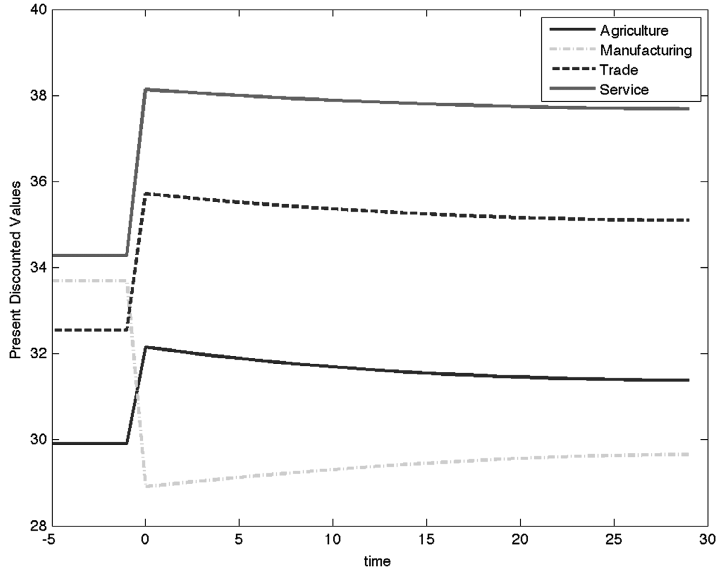

Figure A3.3: *Simulated Trade Liberalization I – Values*

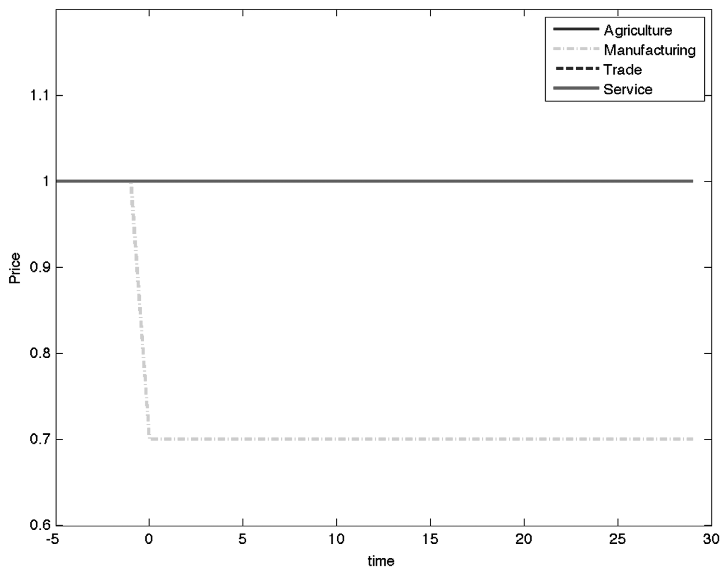

Figure A3.4: *Simulated Trade Liberalization I – Prices*

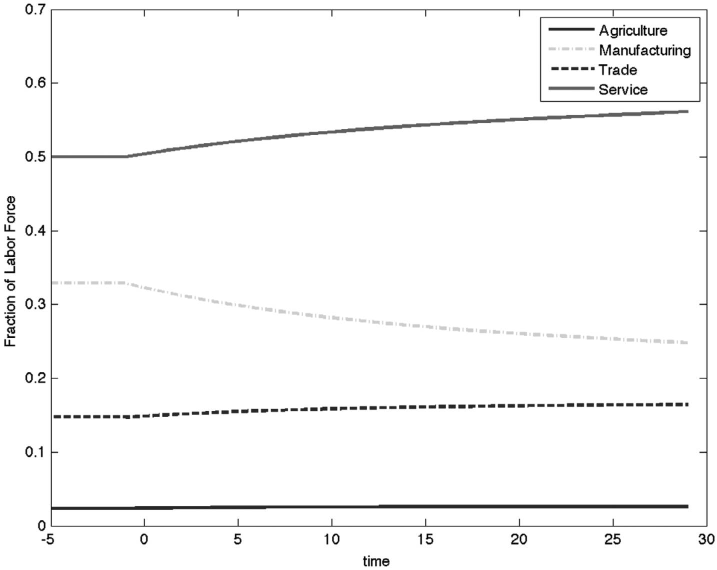

Figure A3.5: *Simulated Trade Liberalization II – Labor Allocation*

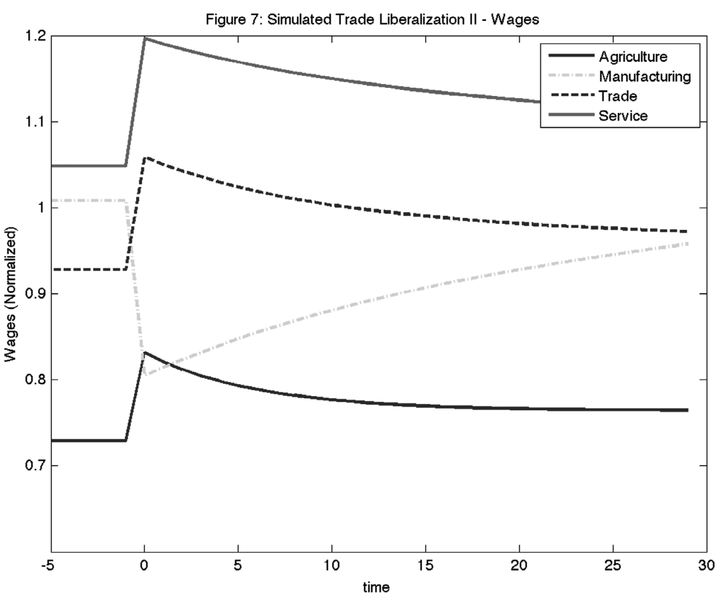

Figure A3.6: *Simulated Trade Liberalization II – Wages*

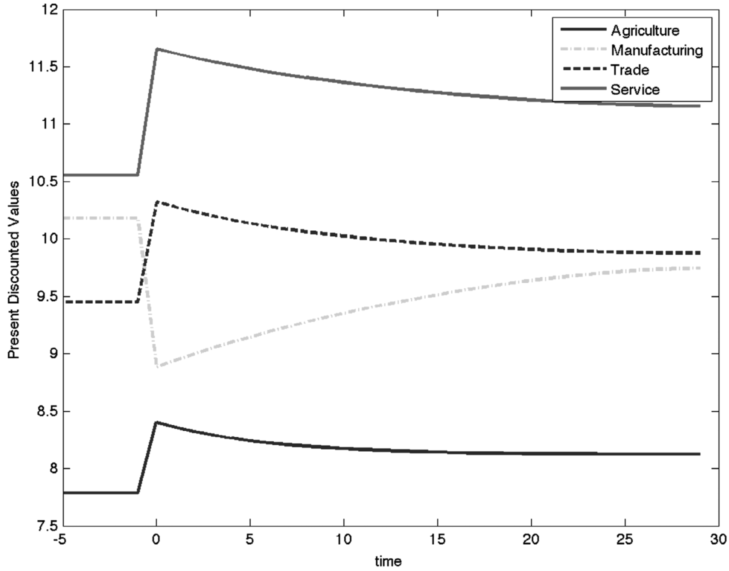

Figure A3.7: *Simulated Trade Liberalization II – Values*

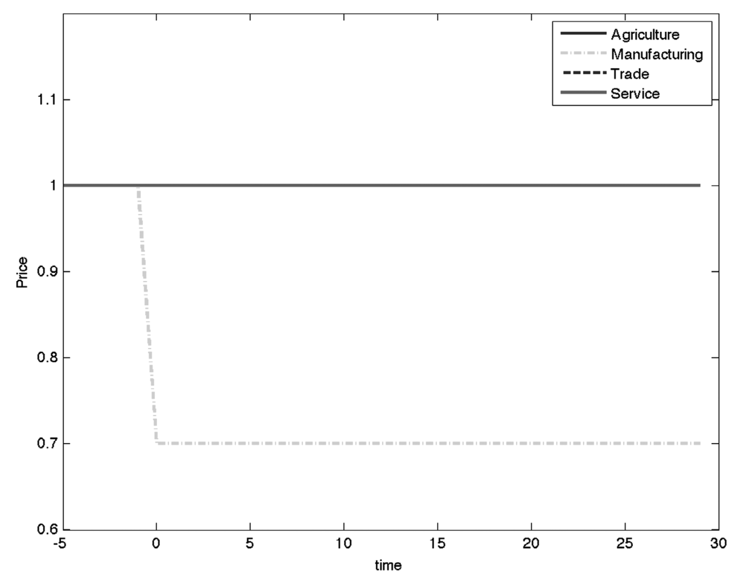

Figure A3.8: *Simulated Trade Liberalization II – Prices*

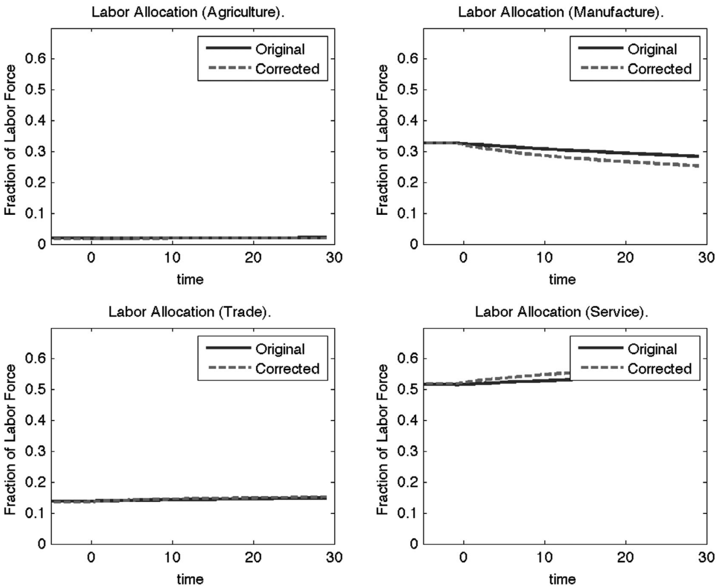

Figure A3.9: *Labor Allocation (β =0.97)*

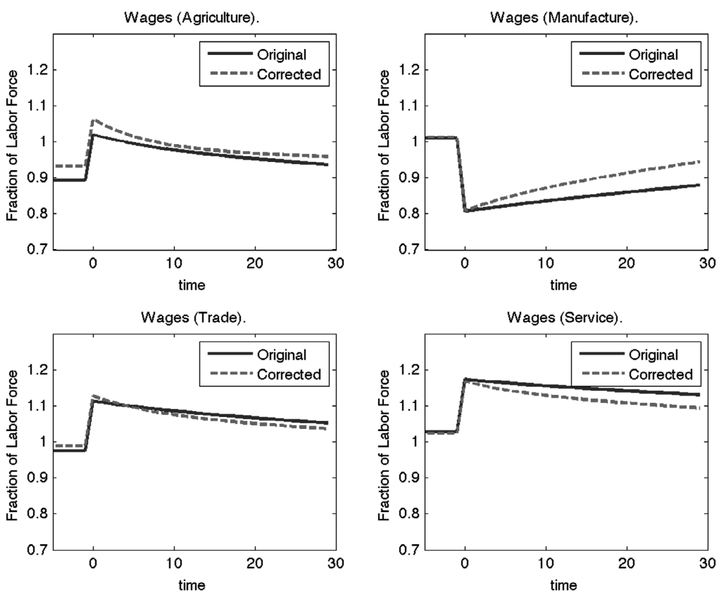

Figure A3.10: *Wages (β =0.97)*

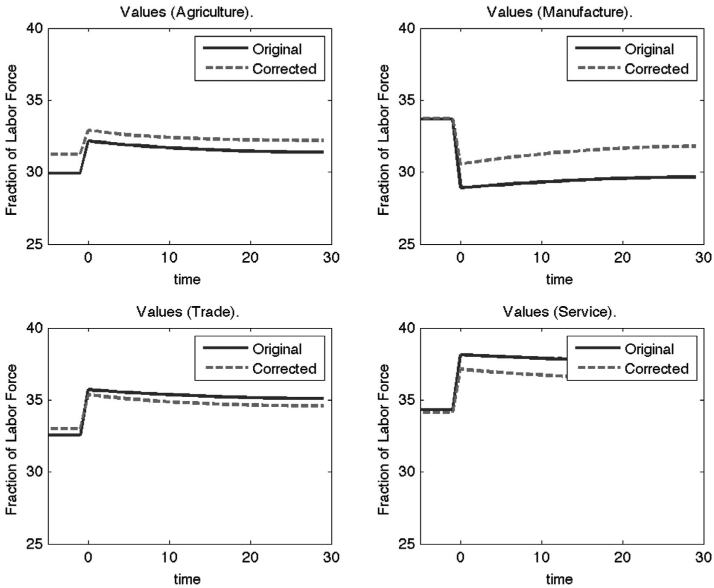

Figure A3.11: *Values (β =0.97)*

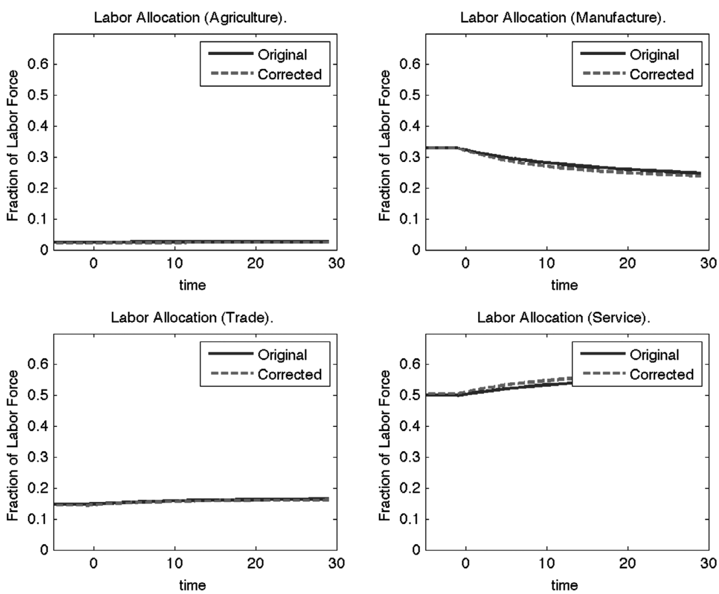

Figure A3.12: *Labor Adjustment (β =0.90)*

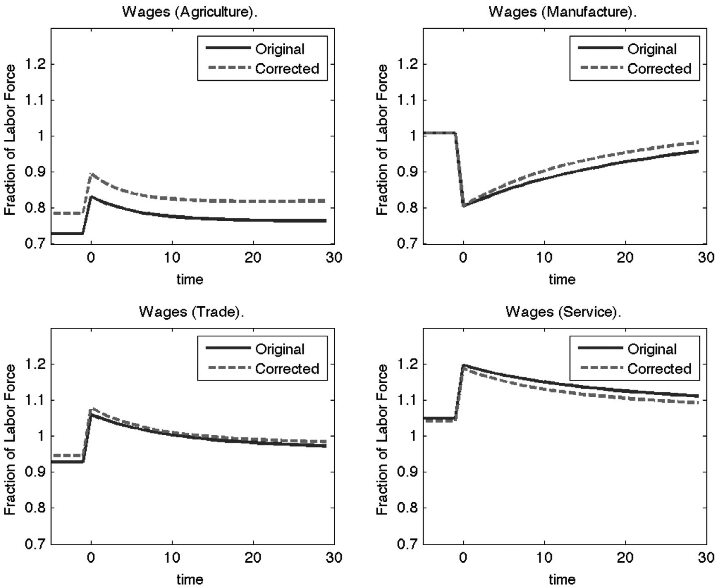

Figure A3.13: *Wages (β =0.90)*

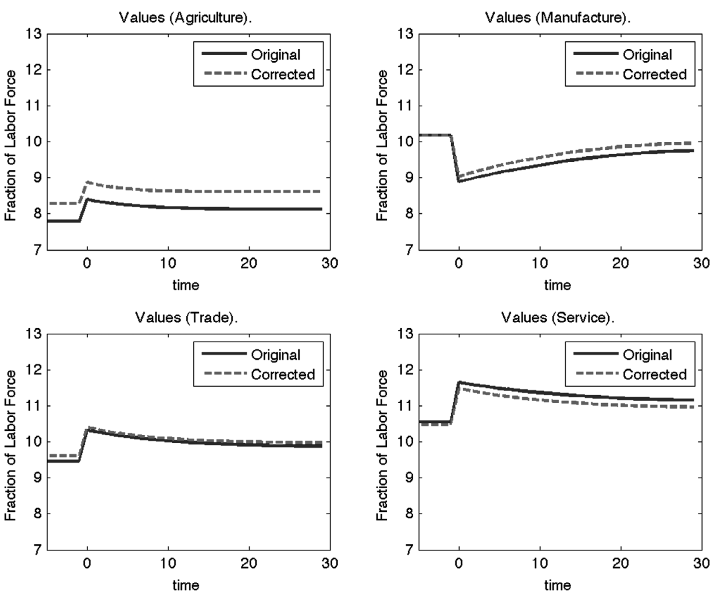

Figure A3.14: *Values (β =0.90)*

4

Reallocation and Adjustment in the Manufacturing Sector in Uruguay

CARLOS CASACUBERTA AND NÉSTOR GANDELMAN

1. INTRODUCTION

Microeconomic analysis of firm performance has boomed in recent years, due to the development of theories stressing firm heterogeneity beyond the former representative agent models, as well as the increased availability of large panel data sets assembling information on relevant firm level indicators. The adjustment costs literature has had a particularly strong development, both on the theoretical side as well as in empirical exercises. Some papers have started to analyze not only industrial economies—but also less developed ones, particularly Latin American countries.[1]

This note is based on our previous and current research on factor adjustment, protection, and openness in Uruguay. Although the main concern is the same there is a first group of papers whose main focus is on factor (employment, capital) creation, destruction and reallocation[2] and a latter group of papers focused on adjustment functions.[3]

Besides characterizing the factor reallocation and factor adjustment process in the Manufacturing sector in Uruguay our research looks at how policies interact with these processes. Much of our effort was devoted to understand how trade liberalization impacted on allocation and (or) reallocation of production factors and firm productivity. Besides changes in protection, we wondered how the factor flows were affected by other relevant variables like unionization, concentration, and the irruption of China and India as new relevant players in international trade.

In this note we report and discuss the main results of our work and suggest directions for continuing research.

2. WHY IS URUGUAY AN INTERESTING CASE STUDY?

After an early financial liberalization, Uruguay started to open its economy in the 1970s. In the 1990s the process was intensified, along with the signature of the

[1] See for instance Eslava et al. (2005)
[2] Casacuberta, et al. (2004a; 2004b; 2005; 2007)
[3] Casacuberta and Gandelman (2006; 2009)

Mercosur treaty with Argentina, Brazil, and Paraguay. Trade liberalization sought to continue and deepen the openness process, and reversing the anti-export bias that characterized previous import substitution policies. In 1991, with the signature of the Mercosur Treaty a program of scheduled tariff reductions began, that ended in 1995 with the establishment of an imperfect trade union. Through signing binding international treaties (Mercosur and World Trade Organization), the government significantly reduced its ability to provide discretionary protection.

Pervasive double digit inflation rates during the 1960s and 1970s eroded the confidence in the peso and dollarized the economy. In the early 1990s a stabilization program based on an exchange rate anchor was launched, which considerably reduced inflation, but was accompanied by a significant real appreciation of the peso, especially vis-à-vis non Mercosur countries.

In the first half of the 1990s, trade liberalization proceeded in the presence of strong—at least initially—unions and different industry concentration levels. Following the loss of democracy in 1973 and until 1984 unions were banned. After that, with the democratic recovery in 1985 and until 1991 tripartite (including worker, entrepreneur, and government representatives) wage bargaining was set at the industry level with mandatory extension to all firms within the sector. This boosted the share of unionized workers. Beginning in 1992–93, there was a shift towards decentralization and firm-specific bargaining that was reversed again in 2005 with new tripartite labor bargaining.

In 2002, Uruguay suffered a profound financial crisis triggered by contagion effects from the run on banks, massive currency devaluation, and gigantic default on sovereign debt that took place in next-door Argentina. In the wake of a run on its own exceedingly dollarized banking system, Uruguay's government was forced by the ensuing loss of international reserves to let the currency depreciate rapidly. It also rescued the banking system and intervened in several failing banks, obtaining massive financial backing from the Washington-based multilateral agencies to that end. Eventually, the government also had to arrange for a market-friendly restructuring of the public debt. By 2004, Uruguay had regained access to the international capital markets and the economy started to recover.

3. DATA

At the time we wrote the aforementioned papers, data availability was restricted to the manufacturing sector. We used annual establishment level data from the Manufacturing Survey conducted by the Instituto Nacional de Estadística (INE) for the period 1982–1995, covering establishments with five or more employees. Constant prices measures of value added, materials, and energy inputs were obtained using sector level price indexes from INE. The survey reports separately blue and white collar employment. Using the 1988 data from the economic census, an annual capital stock series was constructed using the perpetual inventory method.

Recently, new firm level data have become available for research. New data cover the 1997–2005 period and, more importantly, extend the sector coverage.

Besides the manufacturing sector, new data include (according to ISIC Rev. 3): D-Manufacturing, E-Electricity, gas and water, G-Commerce, H-Hotels and restaurants, I-Transportation and communication services, K-Real estate and machine rentals, M-Educational services, N-Health services, and O-Other community, social and personal services. These data have not yet been used to study reallocation and adjustment costs.

4 METHODOLOGY

4.1 Factor reallocation

To measure factor flows, we follow Davis and Haltiwanger (1992) and Davis, et al. (1996) and measure size for establishment *i* at time *t* as the average between factor used in period *t* and *t-1*, $\phi_{it} = \frac{X_{it} + X_{it-1}}{2}$. In turn the rate of growth is defined as $\Delta X_{it} = \frac{X_{it} - X_{it-1}}{\phi_{it}} = \frac{X_{it} - X_{it-1}}{((X_{it} + X_{it-1})/2)}$. These are not the usual rates of growth with the value observed in period t-1 in the denominator, but there is a monotonic relation between both of them. Denoting the traditional growth rate $g_{it} = \frac{X_{it} - X_{it-1}}{X_{it-1}}$, it can be shown that $g_{it} = \frac{2\Delta X_{it}}{2 - \Delta X_{it}}$.

There are several advantages from using these growth rates. They are symmetrical about zero and restricted to finite values. Traditional growth rates attribute a plant birth a growth rate of positive infinity and a plant exit a growth rate of –1. With growth rates defined in relation to average employment (and not past employment), a plant birth and a plant exit correspond to 2 and –2 growth rates respectively.

Using these growth rates we compute summary factor flow statistics. For instance, for employment, Net creation (*Net*) is the change in total jobs, creation (*Pos*) is the sum of all newly created jobs (in the expanding plants), and destruction (*Neg*) is the sum of all destroyed jobs (in contracting plants). Reallocation (*Sum*) summarizes the heterogeneity in plant level outcomes by adding the factor quantity destroyed and created in the period. Note that from these definitions reallocation is the sum of factor creation and factor destruction $Sum_t = Pos_t + Neg_t$.

Formally,

$$Net_t = \sum_i \phi_{it} \Delta X_{it}$$

$$Sum_t = \sum_i \phi_{it} |\Delta X_{it}|$$

$$Pos_t = \sum_i \phi_{it} max(\Delta X_{it}, 0)$$

$$Neg_t = \sum_i \phi_{it} \left| min(\Delta X_{it}, 0) \right|$$

4.2 Adjustment functions

To study adjustment costs we followed the methodological approach of Eslava et al. (2005) to estimating adjustment functions. It is based on the observation that employment and capital are not generally at their desired levels when firms face adjustment costs. At any point in time there is a gap between the actual and the desired levels. The adjustment function of employment (capital) is defined as the percentage of the employment (capital) gap that is actually closed when adjusting employment (capital) and is modeled as a function of the employment and capital gaps.

The desired rate of change ZX_{it} is defined, paralleling the previously defined growth rates, as a fraction of the average between the present desired level and the past observed level:

$$ZX_{it} = \frac{X^*_{it} - X_{it-1}}{\left((X^*_{it} + X_{it-1})/2\right)},$$

where X^*_{it} is the firm's desired level of factor of production X.

The desired rates of growth can be thought of as shortages, that is, when $ZX_{it} > 0$, implying the firm desires a factor level higher this period with respect to that observed the period before (that is, job and (or) capital creation), or surpluses, that is, $ZX_{it} < 0$ implying that the firm is willing to reduce its factor level with respect to last period's (that is, job/capital destruction).

The adjustment function is defined as the fraction of the shortage that is actually closed by the firm. It is the ratio between the change in factor use and the shortage rate. Hence adjustment functions are defined as follows:

$$AX_{it} = \frac{\Delta X_{it}}{ZX_{it}}.$$

A key methodological step in this procedure is the estimation of desired factor levels X^*_{it}. To do so, we first estimate the firm's frictionless factor levels, by a counterfactual profit maximization program. Our general framework is one of monopolistic competition in which firms have certain degree of market power. Our first step is to estimate the firm's frictionless factor demands. Frictionless levels correspond to input levels that the firm would choose in the absence of adjustment costs, and are obtained from the firm's optimization problem.

A Cobb–Douglas technology was assumed, dependent on capital, blue collar employment and hours, white collar employment, energy, materials, and a total factor productivity shock. While demands for all factors are obtained, it is assumed that in the absence of adjustment costs for hours, energy, and materials the frictionless levels of those inputs coincide with the observed levels. This leads us to the three equations of the frictionless levels of capital, blue-collar employ-

ment, and white collar employment as functions of the parameters of the model to be estimated and observed variables.

The productivity shock and production function parameters were recovered using the Levinsohn and Petrin (2003) methodology. We also estimated establishment level demand shocks using the inverse demand equation implicit in our monopolistic competition assumption. The inverse (log) demand function was estimated and the demand shock recovered as a residual.

Frictionless factor levels are not the same as the desired ones. Both differ in that the desired levels are those that would be observed if adjustment costs were momentarily removed, while frictionless levels are those that would hold in the absence of adjustment costs in all periods. To obtain the desired factor levels Eslava et al. (2005) propose to follow Caballero and Engel (1993), who show that even with adjustment costs, under general assumptions the desired and frictionless factor demands are proportional.

Following Caballero et al. (1995; 1997) the proportionality constants can be obtained as the ratio between the actual and frictionless factor levels, for the year where investment and employment growth take the median value for each of them. It implies that for a firm, in the year of the median employment growth and median investment, the desired and the actual adjustment coincide.

4.3 Protection, trade liberalization and other issues of interest

In the case of Uruguay we can reasonably argue that the usual endogeneity critique to the use of tariffs or changes in tariffs as explanatory variables does not apply. Uruguay is a minor player integrated with its much larger neighboring economies in Mercosur. Hence the common external tariff and the changes in Uruguayan tariffs to converge to the trade bloc protection level are basically affected by Argentinean and Brazilian political players and beyond control for local firms. This conclusion can be drawn from Olarreaga and Soloaga (1998), an application of a Grossman and Helpman (1994) 'protection for sale' model to the MERCOSUR common external tariff, in which it is shown that the customs union external tariff follows closely the Brazilian tariff structure.

In our econometric exercises we estimated the effect on reallocation and adjustment functions of various protection and trade liberalization measures. Making use of the particular institutional changes in Uruguay we could estimate the effect of unions in reallocation rates and how they increase or mitigate the effects of trade liberalization. By interacting these variables with reallocation and adjustment functions we also studied the impact of different sector concentration levels as well as the impact of China and India's import competition.

5. RESULTS

5.1 Reallocation

We found that creation and destruction rates for employment and capital were relatively high and pervasive over time, both for white and blue-collar employ-

ment. Exits explained a sizeable part of destruction rates. Capital intensity increased, while the relative capital labor price ratio fell, consistent with firms moving towards more capital intensive technologies.

Most of the excess reallocation (reallocation that was not required to accommodate changes in factor use, that is, $Sum_t - |Net_t|$), was due to movements 'within' rather than 'between' sectors. Thus, reallocation rates seemed to be linked to establishment level heterogeneity rather than to aggregate shocks.

Larger and older firms tend to have more stable use of their factor of production. Most factor reallocation takes place in the smaller and younger firms. Larger firms create and destroy fewer jobs and less capital than the smaller ones. The latter effect is stronger than the former, implying that larger establishments had higher net creation rates.

Higher international exposure implied a slightly higher job creation rate and an important increase in job and capital destruction. Overall the opening of the economy is associated with larger reallocation rates.

There is agreement in the literature on the effects of unions on wages, but their effect on other dimensions is less clear. We found that unions were able to weaken the direct impact of trade liberalization by reducing job and capital destruction with almost no effect on creation. Therefore, we found unionization to be associated with more stable factor use, that is, less reallocation. Industry concentration also was found to mitigate the destruction of jobs but had no effect on job creation or in capital dynamics.

The reallocation of production factors was accompanied by an increase in total factor productivity especially in sectors where tariff reductions were larger and unions were not present. Our result that unions were effective in reducing employment and capital destruction but inhibited productivity growth underscored the tradeoff between short term costs, as jobs are lost and long-term gains made as productivity rises.

Over the 14 years covered in our studies (1982–95), blue and white collar net employment creation rates were similar, but there are two distinctive periods. In the first, under higher protection and government intervention in wage bargaining, blue collar net creation rates were higher than white collar rates. In the second period when substitution of labor for capital was stronger, protection was lower and employment bargaining was set at the firm level without government participation, more blue collar than white collar jobs were destroyed.

It is useful to put our results into perspective. Haltiwanger et al (2005) present evidence on job reallocation for a sample of Latin American countries in the same period experiencing similar tariff reductions and exchange rate appreciation processes than Uruguay. They report an average job reallocation rate of 21 per cent, while Uruguay's 14 per cent remains the lowest among all. Are these relatively lower reallocation rates due to higher adjustment costs?

5.2 Adjustment functions

The estimation of the adjustment functions allows a graphical representation that illustrates our analysis. In the basic setup adjustment functions for white collar

workers, blue collar workers, and capital look like Figure 1. The percentage of adjustment is plotted as a function of the shortage or surplus. Negative desired rates of growth indicate that the past level of the input is above the desired one (there is a factor surplus); hence to close this gap the firm needs to decrease this factor, and finds itself in the job or capital destruction side. Conversely, positive desired rates of growth show a past factor level below the desired one (there is a factor shortage); hence if the firm wants to close the gap it will be in the factor creation side, that is, it will invest or hire.

Figure 4.1 shows an asymmetric behavior in the adjustment process (both in the intercept and the slope of the adjustment functions). At small shortage levels, white and blue collar employment adjustment functions show an upward shift in the positive side. Firms tend to adjust a larger fraction of the gap between the desired and the actual employment when the observed levels are below the desired ones, that is, the firm finds it easier to create labor than to destroy it except when the destructive adjustment is large. Note that for desired rates of growth in absolute value below 1, firms close a larger fraction of the gap of blue collar workers than of the gaps in other factors. A shortage of 1 or −1 corresponds to firms desiring to triplicate or reduce to one-third their actual factor use. Therefore, for most firms, blue collar adjustment costs are lower than are adjustment costs in other factors.

A firm that would like to cut all levels of factor employment by half (factor shortage = −0.66), would only reduce its level of white collar workers by 5 per cent (10 per cent of 50 per cent), its level of blue collar workers by 10 per cent (20 per cent of 50 per cent) and its level of capital by 5 per cent (10 per cent of 50 per cent). On the contrary a firm that would like to increase its factors by 50 per cent (factor shortage = 0.66), would only increase its level of white collar workers by 12.5 per cent (25 percent of 50 percent), its level of blue collar by 15

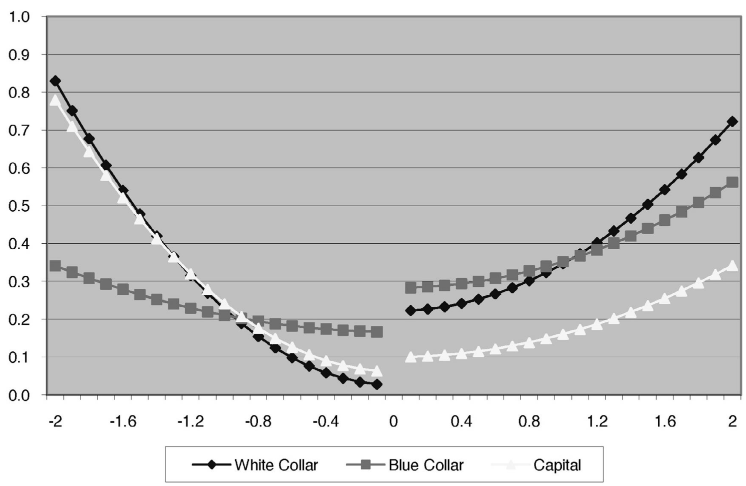

Figure 4.1: *Adjustment functions*

per cent (30 per cent of 50 per cent) and its level of capital by 5 per cent (10 per cent of 50 per cent). This suggests strong adjustment costs on the hiring and firing side for Uruguayan manufacturing firms.

The econometric estimations behind Figure 1 showed that shortages or surpluses of other factors are relevant to understand the adjustment process. The larger the shortages of one factor, the less responsiveness in adjustment in the creation side of other factors, but larger adjustment in the destruction side. If a firm whose desired level of two factors is above the current level, the larger the shortage in one factor the lower the adjustment in the other. For a firm desiring to have a lower level of two factors than their actual value, the larger the surplus in one factor the larger the adjustment in the other. This points against (in favor of) economies of scope in the adjustment cost function in the creation (destruction) direction. When firms want to hire more, it is cheaper to adjust one factor at a time but when they want to reduce employment or scrap capital, it is cheaper to reduce the use of both factors together.

Comparing the different factor adjustment functions, both in the creation and the destruction side, the slopes for white collar labor are larger than for blue collar labor. This may relate to differences in adjustment costs for each factor. Unions tend to be stronger in industries more intensive in blue collar labor, thus inducing higher adjustment costs on the destruction side when employment surpluses are large; that is, there will be lower adjustment when the desired rate of change is large in the destruction side. The white collar adjustment function has a steeper slope than blue collar adjustment on both sides. When the shortage is small in absolute value, adjustment is lower in white collar than in blue collar labor. Conversely, if the shortage is large in absolute value, a larger proportion of the gap is closed for white collars than for blue collars.

This may be because white collar labor includes workers with specific human capital, which is difficult to create. Therefore firms probably may be willing to accept small shortages without adjusting, but the adjustment will be fuller when the shortage becomes large in absolute value. For instance, consider a firm that has more clerks than needed, but they are familiar with the workings of the firm: if this shortage is not too large, the firm may prefer to keep these extra workers. On the other hand, if blue collars have less specific training, they may be more easily disposed. On the creation side, hiring an extra clerk implies higher training costs; hence the firm may prefer to use the existent workers more intensively if the shortage is small. If the shortage becomes large enough, the cost of the extra hours will be higher than the training cost of the newly hired white collar workers. These search and training costs include a fixed cost that can be covered only when the percentage of the gap closed is large enough.

Our analysis of the trade liberalization process, and the impact of China and India imports, was framed in terms of the changes in the intercept and the slope of the previously described adjustment functions.

Overall, we find that our period of analysis is particularly appropriate to evaluate the effect of tariff reductions on firm performance, since in those years tariff reductions were steady and of a significant magnitude. We analyzed the effect

of changes as well as levels of import taxes (defined as simple averages of Harmonized System items by 4-digit ISIC class). The average tariff was reduced 43 per cent to 14 per cent between 1985 and 1995. On average, annual tariff changes accelerated from −2.1 per cent before 1990 to −3.0 per cent thereafter.

In all of our regressions at least one policy interaction was significant. To evaluate the differences in adjustment with varying levels of trade liberalization we simulated the predicted adjustment using the estimated coefficients for different levels of changes in tariffs (0, 2, and 4 point reductions). While we find a negligible effect in capital adjustment, there is a pattern for both types of labor. For white collar and blue collar labor the fraction of the gap actually adjusted decreases in the creation side, while it increases in the destruction side in the presence of tariff reductions. For white collars this is produced by a change in the intercept, while for blue collars it is produced by a change in the slope (both statistically significant).

Firms in sectors that experienced higher tariff reductions could adjust a larger proportion of their surpluses than those not so exposed. For white collars this result is present for low levels of desired rates of growth, while for blue collars the effect shows for higher desired rates of change. The change in the slope for white collars is not statistically significant, therefore changes in this case should be considered as parallel shifts. On the creation side it was the opposite: firms with lower tariff reductions adjusted a larger proportion of their shortages.

When instead of using the import tax change in the firm sector we used the import tax level as a shifter of the adjustment functions, most policy interaction terms were statistically significant for blue collar and capital, while for white collar labor only the constant shifter for the creation side changed significantly. In this case we simulated the adjustment functions for tariff levels of 10, 20, and 30 percent, and the results confirmed our previous finding: the effects of the protection level were stronger for blue collar and capital than for white collar.

Lower tariff levels were associated with higher adjustment on the creation side, especially for blue collar jobs but also for white collar jobs and capital. Adjustment in the destruction side seemed not to change with tariff levels in the case of white collar adjustment functions. For capital and blue collar labor, higher tariff levels were associated with lower adjustments in the destruction side, the opposite of the creation side.

A possible explanation is that protected sectors are typically less competitive industries prone to reduce net employment even in the presence of trade protection. It can also be an indirect way to show that protection may in fact destroy jobs, rather than create them. If shocks to firms are iid, protection will lead to lower levels of employment, and this may have to do with firms' expectations. If there is a generalized positive demand shock, a firm in a highly protected sector will not adjust completely in the presence of adjustment costs and so on (firing workers) unless the government has credibly committed to maintain protection. If there is any risk that the tariff will go down, the firm may be more reluctant to hire more workers than a similar firm in another sector not exposed to the risk of the government reducing tariffs. The same applies to the job destruction side.

A highly protected firm that suffers a negative shock will be more likely to fire workers if the government's tariff is not a credible permanent policy.

Another interpretation is based on the fact that during the sample period there was a trend in tariff reductions in most industries, with the greatest reductions occurring in those industries initially most protected. These tariff changes were for the most part predictable in advance (especially after the signature of the Mercosur Treaty in 1990). The nature of factor adjustment due to such a predictable, and probably permanent, change in the economic environment is likely to be different to factor adjustment in response to fluctuations in demand or input prices over the business cycle. Firms would have incorporated expectations regarding how tariff reductions would affect the product demand curve they face and probably also input prices relevant to their factor employment decisions.

In this line, sectors that were initially more protected and foresaw downsizing would not have created jobs or capital even if they faced a temporary positive shock. The long run trend dominated their decisions and therefore more protection may be associated with less adjustment on the creation side. Similarly, firms in a highly protected sector that predicted lower levels of employment and capital in the medium run would have more likely destroyed factors in the face of a temporary negative shock since they knew that the general trend went that way. With a probably predictable trade liberalization trend, tariff protection is associated with more adjustment on the destruction side and less in the creation side.

With respect to imports from China, the interaction terms for white collar workers with the import share were not significant, suggesting that the adjustment costs associated with this factor of production for firms facing strong competition from China are not different from those for firms that do not. The effect is felt on adjustment costs for blue collar workers. In the case of shortages the coefficient is negative, suggesting that the adjustment is smaller (and adjustment costs larger) for blue collar workers in the presence of surpluses. However in the presence of positive shortages the adjustment is larger: it is easier to hire blue collar workers. For capital, the only significant interaction suggests that firms subject to strong competition from China find it easier to adjust in the presence of capital shortages.

The analysis does not imply causality, since possibly smaller adjustment costs allow for higher import penetration from China. As adjustment costs are smaller in the presence of shortages and larger in the presence of surpluses when firms are exposed to import competition from China. These may result from higher perceived volatility in Chinese imports when compared to imports from other regions. If this is so, firms would be more reluctant to fire workers and more willing to hire workers when exposed to more import competition from China than from more established and better-understood trading partners. Coefficients of variation for imports from China and India are twice the coefficient of variation for imports from the rest of the world.

Results for India suggest that adjustment functions for all factors change in the presence of stronger import competition from India. A number of interactions with India's import shares are significant, but their signs make it difficult to assess the direction of this effect.

6. CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

All in all, our micro data evidence supports some regularities present in previous literature on adjustment functions. Investment and job creation are not the result of smooth and continuous microeconomic decisions. Individual adjustment constraints depart significantly from those implied by quadratic adjustment costs, and there are several sources of irreversibilities (technological, market-induced, increasing returns in the adjustment technology). The evidence provided seems to confirm a pattern that has important nonlinear features, and hence is consistent with such constraints. This impacts the use of all factors of production, particularly employment.

Adjustment costs faced by capital, white and blue collar labor are non trivial in the Uruguayan manufacturing sector, which has consequences in terms of factor unemployment and economic efficiency. The size of adjustment costs may reduce factor reallocation and dampen productivity gains.

Our main results confirmed the asymmetric nature of firms' adjustment process. Large shortages of one factor lead to less responsiveness in adjustment in the creation side of other factors and to larger adjustment in the destruction side.

We also assessed the effects of protection and trade liberalization on firms' adjustment process. The constraints arising from adjustment cost functions may become an important part of policy analysis. Our results point to a significant shift in adjustment functions for all production factors associated with increased liberalization after the Mercosur Treaty. Specifically, trade policy variables measured by tariffs levels and reductions in tariffs significantly shifted adjustment functions. Firms in less protected sectors have shown higher adjustment fractions in the creation side and lower in the destruction side, particularly for blue collar labor. Sectors facing larger tariff changes adjusted less in the creation side, particularly for blue collars, and more on the destruction side. In the context of tariff reductions of Mercosur, more highly protected sectors were probably those that faced the largest tariff reductions. Overall the impact of higher international exposure on factors of production is stronger for blue collar workers than for white collar workers.

Though our work does not provide an empirical identification strategy to pinpoint the causal relationship between trade openness and adjustment costs, we believe our estimation results provide a powerful descriptive insight and valuable suggestions for the mechanisms behind observed firm behavior.

Our analysis also showed that adjustment costs faced by firms subject to strong Chinese and Indian competition seemed to be particularly large for firms that would like to reduce their levels of white collar workers. For firms experiencing factor shortages, however, adjustment costs seem to be smaller when subject to import competition from China and India (except perhaps for small shortages of skilled labor when subject to import competition from India).

Finally, policies should be based on a wide look at the economy. Inference from the manufacturing sector may not be extendable to other sector of production. As new panel data on service sectors become available we should extend our empirical work to cover all sectors of activity.

Néstor Gandelman is the Director of the Department of Economics of Universidad ORT, Uruguay

Carlos Casacuberta is Lecturer at the Department of Economics, School of Social Science, Universidad de la República, Uruguay

REFERENCES

Caballero, R, and E Engel (1993). Microeconomic adjustment hazard and aggregate dynamics. *Quarterly Journal of Economics*, 108(2): 359–83

Caballero, R., E. Engel and J. Haltiwanger (1995). Plant-level Adjustment and Aggregate Investment Dynamics. *Brookings Papers on Economic Activity*, 2: 1–54

—(1997). Aggregate employment dynamics: Building from microeconomic evidence. *American Economic Review* 87(1): 115–37

Casacuberta, C, Fachola, G and N Gandelman (2004a). The Impact of Trade Liberalization on Employment, Capital and Productivity Dynamics: Evidence from the Uruguayan manufacturing sector. *Journal of Policy Reform*, 7(4): 225–48

—(2004b). Employment, Capital and Productivity Dynamis: Evidence from the manufacturing sector in Uruguay. Documento de Trabajo 16, Facultad de Administración y Ciencias Sociales, Universidad ORT Uruguay

—(2005). Creación, destrucción y reasignación de empleo y capital en la industria manufacturera. *Revista de Economía del BCU*, 12(2): 90–124

—(2007). International Exposure, Unionization and Market Concentration: The Effects on Factor Use and Firm Productivity in Uruguay, in E. Aryeetey and N. Dinello (Eds.) *Testing Global Interdependence: Issues on Trade, Aid, Migration and Development*, Edward Elgar Publishing, Northampton, Massachusetts, US

Casacuberta, C and N Gandelman (2006). Protection, Openness and Factor Adjustment: Evidence from the manufacturing sector in Uruguay, World Bank Policy Research Working Paper 3891

—(2009). Factor adjustment and imports from China and India: evidence from Uruguayan manufacturing in D Lederman, M Olarrreaga and G Perry (Eds.) *China's and India's Challenge to Latin America*. The World Bank

Davis, S and J Haltiwanger (1992). Gross job creation gross job destruction and employment reallocation. *Quarterly Journal of Economics*, 107(3): 819–63

Davis, S, Haltiwanger, J and S Schuh (1996). *Job creation and destruction*, MA: MIT Press

Eslava, M, Haltiwanger, J and M Kugler (2005). Factor adjustment after deregulation: panel evidence from Colombian plants, NBER Working Paper No. 11656

Grossman, G and E Helpman (1994). Protection for sale. *American Economic Review*, vol. 84(4): 833–50

Levinsohn, J, and A Petrin (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70(2): 317–41

Olarreaga, M and I Soloaga (1998). Endogenous tariff formation: the case of Mercosur. *World Bank Economic Review*, 12(2): 297–320

5

Trade Reforms in Natural-Resource-Abundant Economies[1]

JAIME DE MELO

1. INTRODUCTION

If trade reform may not have poverty alleviation as its main objective, in low-income countries it is expected that trade reforms would at least raise the incomes of the segments of the rural population engaged in exporting activities. Are farmers engaged in crops likely to be reached by trade reforms, and do they benefit from them? What are the adjustment costs? Reductions in government intervention in these markets should be beneficial for the rural poor since it is well-documented that agriculture is taxed both directly via export taxes and indirectly via protection of manufactures (see for example, Krueger et al. 1988; Schiff and Valdes 1992). Removal of price controls and taxes and the dismantling of non-operational marketing boards and stabilization funds should benefit the rural poor. Moreover, since these reforms are usually extended to most of the agricultural sector, macroeconomic performance should improve, if only because agriculture weighs so heavily in the economy. Yet examination of the record shows that these reforms have been controversial, often yielding disappointing results in the low-income natural-resource-based economies reviewed here.

While the wave of pro-market reforms of the 1990s appeared against a backdrop of widespread failure in government intervention, a growing body of case studies is pointing towards the importance of context-specificity.[2] Export crops are often extracted from a narrow geographic and economic base with few players all along the value chain, a situation ideal for strategic interaction over the appropriation of rents, especially when market and product characteristics result in asymmetric information and marketing externalities.

[1] I thank Olivier Cadot, Céline Carrère, Marcelo Olarreaga and Mario Piacentini for comments on a previous draft.

[2] Two studies illustrate the point. In the case of the allocation of quota rents to Indonesian coffee traders under the ICA, Bohman et al. (1996) give evidence of both rent-seeking activities and the creation of barriers to entry by bureaucrats to increase the level of rent-seeking activities at the national level. Using six crops (cocoa, coffee, cotton, groundnuts, tobacco, and vanilla), McMillan (2001) shows that the self-defeating (that is, high tax policy) is pursued by governments for these cash crops because of time-inconsistency caused by fixed costs associated with tree crops. Using data for these crops for 32 African countries, she finds that tax rates are higher, the higher is the ratio of sunk to total costs, and the higher are expected future earnings (proxied by average past profits).

Variants of the 'resource curse' in which rents conjure with geography, weak public institutions, and dysfunctional political regimes have been invoked to account for the disappointing performance. Starting with Sachs and Werner (1997) slow growth has been found associated with dependence on natural-resource-based export structures. Along the same lines, countries with export structures specialized in 'point-source' (as opposed to 'diffuse') natural resources have been strongly associated with weak public institutions and low growth (Isham et al. 2005). Most recently, Amin and Djankov (2009), using the 'Doing Business' data on specific micro-measured regulatory or legal reforms also find a negative correlation between the share of primary exports in total exports and their reform variable (coded as a dummy variable on the basis of reforms in one of 10 sets of indicators).

The literature surveyed below isolates the channels through which trade reforms operate in economies specialized in perennial crops where barriers to entry along the value chain create rents, but the cross-country evidence is not robust. This lack of robustness is now well-documented and has contributed to the newer diagnostic-oriented approach to policy reform which is suspicious of best-practices, where expectations are based on the traditional presumptive approach to reforms (we know how markets work and here is the list of reforms to be carried out).[3] To illustrate the usefulness of case studies I chose two case studies: the cashew reforms in Mozambique, a cause célèbre for anti-globalists, and vanilla in Madagascar, where the government and market failures produced also contributed to a disappointing outcome. Both case studies help to illustrate the difficulties of assessing the impact of reforms in high-rent environments with strong market failures.

I start in Section 2 with a discussion of the channels emphasized in the cross-section literature, then turn to the two case studies. Section 3 discusses the market characteristics and the reforms in the cashew and vanilla sectors. Section 4 describes the difficulties encountered in measuring the impact of the reforms in each case study, and Section 5 draws lessons from the case studies.

2. TRADE REFORMS IN HIGH-RENT ENVIRONMENTS

The most heralded framework (Acemoglu et al. 2001) describing trade reforms in high-rent environments combines climate (disease vectors, rainfall levels, temperatures) and topography (soil/mineral quality) with the profitability of 'point-source' natural resources as the deep determinants of the institutional set-up chosen by colonizers who chose extractive rather than settler colonies; that is predatory institutions facilitating rent extractions. In turn, these institutions have

[3] The huge success of this agnostic approach is exemplified in the recent outburst of randomized control trials of projects and aid programs. Unfortunately this approach cannot be applied to evaluate the impact of trade policy reforms because of a lack of natural control groups. To take the example of the reforms in the vanilla market in Madagascar discussed below, it is difficult to see how the reforms could have been carried out in the Sambave and Antalaha districts but not in the Andapa and Vohémar districts.

lasted through the inertia that characterizes institutional reform. But, as pointed out by Isham et al. (2005), India is different from Mexico and Russia in spite of similar colonial heritage. Digging further, institutional history and the pattern of policies have been linked to endowments through two channels: first via a rentier effect as easily extracted revenues are straightforwardly controlled. This implies for the state that: (i) it needs less revenue and hence incentives to develop 'checks and balances' are limited[4]; (ii) dissent can be bought off; (iii) if necessary, resources are available to control dissent. Second, modernization is delayed (that is, institutions do not develop), as a small group of elites can extract the surplus, in some cases because they own the land (as in Latin America), or in the case of cash crops grown by small farmers, because they are the intermediaries between the producers and the export stage. This corresponds to the cases of vanilla and cashews. These same groups oppose modernization or reform, as these would create an alternative source of power. By losing political power, the elites would lose control over the rents (as happened between the change of political regime in Madagascar when vanilla policy was altered from phase II to phase III; see Table 5.1). As pointed out by political scientists and evidenced by economists (for example, Gylafson 2001), in addition to the Dutch-disease overvaluation slowing industrialization, resisting modernization also means less increases in literacy and in education generally, and less labor organization. As put by Isham et al. (date), resource abundance simultaneously 'strengthens states' and 'weakens societies', leading to an environment where beliefs about announced policies lack credibility and end up in a lack of a supply response (planting trees involving irreversible costs).

The natural resource curse comes from the two-way causation between institutions and performance: institutions are demanded in high-performance environments, so the rich countries are rich because they ask for institutional quality. However, recent literature (Brunnschweiler and Bulte (2008a; 2008b)) has questioned the evidence on the natural resource curse which is captured by the negative correlation between natural resources and growth, or the negative correlation between natural resources and income per capita. They question the use of the primary export share as an adequate proxy for natural resources, which they argue is a measure of natural resource dependence, rather an appropriate measure of natural resource abundance. When they use instead the discounted value of expected future resource rents over a 25 year period—a better measure of resource abundance—they obtain a positive correlation between resource abundance and growth. This implies that the resource sector becomes the default sector in the absence of decent institutions when no one is willing to invest in alternative forms of capital. Natural resource dependence is endogenous.[5]

[4] In the case of vanilla and cashews, few intermediaries purchase the raw crop from farmers in small holdings.

[5] This was particularly striking in the case of Madagascar when the time-limited preferences under AGOA provided a unique opportunity to invest in better infrastructure to attract some irreversible FDI in the EPZ zone producing apparel.

If resource rents trigger 'rentier state' dynamics, recent literature starting with Collier and Hoeffler (2004) suggests that the search for rents by greedy rebels (helped by inhospitable geography making it easy to set up the rebel groups and elude the states' police and military presence) engenders conflicts over the control of the natural resources (oil for Nigeria, or diamonds for Sierra Leone and Angola). Taking into account the endogeneity of resource dependence in growth and conflict regressions, Brunnschweiler and Bulte (date) find that the curse disappears and that it becomes a symptom rather than a cause of underdevelopment. When using resource abundance as the indicator of natural resources, they obtain a positive correlation between abundance and growth (thus Botswana with high-quality institutions has diamonds and high-growth, but neighboring Angola equally well endowed in diamonds) has low—quality institutions and low growth)

This recent literature highlights two other channels through which natural resources affect policies and outcomes. First, Hodler (2006) and others recognizing that fighting over rents will depend on the number of contestants show that there may be rent over dissipation (in the sense that resource-waste in the search for rents exceeds the value of the rents) if ethnic fractionalization and resource abundance have a negative effect on property rights. Then natural resources intensify rent-seeking and can lead to a decrease in production that exceeds the value of the rents if institutions are weak, leading to a curse. If, on the other hand institutions are strong, resource abundance is a blessing. Second, the political landscape also enters the picture. Collier and Hoeffler (2009) model income (rents) and substitution (public goods) effects under democratic and autocratic regimes. Under democratic regimes there is a greater provision of public goods and more 'checks and balances'. But this comes at the expense of more wasteful lobbying-for-rent-appropriation activity than under autocracy. The interplay of these effects is confronted with the data to check the role of the type of government in determining performance in natural-resource-rich environments. They produce cross-section evidence that autocratic regimes appear to outperform democracy along these two channels in resource-rich countries when institutions are weak.

With all these channels it is no surprise that recent evidence is inconclusive about the existence of a resource curse. In their extensive exploration of the resource curse with the most recent data, Arezki and Van der Ploeg (2008) find a tenuous link between resource dependence or resource abundance, and either growth performance or cross-country variations in per capita income. In regressions exploring the correlates of cross-country per capita income variations, they fail to find robust evidence that good institutions may turn the curse into a blessing, even after controlling for geography, institutions, and openness. While they find that bad trade policies worsen the resource curse, since bad trade policies are strongly correlated with other policies (in particular bad fiscal policies, see Easterly (2005)), in spite of exploring several channels through which geography, institutions, and politics affect trade reforms, this literature cannot be expected to yield robust lessons about the likely effects of trade policies in natural-resource-based economies. The two case studies below illustrate how trade reforms can unfold in a high-rent environment.

3. MARKET CHARACTERISTICS AND REFORMS IN THE CASHEW AND VANILLA MARKETS[6]

3.1 Characteristics and market structure

Both crops share many common characteristics. These are described in Table 5.1 and include: (i) production by small (and poor) farmers and their family on plots of land they own, with production taking place under competitive conditions[7]; (ii) few intermediary buyers who either process the raw cashews or cure the green vanilla; (iii) a regulatory environment in which purchase prices and margins by processor-traders were controlled by the government; and (iv) export taxation of processed vanilla or of raw cashews. I start with the internal market structure and then turn to the external market structure.

Cashews: McMillan et al. Figure 5 show three layers of intermediation between cashew farmers and world markets: (i) local buyers and small traders; (ii) larger wholesale traders; and (iii) exporters and the domestic processing factories. Entry barriers at each level as a result of set-up costs and regulatory restrictions combine to explain the monopsonistic power along the value chain. These barriers to entry contribute to explain the resulting low share of the world price received by producers (see Figure 2b). McMillan et al. (2003) note that there were eight to 10 exporters of raw and processed cashews at the time of the reforms.

Labor is the most important input in cashew with 50 percent of costs associated with curing the trees prior to harvesting, the remainder taking place at harvesting. Processing could take place either in large factories using one of two highly mechanized technologies that depend on a constant flow of calibrated nuts (to yield a high proportion of whole rather than broken kernels which fetch a higher price), or on semi-mechanized more labor-intensive technologies closer to the hand shelling done by Indians at home.[8] There is disagreement about the causes of the failure of the processing industry following its privatization, some arguing that a more labor-intensive technology would have been appropriate.

Vanilla: Cadot et al. (2009) Figure 1 shows the three phases in vanilla production: (i) vanilla growing, which produces the ‘green’ beans; (ii) curing, the stage at which it develops its quality (flavor profile and natural vanillin content); and (iii) packing (sorting, grading, and tying in small homogenous bunches). Each stage requires specific skills.

Growing is highly labor-intensive, as crop husbandry requires 260 man-days per hectare during the first year, and about 460 during the four to eight years

[6] Much of what follows is a summary of McMillan et al. (2003) and Cadot et al. (2009).

[7] Vanilla is a cash crop with all production sold. According to the household survey in Madagascar in which around 500 families in each survey listed vanilla production as their main activity, the share of vanilla in full income was around 30 percent. For cashews, McMillan et al. report that farmers only sell about half of their crop, the remainder being kept for on-farm consumption. These characteristics imply that the income effects of trade reforms will be muted.

[8] The manual technology yields a higher proportion of whole kernels but is hazardous for the workers, because they come into contact with a corrosive and toxic liquid called cashew nut shell liquid.

where plants reach maturity. Pruning and weeding are then supplemented by hand-pollination which means that each flower on the vine has to be pollinated by hand and at different times and harvesting.[9] Most workers engaged in growing belong to the family and very few producers turn to employ workers because of the meticulous work required for vanilla production. With few purchased inputs (producers need very little equipment and pesticides are useless), entry and exit costs are low although plants require over three years to become productive and growing conditions are rather exacting (small tracts of rich soil under the shade of trees).

Curing entails dipping beans in near-boiling water, then triggering an enzymatic reaction by alternate heating and 'sweating' which means boxing the beans and exposing them to sunlight. The process is repeated 10 to 20 times before the beans are left to dry outdoors for two to three months. By then, they possess a uniform dark color and strongly smell of vanilla. Once cured, vanilla beans are prepared, packed, and stored to keep their flavor, a stage that is peculiar to Indian Ocean producers. The storage process, which can last up to two years, is risky, as vanilla can mold and weekly inspections are required. Although it need not be the case, packers often export, and importers from the three main importing countries the United States, France, and Germany keep close and lasting marketing contacts with exporters, as this helps establish confidence and overcome informational failures described below. [10]

Several of a bean's quality characteristics are unobservable (five months after flowering, vanilla beans have reached their optimal size but if harvesting takes place before eight months, the beans will have less than half the full vanillin content), and largely determined by growing conditions, time of harvest, and the curing process. When prices are high, stealing will occur and fringe traders will compete on collection dates.[11] Less than half a dozen packers operated in the packing industry at the time of the reforms. This high concentration arguably resulted as much from government policies and rent-seeking as from economic rationality; this is because marketing externalities and the production—farm, curing, and packing levels—associations among producers would help reduce risks for buyers and risks for sellers. In fact, prior to its elimination, the Vanilla Stabilization Fund (VSF) was the quasi sole purchaser from the packers.

[9] Ecott (2004) is essential reading for anyone interested in the 'story' of vanilla. Chapter 4 describes the discovery of hand pollination in La Réunion in the 19th Century. Blarel and Dolinsky (1995) provide a detailed discussion of market failures in the market for vanilla.

[10] Blarel and Dolinsky (1995) also emphasize the asymmetric information between buyers (the food industry and brokers) and the sellers of vanilla beans. They "...note that confidentiality about the quality and technical characteristics of the vanilla bean and its processed products is of paramount importance in the commercial relationship of the food industrialist and the vanilla broker" (Blarel and Dolinsky (263).

[11] The resulting market failure could in principle be addressed by a variety of market mechanisms, including vertical integration, branding, or industry standards but as a matter of fact, vertical integration between farming and processing is virtually nonexistent. Vertical integration is still limited because the activities require specific skills. As a result, the industry has developed weaker mechanisms to alleviate adverse-selection issues, such as the introduction of identification marks that remain visible after curing.

On the external front, both products face highly non-competitive markets except for processed cashews. For raw cashews, Mozambican exporters directed sales to one country, India, with only a few importers having little bargaining power against the exporters because of a lack of storage facilities and high financing costs. As a result, reducing the export tax on raw cashew probably resulted in a terms-of-trade loss for Mozambican exporters of raw cashews (reportedly Mozambique accounted for 10 percent of the world market for raw cashews). For processed cashews, where Mozambique holds about 5 percent of the world market, the market structure is far more symmetric with about the same degree of concentration on both sides of the market.

Madagascar accounted for between 40 percent and 60 percent of the world market for high-quality 'bourbon' vanilla. As described in Melo et al. (2000), following the high rent-extraction policy during phase II (see Table 5.1) where the export tax reached 82 percent of the FOB price, Indonesians entered the market with their market share equaling that of Madagascar by 1993, just before Madagascar abandoned its extortionist policies (and had just burnt four years worth of stock in 1990). After major fires in Indonesia in 1996–7, Madagascar regained prominence in market share, accounting for around 60 percent of the world market.

The United States accounts for about half of the value of imports of vanilla (the world leader both for high- and low-quality vanilla with the dominance of McCormick in the former and Coca-Cola in the latter)[12], and as a result of re-exports of processed or packaged vanilla by developed countries, less than half the world share of vanilla trade is accounted for by developing countries.[13]

3.2 The reforms

The reforms were straightforward in both cases. For vanilla, following phase I when regulation was cooperative[14], rent-extraction policies were carried out during phase II culminating in the hands of a single packer-exporter, effectively creating a monopsonist buyer of green and cured vanilla on the domestic market, and a single exporter of vanilla in the export market (Blarel and Dolinsky, date 304). When they were abandoned (this coincided with a change of presidential regime and the reforms were not carried out under World Bank conditionality), the 'golden goose' had been killed, with export volumes about half those achieved

[12] About 70 percent of cured natural vanilla is bought by around 10 multinational companies. Another 10 'flavor houses' dominate the flavor compound business; see May and Arnold (date, Section 4.4).

[13] As discussed by May and Arnoldus (date, Section 5.1) the labelling laws are more stringent in the European Union than in the United States with the result that only Madagascar produces vanilla that meets the standards of 1.6 percent vanillin per single fold extract from 100 grams of vanilla extract per 1 liter of extract. As a result, European food producers and flavor houses have had to rely on Malagasy vanilla. Branding the beans to distinguish the grower provides identification that survives the curing stage.

[14] See the discussion in Cadot et al (date). During this phase a vanilla stabilization fund (VSF) was established with a licensing committee overseeing export trade with, at all stages of the process, prices set by a cost-plus formula and the VSF committed to buying the stock at the prevailing price. Curers, packers, and exporters had to obtain a license to operate.

Table 5.1: *Cashews and Vanilla: Background and Reforms*

	Cashews	Vanilla
Background Production	1 million small farmers Stable production, High sunk costs: Trees start producing after 5 years for a productive life of 25 years Labor costs: 50% of time caring for trees prior to harvest and rest harvesting	100,000 small farmers proprietors (2ha size) Highly unstable (c.v. of production of 15% (41%) for 1961–1979 (1979–1990). Low perishability, low substituability across producers. High sunk costs: Trees take 3 years; productive life of 8 years Labor Costs: hand-pollination; no inputs
Regulatory environment	High regulation during and after the Portugese left. Trading and processing under State control; producer prices and trading margins set by government	Prices and margins set by government and licenses required for all stages (curing, packing and exporting) until phase III (see below)
Internal Market Structure	Three layers (small rural traders/retailers co-existing with large traders/wholesalers), processing factories; exporters (for raw and processed cashews)	Two layers: Curing (gives commercial value), and packing (sorting, grading and storing).
International Market Structure	All production exported, raw (5%) or processed (10%) – world market shares in parenthesis India is single buyer of raw cashews with less concentration on the export side; similar levels of concentration on both sides of market for processed cashews	40–60% world market share with few competitors (Indonesia, PNG, Uganda). Protection by labelling laws. Few buyers from the food-processing and fragrance industry.
Policy reforms	1) Export ban on raw cashews lifted in 1992, but 60% export tax and QR of 10,000 tons. 2) In 1995, as part of conditionality, liberalization of cashew marketing (export tax on raw cashews reduced to 14%)to be followed by privatization of processing industry and elimination of rationing of export licenses (but actual sequencing was the opposite leading to strong protests by the industry and reinstatement of the export tax to a level between 18% and 22%). Privatization brought 10% of expected revenues.	Phase I (60-75): Successful vanilla stabilization fund with tri-partite process involving growers, packers-stockers and GOM Phase II (75-93): Replacement by centralised and politically motivated decisions with rent-seeking, inefficicency and extortionary taxation Phase III (93-): GOM abandons the VSF and intervention in the sector except for sanitary/quality inspections and setting date for marketing

under phase I. Trees were largely abandoned, the quality had fallen and rents were dwindling for the Indian Ocean cartel. Putting dynamic considerations aside, Melo et al. (2000) estimate that the export price set by the VSF was about twice the static welfare-maximizing level, and a third above the revenue-maximizing level, implying welfare losses adding to about 1 percent of GDP.[15]

As shown in Table 5.1, the reforms boiled down to a withdrawal of the government except for sanitary and (or) quality inspections and the important role of announcing the date when the harvest can start, to prevent competition among traders to pick vanilla before maturity when vanillin content is low.

For cashews, Mozambique had been the world leader until independence in 1974, and the first African country to process cashews on an industrial scale under the Portugese. In 1974, 150,000 tons of raw cashews were processed, with the decline starting at independence ending at 8,000 tons in 1992 at the end of the 10-year civil war. Even when it was successful, cashew production was highly regulated, much like vanilla, with prices and marketing margins regulated throughout the chain. This continued after independence, with raw cashew exports banned, starting in 1978. Following initial reforms in 1991–2, as part of the adjustment lending operations under the World Bank, conditionality in the reforms involved removing or lowering the export tax on raw cashews in a first step, to be later followed by privatization of large processing plants that were already in deep trouble. As documented in McMillan et al. (2003), privatization took place first (at a much lower price than expected) and was followed by the reduction in export tax. This sequencing of reforms was vehemently opposed by processing plant owners who then managed to get the export tax raised again (see Table 5.1). Unemployment in the industrial processing plants reached 10,000 or 90 percent of the work force, the flambeau for this cause célèbre espoused by the anti-globalization movement.

4. MEASURING THE IMPACT OF THE REFORMS

Lack of adequate household data hampered the analysis of the reforms for both case studies making it difficult to come up with a reasonably accurate diagnosis of the outcome with both papers relying on suggestive simulations based on anecdotal evidence about the increase in the number of traders in the processing chain and data suggestive of increasing margins for farmers. To begin with neither case study had access to reliable price and quantity data to assess credibly the effects of the reforms.[16] In the case of cashews, no household surveys were

[15] The simulations come from an econometrically estimated Stackelberg model with Madagascar the leader and Indonesia the follower, with vanilla demand competing with vanillin substitutes. Demand elasticities were estimated quite precisely, but not supply elasticities for either Madagascar or Indonesia.

[16] In the case of vanilla, discerning the effects of the reforms was further complicated by several exogenous events: a major hurricane in April 2000 in Madagascar; a contested Presidential election in late 2001 that brought the country to a standstill until July 2002; and large-scale fires in Indonesia in 1997 which destroyed a large chunk of the world's supply of natural vanilla.

available, and the trade statistics were questionable (no exports of raw cashew by Mozambique in spite of an estimated world market share of around 10 percent). With lack of firm-level data that would have allowed measurement of efficiency, there was disagreement in the diagnosis of the causes of failure in the cashew processing industry (see McMillan et al, 2003, Section 7).

4.1 Prices, margins and supply response.

In the case of vanilla, four household surveys for 1993, 1997, 1999, and 2001 are available. Unfortunately, although they straddled the reforms with around 500 vanilla-growing households in each survey, they are far from comparable both in terms of data collected and sampling methods. Inspection of the household data does not reveal clear-cut changes in inequality and poverty indices in the vanilla-growing regions. Attempts at estimating supply response from the household survey data met with mixed results.[17] In the case of cashews, no household data were available.

In the end, both case studies relied on government FAO price data and on COMTRADE data, both of questionable quality, which produced, at best, suggestive estimates. For example, in the case of vanilla, there was a large discrepancy in the price series even though the farm-gate prices show the same increasing trend across the series.[18] Nonetheless, it is clear from Figure 5.1 that the producer price share (of the FOB price) was falling for both crops through time indicating rent-extraction from the farmers. In both cases, the producer share of the FOB price rose after the reforms with this increase attributable to greater competition among intermediaries (the case studies give anecdotal evidence of increased competition along the production chain). For vanilla, the fraction of vanilla FOB prices retained by producers was from a low of less than 2 percent in 1991 to a high of close to 34 percent in 2004, while for cashews the fraction of the FOB price retained by producers rose from 30 percent to 50 to 60 percent (as expected by the World Bank when it pushed for the reform).[19]

[17] Figure 5 in Cadot et al (2009). on plantation and yields over a 20 year period show no break around the reforms. Using household survey data, they model farmers' decisions in a two-stage Heckman process with the decision to grow vanilla identified using location and community characteristics and expected profits based on past prices. The model was estimated for 500 households over the two most comparable surveys (1997 and 1999). First stage results are plausible and show a positive supply response to the variable used to capture expected profits. But the aggregate statistics on vanilla production show no increase in supply (there were cyclonic conditions on several occasions) and a report on the Sava region (Monographie de la région de la Sava, GOM, 2003) states that the increase in vanilla-planted area following the elimination of the marketing board corresponds to a better care given to vanilla sprouts and to the renewal of old sprouts rather than to a physical increase in plantation, even though some planting reportedly took place in 2004 when prices rose dramatically. Lack of complementary data at the community level and on the determinants of land allocation to food and vanilla crops precluded trying to disentangle the determinants of the decision to produce vanilla as in e.g. Balat et al.(2008) did for an aggregate of exports crops.

[18] According to the household surveys, farm-gate prices increase over 8-fold in real terms over the period 1993-2001 while the corresponding increase in the FAO series is less than fourfold, and the government figures are flat over the available period (1993-1997).

[19] The abrupt reversal in the squeeze on the intermediation margin for vanilla around 2001 is puzzling (see Cadot et al. (2009) for possible explanations).

Figure 5.1: *Producer Price and Intermediation Margin*

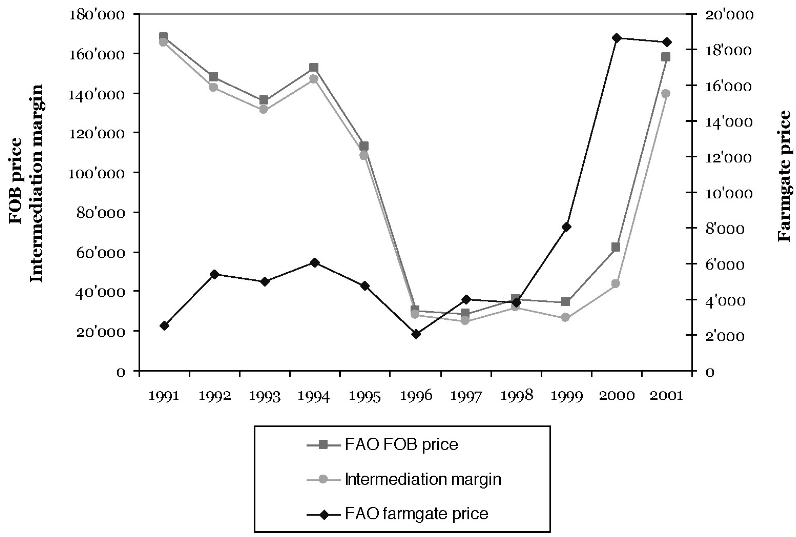

Figure 5.1a: *Vanilla Producer-price (right scale) and FOB and intermediation margins (left-scale)*

Source: Cadot et al. (2009, Table 4). Prices and deflated.

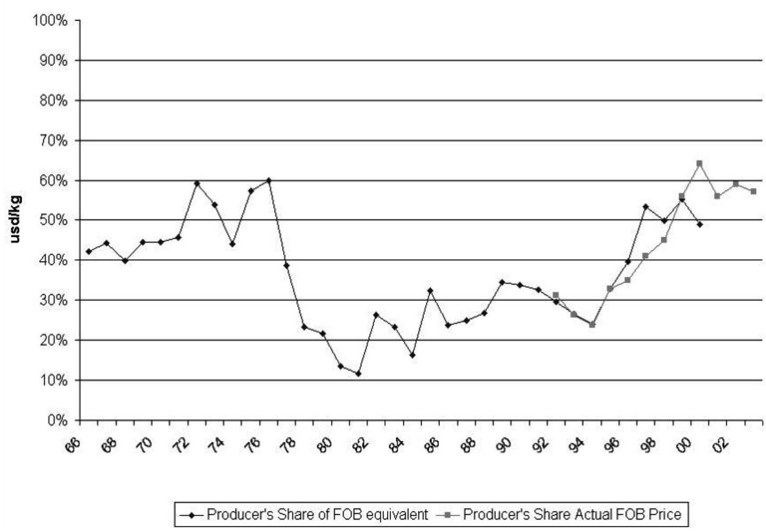

Figure 5.1b: *Cashew producer share of world price*

Source: McMillan et al. (2004, Figure 3)

A last outcome of the reforms was the boom-bust cycle in the vanilla market. Such cycles, common to tree-crops, had occurred before but this one summarized in Figure 5.2 was particularly strong and reflected both the inherent instability in the market (cyclonic conditions, fires in Indonesia), the asymmetric information at several links along the value chain, and also the consolidation in the industry that was getting increasingly concentrated during the period (May and Arnoldus (2009) provide the details). The result was opportunistic behavior all along the value chain, with an end result being a shift away from natural vanilla in the food processing industry wherever possible; natural vanilla content was not protected by required labeling. As put by Ecott (2004, 238):

> "...in Madagascar, Mexico, India and Indonesia the farmers spread rumors about the amount of vanilla they think will be available during the next season. They hope to frighten the curers into offering them a higher price for green beans. In turn, the curers spread rumors among farmers about how much the foreign buyers are willing to pay for the dried product.... Meanwhile the international brokers spread 'market intelligence' to their customers about how much vanilla may or may not be on the market during the forthcoming year."

And, as he got to know the major dealers, Ecott (2004) reports (about the rising prices in 2004) that the major dealers were accusing each other of predatory behavior to drive out the competition.

The overall impression is one of widespread opportunistic behavior in a highly concentrated market. In relation to the denouement of the reforms, market failures which had been successfully addressed in phase I were replaced by government failures in phase II, but were back at center stage by the end of the reforms.

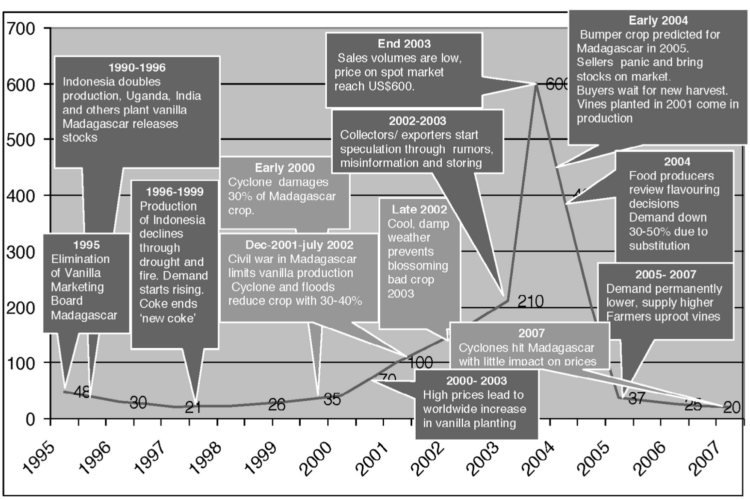

Figure 5.2: *The Vanilla Boom and Bust 1990–2007*

Source: Adapted from May and Arnold (2009, Figure 8)

4.2 Counterfactuals

In the end, with little supply response for both crops, and sustained instability in the natural vanilla market, were there any benefits for the growers from these reforms? Using plausible (but not estimated) supply elasticities, both studies used counterfactual simulations to try to squeeze orders of magnitudes from the stylized facts indicating higher farm-gate prices for both cashew and vanilla producers. In both studies, Cournot competition was assumed among the intermediaries with the number of intermediaries calibrated to the plugged-in intermediation margin (between producer and FOB export prices).

In the cashew case study, there are three layers of intermediaries: small traders, large traders, and processors–exporters, each layer determining price in Cournot–Nash fashion and the inverse of the intermediation margin θ is given by:

$$\theta = \frac{p^P}{P^*(1-t)} = \left(\frac{n_1\varepsilon}{1+n_1\varepsilon}\right)\left(\frac{n_2\varepsilon}{1+n_2\varepsilon}\right)\left(\frac{n_3\varepsilon}{1+n_3\varepsilon}\right) \tag{1}$$

P^P is the producer price and P^* is the (exogenous) world price. It is clear from Equation 1 that the share of the producer price in the world price is an increasing function of the degree of competition at each stage captured by the number of traders at each layer. In this set-up, the deep discount suffered by farmers comes from the multi-tiered nature of the market. Any measure that increases competition along the value chain, such as improving infrastructure, so that each farmer has access to more traders, is helpful for the farmers. Using a supply elasticity of ε=0.25, and actual margins at each stage, McMillan et al (2003) estimate θ=48 percent.

In the Mozambique study, the rising price paid by processors as the export tax is lowered for raw cashew exports creates unemployment (which is costly to society under their assumption that the opportunity cost of workers is zero, which they justify because of the reported unemployment in the range 20 to 50 percent).[20] When this unemployment estimate is subtracted from the gains to farmers, they obtain an overall welfare gain that is negligible.

In the Madagascar study, the counterfactual involved reducing the number of traders to one, resuscitating the marketing board, and re-imposing the taxation at the pre-reform maximum rate of 82 percent during phase II. The set-up is one in which two countries operate where Madagascar competes with Indonesia in the world market for vanilla. As in the cashew model, the share of the world price retained by Madagascar's producers' θ^M is a function of the number of local intermediaries n, and of demand and supply elasticities, yielding a similar profit-maximizing outcome to the one in the case-study:

[20] Mozambique has been growing at 8 percent per year for the past 10 years, so this assumption overestimates the welfare loss from unemployment. A more appropriate assumption would have been to subtract revenue losses evaluated at a fraction of the wage in the processing industry for a period of a few (perhaps five) years from the overall gains to account for these adjustment costs.

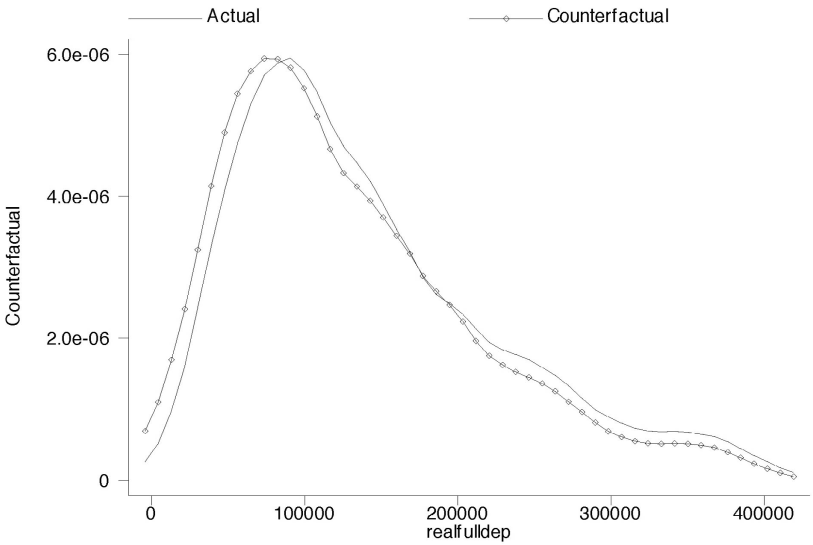

Figure 5.3: Kernel density estimates of income distribution, vanilla region

Source: Cadot et al. (2009) Figure 6.

$$\theta_M = \frac{p^P}{P^*} = \frac{1-\left\{\left(1/\sigma\right)+\left[1-\left(1/\sigma\right)\right]s\right\}}{1+\left[1/\left(n\varepsilon\right)\right]}. \quad (2)$$

The simulation gives new values for prices and quantities of vanilla that were then fed into the household survey data. These simulations, which provide an upper-bound of the effects of the reforms, are shown in Figure 5.3.

In spite of large changes in prices (depending on the values assumed for the elasticities, the simulated margin under the VSF set-up with an 82 percent export tax lies between 2 percent and 11 percent of the FOB price, compared with the actual estimate of 22 percent), effects on the distribution of income are negligible. This is so because most of the effective consumption of Malagasy rural households is self-produced. Cash income represents at most 50 percent of income for the richest decile. Under the central assumption about elasticities, the estimated change in the poverty headcount is that about 20,000 households (representing around a quarter of vanilla-growing households) were lifted out of poverty as a result of the 1995 reforms.

5. DISCUSSION

The two case studies illustrate the difficulties of extracting information on the effects of trade reforms on the basis of limited data but, more importantly, when delving into the characteristics of the markets and of the policies adopted, one

finds clues for the observed outcomes. For vanilla, the characteristics of the vanilla market (highly variable due to unstable climatic conditions), and of vanilla preparation, suggest sufficient externalities and market failures (for example, asymmetric quality of information and externalities in marketing) to justify intervention of the type that was initially set up. So if opportunistic behavior could be controlled, as it was during the early phase I period, cooperation among agents involved in the value chain would help to overcome market failures, to stabilize prices for the producers, and to exploit its (increasingly limited) monopoly power on high-quality (Bourbon) vanilla. In the case of cashews, it is the lack of credibility of government policies around the conditionality for World Bank assistance, together with a wrong sequencing in the reforms that contributed to the disappointing outcome.[21]

For both cashews and vanilla, the low supply response to the reforms is attributable to several factors. First, in both cases, the income effects of any price increase were small because the households were not specialized in the cash crop: in effect the families engaged in the cultivation of these crops are very poor and close to subsistence income (in the case of vanilla, the household surveys show very low annual per capita expenditures of around $30). Second, in both cases high sunk costs dampened any expected supply response from an increase in farm-gate prices, and high sunk-costs required credibility on the part of government policy. This credibility was largely lacking because of the predatory behavior encouraged by the natural resource base. As emphasized by McMillan (2001), high sunk costs involved in tree-crop production imply that farmers have to incur labor costs prior to knowing the price they will get at harvest in markets, which are, furthermore, controlled by a few intermediaries. Farmers can only hope to recoup some of the costs by harvesting, even if the price received does not cover the cost of maintaining the trees or pollinating the vines. For both crops, the sunk costs associated with planting new trees require a credible pricing policy.

Irreversible investments on the part of governments, like improving roads and transport, would help make these crops more profitable leading to greater specialization in export crops, which in turn would likely be associated with less poverty (see the evidence in Barlat et al. 2008 for Ugandan export crops). One might also expect that the reduction in the market power of intermediaries that followed the trade reforms in cashews and in vanilla would have been amplified by complementary investments that would have reduced transaction costs.

Jaime de Melo is a Professor at the University of Geneva and a CEPR Research Fellow.

[21] Sequencing is important. For example, when Mongolia abandoned central planning where the herd was owned and their number controlled by the state, and when pastures were part of the 'commons', the herd was initially privatized without accompanying privatization of the land.

BIBLIOGRAPHY

Acemoglu, D S Johnson, and J Robinson (2001). The Colonial Origins of Comparative Development: An Empirical Investigation, *American Economic Review* 91(5): 1369–401

Amin, M and S Djankov (2009). Natural Resources and Reforms, *CEPR DP #* 7229

Auty, R (2001). The Political Economy of Resource-driven Growth, *European Economic Review*, 45, 939–46

Baffes, J (2005). The Cotton Problem, *The World Bank Research Observer, 20* (1), 109–44

Barlat, J, I Brambilla and G Porto (2008). Realizing the Gains from Trade: Export Crops, Marketing Costs and Poverty, *World Bank Working Paper #* 4488

Blarel, B, and D Dolinsky D (1995). Market Imperfections and Government Failures: the Vanilla Sector in Madagascar, In S Jaffe and J Morton eds., *Marketing Africa's High-value Foods; Comparative Experiences of an Emergent Private Sector* (255–318). Dubuque, IA: Kendall/Hunt Publishing Company

Bohman, M, Jarvis, L, and R Barichello (1996). Rent-Seeking and International Commodity Agreements: The Case of Coffee. *Economic Development and Cultural Change*, *44* (2), 379–404

Brunnschweiler, C and E H Bulte (2008a). Linking Natural Resources to Slow Growth and More Conflict, *Science*, 320(2), 616–7

Brunnschweiler, C and E H Bulte (2008b). The Resource Curse Revisited: A Tale of Paradoxes and Red-Herrings, *Journal of Environmental Economics and Management*, 55(3), 248–64

Cadot, O, Dutoit, L and J de Melo (2009). The Elimination of Madagascar's Vanilla Marketing Board, 10 Years On, *Journal of African Economics*, 18(3), 388–430

Collier, P, and A Hoeffler (2004). Greed and Grievance in Civil War, *Oxford Economic Papers*, 56, 563–95

Collier, P and A Hoeffler (2009). Testing the Neocon Agenda: Democracy in Resource Rich Countries, *European Economic Review*, 53(3), 293–308

Easterly, W E (2005). National Economic Policies and Economic Growth: A Reappraisal, chp. 15 , in P Aghion and S Durlauf eds. *Handbook of Economic Growth*, North-Holland,

Ecott, T (2004), *Vanilla: Travels in Search of the Ice Cream Orchid*. Grove Press, New York

Hodler, R (2006). The Curse of Natural Resources in Fractionalized Countries, *European Economic Review*, 50, 1367–86

Isham, J, Woolcock, M, Pritchett, L, and G Busby, (2005). The Varieties of Resource Experience: Natural Resource Export Structures and the Political Economy of Economic Growth, *World Bank Economic Review, 19* (2), 141–74

Jaffe, S, and J Morton (1995). *Marketing Africa's High-value Foods; Comparative Experiences of an Emergent Private Sector.* Dubuque, IA: Kendall/Hunt Publishing Company

Krueger, A, Schiff, M, and A Valdes (1988). Agricultural Incentives in Developing Countries: Measuring the Effect of Sectoral and Economywide Policies, *World Bank Economic Review, 2* (3), 255–71

May, J and M Arnoldus (2009). *The Vanilla Chain 1997–2007*, mimeo, University of Kwazulu-Natal

McMillan, M (2001). Why Kill the Golden Goose? A Political Economy Model of Export Taxation, *The Review of Economics and Statistics*, *83* (1), 170–84

McMillan, M, Horn K, and D Rodrik (2003). When Economic Reform Goes Wrong: Cashews in Mozambique, *Brookings Trade Forum*, 97–151

Melo, J de, Olarreaga, M, and W Takacs (2000). Pricing Policy under Double Market Power: Madagascar and the International Vanilla Market, *Review of Development Economics, 4* (1), 1–20

Pritchett, L (2005). *Reform is Like a Box of Chocolates: An Interpretative Essay on Understanding the Pleasant and Unpleasant Surprises in Policy Reform*, mimeo, Harvard Kennedy School

Sachs, J and A Warner (1997). Natural Resource Abundance and Economic Growth, NBER Working Paper # 5398

Sachs, J and A Warner (2001). The Curse of Natural Resources, *European Economic Review*, 45, 827–38

Schiff, M, and A Valdes (1992). *The Political Economy of Agricultural Pricing Policy: Volume 4, A Synthesis of the Economics in Developing Countries: A World Bank Comparative Study*. Baltimore and London: Johns Hopkins

6

Barriers to Exit from Subsistence Agriculture

OLIVIER CADOT, LAURE DUTOIT AND MARCELO OLARREAGA

1. INTRODUCTION

Even though market-oriented agriculture consistently provides higher incomes, subsistence agriculture remains prevalent in poor countries. What do we know about the determinants of these modes of production and about the barriers to exit from subsistence agriculture? Recent empirical research has adopted a variety of innovative approaches to the measurement of transaction costs of all sorts—variable, fixed, and sunk. Most of the studies reviewed in this note suggest that those transaction costs are formidable. However the evidence is still fragmentary and needs to be extended both conceptually–as regards the linkages between intermediation markets and farmers' incentives to innovate–and empirically–as regards the precise nature of sunk costs involved to enter commercial circuits, for instance, in order to generate useful policy advice.

2. THE PARADOX OF SUBSISTENCE AGRICULTURE

2.1 What is subsistence agriculture?

Conceptually, subsistence agriculture is easy to define, by analogy with autarky—a situation where the farm household neither sells nor buys, but consumes everything it produces and, consequently, only that. Lack of access to inputs under autarky can be expected to constrain production to particular techniques, like long fallow periods to avoid soil depletion and, in most cases, to endure low productivity levels.

However, empirically, things are not that simple. First, the share of output sold on the market and the share of consumption bought from it are both continuous variables on [0,1]. Just along that dimension, where to draw the line between a 'subsistence farm' and a 'market farm' is a matter of judgment. We will discuss in the following section an econometric method addressing the problem of where to draw the line.

Second, when there is a functioning labor market, farm households may supply labor for off-farm employment (on other farms or in nearby towns), gener-

ating cash income. When that income is used to buy agricultural inputs, even though none of the farm's agricultural output is sold, a key analogy with autarky (not being able to buy inputs because no output is sold) is broken. Thus, a proper understanding of what is subsistence agriculture requires the identification of which markets exist and which don't. Indeed, the modern analysis of the farm–household can be traced to the seminal work of de Janvry et al. (1991) on peasant behavior in the absence of output or labor markets.

Finally, some crops are relatively easy to characterize as cash crops or food crops. For instance, a farm growing cocoa, tea, coffee, cotton, or peanuts is unlikely to be predominantly a subsistence one. The converse is true of a farm growing essentially sorghum or millet. So it may prove convenient to focus, as a shortcut, on the nature of the crops grown instead of the (implied) decision to go to the market or not. The advantage of seeing things this way is that the decision of what to grow can be analyzed fairly naturally as a portfolio-allocation problem, characterized in terms of the risk and return characteristics of food versus cash crops. Indeed a number of classic articles, including *inter alia*, Fafchamps (1992) and Rosenzweig and Binswanger (1993), proposed formal analyses of crop choice under uncertainty. However, it should be kept in mind that what is a cash crop in one country can be, for a variety of reasons discussed below, a food crop elsewhere. For instance, rice is a cash crop in Thailand, but it is a food crop in Madagascar.

All in all, there is little doubt that the prevalence of subsistence agriculture is correlated primarily with low income levels (both across countries and across time) and low population density, albeit without a clear direction of causation between the three. The analysis of peasant production relations and implied social structures by Binswanger and McIntire (1987) highlighted the link between the prevalence of subsistence agriculture, the absence of labor markets, and prohibitive per capita infrastructure costs that characterize land-abundant (low-density) dry zones in Africa. Perhaps the clearest exogenous factor in their analysis is the importance of non-diversifiable weather risk in semi-arid areas.

2.2 Missed opportunities

The prevalence of subsistence agriculture is paradoxical if it is associated with lower incomes than farm households could achieve by participating in commercial exchange, unless they face substantial switching costs. Before we get to switching costs, what does the evidence have to say about income differentials?

Prima facie evidence of income differentials between subsistence and commercial farmers is of course likely to be gravely misleading unless controlling for differences in individual characteristics and endogenous selection. Kennedy (1994) partly overcame the problem by looking at the income effect of participation in a Kenyan government sugarcane out-grower program using a two-period panel of farmers surveyed in 1984–85 and 1985–87. Non-participants and 'switchers' (farmers who took on sugarcane cultivation upon joining the program) had similar initial incomes, but the latter saw theirs grow by 96.2 per cent,

against 40.7 per cent for non-sugar farmers and 30.8 per cent for 'always-sugar' farmers. This indeed suggested large gains from adopting the cash crop, but Kennedy's result had anomalies (for example the fact that switchers had final incomes vastly in excess of that of always-sugar growers), did not quite control for selection (the two may be related), and was partly driven by an artificially high subsidized producer price of sugarcane.

In general, the analysis of income differentials between modes of production—subsistence versus market—requires two ingredients. First, there must be careful control for selection on observables (individual characteristics). Second, it is possible that the mode of production affects not only the level of income, but also the return to factors of production. For instance, when the market takes the form of large foreign buyers offering out-grower contracts, an increasingly prevalent mode of integration into commercial agriculture (see World Bank, 2008, Chapter 5), contractual requirements may be easier to understand and satisfy for farmers with some education. The return to education is thus likely to be higher for contract farmers than for subsistence ones, and this should be taken into account when explaining income on the basis of production regime and individual characteristics.

Taking into account differentials in factor returns in addition to income levels means that in samples including commercial and subsistence farmers, income equations should have the following form. Let *i* index farmers, and let y_{i1} and y_{i2} be income under the market and subsistence regimes respectively. The income equations are

$$y_{i1} = \mathbf{X}_i \beta_1 + u_{i1} \tag{1}$$

$$y_{i2} = \mathbf{X}_i \beta_2 + u_{i2} \tag{2}$$

where X*i* is a vector of individual characteristics. However, the econometrician can never observe both (1) and (2) at the same time. Observed income is equal to either y_{i1} and y_{i2}, depending on the value of a switch variable I_i:

$$y_i = \begin{cases} y_{i1} & \text{if } I_i = 1 \\ y_{i2} & \text{if } I_i = 0. \end{cases} \tag{3}$$

The value *I*, of is itself determined by individual characteristics through a selection equation of the form

$$\Pr\left(I_i = 1\right) = f\left(\mathbf{X}_i ; \mathbf{Z}_i\right). \tag{4}$$

The reader will have recognized a so-called 'switching-regression' problem whose logical structure is close to that of Heckman's selection model, but with an important difference. Namely, here the income of individuals not participating in the market can be considered as observed if self-consumed output is valued at market prices (this may be a trickier assumption than it looks, though—more on this below).

The appropriate estimation technique depends on two aspects of the problem. The first is whether the switch point between regime 1 and regime 2 is observed or not. As we noted earlier, how much a farmer sells is a continuum and it may be hazardous to set an arbitrary switch point. The econometrician may instead want to 'let the data speak' and determine the switch point simultaneously with the model's other parameters. The second aspect is the scope for reverse causality, which is of course unavoidable, as income differentials should be the main driver of selection. Together, these two features of the problem (unknown switch point and endogenous selection) call for a particular ML estimation technique inspired of Heckman's selection model.[1] Consistent estimation of the parameters in (1) and (2) makes it possible to calculate an individual's predicted income in both regimes, including the unobserved one, and hence to estimate the income differential conditional on individual covariates. Applying this technique to farmers surveyed in Madagascar's *Enquête Permanente des Ménages*, Cadot et al. (2006) found a switch-point at zero market sales, which defined subsistence farmers as those that were in true autarky (10 per cent of the sample), and an average income loss of 43 per cent for those farmers, conditional on covariates and controlling for endogenous selection.

Thus, although parametric evidence is fragmentary, it is suggestive of very substantial income differentials after controlling for individual effects, begging the question, what prevents subsistence farmers from exploiting profitable market opportunities? Clearly, if subsistence farmers forsake substantial income by not going to the market, or not producing what the market would buy, they must face formidable barriers to 'going commercial'. What are those barriers?

3. BARRIERS TO EXIT

3.1 Risk

In the absence of properly functioning insurance mechanisms, food self-sufficiency can be seen by farmers as insurance if cash crops are perceived as inherently riskier than food crops. This conjecture is perhaps the oldest in the analysis of subsistence agriculture.

When income-generating production is risky, farmers can, using the terminology of Alderman and Paxson (1992), adopt either (or both) risk-management strategies—for example, diversifying crops whenever possible to reduce income risk—or risk-coping ones—for example, saving in order to reduce the transmission of income risk to consumption. Risk-management strategies have been extensively studied in the literature: see for example, Shahabudin, (1982), Binswanger and Sillers (1983), and Fafchamps (1992) to name but a few. Unsurprisingly, a running theme of that literature is that price uncertainty on cash-crop markets raises the weight of food crops in the optimal allocation of land, relative to what a comparison of returns would suggest.

[1] Dutoit (2006) provides a through survey of switching-regression techniques, together with Stata applications.

Dercon (1996) argued that reliance on risk management should be a decreasing function of a farmer's stock of liquid assets like cattle, since selling them could be used to smooth consumption in periods of negative income shocks. This suggested an obvious identification strategy: regress the share of food crops on the stock of cattle at the farm level—in addition to other individual characteristics, of course. This is what Dercon (ibid.) did on a sample of Tanzanian farmers for whom growing drought-resistant sweet potatoes for food was a low-risk, low-return strategy. When the cattle stock was made endogenous using an asset-accumulation equation, the partial correlation between the share of food crops and the stock of assets was, as expected, negative.[2]

Dercon's and other studies provided empirical support to the view that overreliance on food crops could be understood if one looked not just at the first moment of the distribution of returns, but also at its second moment. The culprit was then the absence of more efficient insurance mechanisms, something that has led to widespread, and largely failed, policy experiments in price stabilization. However, Jayne (1994) observed that nine smallholders out of 10 grow food crops in semi-arid areas of Africa where cash crops are actually more resistant to local conditions. Thus, something else than just risk management must be at play.

3.2 Missing markets

The analysis of peasant behavior when some markets are missing goes back to the seminal work of de Janvry et al. (1991), whose objective was to rebut old claims that peasants are irrational. They showed that feeble supply responses to price signals (see de Melo's contribution in this volume for a recent version of that observation) reflect the dampening effect of induced variations in the shadow price of non-traded goods—either labor or one of the farm's crops.

Missing markets are a limiting case. Between deep and perfectly liquid markets and no markets at all, there is a range of situations characterized by various levels of variable, fixed, and sunk costs of transacting. All three can explain why some potential participants are excluded or why, in extreme cases, nobody participates. Thus, if we want to understand why markets fail, we need to understand which transaction costs are prohibitive, for whom, and why. Of course, transaction costs are rarely observed in practice, so a number of ingenuous empirical strategies have been devised to get, indirectly, a hold on them.

When transaction costs prevent or limit arbitrage, the lack of market integration typically shows up as limited co-movement of prices, and the study of price co-movement, by various techniques, has been a prime vehicle to test for market integration. Recently, Moser et al. (2005) argued that those tests may be of limited validity in the presence of seasonally reversing flows across markets, and they proposed an alternative approach with a typology of situations. When all arbitrage opportunities are taken, price differentials across markets should just equal

[2] Dercon (date) estimated the model recursively, with cattle accumulation not a function of crop choice, even though an argument could be made for making it a full simultaneous model if the share of land allocated to low-risk, low-return food crops affects the pace of cattle accumulation.

the cost of transportation between those markets. When transportation costs are higher than price differentials, arbitrage is prevented. When they are smaller, something else must be preventing arbitrage, like for example, anticompetitive practices. They dub these three regimes 1, 2, and 3 respectively. On a sample of 1,400 communes in Madagascar, they find about two-thirds of them integrated locally (that is, for which the probability of being in regime 1 is highest), but also two-thirds of them in regime 3 vis-à-vis regional cities, suggesting substantial barriers to competition.[3]

Large transaction costs not only prevent market integration, they can also feed back onto crop choices. The argument, due to Jayne (1994), goes like this. Suppose that a large transaction cost τ creates a wedge between the farm-gate price p of a food crop and its buying price $p + \tau$ (say, from neighbors or local dealers if there is a village market). Suppose that one hectare of land produces one unit of the food crop. For a farm that is more than self-sufficient in food grain, the opportunity cost of planting one hectare with a cash crop is p (what it would get by producing the food crop on that hectare); for a farm that is less than self-sufficient, it is $p + \tau$ (the cost of procuring the food crop). This discontinuity means that grain-deficit households may not find it profitable to diversify into cash crops when grain-surplus households do. Indeed, this is what Jayne finds on the basis of prices observed in a 1990 survey of 276 Zimbabwean farmers. Parametric evidence, however, is not so clear-cut, as there is no statistically significant jump in the area planted with cash crops at the point where households reach self-sufficiency.

Consider now fixed transaction costs (that is, independent of volumes transacted). These may include the cost of searching for partners, of enforcing contracts with distant buyers, of establishing quality, and so on. Vakis et al. (2003) provide interesting survey evidence from Peru on these costs as perceived by the farmers themselves (more on this below). Whereas variable transaction costs are supply or demand shifters, fixed costs make supply and demand curves discontinuous, calling for particular estimation techniques. Renkow et al. (2004) estimated simultaneously, by maximum likelihood, a system of three equations looking roughly like this:

$$x_i^s = \begin{cases} \alpha_0 + \alpha_1 p_i + \boldsymbol{\alpha}_2 \mathbf{x}_i + \delta_v + u_{1i} & \text{if } p_i > p_i^A + \tau_i \\ 0 & \text{otherwise} \end{cases}, \tag{5}$$

for supply (where is p_i^A farmer i's autarky price, p_i is the price he receives in the market, x is a vector of individual demand and supply shifters, δ_v are village effects, and τ_i is the *ad valorem* equivalent (AVE) of the fixed transaction cost farmer i faces, assumed to be symmetric between selling and buying);

[3] These are barriers to entry on city markets. Madagascar's informal trucking cartel biases results in favour of regime 2 (high transportation costs) rather than regime 3.

$$x_i^d = \begin{cases} \beta_0 + \beta_1 p_i + \boldsymbol{\beta}_2 \mathbf{x}_i + \delta_v + u_{2i} & \text{if } p_i < p_i^A - \tau_i \\ 0 & \text{otherwise} \end{cases}, \tag{6}$$

for demand, and

$$\ln(1+\tau_i) = \ln\left[1 + e^{\bar{\tau}}\right] + u_{3i}, \tag{7}$$

where τ is the common component of transaction costs (that is, expected transaction costs are identical across farmers). Estimating (5) to (7) on a cross-section of 324 maize-producing farmers in Kenya, Renkow et al. (date) obtain a surprisingly low 15 per cent for the AVE of fixed transaction costs.

Vakis et al. (2003) take a different route and propose an interesting approach where transaction costs are retrieved from the farmers' choice of where to sell. They use a 2001 cross-section survey of small Peruvian farmers, in which 1,096 potato transactions are observed individually on five markets. The problem is to estimate simultaneously a price equation (the price effectively received by a farmer on transaction *i* in market *k*), a transaction-costs equation (also on transaction *i* in market *k*) and a market-choice equation. The price equation is

$$p_{ij} = \mathbf{P}_{ij}\gamma + u_{ij1}, \tag{8}$$

where P_{ij} includes determinants of the price received on transaction *i* in market *k*: the price level on market *j*, the volume sold (which must of course be instrumented), and a vector of farm characteristics. The transaction-cost equation is

$$t_{ij} = \mathbf{T}_{ij}\delta + u_{ij2}, \tag{9}$$

where T_{ij} includes determinants of transaction costs on market *j*, including distance etc., and the market-choice equation is

$$j_i = \arg\max_k \pi_{ik}^*, \tag{10}$$

where profit is given by

$$\pi_{ik}^* = \mathbf{X}_{ik}\beta + q_i\beta_k + u_{ik3}, \tag{11}$$

and

$$\mathbf{X}_{ik} = \left[p_{ik} - t_{ik}; z_{ik}^f\right]. \tag{12}$$

In (12), z_{ik}^f is a proxy for the fixed costs of transaction *i* on market *k*, for which Vakis et al.(date) use the percentage of a village's farmers stating that they know prices in market *k*. The first argument in (12) is the source of problems, because prices p_{ik} and transaction costs t_{ik} can be estimated only for transactions that take place; not for off-equilibrium transactions. The solution is, once more, a Heckman-type two-step approach that goes roughly like this.

First, a version of (10) is estimated with exogenous market prices p_k used in lieu of transaction-specific prices p_{ik}. This yields predicted choices and the estimated Mills ratio $\hat{\lambda}_{ij}$. The latter is introduced into second-stage price and transaction-cost regressions (8) and (9), giving predicted prices and transaction costs $\hat{p}_{ij}$ and $\hat{t}_{ij}$. Those are used to generate an estimate of X_{ij}:

$$\hat{\mathbf{X}}_{ik} = \left[\hat{p}_{ik} - \hat{t}_{ik}; z_{ik}^{f} \right]. \tag{13}$$

The last step consists of re-estimating (10) and (11) using (13) instead of (12). The price equivalent of the fixed transaction costs can be taken as the ratio of the coefficients on the first and second arguments in (13), since their ratio gives the marginal rate of substitution between net prices and fixed costs along a constant probability of choosing market *k*. The result is a whopping 77 per cent of the average sales price for the fixed transaction cost, against about 15 to 30 per cent for the transportation cost. Clearly, fixed costs of that magnitude have very different implications from the 15 per cent of Renkow et al (date).

As for sunk costs, Cadot et al. (2006) used their estimate of earnings differentials to generate an estimate of the sunk cost of leaving autarky. The story is illustrated in Figure 6.1, where the horizontal axis measures a farmer's individual trait, say education, and the vertical one measures lifetime earnings. The $V^S(e)$ curve represents the present value of earnings when the farmer is currently under subsistence (as determined by the switching-regression algorithm described earlier in this paper), and $V^M(e)$ the same thing when he is in 'on the market'.

Notionally, farmers should be in subsistence only up to e_1, where the two curves cross. But they are observed to be in subsistence up to e_2. At that point, $V^M(e)$ is above $V^S(e)$ by an amount C. This is the revealed cost of switching from subsistence to market farming (say, through the introduction of a new crop on the farmer's tract of land), taken as a once-and-for-all sunk cost since it is calculated

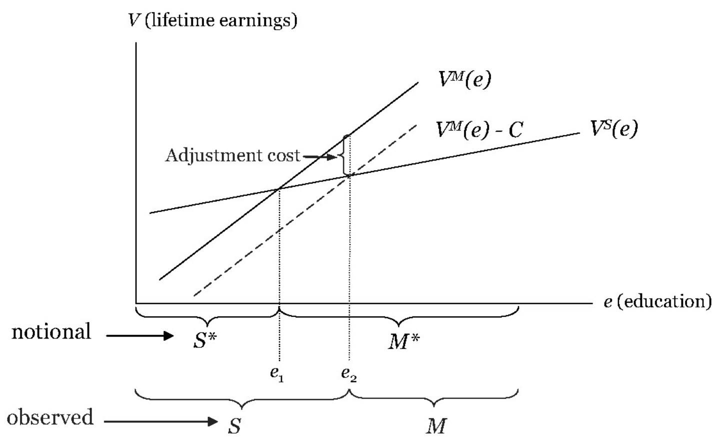

Figure 6.1

from a comparison of lifetime earnings.[4] Cadot et al. (2006) estimate this cost at between 124 per cent and 153 per cent of annual output (valued at market prices). This is a formidable barrier, although the low level of the estimated share of households in subsistence means that the aggregate value of the switching cost is very small relative to GDP.[5]

Thus all in all, the empirical evidence, while still scant, is suggestive of very substantial transaction costs, especially if one thinks of adding up the disparate estimates of variable, fixed, and sunk costs (although adding up figures obtained from different estimation techniques would be hazardous).

4. CONCLUSIONS AND POLICY IMPLICATIONS

4.1 Infrastructure

The variable (per transaction) component of transaction costs is obviously linked to transportation costs. The need to improve rural roads is a cliché in development policy, but it is nevertheless true. Jacoby (2000) found a low elasticity of land prices—taken as the present value of agricultural rents—to distance (about 0.2), but he also found that the distributional effect of road investments is progressive, as remote farmers are typically the poorest. Incidentally, reducing transportation costs does not mean only paving roads, which is sometimes the quick-fix approach for governments that do not want to tackle governance or policy issues seriously. Transportation costs are artificially inflated by informal cartels (as in Madagascar), cartels blessed by regulation (as in West Africa), or irregular payments at roadblocks (as in most of Africa).

The work summarized above has also highlighted the importance of fixed transaction costs; in particular, judging from the results of the Peruvian survey used by Vakis et al. (2003), costs related to search, matching, and bargaining. Those are typically high in the countryside, but the 2008 World Development Report (World Bank, 2008, Chapter 5) suggests a number of initiatives to improve the spread of agricultural information via radios, mobile phones, and other media. If fixed transaction costs are as high as suggested by the estimates, this is a large source of reduction in the barriers preventing farmers from taking up market opportunities.

Large estimated sunk costs of exiting subsistence agriculture are, so far, largely a black box. Although the existence of substantial sunk costs in agriculture has not been questioned since the work of Eswaran and Kotwal (1986), we don't know much about what those costs are; direct, survey-based evidence would be useful to inform policy in this area. It is worth noting again that the estimation exercise on Malagasy farmers suggested that the number of farmers in need of ad-

[4] The figure illustrates the case of a single covariate (education). With many, the technique consists of taking the subsistence farm with the highest propensity score.

[5] Note that the estimation technique can detect only one switch point at a time. It could possibly be repeated in each subsample (say, by distinguishing farmers who sell only at the farm gate from those who sell in more distant markets), generating evidence of further switching costs.

justment assistance to get out of subsistence was small, implying that the level of adjustment assistance required would be modest. It would also be interesting to know whether out-grower contracts with large Northern buyers (for example, supermarkets) reduce the share of those costs borne by farmers.

4.2 Intermediation markets

Minten et al. (2007) analyzed the experience of out-grower contracts for producing French beans for export in Madagascar and showed that farmers who joined the contracts changed their production methods not only for French beans but for other crops as well, in particular rice—which is, as already noted, a food crop in Madagascar. For instance, they resorted to more consistent use of fertilizers and manure. As a result, their productivity rose not just in the part of their plot devoted to the contract crop, but on all their land. This interesting result suggests three remarks. First, it reinforces the point that market-oriented farming generates substantial benefits; here, it leverages complementarities in knowledge. Second, it shows that incentives to innovate, and hence to become capable of switching from subsistence to market agriculture, may come, at least in certain cases, from the buyer side, highlighting the importance of intermediation markets. Third, it brings welcome nuance to the view that food standards are always and everywhere a barrier to trade. Here, tight standards combined with buyer assistance actually improved productivity and hence the ability of farmers to sell to any buyer. However, this comes with a caveat. Maertens and Swinnen (2006) also showed, in the case of Senegal, that out-grower contracts with smallholders progressively gave way to procurement from large plantations. Smallholders were then increasingly driven into those plantations as laborers. The impact effect was a reduction in poverty, but such a fundamental change in the organization of production may prove, in the long run, to have far-reaching—and possibly unwanted—sociopolitical implications.

In other cases, like the Kenyan program studied by Kennedy (1994), government purchases on fixed terms may also provide incentives—albeit artificial when prices are subsidized—for farmers to switch from food to cash crops (see also Goetz, 1993). But the experience with price stabilization funds has been dismal throughout Africa, and so has been the experience with export monopolies acting as sole buyers. By and large, there is nothing to regret from Africa moving away from State buying, even though the experience with privatization has itself been uneven. Brambilla and Porto (2006) showed that when Zambia's cotton export monopoly was privatized in 1994, entry led to a period of failure as farmers would, for instance, seek credit from one intermediary, sell to another, and default on the loans. The market disorganization that followed and lasted until about 2000 led to widespread retreat into subsistence agriculture and a reduction of productivity by half. However, improvements in market organization (essentially through different contract design) led to a subsequent recovery, with productivity ending up 19 per cent above pre-privatization levels. This suggests two concluding remarks: First, the long-term supply response to market signals proved positive,

even though it took time for the right contractual arrangements to emerge; second, even though barriers to exit from subsistence agriculture are formidable, in this case they did not prove insurmountable, perhaps, though, because the retreat into subsistence had been of short duration.

Olivier Cadot is Professor of Economics at the University of Lausanne, Director of the Institut d'Economie Appliquee and a CEPR Research Fellow.

Laure Dutoit is an Economic affairs officer at the Economic Commission for Latin America and the Caribbean (UN-ECLAC).

Marcelo Olarreaga is Professor of Economics, University of Geneva and a CEPR Research Fellow.

BIBLIOGRAPHY

Alderman, Harold, and C Paxson (1992). Do the Poor Insure? A Synthesis of the Literature on Risk and Consumption in Developing Countries. World Bank Policy Research Working Paper 1008

Binswanger, Hans, and J McIntire (1987). Behavioral and Material Determinants of Production Relations in Land-abundant Tropical Agriculture. *Economic Development and Cultural Change* 36: 73–99

Brambilla, Irene, and G Porto (2006). Farm Productivity and Market Structure: Evidence from Cotton Reforms in Zambia; mimeo

Cadot, Olivier, Dutoit, L and M Olarreaga (2006). The Cost of Moving out of Subsistence, mimeo, University of Lausanne

Dercon, S. (1996). Risk, Crop Choice, and Savings: Evidence from Tanzania. *Economic Development and Cultural Change 44*: 485–513

Dutoit, Laure (2006). Switching-regression and Selection Models: A survey; mimeo, University of Lausanne

Eswaran, Mukesh, and A Kotwal (1986). Access to Capital and Agrarian Production Organization. *Economic Journal* 96 :482–98

Fafchamps, Marcel (1992). Cash Crop Production, Food Price Volatility, and Rural Market Integration in the Third World. *American Agricultural Economics Association Journal*

Goetz, Stephan (1993). Interlinked Markets and the Cash Crop–Food Crop Debate in Land Abundant Tropical Agriculture. *Economic Development and Cultural Change* 41: 343–61

Jacoby, Hanan (2000). Access to Markets and the Benefits of Rural Roads. *Economic Journal* 110,: 713–37

de Janvry, Alain, Fafchamps, M and E Sadoulet (1991). Peasant Household Behaviour with Missing Markets: Some paradoxes explained. *Economic Journal* 101: 1400–17

Jayne, T S (1994). Do High Food Marketing Costs Constrain Cash Crop Production? Evidence from Zimbabwe. *Economic Development and Cultural Change* 42: 387–402

Kennedy, Eileen (1994). Effects of Sugarcane Production, in Southwestern Kenya on Income and Nutrition, in J von Braun and E Kennedy (Eds.), Chapter 16 of *Agricultural Commercialization, Economic Development, and Nutrition*, IFPRI

Maertens, Miet, and J Swinnen (2006). Trade, Standards, and Poverty: Evidence from Senegal Leuven: Centre for Transition Economics, LICOS Discussion Paper Series 177/2006

Minten, Bart, Randrianarison, L and J Swinnen (2007). Spillovers from High-value Agriculture for Exports on Land Use in Developing Countries: Evidence from Madagascar. *Agricultural Economics* 37: 265–75

Moser, Christine, Barrett, C and B Minten (2005). Missed Opportunities and Missing Markets: Spatio-temporal Arbitrage of Rice in Madagascar; mimeo, Colgate

Renkow, Mitch, Hallstrom, D and D Karanja (2004). Rural Infrastructure, Transaction Costs and Market Participation in Kenya. *Journal of Development Economics* 73: 349–67

Rosenzweig, Mark, and H Binswanger (1993). Wealth, Weather Risk and the Profitability of Agricultural Investment. *Economic Journal* 103: 56–78

Vakis, Renos, Sadoulet, E and A de Janvry (2003). Measuring Transaction Costs from Observed Behaviour: Market Choices in Peru; mimeo

PART B
ADJUSTMENT IMPACTS

7

Trade Reform, Employment Allocation and Worker Flows[1]

MARC-ANDREAS MUENDLER

1. INTRODUCTION

Two salient workforce changeovers have occurred in Brazil since the late 1980s. Within the traded-goods sector, there is a marked occupational downgrading and a simultaneous educational upgrading by which employers fill expanding low-skill-intensive occupations with increasingly educated jobholders. Between sectors, there is a labor demand shift towards the least and the most skilled, which can be traced back to relatively weaker declines of traded-goods industries that intensively use low-skilled labor and to relatively stronger expansions of non-traded-output industries that intensively use high-skilled labor. Interestingly, and in certain contrast to the experience of other Latin American economies, these observations are broadly consistent with predictions of Heckscher–Ohlin trade theory for a low-skill abundant economy.

To analyze how workforce changeovers come about, actual worker flows need to be observed within and across employers. The research summarized in this chapter uses linked employer–employee datasets for Brazil to show that workforce changeovers are not achieved through worker reassignments to new tasks within employers or by reallocations across employers and traded-goods industries. Instead, trade-exposed industries shrink their workforces by dismissing less-schooled workers more frequently than more schooled workers, especially in skill-intensive occupations, while most displaced workers shift to nontraded-output industries or out of recorded employment. Trade liberalization in Brazil generated worker displacements, particularly from protected industries, but comparative-advantage industries and exporters do not absorb many of the trade-displaced workers. Indeed, these industries displace significantly more workers and hire fewer workers than the average employer. The observed resource reallocation patterns pose a challenge to classical trade theory, including the Heckscher–Ohlin

[1] This chapter synthesizes results of previously circulated research in Menezes-Filho et al. (2008), Muendler (2008); Muendler (2004); and Menezes-Filho and Muendler (2007). The chapter was completed while the author was visiting Princeton University. Financial support from the World Bank is gratefully acknowledged.

framework, as well as to modern trade theories with heterogeneous firms in the absence of endogenous productivity change.

Menezes-Filho and Muendler (2007) combine the linked employer–employee data with data from a Brazilian manufacturing survey to show that labor productivity increases faster than production in comparative advantage industries and at exporters. They conclude that the most plausible explanation seems to be that trade triggers faster productivity growth at exporters and in comparative-advantage industries, because for surviving firms in these industries larger market potential offers stronger incentives to improve efficiency. If productivity increases faster than production, then output shifts to more productive firms but labor does not. This labor market evidence for Brazil is also suggestive of a novel explanation why pro-competitive reforms might be associated with strong efficiency gains at the employer level but not in the aggregate, where idle resources can result.

The empirical literature on trade and resource reallocation has taken mainly three approaches. First, industry-level studies use measures of job creation, destruction, and churning (excess turnover beyond net change), as well as informality. Haltiwanger et al. (2004) show for a panel of six Latin American countries, for instance, that tariff reductions are associated with heightened within-sector churning and net employment reductions at the sector level.[2] Beyond those studies, research with linked employer–employee data documents the direction of factor flows between types of employers, and identifies the incidence of idle resources in the process. In contrast to the United States, where industries with faster productivity growth exhibit higher net employment growth (Davis et al. 1996), more productive employers reduced employment in Brazil during the 1990s. Using sector data, Goldberg and Pavcnik (2003) report no statistically significant relation between informal work and trade in Brazil, whereas household survey data suggest that tariff reductions are related to more transitions out of formal work, especially into self-employment and withdrawals from the labor force (Menezes-Filho and Muendler 2007). In the present chapter the focus is on the formal sector, however.

Second, employer-level studies show that trade reforms are associated with product-market reallocation towards more efficient producers (Tybout 2003). But employer-level studies typically report no detectable relationship between trade and employment.[3] The evidence from Brazil discussed in this chapyer indicates that trade variables are not statistically significant predictors of employment changes at the employer level either (Muendler 2008). But worker-level

[2] Using measures of net employment change, Wacziarg and Wallack (2004) detect no statistically significant labor reallocation in a cross-country cross-sector study of trade-liberalization episodes. Other examples of industry-level studies include Davis et al. (1996) for the United States; Roberts (1996) for developing countries; and Ribeiro et al. (2004) for Brazil.

[3] Roberts (1996) reports no clear effect of time-varying trade exposure on employment changes at plants in Chile and Colombia when sector characteristics are taken into account. Using Chilean plant data, Levinsohn (1999, 342) concludes that, 'try as one might, it is difficult to find any differential employment response' to trade liberalization. Neither do Davis et al. (1996) find a clear effect of trade on gross job flows using US data. An exception is Biscourp and Kramarz (2007) who show that French firm-level trade data exhibit a significant association of job destruction with firm-level imports.

regressions on the same data uncover that additional imports trigger significantly more worker displacements, while there are lasting worker flows away from productive high-output employers. This suggests that unobserved workforce heterogeneity hampers regressions at more aggregate levels, even the employer level, and calls for the use of worker panel data.

Third, a worker level literature studies the experience of displaced workers across sectors and worker groups. Kruse (1988) and Kletzer (2001) compare displaced workers between US industries and find that employment histories are largely explained by differences in workforce characteristics across sectors and vary little by a sector's trade exposure.[4] Time variation in our data, by contrast, identifies a salient impact of Brazil's trade opening on labor turnover. Beyond displaced worker survey data, the linked employer–employee records allow quantification of directions of worker flows across employers for many years and show that the economic burden of joblessness is substantial. As discussed further in what follows, the joblessness is partly trade-induced.

The remainder of this chapter is organized as follows. Section 2 briefly summarizes Brazil's trade reform and compares the country's labor market characteristics to other economies. Section 3 introduces the data (with most details relegated to the Appendix). Section 4 presents labor demand changes over the sample period 1986–2001 and discerns between-sector and within-sector changes using a Katz and Murphy (1992) labor demand decomposition. Section 5 investigates how much of the documented workforce changeover is brought about by task reassignments within firms, worker reallocations across firms and industries, and by worker separations without formal-sector reallocations. Section 6 uses a regression design, controlling for worker heterogeneity in turnover, to identify what share of the reallocation flows during the 1990s is predicted by Brazil's heightened trade exposure. Section 7 discusses evidence that labor productivity changes endogenously in response to trade reform, and presents potential implications of the findings for labor market adjustment costs. Section 8 concludes.

2. BRAZIL AND ITS TRADE REFORM

Since the late 1980s, Brazil's federal government initiated a series of economic reforms that by around 1997 resulted in a considerably more open economy to foreign goods and investments, a stable macroeconomy, and a somewhat smaller role for the state in the economy. In 1988, after decades of import substitution and industry protection, the Brazilian federal government under President Sarney initiated an internal planning process for trade reform and started to reduce *ad valorem* tariffs; but lacking public support, the government took little legislative initiative to remove binding non-tariff barriers so that nominal tariff reductions had little effect (Kume et al. 2003). In 1990, the Collor administration launched

[4] Similarly, Hungerford (1995) finds that short-term trade shocks play a minor role for separation rates in the United States.

a large-scale trade reform that involved both the removal of non-tariff barriers and the adoption of a new tariff structure with lower levels and smaller cross-sectoral dispersion. As a surprise to most observers at the time, Collor abolished all non-tariff barriers by presidential decree on his first day in office. Implementation of these policies was largely completed by 1993.

Figure 7.1 depicts Brazil's product-market and intermediate-input tariff schedules in 1990 and 1997 for the twelve manufacturing industries at the subsector IBGE level. Intermediate input tariff levels are calculated as weighted product tariffs using the economy-wide input-output matrix. Both the level and the dispersion of tariffs drop remarkably between 1990 and 1997. While *ad valorem* product tariffs range from 21 (metallic products) to 63 per cent (apparel and textiles) in 1990, they drop to a range from 9 per cent (chemicals) to 34 per cent (transport equipment) in 1997. Except for paper and publishing in 1990, sectors at the subsector IBGE level receive effective protection in both years, with mean product tariffs exceeding mean intermediate-input tariffs. By 1997, however, the relatively homogeneous tariff structure results in a small rate of effective protections for most industries—with the notable exception of transport equipment.

Brazil underwent additional reforms over the sample period. In 1994, during the Franco administration and under the watch of then finance minister Cardoso, drastic anti-inflation measures succeeded for the first time in decades. A privatization program for public utilities was started in 1991 and accelerated in the mid 1990s, while Brazil simultaneously liberalized capital account restrictions. These measures were accompanied by a surge in foreign direct investment inflows in the mid 1990s. The pro-competitive reforms during the 1990s, mostly targeted at product markets, had been preceded by changes to Brazil's labor market institutions in 1988.

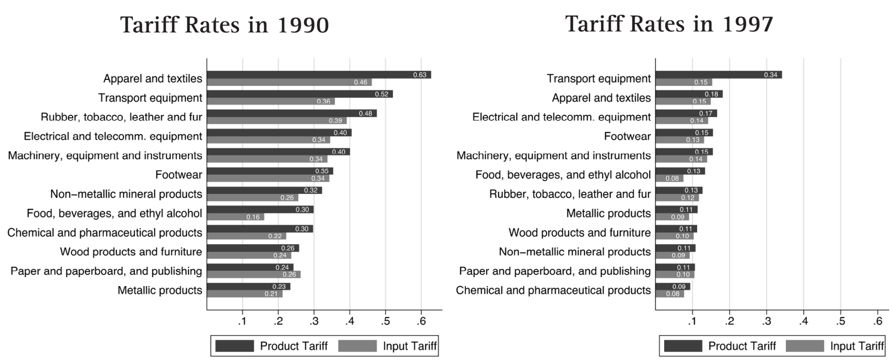

Figure 7.1: *Product-market and intermediate-input tariffs 1990 and 1997*

Sources: Muendler (2008). *Ad valorem* product tariffs at *Nível 80* from Kume et al. (2003).

Note: Intermediate input tariffs are weighted product-market tariffs using national input-output matrices at *Nível 80* (from IBGE). Product-market and intermediate-input tariffs are transformed to the subsector IBGE level using unweighted means over the *Nível 80* classifications.

Table 7.1: *Labor market rigidity comparisons*

	Rigidity and Difficulty Indices				
	Hiring difficulty (1)	Hiring difficulty (2)	Firing difficulty (3)	Employment rigidity (4)	Firing costs[a] (5)
Brazil	67.0	80.0	70.0	72.0	165.0
Trade partners					
weighted by trade volume[b]					
1990	25.2	42.0	22.7	29.9	43.3
1997	28.1	45.3	24.4	32.4	47.6
weighted by source-country imports					
1990	23.2	42.9	21.7	29.1	46.8
1997	27.2	44.3	23.6	31.6	46.0
weighted by destination-country exports					
1990	26.4	41.5	23.4	30.3	41.2
1997	29.1	46.4	25.2	33.4	49.5

[a] In weekly wage equivalents
[b] Country sum of exports from and imports to Brazil.

Source: Botero, Djankov, La Porta, Lopez de Silanes and Shleifer (2004) labor market rigidity measures.

Note: A higher index and a higher rank indicate a more rigid labor market. Trade partner averages weighted by WTF (NBER) bilateral trade data for 1990 and 1997.

Brazil's constitution of 1988 introduced a series of labor market reforms that aimed to increase workers' benefits and the right to organize, thus raising labor costs. Most important, firing costs increased substantially.[5] Given their constitutional status, these labor market institutions remained unaltered throughout the 1990s, the period of chief interest for this chapter. Table 7.1 compares World Bank indices of labor market rigidity for Brazil to its mean trading partner and shows that Brazil's labor market is considerably more rigid than its trading partners' labor markets are. For the World Bank's four rigidity and difficulty indices (hiring difficulty, hours rigidity, firing difficulty, employment rigidity) and its firing-cost measure, Brazil exhibits mean values between 67 and 165, whereas the mean values for Brazil's trading partners vary between 20 and 49 for three choices of trade weighting (considering trade volume, source-country import, and destination-country export weighting using WTF (NBER) data for Brazil). The difference is partly due to the fact that Brazil's largest trade partners are highly flexible economies. Not weighted by trade, however, Brazil still ranks in the rigid tercile of countries.

Among the reforms, trade liberalization played a dominant role for labor market outcomes. Multivariate regressions in Section 6 control for sector and year effects,

[5] The 1988 reforms reduced the maximum working hours per week from 48 to 44, increased the minimum overtime premium from 20 per cent to 50 per cent, reduced the maximum number of hours in a continuous shift from 8 to 6 hours, increased maternity leave from 3 to 4 months, increased the value of paid vacations from 1 to 4/3 of the normal monthly wage, and increased the fine for an unjustified dismissal from 10 per cent to 40 per cent of the employer-funded severance pay account (FGTS). See Heckman and Pagés (2004) and Gonzaga (2003) for further details.

as well as variables related to simultaneous reforms. Results confirm the overwhelming predictive power of trade liberalization and an employer's export status for employment changes. Before an analysis or worker flows in sections 5 and 6, however, I first turn to a conventional decomposition of employment changes in Section 4, and to the data sources in the next Section 3.

3. LINKED EMPLOYER–EMPLOYEE DATA

The focus in what follows is on labor reallocation for prime-age male workers, 25 to 64 years old, in the formal sector anywhere nationwide. The restriction to prime-age workers is meant to reduce the sample to workers after their first labor force entry, highlighting the reallocation of active labor resources. Prime-age male workers are known to exhibit low wage elasticities of labor supply so that the presented results are possibly little affected by labor supply changes. A recent revision to Menezes-Filho and Muendler (2007) shows that results are similar for samples that include both genders and all age groups.

The linked employer–employee data underlying most results reported here derive from Brazil's labor force records RAIS (*Relação Anual de Informações Sociais* of the Brazilian labor ministry MTE). RAIS is a nationwide annual census of workers formally employed in any sector (including the public sector). RAIS covers, by law, all formally employed workers, captures formal-sector migrants,[6] and tracks the workers over time. By design, however, workers with no current formal-sector employment are not in RAIS.

RAIS primarily provides information to a federal wage supplement program (*Abono Salarial*), by which every worker with formal employment during the calendar year receives the equivalent of a monthly minimum wage. RAIS records are then shared across government agencies and statistical offices. An employer's failure to report complete workforce information can, in principle, result in fines proportional to the workforce size, but fines are rarely issued. In practice, workers and employers have strong incentives to ascertain complete RAIS records because payment of the annual public wage supplement is exclusively based on RAIS. The Ministry of Labor estimates that well above 90 per cent of all formally employed workers in Brazil were covered in RAIS throughout the 1990s.

The full data include 71.1 million workers (with 556.3 million job spells) at 5.52 million plants in 3.75 million firms over the 16-year period 1986–2001. Every observation is identified by the worker ID (PIS), the plant ID (of which the firm ID is a systematic part), the month of accession, the month of separation, and the occupation (if a worker holds multiple jobs at the same plant). Relevant worker

[6] Migration among metropolitan workers, for instance, is substantial. Among the prime-age male workers in RAIS with a metropolitan job in 1990, 15 per cent have a formal job outside their 1990 city of employment by 1991 and 25 per cent by 1993. Similarly, among the metropolitan workers in 1994, 17 per cent have a formal job elsewhere by 1995 and 27 per cent by 1997. These statistics also suggest that conventional unemployment rates from household surveys could be exaggerated if migrating households are dropped from the numerator and denominator as missing, thus biasing the unemployment rate in household surveys upwards.

information includes age, gender, educational attainment; job information includes tenure at the plant, occupation, and the monthly average wage; plant information includes sector and municipality classifications. To facilitate tracking, RAIS reports formal retirements and deaths on the job. RAIS identifies the plant and its firm, which in turn can be linked to firm information from outside sources such as exporter data.

Table 7.2: *Employment by employer's sector and export status*

	Traded Goods		Nontraded Output			Overall[a]
	Primary (1)	Manuf. (2)	Comm. (3)	Services (4)	Other (5)	
			Allocation of workers, nationwide			
1990	.021	.238	.128	.280	.333	22,844
1997	.044	.195	.152	.320	.289	24,068
			Allocation of prime-age male workers, nationwide			
1990	.029	.263	.111	.284	.314	10,763
1997	.063	.221	.131	.308	.278	11,483
Nonexporter	.882	.494	.935	.937	.930	.830
Exporter	.118	.506	.065	.063	.070	.170
			Allocation of prime-age male workers, metropolitan areas			
1990	.015	.270	.104	.309	.302	5,965
1997	.024	.213	.125	.363	.275	6,060
Nonexporter	.760	.390	.887	.913	.898	.778
Exporter	.240	.610	.113	.087	.102	.222

[a] Total employment (thousands of workers), scaled to population equivalent.

Sources: Muendler (2008), RAIS 1990–2001 employment on December 31st, and SECEX 1990–2001.

Note: Nationwide information based on 1-percent random sample, metropolitan information on 5-percent random sample. Period mean of exporter and nonexporter workforces, 1990–2001.

The samples behind results reported here chiefly derive from a list of all proper worker IDs (11-digit PIS) that ever appear in RAIS at the national level, from which a 1 per cent nationwide random sample and a 5 per cent metropolitan random sample were drawn. These randomly sampled workers are then tracked through all their formal jobs. Industry information is mostly based on the subsector IBGE classification (roughly comparable to the NAICS 2007 three–digit level), which is available by plant over the full period (see Table A7.1 in Appendix A for sector classifications). For the calculation of separation and reallocation statistics, a worker's separation is defined as the layoff or quit from the highest paying job.[7]

[7] Among the male prime-age workers nationwide, 3 per cent of the job observations are simultaneous secondary jobs. Tables 5.1, 5.2, and 5.3 are based on the so-restricted sample, whereas all aggregate statistics, Katz-Murphy decompositions, and regressions are based on the full sample. The restriction to a single job at any moment in time permits a precise definition of job separation as a layoff or quit from the highest-paying job (randomly dropping secondary jobs if there is a pay tie). Removing simultaneously held jobs does not significantly affect estimates of skill, occupation, and gender premia in Mincer (1974) regressions such as those reported in Table C7.1 in Appendix C.

Table 7.2 shows the allocation of workers across industries in 1990 and 1997 (a detailed employment share breakdown for the RAIS universe can be found in Table A7.1 in Appendix A). The nationwide RAIS records represent 23 million formally employed workers of any gender and age in 1990, and more than 24 million formal workers by 1997. The bulk of Brazil's formal employment is in manufacturing, services, and other industries (which include construction, utilities, and the public sector), with roughly similar formal employment shares between a quarter and a third of the overall formal labor force. Commerce (wholesale and retail) employs around one in eight formal workers, and the primary sector (agriculture and mining) at most one in twenty-five formal workers, partly because of a high informality rate in agriculture.

Prime-age male workers nationwide make up slightly less than half of the total workforce in 1990 and 1997. In both years, prime-age male workers are slightly more frequently employed in the primary and manufacturing sector than the average worker of any gender and age but less frequently in commerce, services, and other sectors. More than half of the RAIS-reported formal employment of prime-age males occurs in the six metropolitan areas of Brazil: São Paulo city, Rio de Janeiro city, Belo Horizonte, Porto Alegre, Salvador, and Recife. Compared to the nationwide average across gender and age, prime-age males in metropolitan areas are slightly less frequently employed in the primary sector, commerce, and other sectors, and somewhat more frequently employed in manufacturing and services. Overall, however, the labor allocation across sectors is broadly similar across regions and gender and age groups, whereas changes over time between 1990 and 1997 are more pronounced. Between 1990 and 1997, there is a marked drop in formal manufacturing employment, which is accompanied by an increase of employment in primary sectors, commerce, and especially services. Overall, between roughly a quarter and a third of the nationwide and metropolitan prime-age male workforces are employed in traded-goods sectors, and two thirds to three quarters in nontraded-output sectors.

Table 7.3 provides a summary comparison of variables for manufacturing industries in different quintiles of comparative advantage, and between exporters and the average employer. Top comparative-advantage industries (in the highest quintile) show a higher labor turnover than the average sector, with both more worker separations and more accessions, whereas exporting firms exhibit below-average turnover with fewer worker separations and fewer accessions than average. Among the separations, reported quits play a minor role.

The average exporter is active in a sector with a slightly lower than average comparative advantage level. Similarly, there are fewer worker observations at exporters in a top comparative-advantage sector than at exporters overall. The reason is that there is a larger number of small-scale exporters in industries without comparative advantage.[8] As expected for a country with a history of import-substitution industrialization Brazil's top comparative-advantage

[8] I control for employment in the regression to capture exports per worker effects.

industries have lower than average tariffs. Comparative-advantage industries also exhibit lower import penetration. Firms in the top comparative-advantage industries and exporters have larger workforces than average (85 and 326 workers more, respectively, than the average formal-sector manufacturing plant with 257 workers). The sample from RAIS is a random draw of workers from the formal sector worker universe, so that larger plants are over-represented. Manufacturing employment drops between 1990 and 1998, and drops faster than average in the highest quintile advantage sectors.

Table 7.3: *RAIS summary statistics for manufacturing*

	All sectors and firms			
	Mean (1)	Std.Dev. (2)	5th comp. adv. Quintile Mean (3)	Exporter Mean (4)
Outcomes				
Indic.: Separation	.282	.450	.314	.260
Quit	.026	.160	.031	.020
Indic.: Accession	.292	.455	.326	.237
Main covariates				
Balassa (1965) Comp. Adv.	1.450	1.047	3.223	1.373
Exporter Status	.495	.500	.439	1.000
Product Market Tariff	.193	.103	.174	.204
Intm. Input Tariff	.146	.077	.105	.154
Import Penetration	.064	.052	.031	.074
Plant-level covariates				
Log Employment	5.148	1.952	5.551	6.210
Log Employment 1998/90	.930		.919	.976
Log Labor Productivity	11.186	.706	11.081	11.233
Log Labor Productivity 1998/90	1.045		1.025	1.047

Sources: Menezes-Filho and Muendler (2007). RAIS 1990–98 (1-percent random estimation sample), male workers nationwide, 25 to 64 years old, with manufacturing job.

Note: Statistics based on separation sample, except for accession indicator (146,787 observations in separation, 112,974 in accession sample). Sector information at subsector IBGE level. PIA 1986–98 for labor productivity information.

To obtain labor productivity, a random extract and three-firm aggregate of the manufacturing firm survey PIA is used (see Appendix B). There are remarkable mean differences in labor productivity between an exporter and an average firm. A reason is that substantial employer heterogeneity prevails within industries, with diverse exporters and nonexporters having shifting mean characteristics. Labor productivity increases between 1990 and 1998. At exporters, labor productivity is higher than average over the whole sample period, but lower than average at firms in comparative-advantage industries. Log labor productivity in 1998 exceeds log labor productivity in 1990 by 4.5 per cent in the estimation sample, and by 4.7 per cent at manufacturing exporters.

4. EMPLOYMENT REALLOCATION

A conventional way to measure employment reallocation is the Katz and Murphy (1992) method. The method decomposes labor demand changes into shifts between industries, associated with variations in sector sizes, given sectoral occupation profiles, and within industries through changing occupational intensities. The former shifts between industries relate to the changing allocation of employment across sectors, whereas the latter shifts within industries reflect the change in relative skill intensities of occupations or alterations to the sectoral production process.

4.1 Between and within industry demand shifts

Applying the Katz and Murphy (1992) method to employment in the Brazilian formal sector over the years 1986–2001 reveals the main patterns of labor market adjustment. The decomposition into between and within sector variation indicates how two important sources of change contribute to workforce changeover. Between-industry shifts are arguably driven by changes in final goods demands, sectoral differences in factor-nonneutral technical change, and changes in the sector-level penetration with foreign imports. Within-industry shifts can be related to factor-nonneutral technical change, factor price changes for substitutes or complements to labor, and international trade in tasks which allocates activities along the value chain across countries.

The Katz and Murphy (1992) decomposition relates back to Freeman's (1980) manpower requirement index and is designed to measure the degree of between-industry labor demand change under fixed relative wages. The decomposition tends to understate the true between-industry demand shift in absolute terms when relative wages change. Though possibly overstating the within-industry effects, the Brazilian evidence suggests that within-industry demand changes are an important source of employment changeover in Brazil especially since 1990. Beyond the Katz and Murphy (1992) framework, I therefore offer statistics that document time variation in the occupational profile within industries, and the skill changeover within occupations.

Under the assumption that the aggregate production function is concave (so that the matrix of cross-wage elasticities of factor demands is negative semi-definite), Katz and Murphy (1992) show that an appropriate between-industry demand shift measure ΔD_k for skill group k is

$$\Delta D_k = \sum_j X_{jk} \frac{w' dX_j}{w' X_j}, \tag{1}$$

where X_{jk} is the employment of skill group k in industry j, w is a $k \times 1$ vector of constant wages, and dX_j and X_j are the $k \times 1$ vectors of employment changes and levels in industry j, respectively. Equation (1) is simply the vector of weighted sums of industry employments for each skill group k, with the weights given by the percentage changes in overall employment in every industry j. The measure is similar to standard labor-requirement indexes Freeman (1980), only that

changes are measured in efficiency units at constant wages rather than in head counts (or hours). Intuitively, skill groups that are intensively employed in expanding sectors experience a demand increase, whereas skill groups intensively employed in contracting sectors face falling demand. Under constant wages, the measure indicates whether the data are consistent with stable labor demands within sectors. Wages change, however, so that there is bias in the measure. Katz and Murphy (1992) show that the bias is inversely related to wage changes if substitution effects dominate the employment decisions, so that measure (1) understates the demand increase for groups with rising relative wages.

In the Brazilian context, the formal-sector economy can be divided into 26 2-digit industries (using the subsector IBGE classification) and five occupations (professional and managerial occupations, technical and supervisory occupations, other white-collar occupations, skill-intensive blue-collar occupations, and other blue-collar occupations). The classification of activities into both sectors and occupations is motivated by the idea that international trade of intermediate and final goods can be understood as trade in tasks along the steps of the production chain. Using the resulting 130 industry–occupation cells, an empirically attractive version of the between-industry demand shift measure (1) is

$$\Delta X_k^{di} = \frac{\Delta D_k}{E_k} = \sum_i \left(\frac{E_{ik}}{E_k}\right)\left(\frac{\Delta E_i}{E_i}\right) = \frac{\sum_i \alpha_{ik}\,\Delta E_i}{E_k}, \tag{2}$$

where E_i is total labor input in sector-occupation cell i measured in efficiency units, and $\alpha_{ik} \equiv E_{ik}/E_i$ is skill group k's share of total employment in efficiency units in sector i in the base period. Equation (2) expresses the percentage change in demand for each skill group as a weighted average of the percentage changes in sectoral employment, the weights being the group-specific efficiency-unit allocations. Following Katz and Murphy (1992), I turn index (2) into a measure of relative demand changes by normalizing all efficiency-unit employments in each year to sum to unity. The base period is the average of the sample period from 1986 to 2001 so that α_{ik} is the share of total employment of group k in sector i over the 1986–2001 period and E_k is the average share of skill group k in total employment between 1986 and 2001.

The overall (industry–occupation) measure of demand shifts for skill group k is defined as ΔX_k^{di} from equation (2), where i indixes the 130 industry–occupation cells. The between-industry component of this demand-shift measure is defined as the group-k index ΔX_k^{di} from equation (2), where $i = j$ now indexes only 26 industries. Accordingly, the within-industry component of demand shifts is $\Delta X_k^{dw} \equiv \Delta X_k^{di} - \Delta X_k^{dj}$.

Table 7.4 presents the nationwide demand decomposition and the overall demand shifts by group of educational attainment for the economy as a whole, and separately for the traded-goods and the nontraded-output sectors. As in Katz and Murphy (1992), the percentage changes are transformed into log changes with the formula $\hat{\Delta X}_k^d = \log(1 + \Delta X_k^d)$. By construction, in the (vertical) sectoral dimension the economy-wide demand shift indices for each skill group are a

Table 7.4: *Industry and occupation based log demand shifts, 1986–2001*

	Between Industry				Within Industry				Overall Industry–Occupation			
(in %)	86–90 (1)	90–97 (2)	97–01 (3)	86–01 (4)	86–90 (5)	90–97 (6)	97–01 (7)	86–01 (8)	86–90 (9)	90–97 (10)	97–01 (11)	86–01 (12)
Economy wide												
Illiterate or Primary Dropout	−0.1	4.7	−0.1	4.5	−2.2	−0.2	−1.0	−3.3	−2.3	4.5	−1.1	1.1
Primary School Graduate	−2.1	−0.1	−1.7	−3.9	−1.4	0.5	−1.5	−2.4	−3.6	0.4	−3.2	−6.4
Middle School Graduate	−1.9	−0.9	−0.5	−3.3	−0.1	1.5	−1.2	0.1	−2.0	0.6	−1.8	−3.1
High School Graduate	0.3	−1.7	−0.8	−2.3	1.1	0.9	0.2	2.2	1.4	−0.9	−0.6	−0.1
College Graduate	3.3	0.4	2.9	6.6	1.3	−2.4	2.5	1.4	4.6	−2.0	5.4	8.1
Traded-goods sectors												
Illiterate or Primary Dropout	−3.0	3.7	−1.7	−0.9	−0.7	−0.2	−0.2	−1.1	−3.7	3.6	−1.9	−2.0
Primary School Graduate	−3.8	−2.0	−2.4	−8.2	−0.4	0.2	−0.6	−0.7	−4.2	−1.8	−3.0	−9.0
Middle School Graduate	−3.9	−4.0	−2.6	−10.6	0.0	0.3	−0.5	−0.2	−3.9	−3.8	−3.1	−10.7
High School Graduate	−3.7	−4.4	−2.4	−10.5	0.5	−0.1	0.2	0.7	−3.2	−4.5	−2.1	−9.8
College Graduate	−3.6	−4.6	−2.1	−10.4	0.5	−0.5	1.4	1.4	−3.1	−5.1	−0.7	−8.9
Nontraded-output sectors												
Illiterate or Primary Dropout	3.6	0.4	1.9	5.9	−1.5	0.1	−0.7	−2.2	2.1	0.4	1.2	3.7
Primary School Graduate	2.8	2.7	1.5	7.0	−1.0	0.3	−0.9	−1.6	1.9	2.9	0.6	5.4
Middle School Graduate	2.3	3.3	2.3	7.9	−0.2	1.2	−0.7	0.3	2.1	4.6	1.6	8.3
High School Graduate	3.2	1.9	1.2	6.3	0.6	0.9	0.0	1.5	3.9	2.8	1.2	7.8
College Graduate	5.2	3.3	3.9	12.4	0.8	−1.8	1.4	0.4	6.0	1.5	5.3	12.8

Source: Muendler (2008). RAIS 1986–2001 (1-percent random sample), male workers, 25 years or older. *Note:* Overall and between-industry demand shift measures for skill group k are of the form $\Delta D_k = \sum_j \alpha_{jk}(\Delta E_j/\Delta E_k)$ where α_{jk} is the average share for group k of employment in cell j over the period 1986–2001, E_j is the share of aggregate employment in cell j, and E_k is the average share of total employment of group k over the period 1986–2001 Katz and Murphy (1992). Reported numbers are of the form $\log(1+\Delta D_k)$. In the overall measure, j indexes 130 industry-occupation cells; in the between-industry measure, $i=j$ indexes 26 industries (14 traded–goods and 12 nontraded-output sectors). The within–industry index for group k is the difference of the overall and between-industry measures. Employment is measured in efficiency units.

weighted sum of the traded and nontraded sector indices (except for occasional rounding errors because of the log transformation), where the weights are the skill groups' shares in the sectors. In the (horizontal) time dimension, the indices are the sum of the time periods for each skill group.

The entries for overall shifts across all sectors summarize Brazil's labor demand evolution (five first rows of column 12). Over the full period from 1986 to 2001, the least and the most skilled prime-age male workers experience a positive relative demand shift of 1 and 8 per cent, respectively, whereas the three intermediate skill groups suffer a labor demand drop. This overall pattern, with demand surges at the extreme ends of the skill spectrum and drops for the middle groups, can be traced back to two overlaying developments. First, before and after the main economic liberalization episode, that is in the periods 1986–90 and 1997–2001, demand for college graduates rises by around 5 per cent while demand drops for all other skill groups in 1997–2001 and for all other skill groups but high-school graduates in 1986–90. Second, during the period of economic liberalization between 1990 and 1997 the reverse labor demand change occurs,

with demand for the least-educated males increasing by roughly 5 per cent and dropping for college graduates by –2 per cent. The demand rise for the least-educated during liberalization more than outweighs the demand drops before and after, so that a net demand increase remains by 2001. For college graduates, demand surges before and after liberalization are so strong that the drop during liberalization is of little importance and a strong net demand remains by 2001. This pattern is consistent with a Heckscher–Ohlin interpretation of the specialization pattern following trade liberalization. Brazil, whose labor force is relatively low-skill abundant, experiences a shift towards low-skill intensive economic activities between 1990 and 1997—against the longer-term trend manifested before (1986–90) and after (1997–2001) by which demand for highly skilled workers increases but drops for lower-skilled workers.

Between and within decompositions, as well as a distinction of traded and nontraded sectors, lend additional support to a Heckscher–Ohlin interpretation of labor demand changes. The decomposition for all sectors (five first rows) into between-industry and within-industry changes indicates that the overall evolution is mostly driven by between-industry changes, with demand surges at the extreme ends of the skill spectrum and drops for the middle groups (column 4). In contrast, the within-industry labor demand changes favor the least skilled the least, with a demand drop of –3 per cent, and the most skilled the most, with a demand increase of 1 to 2 per cent for high-school educated workers and college graduates. The within-industry demand changes are almost monotonically increasing as one moves up the educational attainment ranks (column 8) in the 1986–2001 period, and would indeed monotonically increase if it were not for a within-industry drop in demand for college graduates during the liberalization period. This chapter returns to the within-industry demand changes with additional evidence further below. In fact, the within-industry workforce changeover is found to reinforce a broad Heckscher–Ohlin interpretation of Brazil's experience.

A distinction by sector relates the between-industry demand evolution to differences across traded-goods industries (middle five rows) and nontraded-output industries (last five rows). In the traded-goods sectors, where trade liberalization is expected to exert its impact, Brazil experiences a salient labor demand drop beyond –10 per cent for the three more educated skill groups between 1986 and 2001. Expectedly for a low-skill abundant country, the demand drop is the strongest for the highly skilled and the weakest for the low-skilled workers (column 4). Most notably, during the liberalization episode illiterate workers and primary school dropouts experience a rise in demand due to between-industry shifts, whereas more skilled workers experience demand drops of monotonically larger magnitudes as one moves up the skill ladder (column 2). The nontraded-output sectors exhibit a relatively homogeneous demand increase between 6 and 8 per cent for workers with no college degree, and a strong 12 per cent increase for college graduates (column 4). The demand increase for the least skilled in nontraded-output sectors combined with only a slight demand drop for them in the traded-goods sectors results in an overall positive demand for the skill group from the between-industry

component (column 4). Similarly, the strong demand for college graduates in nontraded-output sectors more than outweighs their demand drop in traded-goods sectors. For intermediate skill groups between these two extremes, the demand drop in the traded-goods sectors outweighs their demand increase in nontraded-output sectors and results in overall negative demand changes.

Within industries there is a clear and pronounced pattern of falling demand for the least skilled, and increasing demand for the more skilled, with monotonically stronger demand changes as one moves up the skill ranks, except only for college graduates (column 8). This pattern is similar across both traded and nontraded sectors and most time periods. The reason for the break in monotonicity at the college-graduate level (column 8) is a demand drop for this skill group during the liberalization period (column 6). A Stolper–Samuelson explanation is consistent with the outlier behavior of college graduates during this period. Note that the Stolper–Samuelson theorem predicts wage drops for more-educated workers in a low-skill abundant economy after trade reform, and Gonzaga et al. (2006) document that skilled earnings differentials indeed narrow over the course of the trade liberalization period. Because labor is measured in current-period efficiency units, a relative drop in wages for college educated workers tends to turn their within-industry demand index negative. With this explanation for the outlier behavior of collage graduates in view, there is a striking monotonicity in the increase in within-industry labor demand change as one moves up the skill ranks.

4.2 Within-industry employment changeovers

The demand decompositions above show a noteworthy within-industry labor demand reduction for low-skilled workers and a demand increase for high-skilled workers, both in traded-goods and nontraded-output sectors. The sources of this

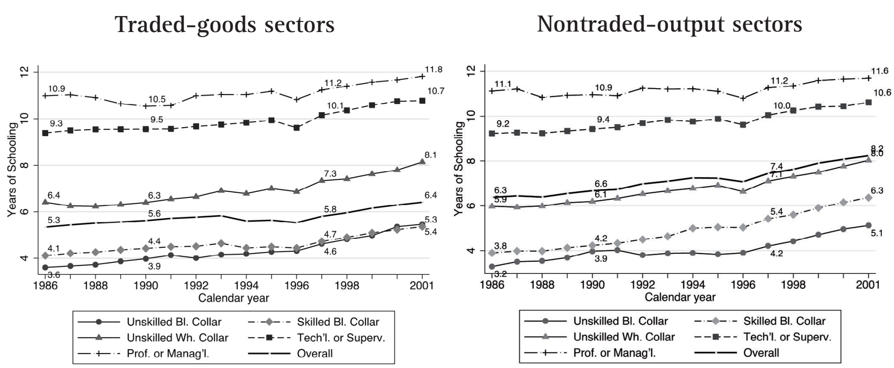

Figure 7.2: *Schooling intensity of occupations*

Source: Muendler (2008). RAIS 1986–2001 (1 per cent random sample), male workers nationwide, 25 to 64 years old, with employment on December 31st.

Note: Traded-goods sectors are agriculture, mining and manufacturing (subsectors IBGE 1–13 and 25), nontraded-output sectors are all other industries. Mean years of schooling weighted by employment within occupations.

change deserve more scrutiny. Abandoning the efficiency-unit perspective on employment in favor of counts of workers to keep wage effects separate, I turn to an assessment of labor allocation to activities by period.[9]

Figure 7.2 shows the evolution of the skill assignment by occupation over time. In both traded-goods and nontraded-output sectors, there is a marked increase across all five occupation categories in the educational attainment of the job holders. From 1986 to 2001, the mean number of years of schooling in unskilled blue-collar occupations rises from below four years to more than five years in both traded and nontraded sectors (in traded sectors schooling in unskilled blue-collar occupations even slightly exceeds the schooling in skilled blue-collar jobs by 2001). The average number of school years increases from around four to more than five years for skilled blue-collars jobs in traded sectors and to more than six years in nontraded sectors by 2001. For unskilled white-collar occupations, the average job holder's schooling goes from around six to more than eight years both in traded and nontraded goods sectors. The shift also extends to technical and supervisory positions, where the average job holder's schooling goes from less than 10 to more than 10 years of schooling both in traded and nontraded sectors, and to managerial positions, where mean schooling rises from 11 to almost 12 years over the period 1986–2001. These largely steady within-industry changeovers in workers' occupational assignments between 1986 and 2001 overlay the shorter-lived between-industry changes with much time variation across the three subperiods: 1986–90, 1990–97 and 1997–2001.

One might suspect that the considerable surge in schooling levels is partly due to labor supply changes such as the entry of increasingly educated cohorts of male workers into the labor force, or relatively more frequent shifts of skilled male workers from informal to formal work status over the sample period. In fact, the sector-wide average schooling level rises from less than six to more than six years in the traded-goods sector, and in the nontraded-output sector from more than six to more than eight years (as the respective overall curves in Figure 7.2 show). To control for overall skill labor supply by sector, I extend the Katz and Murphy (1992) idea to the present context and subtract the mean annual years of schooling in a sector from the occupation-specific means in the sector. For this purpose, I consider all traded-goods industries as one sector, and all nontraded-output industries as another sector. Subtracting the annual mean years of schooling, instead of dividing by the annual total as in Table 7.4 before, preserves the cardinal skill measure of years of schooling and expresses occupation-specific skill demands as deviations from the sector-wide employment evolution in terms of years of schooling.

Figure 7.3 presents average years of schooling by occupation, less the sector-wide mean schooling across all occupations. By this measure, skill demand within every occupation category increases in the traded-goods sector since 1990: from a difference of −1.6 to −0.9 years in unskilled blue-collar occupations, from −1.2

[9] An efficiency-unit based analysis shows broadly the same patterns of workforce changeovers in terms of wage bills as the head-count based analysis that follows.

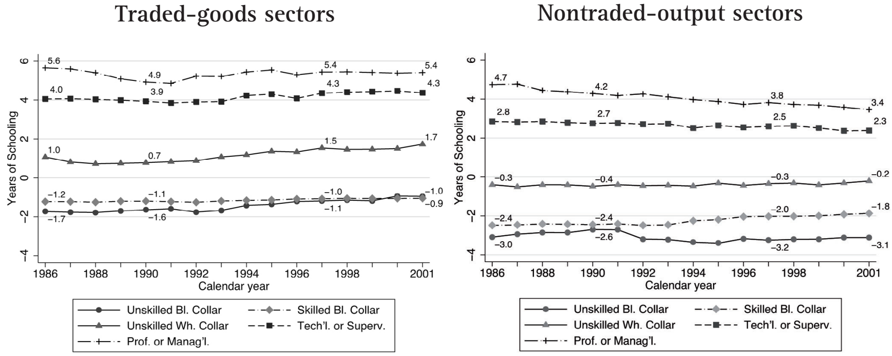

Figure 7.3: *Difference between schooling intensity of occupations and annual mean schooling level*

Source: Muendler (2008). RAIS 1986–2001 (1 per cent random sample), male workers nationwide, 25 to 64 years old, with employment on December 31st.

Note: Traded-goods sectors are agriculture, mining and manufacturing (subsectors IBGE 1–13 and 25), nontraded-output sectors are all other industries. Mean years of schooling weighted by worker numbers within occupations, less mean years of schooling weighted by employment across all occupations.

to −1.1 years in skilled blue-collar occupations, from 0.8 to 1.7 in unskilled white-collar jobs, from 3.9 to 4.4 in technical jobs, and from 4.9 to 5.4 in professional and managerial positions. For all three white-collar occupational categories, the schooling-intensity surge beyond the sector average since 1990 is a reversal of the opposite trend prior to 1990, while schooling-intensity continually increases for blue-collar occupations in the traded sector after 1986. By construction, the persistent occupation-level increases in worker schooling after 1990 go beyond the change in the sector-wide workforce schooling. The puzzling pattern that changes beyond the sector mean are uniformly directed towards higher schooling in every single occupation since 1990 implies that there must be an employment expansion in less skill-intensive occupations—otherwise it would be impossible for every single occupation category to exhibit a faster skill-intensity increase than the average over all occupations. In contrast to the traded sector, nontraded-output industries do not exhibit the uniform pattern of schooling increases across all occupations, but a drop in schooling intensity in the technical and managerial occupations and a rise in schooling intensity in skilled blue-collar occupations.

The evolution of schooling intensity in Brazil's traded-goods sector is reminiscent of a Heckscher–Ohlin interpretation as well—though not for industries but for tasks. Think of production activities in the Heckscher–Ohlin framework not as sectors but as occupations and suppose that Brazil has a relatively less schooled labor force than its main trading partners. Brazil's top five trading partners in total trade volume during the 1990s are, in descending order, the United States, Argentina, Germany, Italy, and Japan. As Brazil's integration into the world economy advances, thus reinterpreted Heckscher–Ohlin trade theory predicts that Brazil increasingly specializes in less-schooling-intensive occupations, but that Brazil employs in these expanding occupations relatively

more high-skilled workers because their relative wage declines. Gonzaga et al. (2006) document that Brazil's skilled earnings differential narrows over the 1990s. Using rich linked employer–employee data that control for unobserved worker characteristics, Menezes-Filho et al. (2008) show, however, that the skill premium in wages only changes slightly between 1990 and 1997 (see Table C7.1 in Appendix A). Of course, more research is required to discern this reinterpretation of classic trade theory from alternative explanations. The simultaneous schooling-intensity increase in every single occupation, above and beyond the sector mean, could also be related to factor-nonneutral technical change or factor price changes for substitutes for labor, and not only to international trade in tasks. Yet, the prediction of reinterpreted classic trade theory that foreign trade expands less schooling-intensive occupations in Brazil's traded-goods sector is fully consistent with the data.

Figure 7.4 depicts the nationwide occupation profile within traded-goods sectors and nontraded-output sectors for the years 1986 to 2001. In traded-goods industries, skilled blue-collar jobs expand markedly with the conclusion of the first wave of trade reforms between 1991 and 1993. The share of skilled blue-collar occupations increases from below 60 per cent in 1990 to 68 per cent in 1994 and to 71 per cent by 2001. Recall from the evidence in Figure 7.2 that the average worker's schooling in both skilled and unskilled blue-collar jobs in the traded-goods sector is roughly the same. The growing importance of skilled blue-collar occupations comes at the expense of all other occupations in the traded-goods industries. At the low-skill intensity end, the share of unskilled blue-collar occupations drops from more than 13 per cent in 1990 to 8 per cent in 1994 (but recovers slightly to close to 9 per cent by 2001). More importantly, the expansion of skilled blue-collar occupations in traded-goods sectors comes at the expense of white-collar occupations, whose total employment share drops from 27 per

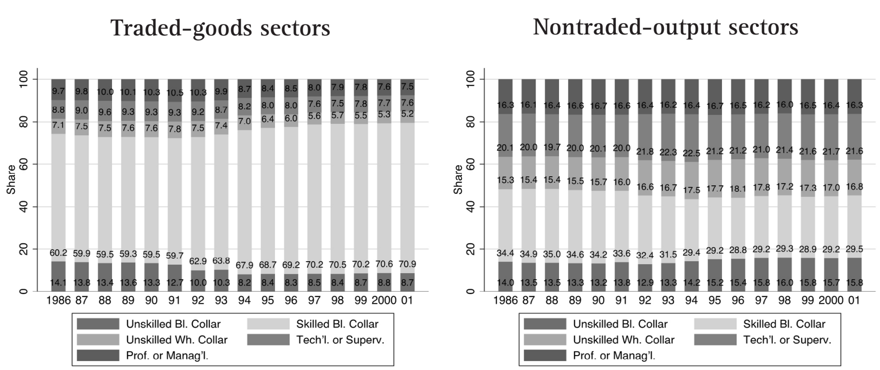

Figure 7.4: *Occupational workforce composition*

Source: Muendler (2008). RAIS 1986–2001 (1 per cent random sample), male workers nationwide, 25 to 64 years old, with employment on December 31st.

Note: Traded-goods sectors are agriculture, mining and manufacturing (subsectors IBGE 1–13 and 25), nontraded-output sectors are all other industries. Shares based on employment.

cent in 1990 to 24 per cent in 1994 and 20 per cent in 2001. In the nontraded-output sectors, in contrast, it is the unskilled blue-collar occupation category that expands the fastest: from 13 per cent in 1990 to close to 16 per cent by 2001, whereas skilled blue-collar jobs are cut back from a share of 34 in 1990 to around 29 per cent by 1997. Similarly, within white-collar occupations it is again the less skill-intensive occupations that exhibit a relative gain: the share of unskilled white-collar workers rises from 16 to 18 per cent between 1990 and 1995 (with a crawling scale-back to 17 per cent until 2001), and the share of technical occupations increases from 20 in 1990 to 21 per cent in 1995. But the share of professional and managerial positions remains roughly constant between 16 and 17 per cent, thus losing in relative importance to less skill-intensive white-collar occupations.

This shift across the occupation profile towards less skill-intensive occupations permits a skill-upgrading workforce changeover, by which less skill-intensive jobs are being filled with more educated workers especially in the traded-goods sector. In practice, employers can achieve this workforce changeover in many ways. Employers can either reallocate workers across tasks in-house, or the economy can reallocate workers across firms and sectors, or there may be no reallocation for extended periods of time if employers pursue the workforce changeover by laying off less-skilled workers from every occupation category in the absence of compensating rehiring within the formal sector. The latter form of workforce changeover would be associated with, arguably, considerable adjustment costs to the economy. As it turns out in the next section, worker separations with little compensating rehiring elsewhere in the formal sector is prevalent.

5. WORKER REALLOCATION FLOWS

Labor demand decompositions so far have shown that there are two main components to the observed workforce changeover in Brazil over the sample period. First, there is a labor demand shift towards the least and the most skilled male workers, which can be traced back to relatively weaker declines of traded-goods industries that intensively use low-skilled labor and to relatively stronger expansions of nontraded-output industries that intensively use higher-skilled labor. Second, there is a within-industry shift towards longer-schooled workers, associated with a skill-upgrading of all occupations in traded-goods industries.

The conventional decomposition leaves unaddressed, however, how the workforce changeover comes about. To analyze how employers achieved the observed workforce changeover, actual worker flows need to be observed and comprehensive, linked employer–employee data are required. Linked employer–employee data for Brazil's economy trace individual workers across their jobs within plants, across plants within sectors, and across firm types and sectors in Brazil's formal sector.

Reallocations across tasks. Employers may choose to reallocate workers across tasks in-house. For this purpose, define an in-house job change as a change in employment between an occupation at the CBO base-group level to another base-group occupation. The 354 CBO base groups roughly correspond to the 4-digit

ISCO-88 occupations at the unit-group level.[10] Table 7.5 shows both continuing and displaced workers and tracks the workers through jobs at the annual horizon between 1986 and 1997. The task assignment pattern is remarkably stable both before and after trade liberalization. Between 86 and 87 per cent of formal-sector prime-age male workers remain in their job at the same employer. Only between 1 and 2 per cent of the workers are assigned to new occupations within the same plant. Less than 1 per cent of the workers switch plants within the same firm. Between 7 and 9 per cent of the workers change employing firm at the annual horizon. So, the bulk of successful reallocations does not take place in internal labor markets but across firms. Reallocations between exporters and nonexporters and across sectors is reported below. The remaining 3 to 4 per cent of workers (not reported in Table 7.5) are unaccounted. Those failed reallocations are also revisited shortly. Overall, the stable and minor percentages of occupation and plant reassignments within employers suggest that the observed workforce changeovers, documented in the preceding section, are not achieved through job reassignments in internal labor markets.

Table 7.5: *Annual occupation continuations and transitions 1986–97*

Year $t+1$	Year t (in %)	1986 (1)	1988 (2)	1990 (3)	1992 (4)	1994 (5)	1996 (6)
Employed in same occupation		86.7	85.9	86.4	85.9	85.0	85.6
at same establishment in new occupation		1.8	1.8	1.9	2.0	2.0	1.3
at same firm but new establishment		0.7	0.6	0.6	0.7	0.6	0.5
at new firm		7.9	8.4	7.4	7.8	8.7	8.3

Source: Muendler (2008). RAIS 1986–97 (1-percent random sample), male workers, 25 years or older.

Note: Frequencies based on last employment of year (highest paying job if many); continuations at same firm exclude continuations at same plant. Occupations are defined at the CBO 3-digit base-group level with 354 categories, which roughly correspond to the 4-digit ISCO-88 unit-group level.

Reallocations across firms and sectors. Between 1990 and 1998, around 6 per cent of the formal-sector workforce nationwide is employed at primary-sector nonexporters, 1 per cent at primary-sector exporters, 11 per cent at manufacturing nonexporters, and 12 per cent at manufacturing exporters. The remaining 70 per cent of the workforce is employed in the nontraded sector. Looking beyond internal labor markets, linked employer–employee data permit an investigation into whether and how the relative expansion of certain traded-goods industries, in the wake of an overall decline of the traded-goods sector, is associated with reallocations of individual workers across firms and sectors. To capture differences in the labor demand responses across subsectors and firms within the traded-goods sector, the following tabulations track individual workers across exporting and nonexporting employers in the primary and manufacturing industries.

[10] For a description of the Brazilian occupation classification system CBO and a mapping to ISCO-88, see Muendler et al. (2004).

Table 7.6 shows worker transitions between firms and sectors over the first year after trade reform, between their last observed formal-sector employment in 1990 and their last observed formal-sector employment in 1991. Only workers who experience a separation from their last employment of the year are included in the transition statistics. Trade theory might lead one to expect a shift of displaced workers from nonexporting firms to exporters following trade reform. Although manufacturing exporters are only about 5 per cent of firms during the 1990s, they employ about half the manufacturing workforce. The dominant share of successful reallocations of former non-exporter workers within the traded-goods industries, however, is to non-exporters again. Among the former non-exporter workers displaced from primary-sector employment, close to 11 per cent are rehired at primary non-exporters and 10 per cent at manufacturing non-exporters, but less than 2 per cent shift to exporters. Among the former non-exporter workers in manufacturing, 19 per cent move to manufacturing non-exporters and 7 per cent to manufacturing exporters, and a very small share to primary-sector firms. Former exporter workers, in contrast, mostly transition to new formal-sector jobs within the sector of displacement and are roughly equally likely to find reemployment at an exporter or a non-exporter. These patterns suggest that reallocations within the traded goods sectors are mostly intra-sector reallocations from exporter to exporter and from non-exporter to non-exporter—contrary to what classic trade theory with full employment and only traded goods might lead us to expect.

Table 7.6: *Year-over-year firm and sector transitions, 1990–91*

		Primary		Manufacturing				
	To:	Nonexp.	Exp.	Nonexp.	Exp.	Nontraded	Failure	Total
From:	(in %)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Nonexporter		10.7	.7	10.3	1.2	40.3	36.8	100.0
Primary Exporter		6.7	6.7	3.3	3.3	45.0	35.0	100.0
Manufact. Nonexporter		1.4	.1	19.3	7.2	34.9	37.1	100.0
Manufact. Exporter		1.2	.1	14.5	15.5	33.5	35.2	100.0
Nontraded		1.3	.0	5.4	2.4	54.8	36.0	100.0
Failure		2.9	.3	13.2	5.6	78.0	.	100.0
Total		2.1	.2	10.1	4.8	59.7	23.2	100.0

Source: Muendler (2008). RAIS 1990–91 (1-percent random sample), male workers nationwide, 25 to 64 years old. SECEX 1990–91 for exporting status.

Note: Frequencies are job accessions in Brazil within one year after separation, based on last employment of year (highest paying job if many). Failed accessions are separations followed by no formal-sector accessions anywhere in Brazil within a year, excluding workers with prior retirement or death, or age 65 or above in earlier job.

In the initial year after trade reform, between one third and two-fifths of displaced traded-sector workers with a successful reallocation end up in nontraded-sector jobs. An equally large fraction, however, fail to experience a successful reallocation to any formal-sector job within the following calendar year (retirements, deaths, and workers at or past retirement age are excluded from the

displaced worker sample).[11] Of the workers with a failed reallocation before year-end 1990, by far the largest fraction (of 78 per cent) with a successful reallocation by year-end 1991 find employment in the nontraded-sector. In summary, at the time of the largest impact of trade liberalization in 1990–91, traded-goods industries exhibit little absorptive capacity for displaced workers compared to nontraded-output industries and compared to the prevalence of failed transitions out of the formal sector. Among those failed reallocations can be transitions to informal work, unemployment, or withdrawals from the active labor force, which are not directly observed in the RAIS records.[12]

In comparison, Table 7.7 tracks annual transitions six years after the beginning of trade liberalization and three years after its conclusion. By 1996–97, more firm and sector reallocations from the primary sector are directed to jobs within the traded-goods sector. In the manufacturing sector, however, the dominant destination sector of displaced workers remains the nontraded sector in 1996–97, both for workers from exporters and for workers from non-exporters. As in the initial period 1990–91, in 1996–97 former non-exporter workers most frequently find reemployment at non-exporter firms, and former exporter workers are roughly equally likely to find reemployment at exporter and non-exporter firms in manufacturing but less likely to transition to an exporter in the primary sector. By 1996–97, an even larger fraction of displaced primary-sector workers than in 1990–91 fail to experience a successful formal-sector reallocation and a roughly equally large share of former manufacturing workers as in 1990–91 fail to find a formal-sector job within the following calendar year.

Table 7.7: *Year-over-year firm and sector transitions 1996–7*

		Primary		Manufacturing				
From:	To: (in %)	Nonexp. (1)	Exp. (2)	Nonexp. (3)	Exp. (4)	Nontraded (5)	Failure (6)	Total (7)
Primary Nonexporter		32.1	2.5	6.0	2.9	15.4	41.1	100.0
Primary Exporter		17.1	13.0	6.5	3.3	18.7	41.5	100.0
Manufact. Nonexporter		5.6	.4	18.9	6.5	32.1	36.5	100.0
Manufact. Exporter		7.2	.7	12.1	13.9	27.3	38.8	100.0
Nontraded		1.3	.2	3.8	2.0	55.8	36.9	100.0
Failure		8.9	.7	12.2	6.1	72.1	.	100.0
Total		6.5	.6	8.8	4.7	56.9	22.5	100.0

Source: Muendler (2008). RAIS 1996–97 (1-percent random sample), male workers nationwide, 25 to 64 years old. SECEX 1996–97 for exporting status.

Note: Frequencies are job accessions in Brazil within one year after separation, based on last employment of year (highest paying job if many). Failed accessions are separations followed by no formal-sector accessions anywhere in Brazil within a year, excluding workers with prior retirement or death, or age 65 or above in earlier job.

[11] The slightly smaller unaccounted percentage in Table 7.5 compared to the reallocation failure rates in Tables 7.6 and 7.7 is largely due a restriction of the initial sample to workers with comprehensive occupation information in Table 7.5.

[12] For evidence on those work status transitions using household survey data, see Menezes-Filho and Muendler (2007).

Together with the evidence on infrequent task reassignments in-house, these labor market transitions suggest that the observed workforce changeovers from the preceding section are neither achieved through worker reallocations within employers nor are they brought about by labor reallocations across employers and sectors. By exclusion, the remaining explanation is that formal-sector employers in the traded-goods industries shrink their workforces by dismissing less-schooled workers more frequently than more schooled workers, while the thus displaced workers fail to find reemployment at least at the annual horizon. In the aggregate, the lacking traded-sector reallocations result in a considerable decline of formal manufacturing employment from 26 to 22 per cent (Table 7.2). The simultaneous expansion of nontraded-output industries can partly be driven by a long-term shift from primary to manufacturing to services activities in the economy, or by trade liberalization if fast productivity change reduces manufacturing employment in favor of non-traded sector employment, or by Brazil's overvalued real exchange rate during the sample period, or by foreign direct investment (FDI) flows in the wake of Brazil's concomitant capital-account liberalization and privatization programme, or by a combination of these changes. The next section turns to the predictive power of these competing explanations and their associated variables, using linked employer–employee data at the job level.

6. TRADE-RELATED WORKER SEPARATIONS AND ACCESSIONS

Employers adjust workforces through worker separations and accessions. A separation is defined as a worker's quit or layoff from the last formal employment in the calendar year. Among the separations, quits are infrequent compared to layoffs (Table 7.3).[13] Conversely, an accession is defined as a worker's hiring into the first formal employment in the calendar year. Separations in turn fill, and accessions empty the pool of workers to be reallocated.

To understand determinants of labor reallocation in the formal sector, regression analysis can simultaneously condition on industry, plant, job, and worker characteristics as explanatory variables for separations and accessions. Consider the probability that an employer–employee match is terminated (a separation) or is formed (an accession), conditional on a worker-fixed component that is observable to the employer and the worker:

$$Pr\left(\sigma_{i,t} \mid \mathbf{x}_{i,t}, \mathbf{y}_{J(i),t}, \mathbf{z}_{S(J(i)),t}\right) = \frac{\exp\{\mathbf{z}_{S(J(i)),t}\beta_z + \mathbf{y}_{J(i),t}\beta_y + \mathbf{x}_{i,t}\beta_x + \alpha_i + \alpha_t\}}{1+\exp\{\mathbf{z}_{S(J(i)),t}\beta_z + \mathbf{y}_{J(i),t}\beta_y + \mathbf{x}_{i,t}\beta_x + \alpha_i + \alpha_t\}}, \tag{3}$$

where $\sigma_{i,j}$ denotes the binary outcome (separation or not, accession or not) for worker i at time t. $\mathrm{z}_{S(J)(i),t}$ is a vector of sector-level covariates of the worker's displacing or hiring sector $S(J)(i)$, including a sector-fixed effect in some specifications; $\mathrm{y}_{J(i),t}$ is a vector of plant-level covariates of worker i's displacing

[13] Separations are treated as a single category for regression analysis, where no marked differences between quits and layoffs for trade-related predictors can be detected.

or hiring plant $J(i)$; x_{it} is a vector of covariates that are worker, job or match specific; β_z, β_y, β_x are coefficient vectors; α_i is a worker-fixed effect and α_i a year effect. There is an unobserved error to terminations and formations of employer–employee matches. For theoretical consistency with random shocks to employer–employee matches, the disturbance is assumed to be logistic and independent across matches. This conditional logit model equation (3) is fit using conditional maximum likelihood estimation (the full maximum likelihood estimator is inconsistent). Identification of worker fixed effects requires restriction of the sample to workers who experience at least one separation or accession. Coefficients on worker and job covariates are identified from time variation within and across employers. Educational attainment changes little among prime-age males, however. Consequently, education categories are dropped from the worker characteristics vector but educational workforce composition shares are kept among the plant-level regressors. When inferring separations and accessions in this and subsequent sections, transfers across plants within the same firm, as well as retirements and reported deaths on the job are excluded.

Table 7.8 presents conditional logit estimates of separations from formal manufacturing jobs, where the conditioning removes worker-fixed effects (worker-FE logit) and year effects. For comparison, the first five columns present regressions without sector fixed effects so that sector-specific variables such as comparative advantage (which varies little over time) can be kept among the regressors. Separations are significantly more frequent in sectors with a stronger comparative advantage and at exporters—contrary to predictions of standard trade theory. Elevated product tariffs predict lower separation rates from formal jobs (though only significant at the 10 per cent level), but high input tariff barriers are associated with significantly higher separation rates. Note that high input tariffs reduce a plant's effective protection from foreign competition (Corden 1966; Anderson 1998) because high input prices exert competitive pressure. Similarly, additional import penetration predicts significantly higher displacement odds. When including observed market penetration with imports to proxy for changing non-tariff barriers and all earlier trade related predictors, point estimates and statistical significance of coefficients are hardly affected as the specification is gradually enriched (moving from column 1 to column 6). FDI inflows into the sector predict a statistically significant reduction in displacement rates. The sectoral real exchange and the Herfindahl concentration index have no significant predictive power after conditioning on year effects.

When year indicators are excluded from the regression (column 5), comparative advantage and exporting status become even stronger predictors of displacements. Tariffs and import penetration coefficients now also reflect the effect of reducing trade barriers over time and predict that reduced barriers both at the input and the output margin, and the arrival of additional imports, are associated with more worker separations. Using further controls—such as the inflation rate in addition to sectoral price levels behind the real exchange rate, FDI stocks in addition to FDI flows, and controls for privatization and outsourcing—beyond the large set of sector- and firm-level variables that already

control for time-varying changes to the competitive environment does not change coefficients in important ways.

Inclusion of sector fixed effects removes unobserved sectoral differences that potentially co-determine separations (column 6). The sector effects control for potential differences in the effect of labor institutions, for instance, whose reform in 1988 precedes trade liberalization in 1990. As expected, inclusion of sector indicators turns the coefficient on comparative advantage, which is highly sector-specific and largely time-invariant, insignificant at the five per cent significance level. For the other trade regressors, however, coefficient estimates increase in absolute value (compared to column 4) and remain highly significant. In subsequent discussion, this chapter emphasizes the more conservative estimates without sector effects.

Before discussing plant and worker-level variables, turn to the opposite margin: Table 7.9 presents conditional logit estimates of accessions into formal manufacturing jobs, controlling for worker-fixed accession effects. Mirroring the signs from separation regressions, accession rates are lower in sectors with stronger comparative advantage, when other trade-related variables are controlled for (column 4). The coefficient is not statistically significant at conventional levels in this regression (but will become statistically significant when controlling for higher-order interactions between trade variables in Table 15). Exporters exhibit significantly lower accession rates, mirroring their higher separation rates. Elevated product tariffs predict significantly more accessions, mirroring the sign from separation regression, whereas higher intermediate-input tariffs predict significantly fewer accessions, also mirroring the sign from separation regression. Import penetration has no statistically significant effect, and neither does the real exchange rate. FDI inflows are associated with significantly more accessions and more concentrated manufacturing industries exhibit fewer accessions.

When year effects are omitted (column 5), comparative advantage and exporting status become even stronger predictors of reduced accessions. Tariffs and import penetration coefficients now also reflect the effect of reducing trade barriers over time. Lower input tariffs, which tend to make competition less fierce, predict more accessions. Lower output tariffs and the arrival of additional imports, which tend to make competition more fierce, are associated with fewer accessions. When conditioning on both year and sector effects (column 6), the largely time-invariant a sector-specific comparative advantage variable does expectedly not turn significant, whereas coefficients for all other trade regressors increase in absolute value (compared to column 4) and remain or become highly significant. As for separations, this chapter therefore bases much of the subsequent discussion on the more conservative estimates without sector effects.

Larger manufacturing plants offer more employment stability: they displace significantly fewer (Table 7.8) and they hire significantly fewer workers (Table 7.9). Plants with less educated workforces and more blue-collar jobs separate from workers significantly less frequently and hire significantly more frequently. Interestingly, workers with a longer tenure at the plant and longer labor market experience suffer significantly more frequent separations at the separation

Table 7.8: *Worker–fixed effect logit estimation of separations*

	(1)	(2)	(3)	(4)	(5)	(6)
Balassa Comp. Adv.	.080			.169	.204	–.094
	(.021)***			(.024)***	(.023)***	(.049)*
Exporter Status		.289		.283	.301	.284
		(.028)***		(.028)***	(.028)***	(.028)***
Product Market Tariff			–.104	–.705	–1.383	–2.361
			(.416)	(.426)*	(.410)***	(.476)***
Intm. Input Tariff			1.601	2.880	–1.420	5.149
			(.633)**	(.678)***	(.553)**	(.748)***
Import Penetration				1.257	6.035	3.227
				(.388)***	(.349)***	(.638)***
Sector–level covariates						
Sector real exch. rate	.733	.843	.353	–.398	.213	–1.224
	(.624)	(.626)	(.640)	(.645)	(.069)***	(.699)*
FDI Flow (USD billion)	–.025	–.012	–.018	–.048	.047	–.039
	(.020)	(.020)	(.020)	(.020)**	(.019)**	(.020)**
Herfindahl Index (sales)	–.371	–.517	–.399	–.354	.929	.881
	(.317)	(.316)	(.329)	(.343)	(.320)***	(.639)
Plant–level covariates						
Log Employment	–.343	–.370	–.341	–.377	–.410	–.383
	(.011)***	(.011)***	(.011)***	(.011)***	(.011)***	(.011)***
Share: Middle School or less	–.750	–.658	–.719	–.663	–.793	–.692
	(.131)***	(.131)***	(.131)***	(.132)***	(.129)***	(.132)***
Share: Some High School	–.444	–.392	–.440	–.393	–.214	–.413
	(.148)***	(.148)***	(.147)***	(.148)***	(.145)	(.148)***
Share: White–collar occ.	.721	.700	.739	.691	.552	.683
	(.075)***	(.074)***	(.074)***	(.075)***	(.073)***	(.075)***
Worker–level covariates						
Tenure at plant (in years)	1.367	1.350	1.362	1.351	1.390	1.351
	(.036)***	(.036)***	(.036)***	(.036)***	(.037)***	(.036)***
Pot. labor force experience	.006	.006	.006	.006	.031	.006
	(.002)**	(.002)**	(.002)**	(.002)**	(.002)***	(.002)**
Unskilled Wh. Collar Occ.	–.256	–.251	–.259	–.262	–.199	–.267
	(.067)***	(.067)***	(.067)***	(.067)***	(.065)***	(.067)***
Year effects	yes	yes	yes	yes		yes
Sector effects						yes
Obs.	145,408	145,408	145,408	145,408	145,408	145,408
Pseudo R^2	.148	.149	.148	.150	.137	.151

Source: Menezes-Filho and Muendler (2007). RAIS 1990–98 (1 percent random sample), male workers nationwide, 25 to 64 years old, with manufacturing job.

Note: Separations exclude transfers, deaths, and retirements. Reference observations are employments with no reported separation in a given year. Sector information at subsector IBGE level. Professional or managerial occupations and skilled blue collar occupations (not reported) not statistically significant at five-percent level. Robust standard errors in parentheses: *significance at ten, **five, ***one percent.

Table 7.9: *Worker–fixed effect logit estimation of accessions*

	(1)	(2)	(3)	(4)	(5)	(6)
Balassa Comp. Adv.	.041			–.016	–.114	–.067
	(.017)**			(.020)	(.019)***	(.048)
Exporter Status		–.449		–.439	–.429	–.438
		(.027)***		(.027)***	(.026)***	(.027)***
Product Market Tariff			1.306	1.246	2.474	1.822
			(.379)***	(.393)***	(.379)***	(.498)***
Intm. Input Tariff			–3.258	–3.073	–3.846	–2.954
			(.540)***	(.598)***	(.514)***	(.750)***
Import Penetration				.198	–3.919	1.764
				(.355)	(.307)***	(.665)***
Sector–level covariates						
Sector real exch. rate	–1.264	–.955	–.953	–.810	.038	–.844
	(.605)**	(.606)	(.626)	(.639)	(.076)	(.718)
FDI Flow (USD billion)	.039	.047	.056	.058	.031	.058
	(.022)*	(.021)**	(.021)***	(.022)***	(.021)	(.022)***
Herfindahl Index (sales)	–.348	–.344	–.795	–.788	–2.335	–.838
	(.268)	(.268)	(.282)***	(.297)***	(.277)***	(.655)
Plant–level covariates						
Log Employment	–.190	–.140	–.189	–.141	–.112	–.138
	(.008)***	(.009)***	(.008)***	(.009)***	(.008)***	(.009)***
Share: Middle School or less	.947	.857	.940	.850	.828	.849
	(.107)***	(.105)***	(.107)***	(.105)***	(.104)***	(.105)***
Share: Some High School	.740	.667	.739	.668	.468	.668
	(.124)***	(.122)***	(.124)***	(.122)***	(.120)***	(.122)***
Share: White–collar occ.	–.675	–.614	–.679	–.621	–.534	–.625
	(.067)***	(.067)***	(.067)***	(.067)***	(.064)***	(.067)***
Worker–level covariates						
Prof. or Manag'l. Occ.	–.801	–.807	–.801	–.807	–.827	–.810
	(.068)***	(.068)***	(.068)***	(.068)***	(.066)***	(.068)***
Tech'l. or Superv. Occ.	–.603	–.610	–.597	–.604	–.623	–.601
	(.064)***	(.064)***	(.064)***	(.064)***	(.062)***	(.064)***
Unskilled Wh. Collar Occ.	–.490	–.497	–.488	–.495	–.519	–.497
	(.061)***	(.062)***	(.062)***	(.062)***	(.060)***	(.062)***
Skilled Bl. Collar Occ.	–.417	–.413	–.413	–.410	–.443	–.410
	(.032)***	(.032)***	(.032)***	(.032)***	(.031)***	(.032)***
Year effects	yes	yes	yes	yes		yes
Sector effects						yes
Obs.	112,974	112,974	112,974	112,974	112,974	112,974
Pseudo R^2	.036	.040	.037	.041	.026	.042

Source: Menezes-Filho and Muendler (2007). RAIS 1990–98 (1 percent random sample), male workers nationwide, 25 to 64 years old, with manufacturing job.

Note: Accessions exclude transfers. Reference observations are employments with no reported accession. Sector information at subsector IBGE level. Robust standard errors in parentheses: *significance at ten, **five, ***one percent.

margin. This fact is consistent with the hypothesis that Brazilian firing costs, which proportionally increase with tenure, lead employers to shorten tenure through displacement. Workers in occupations of intermediate skill intensity experience significantly fewer separations, and workers are significantly less likely to be hired into high-skill-intensive manufacturing occupations (with a monotonic drop in accession odds as an occupation's skill intensity increases). Year effects are significant at the 1 per cent level and show both a strictly monotonic increase in manufacturing separations and a strictly monotonic drop in manufacturing accessions.

Worker heterogeneity is an important predictive component of separations and accessions. A comparison between conditional and unconditional logit estimation (not reported here) shows that regressions are highly sensitive to the omission of worker fixed effects. The relevance of conditional worker effects is consistent with the hypothesis that the termination and formation of employer–employee matches is not random, even after controlling for a comprehensive set of observable worker and employer characteristics.

The evidence so far shows that Brazil's trade reform predicts salient changes to worker separations and accessions. But neither comparative advantage sectors nor exporters exhibit the expected labor absorption; they separate from their workers significantly more frequently than other sectors and firms. Exporters also hire significantly less frequently. There are empirical concerns for these predictions of worker flows such as the potential simultaneity of trade policies and exporting status, and the relevance of Brazil's concomitant reforms. Those issues are addressed in Menezes-Filho and Muendler (2007) and refined subsequent research.

7. EXPLANATIONS OF REALLOCATION FLOWS

A strong candidate explanation for the reverse labor flows away from comparative-advantage sectors and away from exporters is endogenous change in productivity. Several case studies have documented for various countries in the context of trade liberalization episodes and other structural reforms that within-firm productivity rises in response to the removal of trade protection (for example, Levinsohn 1993; Hay 2001; Pavcnik 2003; Schor 2003; Eslava et al. 2004; Fernandes 2007; Muendler 2004). Exporters are more productive than non-exporters, as Table 7.3 has documented. If trade triggers faster productivity growth for exporters and in comparative-advantage industries because for these firms and industries larger market potential offers stronger incentives to improve efficiency, and if productivity increases faster than production, then labor will flow away from comparative-advantage sectors and away from exporters. Production, and market shares, increase less than proportionally with productivity if the elasticity of demand is less than unity in absolute value. As a result, output shifts to more productive firms but labor does not. Menezes-Filho and Muendler (2007) offer detailed evidence on this adjustment process.

7.1 Trade theories

While there is no single workhorse model for unilateral trade reforms and endogenous productivity changes in response, recent trade theories investigate industry dynamics when trade costs drop worldwide in lock step and firms simultaneously engage in innovation and export-market participation.[14] Yeaple (2005) shows in a static model with *ex ante* identical firms and heterogeneous workers, whose skill is complementary to innovative technology, that the firms' binary choice of process innovation induces the sorting of more skilled workers to innovative firms, leading to firm heterogeneity *ex post* and to increased within-firm productivity in equilibrium. As multilateral trade costs drop, more firms in the differentiated-goods sector adopt innovative technology and raise their employment, hiring away the top-skilled workers from differentiated-goods producers with lower technology. Also considering *ex ante* identical firms, Ederington and McCalman (2008) allow for a continuous technology choice in a dynamic industry equilibrium model and show that a drop in foreign trade costs raises the rate of technology adoption at exporters but delays it at non-exporters. Departing from *ex ante* heterogeneous firms, Costantini and Melitz (2008) reintroduce a stochastic productivity component from Hopenhayn (1992) into the Melitz (2003) model and allow firms to choose process innovation. In simulations of the dynamic industry equilibrium, an announced future reduction of multilateral trade costs leads firms to adopt innovation in advance while waiting for export market participation.

The mechanism by which productivity increases in these models is that globally reduced trade costs raise the returns from accessing the export market so that firms, which can both choose export-market participation and engage in innovation, adopt innovative technology because each activity raises the return to the other. Globally reduced trade costs are a carrot. Under a unilateral trade reform, in contrast, expected profits for domestic producers fall, potentially reducing incentives for innovation. So, unilateral trade reform is a stick. But it is a long-standing tenet in economics that product market competition may discipline managers and workers by strengthening incentives in the respective principal–agent relationships. Stronger product market competition may lead principals to become better informed (Hart 1983), induce managers to exert more effort to avert bankruptcy (Schmidt 1997), or lead surviving firms to strengthen incentives because induced exits of other firms raise profit opportunities (Raith 2003). This family of models, though never explicitly embedded in a trade context (or a general equilibrium context for that matter), offers a key explanation of why firms may improve productivity in response to unilateral trade reform.

While endogenous productivity change in response to unilateral trade reform is absent from much of trade theory, classic and recent trade models nevertheless offer numerous predictions that are consistent with Brazil's experience. A particularly attractive model for empirical work is the Bernard et al. (2007)

[14] As Arbache et al. (2004) have argued in the context of relative wage responses to trade before that standard trade theory ignores trade-induced technology adoption and implied relative labor demand changes.

framework, which embeds heterogeneous firms in a classic trade model and derives predictions for labor turnover. Their setting preserves the prediction from classic trade theory that there is net job creation in comparative-advantage industries and net job destruction in disadvantage industries. In the presence of productivity heterogeneity across firms, however, important differences between gross and net job creation and destruction result. In disadvantage industries, where there is net job destruction, high-productivity firms expand to serve the export market and create new jobs. In comparative-advantage industries, where there is net job creation, existing jobs are destroyed at low-productivity firms.[15]

7.2 Potential implications for adjustment costs

The reported estimates owe their generality and robustness to a lean set of identifying assumptions. For the estimates in section 6, no structural assumption was needed other than that unobserved match-specific logistic shocks trigger separations or accessions beyond the observed variables. For the precise measurement of labor market adjustment costs that are associated with trade reform, more explicit structural assumptions are required to model reallocation delays and failures in general equilibrium.

A chief concern of reallocation costs relates to potentially idle labor, and especially to displaced workers who await formal-sector reallocation. Workers awaiting reallocation are not directly observable in formal-sector worker censuses. However, the Brazilian RAIS record changes at two margins that alter the pool of prime-age male workers to be reallocated: separations from formal jobs fill the pool, and accessions into formal jobs empty the pool of workers to be reallocated.

So two important measures for the potential idleness of labor are the rate of failed reallocations within a given time period, such as four years (48 months) following displacement, and the average durations of successful reallocations within the given time period. Numerous economic causes can be responsible for changes to the rate of failed reallocations and changes to the durations of successful reallocations. Menezes-Filho and Muendler (2007) document that the share of displaced workers without reallocation for four years increases from 18 percent to 22 percent between 1989 and 1997 and does not subside again during the 1990s. There is some variation in the failure rate across skill groups within any given year: young and college-educated workers' reallocations fail less frequently than average. Time variation, however, dwarfs the skill-group differences. A similar pattern applies to durations of successful reallocations, which increase from an average of 6.3 months (out of 48 months maximally) in 1989 to 9.5 months in 1997 and also never lastingly drop back to a pre-1990s level. The relatively minor cross-sectional differences between skill groups, compared to major time variation, suggests that studying macroeconomic sources of variation in labor-market performance promises to uncover first-order changes in labor-market outcomes. Refined estimates following Menezes-Filho and Muendler (2007), and tariff-predicted

[15] Formally, existing jobs are destroyed at low-productivity firms that exit. But a firm exit could also be interpreted as a plant closure within a firm, or as the shutdown of a product line within a plant.

changes to separations and accessions, suggest that trade reform is one such major contributor among the macroeconomic changes that alter labor-market outcomes.

8. CONCLUSION

Brazil's labor market adjustment after large-scale trade reform in the early 1990s offers important insights into prospective reallocation shifts and potential adjustment costs from rising reallocation durations and failure rates. A conventional labor demand decomposition documents two salient workforce changeovers. Within the traded-goods sector, there is a marked occupational downgrading and a simultaneous educational upgrading, by which employers fill expanding low-skill-intensive occupations with increasingly educated jobholders. Between sectors, there is a labor demand shift towards the least and the most skilled, which can be traced back to relatively weaker declines of traded goods industries that intensely use low-skilled labor and to relatively stronger expansions of nontraded-output industries that intensively use high-skilled labor. These observations are broadly consistent with predictions of classic trade theory for a low-skill abundant economy.

Actual worker flows, however, reveal a much more nuanced picture. Rich linked employer–employee data show that workforce changeovers are neither achieved through worker reassignments to new tasks within employers, nor are they brought about by reallocations across employers and traded-goods industries. Instead, trade-exposed industries shrink their workforces by dismissing less-schooled workers more frequently than more-schooled workers. Most displaced workers shift to nontraded-output industries or out of formal employment. Brazil's trade liberalization triggers worker displacements particularly from protected industries, as trade theory predicts and welcomes. But neither comparative-advantage industries nor exporters absorb trade-displaced workers. To the contrary, comparative-advantage industries and exporters displace significantly more workers and hire fewer workers than the average employer, and resource reallocation appears to remain incomplete for years. Menezes-Filho and Muendler (2007) argue that these patterns are best explained by relatively fast labor productivity increases at exporters and in comparative advantage industries. Employers in those activities raise productivity in response to heightened competition and market opportunities, and they do so faster than non-exporters and firms in disadvantage industries, because expected exporting activity increases the return to innovation. As a result, product market shares shift to more productive firms. Product market shares grow less than proportionately with productivity, however, so that trade-induced productivity change leads to labor savings at exporters and in comparative advantage industries.

The labor market evidence for Brazil is also tsuggestive of a novel explanation why pro-competitive reforms might be associated with strong efficiency gains at the employer level but not in the aggregate. If idle resources result from sluggish reallocation, their foregone wage bills can be a drain on GDP. It remains to rigorously estimate the adjustment costs implied by foregone GDP. Those adjustment costs will then have to be set against the repeated static gains from

trade. A promising path for future research is the rigorous theoretical modelling and empirical measurement of adjustment cost in the labor market in response to trade integration.

Marc-Andreas Muendler is an Associate Professor in the Department of Economics of the University of California, San Diego.

APPENDIX A. LINKED EMPLOYER-EMPLOYEE DATA

The main data source underlying statistics in this chapter is linked employer–employee data. Brazilian law requires every Brazilian plant to submit detailed annual reports with individual information on its employees to the ministry of labor (*Ministério de Trabalho*, MTE). The collection of the reports is called *Relação Anual de Informações Sociais*, or RAIS, and typically concluded at the parent firm by late February or early March for the preceding year of observation. RAIS primarily provides information to a federal wage supplement program (*Abono Salarial*), by which every worker with formal employment during the calendar year receives the equivalent of a monthly minimum wage. RAIS records are then shared across government agencies. An employer's failure to report complete workforce information can result in fines proportional to the workforce size; but fines are seldom issued. A strong incentive for compliance is that workers' benefits depend on RAIS, so that workers follow up on their records. The payment of the worker's annual public wage supplement is exclusively based on RAIS records. The Ministry of Labor estimates that currently 97 per cent of all formally employed workers in Brazil are covered in RAIS, and that coverage exceeded 90 per cent throughout the 1990s.

Observation screening. In RAIS, workers are identified by an individual-specific PIS (*Programa de Integração Social)* number that is similar to a social security number in the United States (but the PIS number is not used for identification purposes other than the administration of the wage supplement program *Abono Salarial*). A given plant may report the same PIS number multiple times within a single year in order to help the worker withdraw deposits from that worker's severance pay savings account (*Fundo de Garantia do Tempo de Serviço*, FGTS) through spurious layoffs and rehires. Bad compliance may cause certain PIS numbers to be recorded incorrectly or repeatedly. To handle these issues, I screen RAIS in two steps: (1) observations with PIS numbers shorter than 11 digits are removed. These may correspond to informal (undocumented) workers or measurement error from faulty bookkeeping; (2) for several separation statistics, I remove multiple jobs from the sample if a worker's duplicate jobs have identical accession and separation dates at the same plant. For a worker with such multiple employments, I only keep the observation with the highest average monthly wage level (in cases of wage ties, I drop duplicate observations randomly).

Experience, education and occupation categories. For the years 1986–93, RAIS reports a worker's age in terms of eight age ranges. For consistency, age in years is categorized into those eight age ranges also for 1994–2001 when precise age

would be available. I construct a proxy for potential workforce experience from the nine education categories and the mean age within a worker's age range. For example, a typical Early Career worker (34.5 years of age) who is also a Middle School Dropout (left school at 11 years of age) is assigned 23.5 years of potential workforce experience.

The following tables present age and education classifications from RAIS, along with the imputed ages used in construction of the potential experience variable. I use the age range information in our version of RAIS to infer the 'typical' age of a worker in the age range as follows:

RAIS Age Category		Imputed Age
1.	Child (10–14)	excluded
2.	Youth (15–17)	excluded
3.	Adolescent (18–24)	excluded
4.	Nascent Career (25–29)	27
5.	Early Career (30–39)	34.5
6.	Peak Career (40–49)	44.5
7.	Late Career (50–64)	57
8.	Post Retirement (65–)	excluded

For regression analysis, our education variable regroups the nine RAIS education categories into four categories as follows:

Education Level		RAIS Education	Years
1.	Illiterate, or Primary or Middle School Educated	1–5	0–8
2.	Some High School or High School Graduate	6–7	8–12
3.	Some College	8	12+
4.	College Graduate	9	16+

Occupation indicators derive from the 3-digit CBO classification codes in our nationwide RAIS data base, and are reclassified to conform to the ISCO-88 categories.[16] I map ISCO-88 categories to RAIS occupations as follows:

Isco-88 Category		Occupation Level
1.	Legislators, senior officials, and managers	Professional & Managerial
2.	Professionals	Professional & Managerial
3.	Technicians and associate professionals	Technical & Supervisory
4.	Clerks	Other White Collar
5.	Service workers and shop and market sales workers	Other White Collar
6.	Skilled agricultural and fishery workers	Skill Intensive Blue Collar
7.	Craft and related workers	Skill Intensive Blue Collar
8.	Plant and machine operators and assemblers	Skill Intensive Blue Collar
9.	Elementary occupations	Other Blue Collar

[16] See documentation at URL www.econ.ucsd.edu/muendler/brazil

Employment

Table A7.1 shows the employment allocation by industry in the universe of RAIS workers in 1986, 1990 and 1997.

Table A7.1: *Employment allocation by subsector*

Sector and subsector IBGE		Employment share 1986 (1)	1990 (2)	1997 (3)
Primary				
1	Mining and quarrying	.007	.006	.004
25	Agriculture, farming, hunting, forestry and fishing	.015	.016	.041
Manufacturing				
2	Manufacture of non-metallic mineral products	.016	.013	.011
3	Manufacture of metallic products	.030	.024	.021
4	Manufacture of machinery, equipment and instruments	.020	.016	.011
5	Manufacture of electrical and telecommunications equipment	.016	.014	.008
6	Manufacture of transport equipment	.019	.016	.013
7	Manufacture of wood products and furniture	.019	.015	.015
8	Manufacture of paper and paperboard, and publishing	.014	.014	.013
9	Manufacture of rubber, tobacco, leather, and products n.e.c.	.019	.016	.009
10	Manufacture of chemical and pharmaceutical products	.024	.022	.020
11	Manufacture of apparel and textiles	.042	.035	.026
12	Manufacture of footwear	.012	.010	.008
13	Manufacture of food, beverages, and ethyl alcohol	.040	.039	.041
Commerce				
16	Retail trade	.106	.103	.127
17	Wholesale trade	.024	.025	.027
Services				
18	Financial intermediation and insurance	.038	.034	.025
19	Real estate and business services	.074	.073	.079
20	Transport, storage and telecommunications	.050	.044	.057
21	Hotels and restaurants, repair and maintenance services	.101	.101	.084
22	Medical, dental and veterinary services	.014	.017	.039
23	Education	.008	.009	.036
Other				
14	Electricity, gas and water supply	.013	.014	.014
15	Construction	.045	.041	.049
24	Public administration and social services	.209	.206	.224
26	Activities n.e.c.	.025	.077	.001
Total employment (thousands of workers)		22,164	23,174	24,104

Source: RAIS 1986, 1990 and 1997, universe of workers.

Note: Employment on December 31st. Slight differences to Table 3.1 are due to random sampling errors.

APPENDIX B. ADDITIONAL DATA SOURCES

Throughout this chapter, I draw on additional data sources beyond RAIS. I use productivity measures from Brazil's annual manufacturing firm survey PIA (*Pesquisa Industrial Anual*) for 1986–98. PIA is a representative sample of all but the smallest manufacturing firms, collected by Brazil's statistical bureau IBGE. I first obtain log TFP measures from Olley and Pakes (1996) estimation at the *Nível 50* sector level under a Cobb-Douglas specification (Muendler 2004). I then convert log TFP to log labor productivity by adding the production-coefficient weighted effects of capital accumulation and intermediate input use. Labor productivity is denominated in BRL-deflated USD-1994 output equivalents per worker. IBGE's publication rules allow data from PIA to be withdrawn in the form of tabulations with at least three firms per entry. I construct random combinations of three firms by drawing from sector-location-year cells. A cell is defined by the firm's *Nível 50* sector, headquarters location, and pattern of observation years. I assign every PIA firm to one and only one multi-firm combination. For each three-firm combination, I calculate mean log productivity but retain the firm identifiers behind the combination—permitting the linking to RAIS (see also Menezes-Filho, Muendlerand Ramey (2008)). I infer a firm's export status between 1990 and 2001 from Brazil's customs records SECEX.

I use data on *ad valorem* tariffs by sector and year from Kume et al. (2003). The tariffs are the legally stipulated nominal rates for Brazil's trade partners with no preferential trade agreement, and not weighted by source country. I combine these tariff series with economy-wide input–output matrices from IBGE to arrive at intermediate input tariff measures by sector and year. I calculate the intermediate-input tariff as the weighted arithmetic average of the product-market tariffs, using sector-specific shares of inputs for the input–output matrix as weights. I use Ramos and Zonenschain's (2000) national accounting data to calculate market penetration with foreign imports. Arguably, domestic firms find the absorption market, corresponding to output less net exports, the relevant domestic environment in which they compete. I define the effective rate of market penetration as imports per absorption. Foreign direct investment (FDI) and annual GDP data are from the Brazilian central bank.

I construct sector-specific real exchange rates from the nominal exchange rate to the US dollar E, Brazilian wholesale price indices P_i, and average foreign price series for groups of Brazil's main trading partners P_i^* by sector i, and define the real exchange rate as $q_i \equiv EP_i^*/P_i$ so that a high value means a depreciated real sector exchange rate. I rebase the underlying price series to a value of 1 in 1995. I use Brazil's import shares from its major 25 trading partners in 1995 as weights for P_i^*. I obtain sector-specific annual series from producer price indices for the 12 OECD countries among Brazil's main 25 trading partners (sector-specific PPI series from *S*ourceOECD; US PPI series from Bureau of Labor Statistics). I combine these sector-specific price indices with the 13 annual aggregate producer (wholesale if producer unavailable) price index series for Brazil's remaining major trading partners (from Global Financial Data), for whom sector-specific PPI are not available.

APPENDIX C WAGE STRUCTURE IN MANUFACTURING

Table C7.1 presents Mincer (1974) regressions of the log wage on individual compensation components. Following Abowd et al. (2001), individual compensation in a given year is given by

$$1\mathrm{n} w_i = x_i + \psi_{J(i)} + \varepsilon_i, \tag{C1}$$

where w_i is worker i's annual wage, x_i is a vector of observable worker characteristics including gender, experience, education and occupation, β is a vector of parameters to be estimated, $\psi_{J(i)}$ is an plant effect ($j = J(i)$ being the plant that employs worker i), and ε_i is an error term. The plant effect combines a pure plant effect with the plant average of pure worker effects:

$$\psi_j = \phi_j + \overline{\alpha}_j, \tag{C.2}$$

where ϕ_j is the pure plant effect and $\overline{\alpha}_j$ is the average of pure worker effects α_i over workers employed at plant j. The plant effect controls for unobservable worker and plant characteristics. Abowd and Kramarz (1999) show that omitting this effect leads to bias in the estimation of β in general.

Regressors are potential worker experience and indicator variables for gender, education, and occupation as measures of individual characteristics. Quadratic, cubic, and quartic terms for potential experience are included. Gender is interacted with all other variables. Table C7.1 presents results for the manufacturing sector in São Paulo state in 1990 and 1997. Comparable estimates for manufacturing workers in France in 1992 and the United States in 1990 drawn from Abowd et al. (2001) are also reported.[17]

[17] Data for France derive from the Enquête sur la Structure des Salaires (ESS), which samples responses to an annual administrative census of business enterprises. Data for the United States derive from the Worker–Establishment Characteristic Database (WECD), which links individual census responses to manufacturing plants surveyed in the Longitudinal Business Database (LBD). See Abowd et al. (2001) for further details.

Table C.7.1: *Manufacturing wages in Brazil, France and the U.S.*

	Brazil 1990 (1)	Brazil 1997 (2)	France 1992 (3)	U.S. 1990 (4)
Primary School Education (or less)	−1.075	−1.000	−.338	−.526
	(.002)	(.002)	(.009)	(.008)
Some High School Education	−.923	−.881	−.256	−.404
	(.002)	(.002)	(.009)	(.007)
Some College Education	−.339	−.316	−.200	−.334
	(.003)	(.003)	(.009)	(.007)
College Graduate			−.064	−.123
			(.016)	(.007)
Professional or Managerial Occupation	.856	.912	.760	.359
	(.002)	(.002)	(.009)	(.004)
Technical or Supervisory Occupation	.600	.632	.401	.206
	(.002)	(.002)	(.007)	(.004)
Other White Collar Occupation	.262	.249	.169	−.039
	(.002)	(.002)	(.011)	(.005)
Skill Intensive Blue Collar Occupation	.239	.225	.155	.083
	(.001)	(.001)	(.007)	(.003)
Potential Labor Force Experience	.095	.082	.069	.083
	(.0005)	(.0007)	(.003)	(.002)
Quadratic Experience Term	−.003	−.003	−.004	−.003
	(.00005)	(.00007)	(.0002)	(.0001)
Cubic Experience Term	.00005	.00008	.0001	.00007
	(2.29e−06)	(2.86e−06)	(1.00e−05)	
Quartic Experience Term	−3.01e−07	−7.64e−07	−1.20e−06	−4.70e−07
	(3.24e−08)	(3.89e−08)	(1.00e−07)	(3.00e−08)
Female	.060	.070	.052	−.078
	(.005)	(.006)	(.024)	(.019)
Female × Primary School Education (or less)	.106	.051	−.0006	.041
	(.004)	(.004)	(.021)	(.016)
Female × Some High School Education	−.016	−.058	−.016	−.009
	(.004)	(.004)	(.021)	(.015)
Female × Some College Education	.018	−.005	.025	−.019
	(.005)	(.005)	(.021)(.015)	
Female × College Graduate			−.062	−.022
			(.029)	(.015)
Female × Professional or Managerial Occupation	−.101	−.058	−.049	−.086
	(.004)	(.005)	(.016)	(.007)
Female × Technical or Supervisory Occupation	−.173	−.250	−.006	.037
	(.003)	(.004)	(.011)	(.008)
Female × Other White Collar Occupation	.088	.071	.033	.046
	(.003)	(.003)	(.013)	(.006)
Female × Skill Intensive Blue Collar Occupation−.208	−.167	−.045	−.043	
	(.002)	(.003)	(.010)	(.008)
Female × Potential Labor Force Experience	−.056	−.036	−.047	−.016
	(.0008)	(.001)	(.004)	(.003)
Female × Quadratic Experience Term	.002	.002	.004	.0003
	(.0001)	(.0001)	(.0003)	(.0002)
Female × Cubic Experience Term	−.00006	−.00005	−.0001	.00000
	(4.35e−06)	(5.63e−06)	(1.00e−05)	
Female × Quartic Experience Term	7.06e−07	5.40e−07	1.20e−06	1.80e−08
	(6.32e−08)	(7.78e−08)	(1.10e−07)	(4.00e−08)
R^2 (within)	.508	.468	.817	.617
Residual degrees of freedom	2,326,428	1,828,049	23,92	148,992

Sources: Menezes-Filho et al. (2008). RAIS Sao Paulo state manufacturing 1990 and 1997 (prime age workers in their highest-paying job on December 31st), Abowd et al. (2001) for France and the U.S., controlling for plant fixed effects.

BIBLIOGRAPHY

Abowd, John M and Francis Kramarz (1999). The Analysis of Labor Markets Using Matched Employer-Employee Data, in Orley Ashenfelter and David Card, (Eds.), *Handbook of Labor Economics*, Vol. 3B of *Handbooks in Economics*, Amsterdam: Elsevier, Chapter 40, 2629–710

Abowd, John M, Francis Kramarz, David N Margolis and Kenneth R Troske (2001). The Relative Importance of Employer and Employee Effects on Compensation: A Comparison of France and the United States. *Journal of the Japanese and International Economies*, December, *15* (4): 419–36

Anderson, James E (1998). Effective Protection Redux. *Journal of International Economics*, *44* (1): 21–44

Arbache, Jorge Saba, Andy Dickerson and Francis Green (2004). Trade Liberalisation and Wages in Developing Countries. *Economic Journal*, February, 114 (493): F73–96

Balassa, Bela (1965). Trade Liberalization and Revealed Comparative Advantage. Manchester School of Economic and Social Studies, May, 33: 99–123

Bernard, Andrew B, Stephen J Redding and Peter K.Schott (2007). Comparative Advantage and Heterogeneous Firms. *Review of Economic Studies*, January, 74 (1): 31–66

Biscourp, Pierre and Francis Kramarz (2007). Employment, Skill Structure and International Trade: Firm Level Evidence for France. *Journal of International Economics*, 72(1): 22–51

Botero, Juan C, Simeon Djankov, Rafael La Porta, Florencio Lopez de Silanes and Andrei Shleifer. (2004) The Regulation of Labor. *Quarterly Journal of Economics*, 119 (4): 1339–82

Corden, W Max (1966). The Structure of a Tariff System and the Effective Protective Rate. *The Journal of Political Economy*, 74 (3), 221–237

Costantini, James A and Marc J Melitz (2008). The Dynamics of Firm-Level Adjustment to Trade Liberalization, in Elhanan Helpman, Dalia Marin, and Thierry Verdier (Eds.), *The Organization of Firms in a Global Economy*, Cambridge. MA: Harvard University Press, November, Chapter 4, 107–141

Davis, Steven J, John C Haltiwanger and Scott Schuh (1996). *Job creation and destruction*. Cambridge and London: MIT Press

Ederington, Josh and Phillip McCalman (2008). Endogenous Firm Heterogeneity and the Dynamics of Trade Liberalization. *Journal of International Economics*, *74* (2): 422–40

Eslava, Marcela, John Haltiwanger, Adriana Kugler and Maurice Kugler (2004). The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia. *Journal of Development Economics*, 75 (2): 333–71

Fernandes, Ana M (2007). Trade Policy, Trade Volumes and Plant-Level Productivity in Colombian Manufacturing Industries. *Journal of International Economics*, 71 (1): 52–71

Freeman, Richard B (1980). An Empirical Analysis of the Fixed Coefficient 'Manpower Requirements' Model, 1960–1970, *Journal of Human Resources*, 15 (2): 176–99

Goldberg, Pinelopi K and Nina Pavcnik (2003). The Response of the Informal Sector to Trade Liberalization. *Journal of Development Economics*, Special Issue December, 72 (2): 463–96

Gonzaga, Gustavo (2003). Labor Turnover and Labor Legislation in Brazil. *Economia: Journal of the Latin American and Caribbean Economic Association*, 4 (1): 165–207

Gonzaga, Gustavo, Naércio Aquino Menezes-Filho and Maria Cristina Terra (2006). Trade Liberalization and the Evolution of Skill Earnings Differentials in Brazil. *Journal of International Economics*, 68 (2): 345–67

Haltiwanger, John C (1997). Measuring and Analyzing Aggregate Fluctuations: The Importance of Building from Microeconomic Evidence. *Federal Reserve Bank of St. Louis Review*, May–June, 79 (3): 55–77

Haltiwanger, John C, Adriana Kugler, Maurice Kugler, Alejandro Micco and Carmen Pagés (2004). Effects of Tariffs and Real Exchange Rates on Job Reallocation: Evidence from Latin America. *Journal of Policy Reform*, Special Issue 7 (4): 191–208

Hart, Oliver D (1983). *The Market Mechanism as an Incentive Scheme. Bell Journal of Economics,* 14 (2), 366–82

Hay, Donald A (2001). The Post-1990 Brazilian Trade Liberalisation and the Performance of Large Manufacturing Firms: Productivity, Market Share and Profits, *Economic Journal*, 111 (473): 620–41

Heckman, James J. and Carmen Pagés, (2004) Law and Employment: Introduction, in James J. Heckman and Carmen Pagés (Eds.), *Law and employment: Lessons from Latin America and the Caribbean*, NBER Conference Report series, Chicago and London: University of Chicago Press, 1–107

Hopenhayn, Hugo A (1992). Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica*, 60 (5): 1127–50

Hungerford, Thomas L (1995). International Trade, Comparative Advantage and the Incidence of Layoff Unemployment Spells, *Review of Economics and Statistics*, 77(3): 511–21

Katz, Lawrence F and Kevin M Murphy (1992). Changes in Relative Wages, 1963–1987: Supply and Demand Factors. *Quarterly Journal of Economics*, 107 (1): 35–78

Kletzer, Lori G, (2001). *Job Loss from Imports: Measuring the Costs* Globalization balance sheet series, Washington, DC: Institute for International Economics, September

Kruse, Douglas L (1988). International Trade and the Labor Market Experience of Displaced Workers. *Industrial and Labor Relations Review*, 41 (3): 402–17

Kume, Honório, Guida Piani and Carlos Frederico Bráz de Souza (2003). A política brasileira de importação no período 1987–98: Descrição e Avaliação, in Carlos Henrique Corseuil and Honório Kume, (Eds.) 2003. *A abertura comercial brasileira nos anos 1990: Impactos sobre emprego e salários*, Rio de Janeiro: MTE and IPEA, Chapter 1, 9–37

Levinsohn, James (1993). Testing the Imports-as-Market-Discipline Hypothesis, *Journal of International Economics*, 35 (1–2): 1–22

—(1999). Employment Responses to International Liberalization in Chile. *Journal of International Economics*, 47 (2): 321–44

Melitz, Marc J (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71 (6): 1695–1725

Menezes-Filho, Naércio Aquino and Marc-Andreas Muendler (2007). Labor Reallocation in Response to Trade Reform. *CESifo Working Paper*, March, *1936*

Menezes-Filho, Naércio Aquino, Marc-Andreas Muendler and Garey Ramey (2008). The Structure of Worker Compensation in Brazil, With a Comparison to France and the United States. *Review of Economics and Statistics*, 90 (2): 324–346

Mincer, Jacob (1974). *Schooling, experience, and earnings*. New York: Columbia University Press for the National Bureau of Economic Research

Muendler, Marc-Andreas (2004). Trade, Technology, and Productivity: A Study of Brazilian Manufacturers, 1986–1998. *CESifo Working Paper*, March, *1148*

—(2008). Trade and Workforce Changeover in Brazil, in Stefan Bender, Julia Lane, Kathryn Shaw, Fredrik Andersson, and Till von Wachter (Eds.), *The Analysis of Firms and Employees: Quantitative and Qualitative Approaches*, NBER Conference Report, Chicago and London: University of Chicago Press, Chapter 9, 267–308

Muendler, Marc-Andreas, Jennifer Poole, Garey Ramey and Tamara Wajnberg (2004). Job Concordances for Brazil: Mapping the Classificação Brasileira de Ocupações (CBO) to the International Standard Classification of Occupations (ISCO-88), April, University of California, San Diego, unpublished manuscript

Olley, G Steven and Ariel Pakes (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64 (6): 1263–97

Pavcnik, Nina, (2003).What Explains Skill Upgrading in Less Developed Countries? *Journal of Development Economics*, 71 (2): 311–28

Raith, Michael (2003). Competition, Risk, and Managerial Incentives, *American Economic Review*, 93 (4): 1425–36

Ramos, Roberto Luís Olinto and Claudia Nessi Zonenschain (2000). The Performance of the Brazilian Imports and Exports Based on the System of National Accounts: 1980–1998, August, IBGE Rio de Janeiro

Ribeiro, Eduardo Pontual, Corseuil, Carlos Henrique, Santos, Daniel, Furtado, Paulo, Amorim, Brunu, Servo, Luciana and André Souza (2004). Trade Liberalization, the Exchange Rate and Job Flows in Brazil, *Journal of Policy Reform*, Special Issue 7 (4): 209–23

Roberts, Mark J (1996). Employment Flows and Producer Turnover, in Mark J Roberts and James R. Tybout, (Eds.), *Industrial evolution in developing countries: Micro patterns of turnover, productivity,and market structure*, Oxford and New York: Oxford University Press for the World Bank, 18–42

Schmidt, Klaus M (1997). Managerial Incentives and Product Market Competition. *Review of Economic Studies*, 64 (2): 191–213

Schor, Adriana (2004). Heterogenous Productivity Response to Tariff Reduction: Evidence from Brazilian Manufacturing Firms. Journal of Development Economics 75(2): 373–96

Tybout, James R (2003). Plant- and Firm-level Evidence on 'New Trade Theories,' in James Harrigan and E. Kwan Choi (Eds.), *Handbook of International Trade*, Vol. 1 of *Blackwell Handbooks in Economics*, New York: Blackwell Publishers, July, chapter 13, 388–415

Wacziarg, Romain and Jessica Seddon Wallack (2004). Trade Liberalization and Intersectoral Labor Movements. *Journal of International Economics*, 64 (2), 411–39

Yeaple, Stephen Ross, (2005). A Simple Model of Firm Heterogeneity, International Trade, and Wages, *Journal of International Economics*, 65 (1): 1–20

8

Adjustment to Trade Policy in Developing Countries

GORDON H HANSON

1. INTRODUCTION

How do developing countries adjust to changes in trade policy? Until the last decade, most trade economists would probably have used the Heckscher–Ohlin (HO) model, or some variant of it, to answer the question.[1] The thinking was that trade by developing countries was largely based on their comparative advantage in industries that were intensive in the use of unskilled labor, agricultural land, or supplies of natural resources. Models of intra-industry trade were viewed as best suited for analyzing trade among developed countries. Developments in both empirical and theoretical research suggest that factor proportions alone are insufficient to understand how a developing economy will respond to trade liberalization at home or abroad.

On the empirical side, the Stolper–Samuelson Theorem, which uses HO logic to predict how changes in goods prices will affect factor prices, has not found much empirical support (Goldberg and Pavcnik, 2007). Applying a simple HO model implies that trade liberalization will tend to reduce income inequality in developing countries, as factors move into labor-intensive industries, causing the relative demand for and income of capital or skilled labor to decline. In fact, trade liberalization is often accompanied by an increase in the relative demand for skill and a rise in wage inequality (Feenstra and Hanson, 2003). Another important development in the literature is recognition of significant differences between exporting and non-exporting firms, as well as between multinational and domestic enterprises. Relative to firms that sell solely on the domestic market, exporters tend to be larger, more skill-intensive, more capital-intensive, and more productive, and to pay higher wages (Bernard et al. 2007), a finding that holds in both developed and developing economies (Tybout, 2003). Multinational firms are larger and more productive, still. When trade barriers fall, exporters expand at the expense of smaller, less skill- and capital-intensive domestic establishments. The HO model is silent about why firms differ, ascribing them little role in adjustment.

[1] See, for instance, Hanson's (2004) discussion of literature on the analysis of trade reform in Mexico.

On the theory side, perhaps the most influential development in trade during the last two decades has been the framework put forth by Melitz (2003), which explicitly allows for firms to be heterogeneous in terms of their productivity and makes fixed costs of exporting central to international commerce. The Melitz model accounts for why exporters are better than non-exporters along most performance dimensions and explains why average industry productivity rises as trade barriers fall. Building on the earlier empirical literature which documented the effects of trade liberalization on industrial productivity, the Melitz model has helped place firms at the center of analysis of how an economy adjusts to changes in trade barriers.

Trade economists now appreciate that the world is more complicated—and more interesting—than the textbook HO model would imply. As a result, the task of empirically identifying the mechanisms through which trade affects wages, employment, and industry structure is commensurately more challenging.

In this paper, I review recent empirical research on trade policy in developing economies. Along the way, I also address relevant theoretical research, which provides context for findings that are hard to reconcile with classical trade theory. I organize the discussion around three topics: how trade affects firms and industries, how trade affects wages and employment, and how trade affects incomes and poverty. I will address mainly trade reform in manufacturing, which has been the focus of the literature.

2. TRADE AND FIRMS

In the Melitz (2003) model, a reduction in tariffs or other variable trade costs changes the composition of firms within an industry. Given fixed costs to exporting, only more productive firms find it profitable to export. Less productive firms sell exclusively on the domestic market (or not at all). If barriers to selling goods abroad fall—due, say, to bilateral or multilateral trade liberalization—more productive firms expand their production, as firms that were already exporting increase their sales abroad and new firms, which were not quite productive enough to break into foreign markets before, begin to export. The expansion of more productive firms comes at the expense of less productive establishments, some of which are forced to exit production. These changes in the composition of firms cause average industry productivity to rise.

There is a substantial empirical literature that documents a correlation between industry productivity and trade barriers, in a manner consistent with Melitz (with much of this work preceding Melitz's theoretical analysis). Harrison's (1994) and Levinsohn's (1993) classic studies of Cote d'Ivoire and Turkey, respectively, were the first to find such a relationship. In more recent work, Pavcnik (2002) found that during Chile's trade liberalization in the 1970s and 1980s aggregate productivity rose in manufacturing, due in part to the reallocation of resources from less productive to more productive plants. Muendler (2008) found similar evidence for Brazil, where falling trade barriers raised the probability of exit for less productive firms and led to higher plant and industry level efficiency. These stud-

ies are emblematic of a larger literature on episodes of trade reform in developing countries, which tend to find a negative correlation between trade barriers and productivity for both plants and industries. Tybout (2002) contains a detailed review of this work, concluding that there is a robust positive correlation between industry productivity and trade reform. The mechanisms underlying the correlation include resource reallocation from less to more productive plants (consistent with Melitz) and productivity improvements within plants (whose source is not well understood). There is also evidence that a reduction of trade barriers is followed by a fall in markups charged by firms and lower dispersion in productivity across firms.

What is notable about the mechanisms described in the literature is that they involve a reallocation of resources within industries—a finding documented using plant level data for several countries in Latin America by Haltiwanger et al. (2004). Absent are the between-industry employment shifts induced by trade, which are important for the Stolper–Samuelson Theorem. In Mexico, for instance, Revenga (1997) found little evidence of between sector employment shifts in manufacturing, following the country's 1980s trade reform. Pavcnik and Goldberg's (2007) survey of how developing countries respond to globalization cited few studies that emphasize the movement of resources from import-competing industries to export-oriented industries. A simple explanation is that specialization occurs not only across industries, but within industries. Schott (2004) found that even within very narrow product categories (10–digit Harmonized System level) the United States imports goods from both high wage and low wage countries. Within individual categories, unit values for goods from high wage countries are substantially higher than unit values for goods from low wage countries, suggesting that products are differentiated by quality and that countries specialize within product categories according to quality.

While some of the factors released by less productive firms are absorbed by more productive firms in the same industry, other factors may leave manufacturing altogether. A common perception in the popular press is that trade reform is associated with growth in the size of the informal sector. However, empirical evidence on the relationship between trade and informality is mixed. For Brazil and Colombia, Goldberg and Pavcnik (2003) found no evidence of a long-run expansion in the informal sector after a fall in trade barriers. In more recent work on Brazil, Menezes-Filho and Muendler (2008) found that workers leaving formal manufacturing sectors in response to reduced trade barriers often moved to the informal sector. Whether these workers tend to take up jobs in informal manufacturing shops or informal service establishments is unknown.

For firms, trade does more than change the opportunity to sell abroad and the intensity of import competition from foreign rivals; it also improves access to imported inputs, thereby enhancing efficiency. Greater product variety can be an important source of welfare gains. For consumers, Broda and Weinstein (2006) have documented that the United States enjoys substantial welfare gains from the increase in imported product varieties and an associated reduction in effective consumer prices. For firms, a related set of gains appear to exist. In India, Gold-

berg et al. (2008) found that lower tariffs in imported inputs led to an increase in the variety of inputs available on the Indian market, which in turn lowered effective input prices for firms. Here again, the evidence on the effects of trade is mixed. Using data for Brazil, Muendler (2008) takes a different approach, examining whether plant productivity is associated with access to foreign inputs. He found that the effect, while present, accounts for only a small part of the post trade reform growth in Brazilian productivity.

Increased trade in intermediate inputs has other important effects on industrial structure. Input trade arises in part because of global production networks, in which multinational firms divide the manufacturing process into stages and locate each stage in the country where it can be performed at least cost (Feenstra and Hanson, 2003). The expansion of US owned *maquiladoras* (export assembly plants) in Mexico (Feenstra and Hanson, 1997), Hong Kong export processing establishments in China (Hsieh and Woo, 2005), Japanese assembly plants in Southeast Asia (Head and Ries, 2002), and European subcontractors in Eastern Europe (Marin, 2008) are all examples of the global fragmentation of manufacturing. Skill and capital intensive stages of production remain in high wage countries, while labor-intensive stages move to low wage countries. Global production networks appear to be based on comparative advantage, but in an environment of extreme specialization. In the HO model, extreme specialization arises only with the absence of factor price equalization (FPE). While the lack of FPE may seem a natural assumption to make for the analysis of trade between developed and developing economies, it is only recently that trade economists have begun to apply empirically the extensions of the HO model that allow for unequal factor prices.

Feenstra and Hanson (1997) provide a model of the globalization of production in which firms in a skill-abundant North use firms in a non skill-abundant South to produce intermediate inputs. Assuming wages differ between nations, the North specializes in high skill tasks and the South specializes in low skill tasks. While a reduction in trade barriers has effects qualitatively similar to the Stolper–Samuelson theorem, the movement of capital or the transfer of technology from North to South has effects that are quite different. If Northern firms use foreign direct investment to move production to the South, they will logically choose to move the least skill-intensive activities that they perform. By moving these activities to the South, the average skill intensity of production rises in the North. The same also happens in the South, since the South initially specializes in the least skilled tasks. When the North 'offshores' production to the South, it turns out that the relative demand for high-skilled workers rises in both countries. Naturally, trade costs determine the magnitude of offshoring from North to South. For data on US multinational firms, Hanson et al. (2005) found that export of intermediate inputs to their foreign affiliates for further processing is strongly negatively correlated with tariffs and shipping costs.

In the next section, I discuss the implications of offshoring for wages. In the remainder of this section, I focus on what offshoring means for how Northern and Southern firms adjust to changes in macroeconomic conditions. If we assume

that the skill-intensive tasks performed in the North include fixed cost activities, such as management, research and development, and marketing, while the tasks performed in the South involve only variable cost production activities, offshoring will alter the relative volatility of output in the two countries. Bergin et al. (2009a) showed theoretically that a shock to, say, demand in the North will lead to greater changes in employment in the South, meaning that offshoring is associated with the South having higher volatility. Suppose the North has a positive demand shock, which causes local production and wages to expand. With higher wages in the North, firms there that previously did not offshore any production to the South now find it profitable to do so. Adjustment in the extensive margin of offshoring transmits the shock to the South in a powerful manner, such that employment volatility is higher in the South than in the North. Bergin et al. (2009b) documented that employment volatility for *maquiladoras* in Mexico is greater than for the corresponding manufacturing industries in the United States, even after controlling for overall differences in the volatility of industrial production between the two countries. Through offshoring, shocks to US manufacturing have a disproportionately large effect on Mexico.

3. TRADE AND WAGES

In developing countries, falling trade barriers are associated with the exit of less productive firms, rising average industry productivity, greater fragmentation of production, greater volatility of employment, and possibly more informality. What do these changes mean for wages of developing country workers? One downside of the failure of the simple HO model is that few alternative models give much in the way of general results about how trade shocks affect wages. Theoretically, a wide range of outcomes are possible; empirically, a wide range of outcomes have been observed. Rather than attempting to catalogue all of these outcomes, I focus on those that are most relevant to our discussion.

It is perhaps useful to begin with what we don't know. One may expect that changes in the composition of firms due to trade liberalization would affect the level and structure of wages. However, there has been little connection between the theoretical literature on firm heterogeneity and the empirical literature on trade and wages. What might we expect to happen? As less productive and skill intensive firms exit production and more productive and skill intensive firms expand, workers who lose their jobs may see a drop in their earning power associated with the destruction of firm-specific human capital. In data for Mexico in the 1980s, Revenga (1997) found that wages for manufacturing workers are positively correlated with industry tariffs. As industry tariffs fall, industry wage premia do as well. Attanasio et al. (2004a) found similar evidence for Colombia; however, in a separate paper on Brazil (Attanasio et al. 2004b) they found no evidence of changes in industry wage premia after trade reform. In the instances where it does occur, declining industry average wages could reflect dislocation effects by workers in the industry or the loss of rents, both of which could result from a decline in trade protection.

The expansion of more efficient, skill intensive plants may increase the relative demand for skilled labor. Theoretical research has only recently begun to address the issue. Helpman et al. (2008) developed a model in which trade can lead to higher wage inequality and greater unemployment in all economies (when moving from autarky to free trade). Their framework depends on firms being imperfectly informed about worker ability, and there being costly matching of workers to firms, which together generate residual wage inequality that increases as more productive firms increase their market share following trade reform. In Indonesia, Amiti and Davis (2008) found that after trade reform, average wages fell in import-competing industries and rose in exporting industries, which they suggest is consistent with firm heterogeneity. Their results, however, depend on firms setting wages according to a 'fairness' principle, which is untested. The bottom line is that we have strong reason to expect that firm heterogeneity will mediate the impact of trade shocks on wages, but we do not yet know which mechanisms for transmission of the shocks are most relevant empirically.

There has been considerably more research on how the global fragmentation of production affects the wage structure (see the survey in Feenstra and Hanson, 2003). Mexico's 1980s trade reform also liberalized foreign investment, which was followed by an increase in the relative wage of skilled labor (Hanson and Harrison, 1999). FDI in Mexican manufacturing was concentrated in *maquiladoras*, many of which were created by US firms moving unskilled labor-intensive production activities to Mexico. Feenstra and Hanson (1997) found that the shift in Mexican manufacturing toward export assembly plants can account for nearly half of the observed increase in the country's demand for skilled labor (nonproduction workers). Hsieh and Woo (2005) documented a similar phenomenon in Hong Kong and China. As Hong Kong moved labor-intensive production activities to China in the 1980s and 1990s, the country saw an increase in the relative demand for skill. Across Hong Kong manufacturing industries, Hsieh and Woo (2005) found a positive correlation between the nonproduction wage share and imports from China, which accounted for over half of the increase in the relative demand for skilled workers that occurred in Hong Kong over the period.

In Mexico, the effects of FDI on the wage structure appear to differ from those of tariff changes. Chiquiar (2008) found that during the 1990s, when the North American Free Trade Agreement was implemented, regions of Mexico closer to the United States enjoyed higher wage growth and a decline in the return to schooling, meaning wage inequality within these regions fell. In theory, as shown by Feenstra and Hanson (1997), it is possible for FDI and tariff reductions to affect wage inequality in an opposing manner. In related work, Hanson (2007) found that regions of Mexico with more initial FDI, more initial trade with the United States, and greater emigration opportunities to the United States enjoyed higher wage growth during the 1990s.

In the developing countries examined by Goldberg and Pavcnik (2007)–Argentina, Brazil, Chile, Colombia, Hong Kong, India, and Mexico—all experienced an increase in wage inequality during the 1980s and 1990s. Which of these experiences can be explained by trade? While offshoring appears important in Hong

Kong and Mexico, it has little relevance for labor market outcomes in the other countries, owing to the fact that they do not participate very extensively in global production networks. It appears that for these countries Stolper–Samuelson effects are not present (given that wage inequality rises following trade reform). Work by Pavcnik (2003) and Muendler (2008) suggests that imported intermediate inputs and skill-biased technical change are also unimportant. The literature has done an admirable job of ruling out explanations for how trade affects wages, but, beyond the countries engaged in global production networks, has not provided strong evidence of a link between wage inequality and trade.

Verhoogen (2008) suggests one alternative mechanism linking trade and wages is that greater openness leads firms to improve the quality of goods they produce, which in turn requires them to upgrade the skill level of their workforce, leading to an increase in the demand for skill and greater wage inequality. He found evidence consistent with this story in Mexico during adjustment to the 1994–95 devaluation of the peso. Kugler and Verhoogen (2009) presented similar evidence for Colombia. Related to the logic of Helpman et al. (2008)—whose work emphasizes sorting of workers by ability rather than quality differences across firms—it is changes inside firms and industries induced by trade reform that lead to wages in the wage structure. This line of work, while intriguing, is still limited to a handful of countries.

In many developing countries, wage inequality rose following periods of trade liberalization (and other economic reforms). While there is evidence in support of particular hypotheses (offshoring, quality upgrading) in particular countries (Colombia, Mexico, Hong Kong), in most developing economies there is no clear empirical relationship between greater economic openness and the structure of wages.

4. TRADE, INCOME AND POVERTY

Beyond concerns about whether trade increases the dispersion of wages, how does it affect income levels? Trade changes household well-being through its impact on wages and goods prices. Identifying the impact of trade on household income thus requires estimating its effects on these price outcomes. Even if trade increases wage inequality, it could still lead to an increase in average incomes and even in incomes of the poor. Is there evidence that trade raises living standards in developing countries?

In one of the few studies to take a general equilibrium approach to trade and living standards, Porto (2006) examined the effect of Argentina's trade reform on household welfare. Trade barriers affect households through their effect on the relative prices of goods, which in turn affect labor income and consumption. Households differ in terms of their consumption patterns and level of educational attainment, meaning that price changes will have differential impacts across families. Porto's (date) approach involves estimating the impact of trade policy changes on goods prices, estimating how changes in goods prices affect wages, and then simulating changes in household welfare, given data on the tariff reductions associated with Argentina's entry into Mercosur, and household budget

shares and factor supplies. Compared to rich households, poor households spend a higher share of their budget on food and other basic items and have low schooling. Porto (date) found that tariff cuts related to Mercosur led to an increase in the prices of goods intensive in low-skill labor, such as food and beverages, to which poor households allocated more of their spending. He also found that the relative price fell for non-traded goods, such as health, education, and leisure goods, to which rich households allocate more of their spending. Together, these results imply Mercosur's tariff cuts were associated with a rise in the inequality of household welfare in Argentina.

In related work, Nicita (2004) extended Porto's (date) framework to Mexico. He found that during the 1990s, tariff changes in Mexico led to an increase in real disposable income for all households, with richer households seeing a 6 per cent increase and poorer households seeing a 2 per cent increase. As a consequence of these income gains, there was a 3 per cent reduction in the number of households in poverty. While Mexico's tariff cuts appeared to lower poverty, they also appeared to increase inequality in income.

Taking a reduced form approach, Topalova (2004) examined the differential exposure of Indian districts to trade liberalization to identify the effects of trade on poverty. Over the period she studied, poverty rates were falling sharply throughout India (the reasons for which her approach cannot address). She found that districts more exposed to trade reform had smaller decreases in poverty; her results on inequality were not precisely estimated. We again see examples of trade having different effects in different countries. In one country trade appears to reduce poverty (Mexico), while in another it appears to slow its decrease (India). Porto (2006), Nicita (2004), and Topalova (2004) are notable for using consumption based measures of well-being to examine the effects of trade reform, rather than much of the rest of the literature, which focuses on wages, whose relation to welfare is less clear cut.

5. DISCUSSION

During the last decade and a half, there has been an explosion in research on how changes in trade policy affect developing countries. While there are a number of robust findings in the literature, the mechanisms behind many of the outcomes are not well understood. Following the liberalization of trade, less productive firms become more likely to exit production, average industry productivity rises, and firms increase the fragmentation of production across borders. The literature is just beginning to assess how the impact of trade on firm and industry productivity translates into changes in wage and employment outcomes in an economy. Recent evidence suggests that the fragmentation of production is associated with greater employment volatility. Interestingly, changes in industry productivity are largely associated with the reallocation of resources among firms within a sector. Between-sector employment shifts do not appear to be a general outcome of trade reform, though in some instances post trade reform churning in the distribution of firms has been associated with greater informality.

In Latin America and some parts of Asia, greater economic openness has come with increases in the dispersion of wages and income. While the coincidence of these outcomes is well established, identifying the channels through which trade affects wages has proven to be more difficult. In some instances, greater wage inequality appears to be a result of developed country firms outsourcing production to developing country establishments, or developing country firms upgrading the quality of output they produce. However, the empirical relevance of these mechanisms is much stronger in some countries (Colombia, Mexico, Hong Kong) than in others (Argentina, Brazil, Chile, India). There are not general empirical results on how trade affects the structure of wages in developing countries, which may be due to the wide variation across countries in industry structures, resource supplies, and reform episodes. Because trade does not have a uniform affect on wages, it does not have a uniform affect on household incomes and poverty either. The impact of trade on poverty depends on how falling trade barriers affect the relative prices of goods consumed by the poor and the demand for factors controlled by poor households. In some countries, trade reform appears to have triggered price changes that help poor households, which in others it has not. While one may want the literature to offer more conclusive results, the data seem unwilling to cooperate.

Gordon Hanson is the director of the Center on Pacific Economies and is a professor of economics at UC San Diego.

BIBLIOGRAPHY

Amiti, Mary, and Don Davis (2008). Trade, Firms, and Wages: Theory and Evidence. NBER Working Paper no. 14106

Attanasio, Orazio, Goldberg, *Pinelopi* and Nina Pavcnik (2004a). Trade Reforms and Income Inequality in Colombia. *Journal of Development Economics*, 74*:* 331–66

—(2004b). Trade Policy and Industry Wage Structure: Evidence from Brazil. *World Bank Economic Review* 18(3): 319–344

Bergin, Paul R, Feenstra, Robert C and Gordon H Hanson (2009a). Volatility due to Outsourcing: Theory and Evidence. Mimeo, UC Davis and UC San Diego

—(2009b). Offshoring and Volatility: Evidence from Mexico's Maquiladora Industry. *American Economic Review*, forthcoming

Bernard, Andrew, Stephen Redding, and Peter Schott (2007). Firms in International Trade. *Journal of Economic Perspectives* 21(3): 105–30

Broda, Christian, and Dave E Weinstein (2006). Globalization and the Gains from Variety. *Quarterly Journal of Economics*, 121(2): 541–585

Chiquiar, Daniel (2008). Globalization, regional wage differentials and the Stolper–Samuelson Theorem: Evidence from Mexico. *Journal of International Economics,* 74(1): 70–93

Feenstra, Robert C and Gordon H Hanson (1997). Foreign Direct Investment and Relative Wages: Evidence from Mexico's Maquiladoras. *Journal of International Economics,* 42: 371–94.

—(2003). Global Production and Inequality: A Survey of Trade and Wages, in James Harrigan, ed., *Handbook of International Trade,* Malden, MA*:* Basil Blackwell

Goldberg, Pinelopi, Khandelwal, Amit Pavcnik, Nina and Petia Topalova (2008). Imported Intermediate Inputs and Domestic Product Growth: Evidence from India. NBER Working Paper no. 14416

Goldberg, Pinelopi and Nina Pavcnik (2003). The Response of the Informal Sector to Trade Liberalization. Journal of Development Economics, *72: 463–96*

Goldberg, Pinelopi, and Nina Pavcnik (2007). Distributional Effects of Globalization in Developing Countries. *Journal of Economic Literature*, 45(1): 39–82.

Hanson, Gordon H (2004).What Has Happened to Wages in Mexico since NAFTA? in Toni Estevadeordal, Dani Rodric, Alan Taylor, Andres Velasco, Eds., *FTAA and Beyond: Prospects for Integration in the Americas*, Cambridge: Harvard University Press

–(2007). Globalization, Labor Income, and Poverty in Mexico. In Ann Harrison, ed., *Globalization and Poverty*, University of Chicago Press and the NBER, 417–56

Hanson, Gordon H, Mataloni, Raymond and Matthew J Slaughter (2005). Vertical Production Networks in Multinational Firms. *Review of Economics and Statistics*, 87: 664–78

Harrison, Ann (1994). Productivity, Imperfect Competition and Trade Reform: Theory and Evidence. *Journal of International Economics,* 36 (1–2): 53–73

Haltiwanger, John, Kugler, Adriana Kugler, Maurice, Micco, A and C Pages (2004). Effects of Tariffs and Real Exchange Rates on Job Reallocation: Evidence from Latin America. *Journal of Policy Reform*, 7(4): 191–208

Head, Keith, and John Ries (2002). Offshore production and skill upgrading by Japanese manufacturing firms. *Journal of International Economics,* 58: 81–105

Helpman, Elhanan, Itskhoki, Oleg and Stephen Redding (2008). Inequality and Unemployment in a Global Economy. NBER Working Paper no. 14478

Hsieh, Chang-Tai and Keong T Woo (2005). The Impact of Outsourcing to China on Hong Kong's Labor Market. *American Economic Review*, *95*(5): 1673–87

Krishna, Pravin and Devashish Mitra (1998). Trade Liberalization, Market Discipline and Productivity Growth: New Evidence from India. *Journal of Development Economics*, 56(2): 447–62

Kugler, Maurice, and Eric Verhoogen. (2009). Plants and Imported Inputs: New Facts and an Interpretation. *American Economic Review Papers and Proceedings*, forthcoming

Levinsohn, James (1993). Testing the Imports-as-Market-Discipline Hypothesis, *Journal of International Economics*, 35: 1–22

Marin, Dalia (2008). A New International Division of Labor in Eastern Europe: Outsourcing and Offshoring to Eastern Europe. *Journal of European Economic Association,* 4(2-3): 612-622

Melitz, Marc J (2003). The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6): 1695–1725.

Menezes-Filho, Aquino, Naércio and Marc Muendler. (2008). Labor reallocation in response to trade reform. Mimeo, UC San Diego

Muendler, Marc (2008). Trade, technology, and productivity: A study of Brazilian manufacturers, 1986–1998. Mimeo, UC San Diego

Pavcnik, Nina. (2002). Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants. *Review of Economic Studies*, 69: 245–76

–(2003). What explains skill upgrading in less developed countries? *Journal of Development Economics,* 71: 311–28

Porto, Guido G. (2006). Using survey data to assess the distributional effects of trade policy. *Journal of International Economics,* 70(1): 140-160.

Revenga, Anna L (1997). Employment and Wage Effects of Trade Liberalization: The Case of Mexican Manufacturing. *Journal of Labor Economics* 15(3): S20–43.

Schott, Peter (2004). Across-product versus within-product specialization in international trade. *Quarterly Journal of Economics*, 119: 647–78

Tybout, James (2003). Plant and Firm Level Evidence on the "New" Trade Theories. In E Kwan Choi and James Harrigan, Eds., *Handbook of International Trade*, Blackwell: Malden, MA, 388–415

Verhoogen, Eric (2008). Trade, Quality Upgrading and Wage Inequality in the Mexican Manufacturing Sector. *Quarterly Journal of Economics*, 123(2): 489–530

9

Production Offshoring and Labor Markets: Recent Evidence and a Research Agenda

MARGARET S MCMILLAN

This note reviews the evidence on the labor market effects of production offshoring for developed and developing countries, highlighting the outstanding issues. Production offshoring involves the relocation of physical manufacturing processes outside a country's borders. Examples of production offshoring include the manufacturing of electronic components by Intel in Costa Rica or the production of personal computers by Lenovo in Vietnam. Although services offshoring is a growing phenomenon, its relative importance for determining labor market outcomes is likely to be limited. Services offshoring to China and India account for a tiny fraction of aggregate US activity in services; in contrast, offshoring accounts for a substantial share of US manufacturing activity.

We begin this note with a review of the theoretical models of offshoring. Next, we review the empirical evidence for developed countries drawing heavily on studies done for the United States. While these studies focus primarily on the impact of offshoring on domestic labor markets, there are some important insights to be gleaned from these studies regarding labor markets in developing countries. For example, Brainard and Riker (1997) find evidence consistent with the notion of a 'race to the bottom'. We then turn to the evidence for developing countries. To date, almost all of the evidence for developing countries is based on single country studies that examine the effect of foreign direct investment by developed countries on labor market outcomes in developing countries.

The final section of this note summarizes the existing evidence and outlines a research agenda. In particular, we identify the following questions as deserving of more attention: (1) Does offshoring increase income volatility? (2) Does offshoring increase wage inequality? (3) How substantial are the wage and employment effects of offshoring? (4) What are the general equilibrium effects of offshoring? (5) What are the implications of offshoring by China?

1. OFFSHORING AND DOMESTIC LABOR MARKETS: THE THEORY

The theoretical literature on the linkages between multinational activity, labor demand, and wages does not yield clear predictions on the relationship between offshore activities and home labor market outcomes. For example, in the Helpman (1984) model, the motivation for foreign investment is based on factor price differences which exist outside the endowment allocation in the presence of factor price equalization. Consequently, in that alternative equilibrium, factor price differences follow from different relative endowments, and foreign investors will be drawn to countries where they could pay (for example) lower wages for a homogeneous type of good. Such a framework implies that, under some initial relative endowments, offshoring for vertically oriented multinationals can be associated with intra-firm imports of low-wage goods, largely invisible exports from headquarters of intangibles such as management skills, falling domestic demand for unskilled labor, and falling domestic wages.

In stark contrast, Grossman and Rossi-Hansberg (2006) show that offshoring tasks can confer a productivity gain that can boost domestic wages. Grossman and Rossi-Hansberg (ibid.) draw on insights from Autor et al. (2002) to develop a general equilibrium framework in which falling costs of offshoring can lead to wage gains for both skilled and unskilled workers at home. Grossman and Rossi-Hansberg (ibid.) use the Autor et al. (date) differentiation between routine and non-routine tasks to build a theoretical model of trade in tasks. Advances in technology (such as improvements in communication) make offshoring of routine tasks less costly, leading firms to increase production abroad. What is surprising is that offshoring of routine tasks for vertically motivated multinationals—there is no horizontal motive for foreign investment here—leads to ambiguous predictions for domestic wages. The intuition behind this result is that falling costs of offshoring act like a positive productivity shock. Although the primary motivation for offshoring is to save labor costs, low-skill workers at home may still gain if terms of trade effects and labor supply effects are not too large—offshoring acts like an increase in the labor supply, which puts downward pressure on domestic wages.

Other general equilibrium models of offshoring also predict benefits from offshoring for domestic workers. For example, Mitra and Ranjan (2007) study the effects of offshoring on unemployment to show that the general equilibrium effects of offshoring can be paradoxical and quite beneficial for domestic workers. In contrast, Antras et al. (2006) employ a matching model with heterogeneous workers to show that offshoring increases wage inequality in poor countries. Spencer (2005) provides a survey of the theoretical work on offshoring.

A separate but related strand of literature considers the impact of offshoring on income volatility. For example, Rodrik (1997) points out that globalization can increase the elasticity of demand for labor, thereby increasing wage volatility. Bermin et al. (2009) show that offshoring by the United States to Mexico can increase income volatility in Mexico (and indeed has), while Karabay and McLaren (2009) show that offshoring increases the volatility of the wages of domestic workers.

Horizontal foreign direct investment (FDI) has been largely left out of the most recent discussions about offshoring. Yet, most of FDI is still primarily market-seeking and not necessarily part of an international production network. There is currently no agreement in the theoretical literature on whether horizontally integrated foreign investment (H-FDI) or vertically integrated foreign investment (V-FDI) is more likely to lead to domestic job losses. An early version of the V-FDI model is presented in Helpman (1984). In the Helpman (1984) model, the motivation for foreign investment is based on factor price differences which exist outside the endowment allocation where there is factor price equalization. Consequently, in that alternative equilibrium, factor price differences follow from different relative endowments, and foreign investors will be drawn to countries where they could pay (for example) lower wages for a homogeneous type of good.

In the Helpman (1984) framework, there is an equilibrium where the parent (headquarters) imports low-wage goods and exports headquarters' services. In such a world, domestic demand for labor to produce the homogeneous good in the headquarters country would fall and wages would continue to decline until factor price equalization is eventually achieved. Such a framework implies that, under some initial relative endowments, V-FDI can be associated with intra-firm imports of low-wage goods, largely invisible exports from headquarters of intangibles such as management skills and knowledge arising from product-specific research and development (R&D) conducted at home, falling domestic demand for unskilled labor, and falling domestic wages.

Other approaches, however, suggest that V-FDI could be associated with rising labor demand at home. In Markusen (1989), domestic and foreign specialized inputs are complements by design, and trade generates welfare gains by increasing the number of specialized inputs available (which are produced under increasing returns to scale technology). There are also models which focus on the implications for labor demand of V-FDI versus H-FDI. Markusen and Maskus (2001) show how different incentives for foreign investment lead to different organizational structures, which in turn should produce different degrees of substitution between employment at home and abroad. Horizontal multinationals, which are defined as firms which produce the same products in different locations, are primarily motivated by trade costs to locate abroad.[1] For H-FDI, investment abroad substitutes for parent exports and foreign affiliate employment should substitute for home employment. In their framework, V-FDI leads to complementarity between trade and foreign investment. Vertically integrated enterprises are motivated by factor endowment differences—and consequently factor price differences in a world where there is no factor price equalization—to locate different components of production in different locations. As pointed out by Brainard and Riker (1997), one implication of this type of modeling approach would be that parent and affiliate employment would be complementary.

[1] For the purpose of simplicity, we will occasionally refer to horizontally integrated firms as horizontal firms, and vertically integrated firms as vertical.

To summarize, both the literature on offshoring and the literature on trade and foreign direct investment come to the same conclusion: the impact of these different forms of doing business on host country employment and wages are theoretically ambiguous. On the other hand, it is typically assumed that recipient countries benefit from these sorts of transactions both in terms of jobs and higher wages. There appears to be no theoretical work on the impact of offshoring by developing countries on host or recipient country labor markets. According to Sachs (2009), the economies of China and India are not dual as in Lewis (1954) but rather triple economies. The three sectors are: high-tech, world class R&D, low-tech standardized mass production of a lot of manufacturing, and hundreds of millions of poor people in the countryside. Sachs argues that the complexity of these markets warrant a new set of models.

2. OFFSHORING AND DOMESTIC LABOR MARKETS: THE EVIDENCE FOR DEVELOPED COUNTRIES

The evidence on offshoring and domestic employment is decidedly mixed. Brainard and Riker (1997) showed that employment across high and low wage affiliate locations of US multinationals is complementary for manufacturing activities. Borga (2005), Desai et al. (2005), and Slaughter (2003) also find that expansion of US multinationals abroad stimulates job growth at home. Slaughter (2003) reports the largest positive effects of offshoring: for every new job created abroad, US employment increases two-fold.[2] Reviewing these studies, Mankiw and Swagel (2006, page) conclude that 'foreign activity does not crowd out domestic activity; the reverse is true.'

Another set of studies on this topic (Brainard and Riker, 2001), Hanson et al. (2003), Muendler and Becker (2006), Harrison and McMillan (2007), and Harrison et al. (2007)) reaches the opposite conclusion: jobs abroad do replace jobs at home, but the effect is small. Moreover, Brainard and Riker (2001) use a factor demand approach to show that labor employed by affiliates overseas substitutes at the margin for labor employed by parents at home, but they emphasize that the results differ depending on geographic location. In particular, they emphasize that there is strong substitution between workers at affiliates in developing countries, with workers in countries like Mexico and China competing for the same jobs. Borga (2005) and Desai et al. (2009) ask a different set of questions; they focus on the correlation between expansion in activity at home and abroad. They show that there is a positive association between growth in domestic investment, assets, employment, and total compensation for multinational parents and their foreign affiliates.

Second, previous studies have used a variety of different methods. While Desai et al. (2009) adopt an instrumental variable approach to estimate the association between growth in employment at home and abroad for US multinationals,

[2] Slaughter's (date) estimates are presented in a recent high profile report released by the government on the consequences of offshoring for the US economy.

Muendler and Becker (2006) and Brainard and Riker (1997) estimate translog factor share equations. Using German multinational data, Muendler and Becker (2006) also explore the importance of selection into affiliate locations for the consistency of their estimates.

Third, previous empirical studies on employment and offshoring have not distinguished between the different motivations for foreign investment. As noted above, the motivation for offshoring has important implications for labor. Harrison and McMillan (2009) develop an empirical framework which is flexible enough to allow substitution or complementarity between home and affiliate employment for firms that have different motivations to engage in foreign activities. With this framework, they identify the separate effects of horizontal versus vertical foreign investment on home employment, and also allow for different degrees of substitution (or complementarity) in high- and low-income affiliate locations. To address the possibility that methodological differences might be driving the conflicting results described above, they adopt a variety of different approaches to estimating labor demand and a range of econometric techniques.

They find that the insights derived from trade theory go a long way towards explaining the apparently contradictory evidence on the relationship between offshoring and domestic manufacturing employment. For US parents primarily involved in horizontal activities, affiliate activity abroad substitutes for domestic employment. For vertically-integrated parents, however, the results suggest that home and foreign employment are complementary. Foreign wage reductions are associated with an increase in domestic employment. The results differ across high- and low-income affiliate locations, in part because factor price differences relative to the United States are much more important in low-income regions. In low-income affiliate locations, a 10 percentage point reduction in wages is associated with 2.7 percentage point reduction in US parent employment for horizontal parents, and a 3.1 percentage point increase in parent employment for vertical firms.

They also show that offshoring is not the primary driver of declining domestic employment of US manufacturing multinationals between 1977 and 1999. In fact, the evidence suggests that operating in low-income affiliate locations preserves jobs (for vertically integrated parents), instead of destroying them. They show that declining domestic employment of US multinationals is primarily due to falling prices of investment goods (such as computers, which substitute for labor), falling prices of consumption goods, and increasing import competition. This research highlights both the importance of heterogeneous firm responses to opportunities for direct investment abroad and the need to account for other avenues through which international competition affects US labor demand.

Regardless of the reasons for discrepancies in results (see Harrison and McMillan, 2009) for a discussion), all of the studies that analyze outcomes within firms registered with the Bureau of Economic Analysis share an important limitation. Since there are no details available on worker characteristics in these data, this research has been generally limited to exploring employment shifts between a US parent and its foreign affiliate.

There is a smaller but growing literature on offshoring and wages. Using data for the US manufacturing sector between 1979 and 1990, Feenstra and Hanson (1999) found that the real wages of production workers were probably unaffected by offshoring activities, while the real wages of non-production workers increased by 1 to 2 percentage points. Feenstra and Hanson use a two-step procedure first to identify the impact of outsourcing and high technology investments on productivity and prices, and then to trace through the induced productivity and price changes to production and non-production wages. Another study by Liu and Trefler (2008) finds that there are small or insignificant effects of offshoring on US wages. They measure the impact of services offshoring to China and India on labor outcomes of service sector employees.

What is most surprising about the growing literature on trade, offshoring, and wages is the lack of studies that use individual level data to explore the linkages between manufacturing wages, offshoring, and international trade. Liu and Trefler (2008) are an exception, but focus purely on offshoring in the services sector to China and India. While they find no impact of services offshoring on wages, it is much more likely that there would be important consequences for US wages from increasing international trade, as well as offshoring of manufacturing activity. Services offshoring to China and India account for a very small fraction of aggregate US activity in services; in contrast, import competition as a share of sales in manufacturing has doubled in the last 20 years and offshoring has also increased significantly. In Feenstra's (2000) book exploring the impact of trade on wages, only one study uses individual level data to explore the linkages. That study by Lovely and Richardson (2000) relies on the PSID data and cannot identify significant effects of trade on US wages, in part due to the fact that they follow a small sample of individuals over time.

In recent work, both Feenstra (2009) and Krugman (2008) suggest that the effects of trade and offshoring on US wages may be more important than these previous studies would suggest. Krugman challenges conventional wisdom by arguing that published research on trade and wages is largely outdated. He theorizes that the dramatic increase in manufactured imports from developing countries since the early 1990s could be responsible for the increase in wage inequality in the United States and other advanced countries. Feenstra (2008, page) in his Ohlin lectures writes that 'my own views have always favored a trade-based explanation [for the shift in labor demand toward more-skilled workers], and that the views of Krugman and others may be changing'.

Ebenstein et al. (2009) examine both the impact of trade and offshoring on US labor market outcomes by combining information on wages and worker characteristics from the March Current Population Surveys (CPS) with data on trade and offshoring across industries over time. Their data on offshoring activities by US multinational firms come from the Bureau of Economic Analysis and provide the only comprehensive coverage of the offshore activities of US firms. Their data on international trade include both export shares and import penetration. Following Autor et al. (2006), they also test whether the impact of offshoring or trade on US wages is more pronounced for occupations which can be character-

ized as routine. They also include a rich set of control variables; in particular, they control for total factor productivity growth and changing investment goods prices.

The paper represents an important break from previous papers by allowing both the effects of trade and offshoring to operate across sectors and within sectors. The standard approach to identifying effects of import competition or offshoring on wages is to use differences in import penetration across industries. This approach has been used to measure industry wage differentials, as well as to measure the effects of sector-specific import competition and offshoring on wages and employment. Their results confirm that wage effects due to either inter-industry differences in import competition or offshoring are not very significant. They find that the impact of offshoring on wages between 1982 and 2002 is also quantitatively small among those who remain in a specific manufacturing sector, consistent with the notion that inter-industry wage differentials are not large. For example, a 10 per cent increase in offshoring to low-wage countries has virtually no impact on wages across all educational categories. A 10 per cent increase in offshoring to high-wage countries is associated with an increase in wages for less educated workers of between 0.6 and 1 per cent. In contrast, we find that workers who leave manufacturing lose on average 3 to 9 per cent in real wages.

They also find small effects of offshoring on employment and only positive effects of offshoring on wages. Consistent with Harrison and McMillan (2006) and Harrison et al. (2007), they find that these small effects on employment depend on the location of offshore activities. A 10 percentage point increase in offshoring to low-wage countries reduces employment in manufacturing by 0.2 per cent while offshoring to high-wage countries increases employment in manufacturing by 0.8 per cent.

While they find significant employment reallocation in response to import competition and smaller employment responses to offshoring, we find almost no industry-level wage effects. Yet inter-industry wage differentials may not be important in a fluid labor market such as the US market, where workers find it easy to relocate either to another manufacturing sector or are driven to leave manufacturing altogether. If most of the downward pressure on wages occurs in general equilibrium, whereby wages equilibrate across manufacturing sectors very quickly as workers relocate, then industry-level analyses miss the important effects of international trade on wages.

They address this problem by calculating an occupation-specific measure of offshoring, import competition, and export activity. If workers find it easy to relocate within manufacturing sectors or leave manufacturing altogether, but are more likely to remain in the same occupation when they switch jobs, then occupation-specific measures of international competition are more appropriate for capturing the effects of trade and offshoring on wages. Their results suggest that this is indeed the case, and that international trade has had large, significant effects on occupation-specific wages. These large wage effects are consistent with our results, showing significant reallocation of employment across industries in

response to import competition. The downward pressure on wages due to import competition has been overlooked because it operates between and not within industries. Their results suggest that a 1 percentage point increase in occupation-specific import competition is associated with a 0.25 percentage point decline in real wages. While some occupations have experienced no increase in import competition (such as teachers), import competition in some occupations (such as shoe manufacturing) has increased by as much as 40 percentage points.

Finally, a recent study by the OECD (2007) finds that offshoring may have contributed to a rise in the elasticity of labor demand in OECD countries. The study shows that the textiles industry, which is known for the relative importance of offshoring, has the most elastic labor demand. By contrast, the study shows that the elasticity of labor demand is low in most service industries where offshoring is less common.

3. OFFSHORING AND DOMESTIC LABOR MARKETS: THE EVIDENCE FOR DEVELOPING COUNTRIES

Unlike the developed countries, most of the evidence concerning the impact of offshoring on developing country labor markets is centered on estimating the impact of developed country FDI on developing country labor markets. This is because the bulk of offshoring has been by developed countries. We begin this section by reviewing that evidence. We then turn to the indirect evidence regarding the impact of offshoring on developing country labor markets. Finally, we note that multinationals from developing countries are increasingly important players in the global economy and this trend is expected to accelerate as China and India continue to grow. More theoretical advances and empirical evidence will be required to understand the implications of offshoring by developing countries: we highlight these issues in the final section of this note.

In a chapter on trade and foreign direct investment for the forthcoming *Handbook of Development Economics*, Harrison and Rodriguez-Clare (2009) review the literature on the impact of FDI on factor markets in developing countries. They report that almost all studies find that workers in foreign firms are paid higher wages presumably because labor markets in developing countries are not perfectly competitive, and because foreign firms tend to be more productive. Before controlling for firm and worker characteristics, the wage gap tends to be large. For example, Martins and Esteves (2007) report a wage gap of 50 per cent for Brazil, and Earle and Telegdy (date) report a wage gap of 40 per cent for Hungary.

However, when researchers control for firm and worker characteristics, the wage premium paid by foreign firms drops significantly. For example, Martins and Esteves (2007) follow workers who move to or leave foreign enterprises using a matched worker and firm panel data set for Brazil for the period 1995 through 1999. They find that workers moving from foreign to domestic firms typically take wage cuts, while those who move from domestic to foreign firms experience wage gains. However, the wage differences are relatively small ranging from 3 to 7 per cent. The authors conclude that their results support a positive view of the role of foreign investment on labor market outcomes in Brazil.

Harrison and Rodriguez-Clare (2009) conclude that there is no evidence to support the view that foreign firms unfairly exploit foreign workers paying them below what their domestic counterparts would pay. Further evidence supporting this view comes from Harrison and Scorse (2008) who find evidence that foreign firms are more susceptible to pressure from labor advocacy groups, leading them to exhibit greater compliance with minimum wages and labor standards. They find that foreign firms in Indonesia were much more likely than domestic enterprises to raise wages and adhere to minimum wage requirements as a result of anti-sweatshop campaigns. They also find that the employment costs of anti-sweatshop campaigns were minimal, as garment and footwear subcontractors were able to reduce profits to pay the additional wage costs without reducing the number of workers.

Harrison and Rodriguez-Clare (2009) do not consider the employment effects of FDI. This is not surprising, since their chapter is primarily about trade, and most analyses of trade reform take as given the long-run level of employment. This 'exogenous employment' assumption, which asserts that in the long run employment reverts to its initial level, has been criticized on the grounds that there are typically short to medium term adjustments that take place as a result of liberalization, which can entail long spells of unemployment for displaced workers. However, according to Hoekman and Winters (2005), there is surprisingly little evidence on the nature and extent of transitional unemployment in developing countries.

Understanding the employment effects of offshoring for developing countries is particularly important since unemployment in many of these countries tends to be very high. Indeed, the promise of job creation is one of the reasons developing countries set up investment offices and provide all sorts of tax breaks to multinational corporations. Yet, we still know very little about the numbers and types of jobs created. The assumption is typically that jobs will be created and that this is a good thing. But this is not always the case. Take for example Chinese investors in Africa. Chinese construction projects in Africa are primarily carried out by state owned enterprises that typically employ imported Chinese workers. The lack of Africans employed in Chinese firms is causing increasing resentment in countries suffering from extreme poverty and high rates of unemployment (Ash, 2007).

In an exception, Feenstra and Hanson (1997) consider the effects of relocating manufacturing activities from the United States to Mexico on the demand for labor in Mexico. For nine industries located across multiple regions in Mexico, they find that the demand for skilled labor is positively correlated with the change in the number of foreign affiliate assembly plants, and that FDI increases the wage share of skilled labor relative to unskilled labor. While this might seem counterintuitive, the reason for this is that tasks performed by unskilled labor in the United States are performed by relatively skilled labor in Mexico. In a separate piece (Feenstra and Hanson, 2009), they find that offshoring by the United States increases wage inequality in the United States. They do not consider wage inequality in Mexico but the implications are clear. To the extent that offshoring increases the demand for skilled labor in Mexico, it would also increase inequal-

ity in Mexico. Feenstra indeed confirms this pattern in a recent lecture on globalization and labor (Feenstra, 2008).

In more recent work, Bergin et al. (2009) tackle the important issue of offshoring and job security. They argue that offshoring increases volatility because it allows the home country to export its business cycle fluctuations. They focus on Mexico's *maquiladora* sector which has displayed more volatility than the overall manufacturing sector in Mexico and any other industry in the United States. They examine the apparel, electronic materials, machinery, and transport equipment sectors, which are Mexico's four main offshoring industries. By using a three-good model that includes a homogenous good exported by each country, as well as the offshored good, they find that the standard deviation of Mexican employment, on average, is twice as high for each industry in the United States. They then compare the Mexican industries to their counterparts in California and Texas to minimize potential size disparities that may explain the higher volatility. But even after correcting for size differences, they still find that the *maquiladora* industries are more volatile.

Their theoretical model suggests that changes in employment by offshoring industries are driven partly by adjustment at the extensive margin. They use employment data and the number of firms in the *maquiladora* industries to find evidence for the adjustment at the extensive margin. The estimates reveal that an increase in the share of aggregate employment in an offshoring industry results in over one-third of the adjustment at the extensive margin, demonstrating that plant entry and exit is an important means by which Mexico's offshoring industry adjusts to aggregate shocks. Providing further evidence for the adjustment at the extensive margin, they use the Harmonized System import data for the United States to reveal a positive correlation between the number of distinct products crossing the border and US manufacturing employment.

Unlike Berman et al. (1998) who argue that productivity shocks are the primary source of international transmissions of business cycles, Feenstra finds that demand shocks are more important. Feenstra uses monthly government expenditure data in the United States and Mexico to calibrate demand shocks and monthly Solow residual data to calibrate supply shocks. His results indicate that home demand shocks are the most important driver of volatility in the Mexican offshoring sector, while productivity shocks generate much less volatility in employment. He states that the fact that the *maquiladora* industries are more volatile reveals that the United States is exporting its cyclical fluctuations to Mexico's economy. This demonstrates that offshoring is important in explaining amplified volatility as firms rapidly shift production across borders.

4. SUMMARIZING THE EVIDENCE AND OUTLINING A RESEARCH AGENDA

Offshoring's impact on labor markets in developed countries is a relatively new and growing area of research, spurred on by the combination of increased offshoring and job losses in the manufacturing sectors of developed countries. Until

very recently, most of the evidence on offshoring indicated that the impact of offshoring on wages in developed countries was negligible. Similarly, a number of studies found that firms that offshore do so at the expense of domestic jobs, but the extent of these effects so far seems to be negligible. Moreover, the counterfactual has not been adequately addressed: what would have happened to jobs and wages in the absence of offshoring? Recent work by Ebenstein et al. (2009) points out that part of the problem with past studies is that they all look for effects within manufacturing. They show that offshoring causes displacement of workers, leading to significant wage declines. They also show that offshoring has significant economy-wide effects on wages measured at the occupational level. They interpret this latter finding as evidence that workers are mobile across sectors but not across occupations.

By contrast, the impact of offshoring by developed countries on developing country labor markets has a long tradition in the literature. Most researchers find that foreign firms pay higher wages and conclude that FDI has beneficial effects on host country labor markets. However, the magnitude of these effects varies substantially. The employment effects of FDI in developing countries are less well understood. Recent work by Bergin et al. (2009) suggests that these effects are important. Similarly, the impact of offshoring by developing countries on domestic labor markets is a growing phenomenon that has received very little attention.

For the most part, the theoretical literature on offshoring has outpaced the empirical research, leaving room for an exceptionally rich and fruitful research agenda. However, practically all of the recent theoretical literature on offshoring applies to developed countries offshoring to underdeveloped countries. Moreover, the bulk of the theoretical literature is primarily concerned with the impact of offshoring on domestic labor markets. As Sachs (2009) points out, new theoretical models will be needed to understand the impact of offshoring by China and India fully. Taking these points into consideration, we highlight below what we believe to be the most promising areas for future research.

4.1 Does offshoring increase income volatility?

Limited evidence for Mexico and the United States suggests that offshoring industries in Mexico experience job volatility twice that of corresponding industries in the United States. Earlier work on offshoring by US multinationals found that workers in countries with similar levels of wages compete with one another for the same jobs. One implication of these earlier findings is that workers in low-wage countries that previously attracted a lot of offshoring may be particularly vulnerable to losing jobs to China. More recent work suggests that this is likely to be the case for workers with similar skill sets across developed and developing countries. Finally, workers who lose jobs in developed countries as a result of offshoring tend to move to lower paying jobs with less job security. All of this suggests that offshoring could indeed increase income volatility, particularly at the low end of the income distribution. Further research documenting the extent and magnitude of these effects is warranted.

4.2 Does offshoring increase wage inequality?

Early evidence has suggested that offshoring could explain as much as 25 per cent of the increase in wage inequality in the United States between 1977 and 1990. The boom in offshoring following this study has been significant, suggesting that a reevaluation of the data is in order. Evidence for developing countries is much more limited. The fact that foreign firms in developing countries tend to pay higher wages and hire relatively skilled labor suggests that offshoring may play an important role in determining wage inequality in developing countries. Country studies that combine sectoral data on offshoring with individual wage and occupation data could help shed light on this issue.

4.3 Much ado about nothing?

Even if offshoring has had an effect on employment and wages, the magnitude of these effects is not well understood. Most studies for developed countries have found that offshoring comes at the expense of domestic jobs, but the magnitude of these effects is limited. Similarly, most studies have found little or no effect of offshoring on wage levels in developed countries. There is surprisingly little evidence on the employment effects of offshoring for developing countries. Most of the evidence is for wages and is that foreign firms pay a wage premium, but the magnitude of this premium is smaller than previously thought. There are a number of reasons to believe that these effects could be much larger for developing countries (see for example Bergin et al., 2009). More evidence documenting the employment and wage effects of offshoring for developing countries would help us to understand whether, in fact, we are making much ado about nothing.

4.4 What are the general equilibrium effects of offshoring?

Studies that look for the impact of offshoring by looking at within-industry trends in wages and employment miss two potentially important effects of offshoring. First, they do not adequately capture the wage losses or gains accruing to individuals who shift from manufacturing to other sectors of the economy. The associated distributional implications are likely to be important, given the magnitude of the reallocation and an historically important wage premium paid to manufacturing workers in the United States and elsewhere. In addition to distributional consequences, there may also be efficiency consequences associated with the reallocation of labor from high to low wages industries—see for example Katz and Summers (1989). Finally, these studies do not capture the cumulative impact of import competition on workers who are easily able to relocate across sectors but cannot easily shift across occupational categories. The most recent work on offshoring indicates that these effects are important in the US labor market. Not only are these issues likely to be important for other countries, but their importance calls into question results of previous studies that looked only for the labor market effects of trade and offshoring within industries. The challenge here is to identify datasets that are rich enough to enable researchers to cap-

ture the general equilibrium effects of offshoring. This is likely to involve gathering data from a variety of sources. The payoffs to this type of research are likely to be large, given the paucity of empirical work.

4.5 What are the implications of offshoring by China?

Between 1990 and 2005, outward direct investment from China grew from $0.8 billion to more than $12 billion (Buckley et al., 2008). Yet, if one does a literature search for offshoring and China, the only thing that comes up are articles and reports touting the benefits of China as a destination for offshore activity. Anecdotal evidence suggests that offshoring by China is different from offshoring by the United States. For example, much offshoring by China is done by state-owned enterprises, often importing labor from China. The implications and labor market consequences of offshoring by China warrant further attention from both theoreticians and empiricists.

Margaret McMillan is an associate professor of economics at Tufts University.

BIBLIOGRAPHY

Antras, Pol, Luis Garicano, and Esteban Rossi-Hansberg (2006). "Organizing Offshoring: Middle Managers and Communication Costs," NBER Working Paper 12196.

Ash, Lucy (2007). "China in Africa: Developing Ties," BBC News, 4 December 2007. Available at *http://news.bbc.co.uk/2/hi/africa/7047127.stm*.

Autor, David H., Frank Levy and Richard J. Murnane, (2003) "The Skill Content of recent Technological Change: An Empirical Exploration," *Quarterly Journal of Economics*.

Bergin, Paul, Robert C. Feenstra, and Gordon H. Hanson (2009). "Outsourcing and Volatility," NBER Working Paper 13144.

Berman, Eli, John Bound, and Stephen Machin (1998). "Implications of Skill-Biased Technological Change: International Evidence," *Quarterly Journal of Economics,* 113: 1245–1280.

Brainard, Lael and David Riker (1997) "US Multinationals and Competition from Low Wage Countries," NBER Working Paper 5959.

Brainard, Lael and David Riker (2001). "Are US Multinationals Exporting US Jobs?," in *Globalization and Labour Markets,* ed. By D. Greenaway, and D.R. Nelson, Vol. 2, Elgar, Cheltenham, UK and Northhampton, MA: pp. 410–26

Buckley, Peter, Adam R. Cross, Hui Tan, Hinrich Voss and Xin Liu (2008). "Historic and Emergent Trends in Chinese Outward Direct Investment", *Management International Review* 48(6): 715-748

Desai, Mehir, Fritz Foley and James Hines (2005). "Foreign Direct Investment and Domestic Economic Activity," Harvard Business School Working Paper.

Desai, Mehir, Fritz Foley and James Hines (2009). "Domestic Effects of the Foreign Activities of U.S.
Multinationals," *American Economic Journal*, 1(1): 181–203.

Ebenstein, Avi, Ann E. Harrison, Margaret McMillan, and Shannon Phillips (2009) "Estimating the Impact of Trade and Offshoring on American Workers Using the Current Population Surveys," NBER Working Paper No. 15107

Feenstra, Robert C. (2000). "The Impact of International Trade on Wages," NBER Books, National Bureau of Economic Research, Inc, number feen00–1.

Feenstra, Robert C. (2008). "Offshoring in the Global Economy," Ohlin Lectures, presented at the Stockholm School of Economics, September 2008. Available at *http://www.econ.ucdavis.edu/faculty/fzfeens/pdf/Feenstra_Ohlin_Lecture_2008.pdf.*

Feenstra, Robert C. and Gordon H. Hanson (1999) "The Impact of Outsourcing and High-Technology Capital on Wages: Estimates for the U.S., 1979–1990," *Quarterly Journal of Economics*, 114(3): 907–940.

Feenstra, Robert C. and Gordon H. Hanson (1997) "Foreign Direct Investment and Relative Wages: Evidence from Mexico's Maquiladoras," *Journal of International Economics,* 42(3/4): 371–393.

Grossman, Gene and Esteban Rossi-Hansberg (2006) "Trading Tasks: A Simple Theory of Offshoring," Working Paper.

Hanson, Gordon, Ray Mataloni and Matthew Slaughter (2003). "Expansion Abroad and the Domestic Operations of US Multinational Firms," mimeo.

Harrison, Ann E., and Margaret McMillan (2007) "Outsoucring Jobs? Multinationals and U.S. Employment," NBER Working Paper.

Harrison, Ann E., Margaret McMillan, and Null, Clair (2007), "U.S. Multinational Activity Abroad and U.S. Jobs: Substitutes or Complements?" *Journal of Industrial Relations*, 46(2): 347–365.

Harrison, Ann E. and Andres Rodriguez-Clare (2009) "Trade, Foreign Investment, and Industrial Policy for Developing Countries," forthcoming, *Handbook of Development Economics*. Available at *http://are.berkeley.edu/~harrison/HandbookFeb200911.pdf*

Harrison, Ann E. and Jason Scorse (2008). "Improving the Conditions of Workers? Minimum Wage Legislation and Anti-Sweatshop Activism," *California Management Review*, 48(2).

Helpman, Elhanan (1984) "A Simple Theory of International Trade with Multinational Corporations," *The Journal of Political Economy*, 92(3): 457–471.

Hoekman, Bernard and Alan L. Winters (2005) "Trade and Employment: Stylized Facts and Research Findings," Working Papers 7, United Nations, Department of Economics and Social Affairs. Available at *http://www.un.org/esa/desa/papers/2005/wp7_2005.pdf*

Karabay, Bilgehan and John McLaren (2009) "Trade, Offshoring, and the Invisible Handshake," NBER Working Paper 15048.

Katz, Lawrence F. and Lawrence H. Summers (1989) "Can Inter-Industry Wage Differentials Justify Strategic Trade Policy?" NBER Working Paper 2739.

Krugman, Paul R. (2008). "Trade and Wages, Reconsidered," Brookings Panel on Economic Activity. Available at *http://www.princeton.edu/~pkrugman/pk-bpea-draft.pdf.*

Liu, Runjuan and Daniel Trefler (2008). "Much Ado About Nothing: American Jobs and the Rise of Service Offshoring to China and India," NBER working paper 14061.

Lovely, Mary E. and J. David Richardson (2000). "Trade Flows and Wage Premiums: Does Who or What Matter?," NBER Chapters in The Impact of International Trade on Wages, National Bureau of Economic Research: 309–348.

Mankiw, Gregory, and Phillip Swagel (2006). "The Politics and Economics of Offshore Outsourcing," NBER Working Paper 12398..

Markusen, James R. (1989)."Trade in Producer Services and in Other Specialized Intermediate Inputs," *American Economic Review*, 79(1): 85–95.

Markusen, James R. and Keith Maskus (2001). "General-Equilibrium Approaches to the Multinational Firm: A Review of Theory and Evidence," NBER Working Paper 8334.

Martins, Pedro and Luiz Alberto Esteves (2007). "Is There Rent Sharing in Developing Countries? Matched-Panel Evidence from Brazil," Working Papers 0017, Universidade Federal do Paraná, Department of Economics, 2007. Available at *http://www.economiaetecnologia.ufpr.br/textos_discussao/texto_para_discussao_ano_2007_texto_08.pdf*

Mitra, Devashish and Priya Ranjan (2007). "Offshoring and Unemployment," NBER Working Paper 13149.

Muendler, Marc-Andreas and Sascha O. Becker (2006). "Margins of Multinational Labor Substitution," UCSD Working Paper, April 15.

OECD (2007). OECD Workers in the Global Economy, Chapter 3 in the 2007 OECD *Economic Outloook*

Rodrik, Dani (1997). "Trade, Social Insurance, and the Limits to Globalization," NBER Working Paper 5905.

Sachs, Jeffrey D. (2009). "The Rise of TNCs from Emerging Markets: the global context" in *The Rise of Transnational Corporations from Emerging Markets: Threat or Opportunity?* ed. Karl Sauvant, Edward Elgar.

Spencer, Barbara J. (2005). "Presidential Address: International Outsourcing and Incomplete Contracts." *Canadian Journal of Economics*, 38(4): 1107–1135.

10

Trade Adjustment and Labor Income Risk

PRAVIN KRISHNA AND

MINE ZEYNEP SENSES

1. TRADE AND INCOME RISK – AN INTRODUCTION

Enhanced efficiency in production is an important channel through which openness to international trade can bring economic gains. For instance, it is expected that, with trade, countries will tend to specialize production in those goods in which they have comparative advantage. This improved allocation of productive resources and the rationalization of industry structure caused by trade may lead to an increase in aggregate productivity as low-productivity firms are displaced by their more efficient counterparts.

While the benefits of trade through more efficient production are generally well understood, it is also true that the process of reallocation of resources necessary for this efficiency enhancement may not be an orderly or costless one and that it may not result in greater earnings for all factors of production – an apprehension that underlies much of the public concern regarding trade and globalization more broadly.

Opening up to international trade may impact workers in a number of ways. In a benchmark theoretical economy, where workers posses no sector-specific skills and are able to move between sectors costlessly, trade liberalization is predicted simply to increase the rewards to abundant factors and lower returns to scarce factors – countries with a relative abundance (scarcity) of skilled workers are predicted to see a widening (compression) of their earnings distribution. An alternative characterization accounts for the fact that workers often posses experience and skills that are sector-specific. For instance, workers with long years of experience in the automobile sector may not be able to transfer their skills over to other sectors that are expanding. Here, trade liberalization will expose import-competing sectors to greater pressure from imports and workers with skills specific to these sectors will experience reduced earnings. Of course, many factors of production are not always wholly specific to a sector. Following a decline in their sector caused by trade liberalization, these factors may move out to other sectors. However, the adjustment process may still be a costly one. For instance,

these workers may find jobs that only partially reward the experience they have earned in the previous sector of employment.[1]

In studying the impact of trade liberalization, the economics literature has traditionally focused on the average (mean) effect of openness on the labor force. In recent research (Krishna and Senses (2009)), we have studied instead what we believe to be an important but underemphasized aspect of the adjustment that takes place in labor markets in response to increased openness – the risk that workers are exposed to due to the fact that similar workers may experience heterogeneous labor market outcomes with openness. (See also the earlier analysis by Krebs, Krishna, and Maloney (2009) upon which this analysis builds).

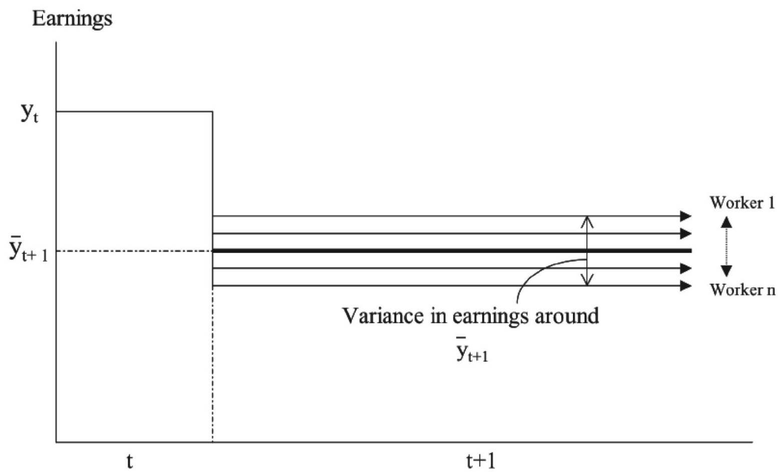

Figure 10:1: *Variance in Wage Outcomes*

Figure 10.1 illustrates the main point. Here we depict income paths for a group of workers whose incomes in time period t are identical and equal to y_t. Assume that the economy opens up to trade at the end of this period. In time period t+1, we see that the average income for this group of workers changes to $\bar{y}_{t+1}$. However, around this mean change in incomes there is a variance in individual outcomes. To the extent that individual outcomes are unpredictable beforehand, this process is risky and workers exposed to risk would find it to be costly. It is this variance around y_{t+1} that we are interested in – while the prior literature has largely examined the mean income gap (y_t – $\bar{y}_{t+1}$).

It is important to recognize that the unanticipated changes in income that we measure may be of a transitory or persistent nature. For example, during an adjustment process following trade liberalization, workers may experience tempo-

[1] The adjustment processes above have been discussed as one-time responses to trade liberalization. However, a more open economy may continually expose import-competing sectors to a more variable international economic environment, with changing international patterns of comparative advantage inducing reallocations of capital and labor across firms within and between sectors on an ongoing basis.

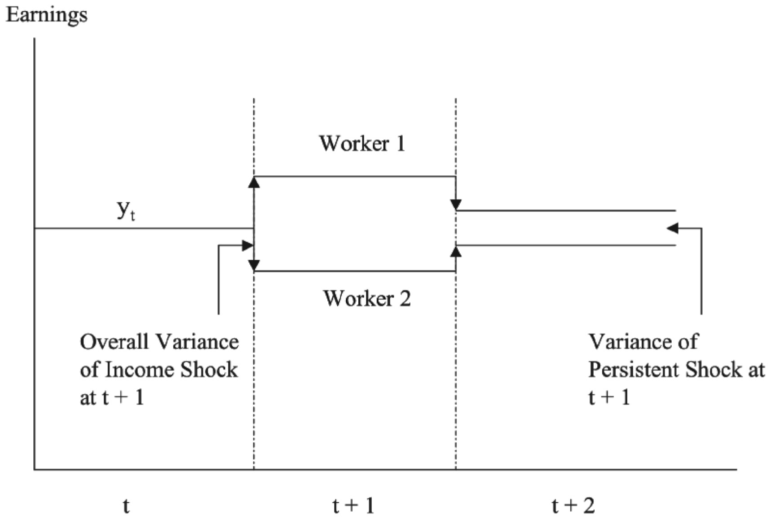

Figure 10.2: *Variance in Wage Outcomes*

rary job loss resulting in a temporary loss of income. Alternately, individuals moving across sectors may see persistent income losses if the work experience they have accumulated in their previous sectors of employment is not valued and thus is not suitably rewarded in their new sector.

Figure 10.2 presents a heuristic illustration to clarify the differences between the risk associated with transitory and persistent income changes. Specifically, Figure 10.2 illustrates the difference between transitory and persistent income shocks for a group consisting of two identical individuals whose incomes in time period t are both equal to y_t. At t+1, they experience shocks to income (some part transitory and some part persistent) that separate their incomes as indicated. By t+2, the transitory components of the income changes they experienced at t+1 expire and incomes for both workers move closer to their initial levels and stay at these levels for the rest of time. In this case, the magnitude of the variance of the persistent shock experienced at t+1 is measured by the spread in incomes at t+2 (and beyond). The spread in incomes at t+1 measures the sum of the variance of the transitory shock as well as the permanent shock experienced at t+1. Importantly, while in the face of a transitory income shock, workers may borrow or use their own savings to smooth consumption, this is clearly not feasible when income shocks are persistent. Thus, highly persistent income shocks have a large effect on individual well-being whereas the effect of transitory shocks is relatively small. Since it is the persistent income shocks that matter the most, it is on these shocks that we focus our attention.

The central analytical challenges addressed in this note concern the measurement of income risk faced by workers in a sector by using longitudinal data (where workers are followed over time) and, separately, the extent to which risk faced by workers in different sectors varies with the sector's exposure to inter-

national trade. In the sections that follow, this note describes the analytical approach that we have taken to study the issue of trade openness and labor income risk and outlines our main findings (See again Krebs, Krishna, and Maloney, 2009).

2. DATA AND ECONOMETRIC ANALYSIS

For the estimation of individual income risk, longitudinal data capturing individual income changes is desirable. It is generally not sufficient to use information on changes in the aggregate distribution of income to make inferences about the extent of income risk faced by individuals. For instance, while the aggregate distribution of income may stay the same across different time periods there still may be stochastic (risky) transitions taking place underneath, with some individuals at the top of the distribution exchanging places with others at the bottom end of the distribution. To capture the risk in incomes faced by these individuals, longitudinal data tracking these individual transitions, is clearly useful to have.

In Krishna and Senses (2009), we use longitudinal data on individuals from the 1993–1995, 1996–1999 and 2001–2003 panels of the Survey of Income and Program Participation (SIPP). Each panel of the SIPP is designed to be a nationally representative sample of the US population and surveys thousands of workers. The interviews are conducted at four-month intervals over a period of three years for the 1993 panel, four years for the 1996 panel, and three years again for the 2001 panel. In each interview, data on earnings and labor force activity are collected for each of the preceding four months. SIPP has several advantages over other commonly used individual-level datasets in that it includes monthly information on earnings and employment over a long panel period for a large sample. Although the Current Population Survey (CPS) provides a larger sample, individuals are only sampled for eight months over a two-year period in comparison to 33 months in the SIPP. While the Panel Study of Income Dynamics (PSID) provides a much longer longitudinal panel, it has a significantly smaller sample size compared to the SIPP and therefore does not support the estimation of risk at the industry level.

Our interest is in estimating labor income risk experienced by workers. Since labor income risk is defined as the variance of *unpredictable* changes in earnings, it is essential that predictable income changes are filtered out. To do this, we assume that the log of labor income of individual i employed in industry j in time period (month) t, $log\ y_{ijt}$, is given by:

$$log\ y_{ijt} = \alpha_{jt} + \beta_t \cdot x_{ijt} + u_{ijt}\,. \qquad (1)$$

In (1) α_{jt} and β_t denote time-varying coefficients, x_{ijt} is a vector of observable characteristics (such as age, age-squared, education, marital status, occupation, race, gender, and industry), and u_{ijt} is the stochastic component of earnings. Changes in the stochastic component u_{ijt} represents individual income changes that are *not* due to changes in the return to observable worker characteristics. In this sense, changes in u_{ijt} over time measure the unpredictable part of changes in individual income.

We assume that the stochastic term is the sum of two (unobserved) components, a permanent component ω_{ijt} and a transitory component η_{ijt}:

$$u_{ijt} = \omega_{ijt} + \eta_{ijt} \cdot \qquad (2)$$

Permanent shocks to income are fully persistent in the sense that the permanent component follows a random walk:

$$\omega_{ij,t+1} = \omega_{ijt} + \varepsilon_{ij,t+1}\,, \qquad (3)$$

where the innovation terms, $\{\varepsilon_{ijt}\}$, are independently distributed over time and identically distributed across individuals, $\varepsilon_{ijt} \sim N(0,\sigma^2_{\varepsilon js})$. In this basic specification, transitory shocks have no persistence, that is, the random variables $\{\eta_{ijt}\}$ are independently distributed over time and identically distributed across individuals, $\eta_{ijt} \sim N(0,\sigma^2_{\eta js})$. Note that the parameters describing the magnitude of both transitory and persistent shocks are assumed to depend on the sector j and the SIPP panel s, but do not depend on t. That is to say, they are assumed to be constant within a SIPP panel, but allowed to vary across panels. Estimation of $\sigma^2_{\varepsilon js}$ and $\sigma^2_{\eta js}$ therefore gives us industry-specific, time-varying estimates of transitory and permanent income risk faced by individuals.

This specification of the labor income process (Equations (1)–(3)) describes shocks to income to be either purely transitory or purely persistent. However, this specification does not capture shocks that have duration greater than one period (that is., are not purely transitory) but that are also not permanent (that is, last for a finite length of time). Estimation of permanent income risk in this case requires us to filter out such shocks of longer duration. To achieve this, we admit into the specification some moving average terms which filter out shocks that last up to 12 months (See Krishna and Senses (2009) for details).

Table 10.1 presents estimates of persistent income risk obtained using the estimation procedure described above. The mean estimate of the monthly variance of the persistent shock ($\sigma^2_{\varepsilon js}$) is 0.0014, 0.0025 and 0.0031 for the 1993, the 1996 and the 2001 panels (with corresponding annualized values of 0.0168, 0.03 and 0.0372), respectively. The annualized standard deviations of the reported estimates of the variance of permanent income risk are 0.13, 0.17 and 0.19 for the 1993, 1996 and 2001 panels, respectively. Clearly, income risk is rising over time.

Table 10.1: *Individual Labor Income Risk Estimates*

	Mean	Median	Std Dev
1993–1995	0.0014	0.0014	0.0019
1996–1998	0.0025	0.0026	0.0018
2001–2003	0.0031	0.0032	0.0025

Reported mean, median and standard deviations are calculated across point estimates for eighteen 2-digit SIC industries.

3. OPENNESS AND INCOME RISK

We are interested in evaluating the relationship between trade openness and income risk. To get to this, we examine the association between industry-level, time-varying estimates of the persistent component of labor income risk and measures of industry exposure to international trade using regression analysis.

We find that within-industry changes in income risk are strongly related to changes in import penetration over the corresponding time-periods. Figure 10.3 plots the changes in estimated permanent income risk, against changes in import penetration calculated at the beginning of each panel. More specifically, Figure 10.3 plots changes in risk and import penetration between the 1993 and 1996 panels and between the 1996 and 2001 panels. As indicated in the plot, the relationship appears to be strongly positive, suggesting that an increase in import penetration is associated with an increase in income risk for the workers in that industry.

Regression analysis confirms the statistical significance of this relationship between income risk and import penetration. Furthermore, this relationship holds for the full sample of workers as well as various sub-samples we consider, such as workers who switch industries within the manufacturing sector and those that switch out of manufacturing altogether. This result is robust to controlling for other time varying industry specific factors (such as exports, skill-biased technological change, offshoring, unionization, and productivity) that are potentially correlated with both income risk and import penetration.

We should emphasize that our analysis focuses *exclusively* on the link between trade and individual income risk. Hence, our results should be taken together with

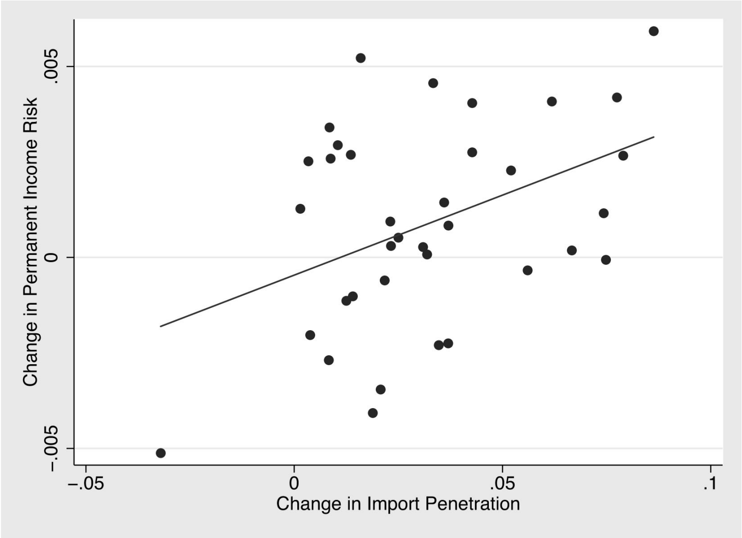

Figure 10.3: *Changes in Permanent Income Risk and Changes in Import Penetration*

the findings of a large literature on international trade exploring the many ways in which trade may affect the economy positively, through improved resource allocation, access to greater varieties of intermediate and final goods, greater exploitation of external economies and by possibly raising growth rates, *inter alia*. Specifically, the results presented here should not be interpreted as suggesting that exposure to trade has negative consequences overall, but instead as evidence that the costs of increased labor income risk ought to be taken into account when evaluating the total costs and benefits of trade and trade policy reform.

Pravin Krishna is the Chung Ju Yung Professor of International Economics and Professor of Economics at Johns Hopkins University.

Mine Zeynep Senses is Assistant Professor of International Economics at Johns Hopkins University.

REFERENCES

Krishna, P and Senses, Z. M. (2009). International Trade and Labor Income Risk in the United States, National Bureau of Economic Research, Working Paper Number 14992, NBER, Cambridge MA.

Krebs T, Krishna, P and Maloney, W (2009). Trade Policy, Income Risk and Welfare, *Review of Economics and Statistics*, forthcoming.

11

Trade, Child Labor, and Schooling in Poor Countries

ERIC V EDMONDS

An increase in international trade affects poor children in low-income economies by changing relative prices and altering living standards. The purpose of this note is to review the existing evidence on how trade influences child time allocation, including child labor and schooling.

Trade's effect on the living standards of the poor is generally found to be the dominant channel through which trade influences child time allocation and schooling. Trade can influence the living standards of the poor by changing consumption prices and through altering labor income and family asset income. It is this latter channel, changes in labor and asset incomes, which researchers have highlighted as the primary pathway through which trade changes child time allocation.

Relative price changes could be important for the link between trade and child labor. This essay reviews the potential ways that trade influences child labor through relative price changes, and discusses why there is little evidence in empirical research of these channels operating. It is useful to distinguish between direct channels—trade changes child labor demand—and indirect pathways, where changes in child time allocation result as a collateral consequence of trade's impact on the overall economy. The direct effects occur when trade directly impacts production and thereby labor demand in the traded sector. Direct effects can alter the types of employment opportunities available to children or the child's potential economic contribution from working, which this essay will refer to as the 'child's wage' even though few children work for wages. Indirect effects occur when trade affects sectors beyond that in which there is trade, perhaps through inputs, or when trade has important general equilibrium effects. These indirect effects can also alter wages and employment opportunities. Other important indirect effects might come from perceptions of changes in returns to education or by influencing occupational choice. Through impacting consumer prices, trade can alter the implicit cost of leisure. Despite many possible channels, there is very little evidence supporting any connection between trade and child time allocation, other than through the impact of trade on the living standards of the very poor.

This observation of the primacy of the living standards of the poor is for two reasons. First, among the poor, the standard of living is one of the most impor-

tant determinants of child time allocation. Second, children are rarely engaged in work that will be easily connected to international trade. Children in poor countries mostly work in agriculture. Outside agriculture, children are not intensively involved in traded sectors in general, and within manufacturing, firms involved in trade tend to be relatively more skill-intensive. Hence, the direct effects of trade on child labor through wages paid in the traded sector, for example, will be minimal. Part of the reason for the lack of evidence of trade influencing child time allocation through factors related to labor demand or returns to education owes to the problem of identifying the indirect and diffuse effects of trade. Most studies of the impact of trade on child labor and schooling exploit some within-country heterogeneity in direct exposure to trade. The diffuse impact of trade, even if substantive, will not necessarily be correlated with direct exposure. Such studies are poorly designed to measure the collateral effects of changes in the economy from trade.

The next section discusses how children work in poor economies and reviews the evidence we have on a direct connection between trade and the employment opportunities available to children. The subsequent section considers the evidence on how and whether trade might influence child time allocation indirectly. The third section reviews the evidence on the impact of trade working through the living standards of the poor. The final section considers priorities for future research.

1. MOST WORKING CHILDREN ARE NOT IN JOBS THAT ARE DIRECTLY CONNECTED TO TRADE OTHER THAN IN AGRICULTURE

There is little evidence of a direct impact of trade on child labor or schooling working through changes in the employment opportunities available to children under 15. This is because most working children in poor economies are not involved directly in international trade. Most children work outside of traded sectors or in non-traded subsectors of traded sectors. When they work in traded sectors, typically they are not engaged in enterprises that export. There are many important exceptions to these generalities, and this section concludes with a discussion of some of the most infamous exceptions.

The industrial composition of child employment does not lend itself to a direct connection between trade and child time allocation. Table 11.1 provides nine examples of the industrial composition of child employment from low-income countries. All of the countries in the table have a GDP per capita in 2005 that is below $5,000 (PPP, international dollars). The countries have been selected because they have detailed, nationally representative surveys that allow us to examine the industrial composition of employment (economic activity) for children aged five to 14. The countries differ in how important trade is in their economies (from Cambodia to Brazil) and the prevalence of economically active children (from Ethiopia to Honduras). In all nine countries, the majority of working children are in agriculture. Wholesale and retail trade, a service sector, is the next most important sector for child employment in all countries but Kenya, where

Table 11.1: *Industrial Composition of Economically Active Children 5 to 14, Selected Countries*

	Cambodia	Honduras	Philippines	Nicaragua	Senegal	Kenya	Ethiopia	Burkina Faso	Brazil
Trade (as a % of GDP) in 2005	137	136	99	87	69	64	51	36	27
Year of data	2001	2004	2001	2005	2005	1999	2005	2003	2004
Economic Activity Rate	44.8	5.4	11	8.4	15.6	6.1	50.1	47	5.7
% in Paid Employment	4.6	19.9	22.8	13.6	4.2	11.9	2.1	0.4	21.2
Industrial Composition of Economically Active Children (as a % of total Economically Active Children)									
Agriculture	76.5	63.3	64.1	70.7	80.6	82.3	95.2	97.4	61.2
Mining	0.3	0.2	0.5	0.0	*	0.5	*	0.1	0.0
Manufacturing	4.9	8.3	4.0	9.6	4.7	1.6	1.3	0.4	6.5
Construction	0.5	1.7	0.5	0.5	1.2	0.2	0.2	0.1	1.4
Wholesale and Retail Trade	15.1	15.6	22.2	15.5	7.6	1.6	1.7	1.1	16.5
Hotel & Restaurant	0.3	3.5	1.8	1.8	0.6	0.5	0.7	0.1	3.7
Transportation and Communication	0.3	1.1	1.5	0.4	0.6	0.2	0.1	*	1.9
Domestic Services	0.9	4.7	3.6	0.5	4.3	6.2	0.3	0.8	5.0
Other	1.2	1.6	1.9	1.0	0.4	6.9	0.5	0.1	3.8

*Not available as separate category—included within 'other'. Trade from World Development Indicators online, May 2009. All other data from UCW Child Labor Statistics by Country: http://www.ucw-project.org, May 2009. Countries selected from UCW country data by following criteria: GDP per capita below $5,000 (2005 PPP International dollars); available trade data; available data for children aged 5 to 14; and industrial composition of child employment available at the level of detail above.

working as a domestic in private households is more prevalent. Manufacturing draws the most academic attention in empirical trade studies and the policy debate. Less than one in 10 working children participates in the manufacturing sector in all nine of the listed countries.

Even within traded sectors, children are often in the non-traded parts of the sector. For example, 82 per cent of employed Kenyan children are in agriculture (Table 11.1), but the Kenyan Central Bureau of Statistics (2001) reports that 1.5 children are in subsistence agriculture for every child engaged in market-oriented agriculture. Bangladesh is another interesting example because of the unique detail available about type of industry and occupation (Edmonds 2008). Of children under 18, 55 per cent are engaged in agriculture (Central Bureau of Statistics 2003). Seventy-one per cent of these youths employed in agriculture work in the cereal sector. Bangladesh imports less than 5 per cent of its cereal consumption (Dorosh 2009). Among the other agricultural sectors that are important employers of Bangladeshi children are vegetable farming (5 per cent of working children) and poultry farming (4 per cent of working children). Neither sector is a substantive source of imports or exports for Bangladesh. Overall, child workers are typically outside the formal cash economy. This is evident in Table 1 where fewer than one in five children work for wages in every country listed except the Philippines and Brazil. Edmonds and Pavcnik (2005a) tabulate data from 36 poor countries representing 124 million children and observe that 2.4 per cent of children aged from five to 14 work in paid employment; 20.8 per cent are unpaid, working in their family farm or business.

Most studies of who the exporters are focus on manufacturing (where child employment is rare to begin with), but those studies typically document that exporting firms, and firms that use imported imports, employ more skilled workers (see Wagner 2007 for a review of 54 studies). Explanations for this phenomenon come from studies that attempt to understand why growing trade in developing countries seems to be associated with greater demand for skilled labor. Two popular explanations imply that trading firms will have workers with higher levels of education. First, compositional changes within industries may favor more skilled labor-intensive producers (see for example Verhoogen 2008). Exports may require a more uniform, high-quality product that requires skilled labor to produce it. Second, imported intermediate goods and FDI may be complementary to skilled labor (Feenstra and Hanson 1996). Thus, at least for exports, children are not likely to be directly involved in exports even when they are involved in export sectors.

The implication of this discussion is that children should not be more likely to work in countries that trade more. This is evident in Figure 11.1 which plots economic activity rates for children aged five to 14 against the importance of trade in the economy, measured by openness (trade—exports plus imports—as a share of GDP). [1] The economic activity rates in this figure are all taken from nationally representative household surveys and are reported on the Understanding Chil-

[1] Pictured economic activity rates are children aged 5-14 for all countries except Morocco, Bolivia, and Guatemala (all aged 7 to 14) and Kenya, India, Namibia, Peru, Columbia, and Turkey (all aged 6 to 14).

dren's Work (UCW) website (http://www.ucw-project.org/).[2] These surveys are conducted in different years. Each survey year has been assigned its country's openness from the World Development Indicators for that year. The modal year of the data is 2005, but survey years range from 1996 to 2007.

There is not any obvious relationship between openness and the economic activity rates of children.[3] Sudan and Vietnam have similar economic activity rates for children, but differ greatly in the importance of trade to their economies. A

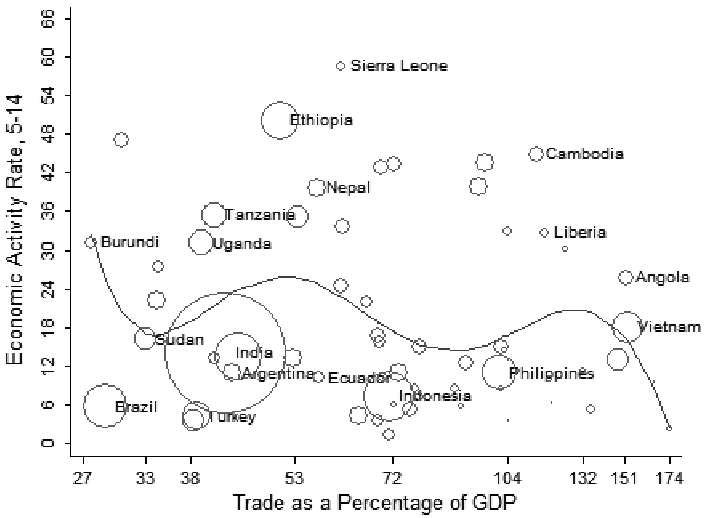

Figure 11.1: *Economic Activity and Trade*

Trade as a percentage of GDP is from the World Development Indicators (series NE.TRD.GNFS.ZS) online in May 2009. Economic Activity rates are computed from nationally representative household survey data by the UCW project as reported on their website in May 2009. When multiple years of data were available, the most recent data are pictured. Trade is taken from the same year as the household survey used to compute economic activity rates.

[2] All tabulations based on the UCW data use data for all countries available as of May 2009 that include economic activity rates for children under 10. The set of possible countries are: Angola, Argentina, Azerbaijan, Bangladesh, Belarus, Belize, Benin, Bolivia, Brazil, Burkina Faso, Burundi, Cambodia, Chile, Colombia, Costa Rica, Cote d'Ivoire, Ecuador, El Salvador, Ethiopia, Ghana, Guatemala, Honduras, India, Indonesia, Jamaica, Kenya, Lesotho, Liberia, Macedonia FYR, Madagascar, Malawi, Mali, Moldova, Mongolia, Montenegro, Morocco, Namibia, Nepal, Nicaragua, Niger, Peru, Philippines, Romania, Rwanda, Sao Tome and Principe, Senegal, Serbia, Sierra Leone, Somalia, Sri Lanka, Sudan, Swaziland, Syrian Arab Republic, Tajikistan, Tanzania, Thailand, Togo, Trinidad and Tobago, Turkey, Uganda, Ukraine, Uzbekistan, Vietnam, Yemen Rep., Zambia, and Zimbabwe. Countries with missing openness information in the World Development Indicators for the survey year of the economic activity data are not included in Figure 1.

[3] Edmonds and Pavcnik (2006a) present a similar picture using different data. They plot economic activity rates for children aged 10 to 14 from the ILOSTAT database in 2000 against openness from Penn World Tables. The ILOSTAT database no longer publishes economic activity rates for children aged 10 to 14. The ILOSTAT data include many countries for which there are no known nationally representative household surveys that would permit the estimation of economic activity rates of children. Hence, the figures differ in the time period covered, age range, data source, and openness measure.

regression of economic activity rates on openness leads to a coefficient that is small and not statistically significant with an R^2 of 0.01.

Children participate in a variety of different activities across countries (for example Table 11.1). It is possible that this heterogeneity masks the fact that the types of economic activity differ fundamentally with trade. To examine this, Figure 11.2 plots participation rates in wage employment (from the same UCW database) against the same measure of exposure to trade. The incidence of wage employment appears uncorrelated with the importance of trade in the economy. Trade is similarly important in the Tanzanian and Indian economies, but wage work is nearly five times more prevalent in India. A regression of wage work-participation rates on openness leads to a coefficient that is small and not statistically significant with an R^2 of 0.01.

The lack of an association between trade and child time allocation in the aggregate is not just a feature of work. Schooling also seems unrelated to the importance of trade in a country's economy. Figure 11.3 plots net primary school enrollment rates from the World Development Indicators against openness for the same year of data used in Figures 11.1 and 11.2 (for comparability).

Figure 11.3 shows the data do not reject the hypothesis that trade is uncorrelated with net primary school enrollment. In fact, a regression of enrollment rates on openness yields a regression with an R^2 that is 0.0006 with a small and insignificant coefficient. While Figure 11.3 is limited to countries pictured in Figures 11.1 and 11.2, the lack of an association is not just a facet of the country

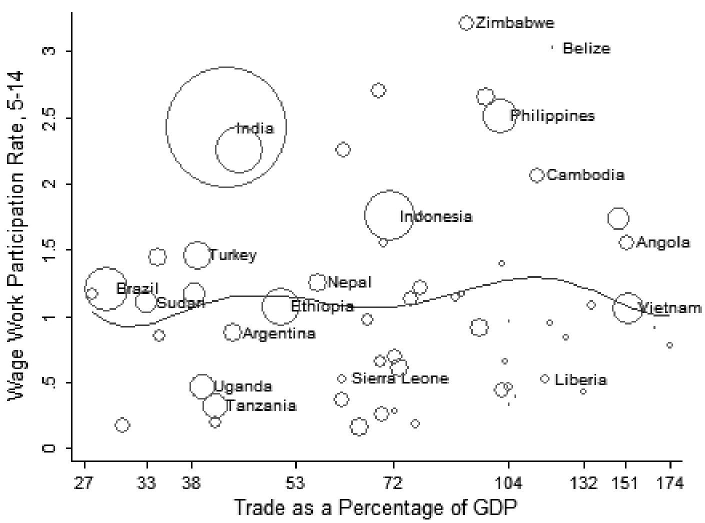

Figure 11.2: *Wage Work Involvement of Children and Trade*

Wage work participation rate is the fraction of children aged 5 to 14 (except as described in footnote 1 of the text) participating in work for pay. See notes from Figure 11.1. Not every country in Figure 11.1 appears in this figure because of missing information on the incidence of wage employment.

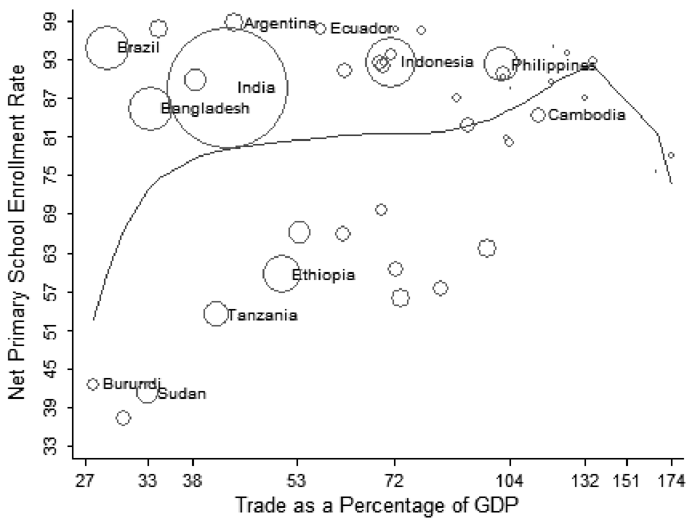

Figure 11.3: *Net Primary School Enrollment and Trade*

Net Primary School Enrollment from World Development Indicators. See Figure 11.1 for country list and the trade measure. Some countries from Figure 11.1 are missing because of missing net primary school enrollment rates in the World Development Indicators. Only countries included in Figure 11.1 are included here for comparability.

selection, as a similar picture emerges in the full World Development Indicators country list.

This observation that the importance of trade in an economy appears uncorrelated with child time allocation is consistent with the hypothesis that children work outside export and import-competing sectors. However, child time allocation is correlated with the total quantity of trade in an economy. Figure 11.4 plots economic activity rates against the log of total trade in the country, the year. This differs from Figure 11.1 in that it does not scale trade by GDP.

Children work less in countries that trade more. In fact 15 per cent of the cross-country variation in economic activity rates can be explained by the volume of trade. A line fitted to the data in Figure 11.4 implies an elasticity of –0.14. A doubling of trade is associated with a 14 per cent reduction in economic activity, on average, across countries. This observation that children work less in countries that trade more is true if one considers exports or imports alone, manufacturing trade, or agricultural trade. The reason for this robust negative correlation between economic activity and the volume of trade is that countries that trade more are richer, and richer countries have fewer working children.

While children are not involved in trade or traded sectors in general, there are exceptions. Any country with large exports of staple crops has the potential to have large-scale involvement of child labor in exports. However, academic studies identifying large-scale participation of children in export or import-compet-

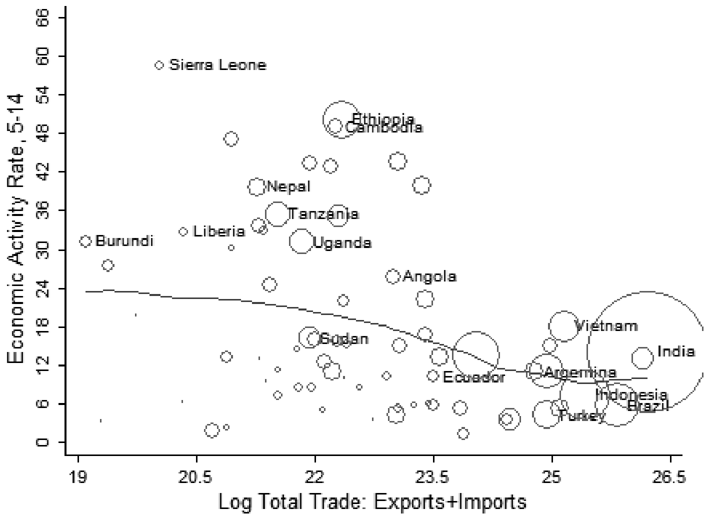

Figure 11.4: *Economic Activity and the Volume of Trade*

The volume of trade is the log of the sum of total merchandise imports and exports from the World Development Indicators. See Figure 11.1 for a description of economic activity rates.

ing crops are rare. The public policy debate and popular discourse is filled with examples where children have been directly involved in the production of exported goods. Some of the more prominent recent examples are:

- Uzbekistan was the third largest exporter of cotton in the world in 2008, and there appears to be forced child labor in its cotton exports. News reports assert that as many as 450,000 children were forced, with the help of police and government officials including school teachers, away from schools to harvest cotton with little or no compensation (for example, Newsnight on BBC Two, October 10, 2007).
- West Africa (especially Cote d'Ivoire) accounts for 70 per cent of world cocoa production, and there are many reports of forced child labor, trafficking, and abuse on cocoa farms (Salaam-Blyther et al. 2005). Estimates of the number of children involved can be as high as 300,000 with approximately 200,000 in Cote d'Ivoire alone.
- Handmade carpets are a traditional craft in South Asia, but the boom in the 1990s of exports to the United States and Europe may have lead to increased child involvement in the sector. As many as 300,000 children are reported to be involved in the export sector in India, Nepal, and Pakistan; there are accounts of trafficking of children as young as four, bondage, and abuse.

These are a few examples. There are many others. However, specific examples are difficult to isolate and identify in empirical work that focuses on population

averages rather than specific cases. The 200,000 children in cocoa farms in Cote d'Ivoire are less than 7 per cent of economically active children in Cote d'Ivoire. The 300,000 children in the carpet sector in the Indian subcontinent is small compared to the 8.4 million economically active children aged 10 to 14 in India alone. Job-specific studies may be able to identify an impact of trade on some types of child employment; but the more specific the study focus, the more difficult it is to know what children would be doing if it were not for their involvement in that specific export job. Hence, focusing on specific types of employment creates a whole new set of inference problems.

When we start to focus on older youths and young adults, employment in formal sector work and export-oriented manufacture becomes more prevalent. Hence, the possibility of finding an impact of trade on the time allocation of older youths is plausible. David Atkin (2009) notes that export sector jobs pay higher wages than non-export sector jobs in Mexico. He shows that the growth in export sector jobs between 1986 and 2000 induced youths to leave school earlier than they would have done if the jobs had never existed. These export jobs pay more initially but offer a flatter growth profile than these youths would have had if the export jobs had not come, if the students then stayed in school, accumulated more education, and entered other formal jobs. His magnitudes are large. For every 10 new jobs created, he finds that one student drops out of school at Grade 9 rather than at Grade 12.

In general, children under 15 are not working directly in trade or traded sectors. Hence, there is little evidence of a direct effect of changes in employment opportunities associated with trade on child labor or schooling, although there is some recent evidence of an impact of trade on youths over 14 in Mexico. There are certainly some sectors where children are involved in export or import-competing tasks, these sectors are sufficiently narrow that they are difficult to capture with nationally representative or aggregate data. Narrower studies are feasible, but are largely absent from existing research because of the problem of drawing inference from unique populations.

2. INDIRECT EFFECTS OF TRADE ON CHILD EMPLOYMENT OPPORTUNITIES CAN BE IMPORTANT

The indirect effects of trade on relative prices can be very important. A large literature in trade debates the relative important of trade's influence on firm size, market structure, firm productivity, input choice, technology, and factor (especially skill) intensity, and thereby inequality and returns to education. Diffuse and general equilibrium effects are notoriously difficult to identify in the data. However, there are several papers that point to a role for indirect effects of trade on child labor and schooling.

One study in Brazil suggests an important role for an indirect effect of trade, even though it does not identify how the indirect effect is working. Kruger (2007) compares changes in child labor and schooling in Northeast Brazil during a coffee boom to areas that do not export coffee. Her study is intriguing, because chil-

dren are not generally involved in coffee cultivation in Brazil. Thus, her study compares the localized, indirect effects of changes in the value of Brazil's coffee exports with changes that occur elsewhere in Brazil. Krueger (ibid.) finds that children work more and go to school less when the value of coffee exports is temporarily high. Her interpretation is that the transitory positive shock to the value of labor's output induces families to take advantage of higher wages in the local labor market, while they are high. In fact, she supposes, it is precisely the transitory nature of the price effects that are important for her results. Permanent income is largely unchanged by a transitory rise in coffee prices, so families seek to take advantage of the transitory opportunity while they can.

There is also some evidence that trade might affect child time allocation through increased opportunities for specialization. Edmonds and Pavcnik (2005b) look at how child labor in Vietnam was impacted by its liberalization of rice trade. Edmonds and Pavcnik (2006b) show that part of the effect of the growing rice trade in Vietnam was increased household specialization. Increased rice exports from rural Vietnam brought cheaper consumption goods in return that could substitute for goods previously produced by the household. Edmonds and Pavcnik (2006b) speculate that this increased household specialization is important in understanding the parts of the decline in child labor in Vietnam that cannot be explained by rising income.

Fafchamps and Wahba (2006) show that in Nepal children work less and attend school more with proximity to urban centers. Their interpretation of this correlation is that in rural areas, subsistence households cannot specialize, so they rely on family labor for much of their consumption basket. Cities bring trade and opportunities for specialization. Hence, household production becomes less important and, as proximity increases, households can buy substitutes for goods previously produced at home. They also note that specialization can also mean that children tend to specialize more. While overall children work less and attend school more with proximity to urban centers (and hence more trade opportunities), they also observe more children who work without attending school, as well as school without work. While the Fafchamps and Wahba (ibid.) study is about internal trade, the same types of channels may be facilitated by international trade as well.

While this specialization channel has received considerable attention, there are many studies that link education to returns to education. Trade appears to increase returns to education. Much of the literature discussed above on why exporting firms are more skill-intensive has this implication. Children go to school more when the returns to education are higher (Strauss and Thomas 1995 is a survey).

Few studies directly link trade, returns to education, and schooling. The Shastry (2008) study of education participation responses to rising returns to speaking English in India is one exception. Trade in services in India has lead to a rise in the return to speaking English. She shows that the parts of India where it is easier to learn English (measured by the similarity of the indigenous language to Hindi, and nationalist pressure about speaking Hindi) experience faster growth in information technology jobs and school enrollment. She also notes that there is a falling skill premium associated with this rise in schooling in areas where it is less costly to learn English.

There does not appear to be any evidence linking trade and child labor through trade's effects on returns to education. In fact, evidence of a link between child work and returns to education is relatively rare (see Edmonds 2008). This is one of the many indirect channels meriting further study. Despite the many possible mechanisms through which trade can indirectly affect child time allocation, it is surprising how little evidence there is of an impact of trade working through anything other than specialization or family incomes.

3. TRADE PRINCIPALLY AFFECTS CHILDREN THROUGH ITS IMPACT ON POVERTY

Most studies that map a connection directly between trade and child time allocation find the effect of trade on living standards to be of critical importance. This is probably because most studies are not designed to capture indirect effects, the potential for direct effects is minimal given how children spend their time, and poverty is one of the strongest correlates of child labor.

Figure 11.5 plots the economic activity rates from Figure 11.1 against GDP per capita (expressed in PPP-adjusted 2005 international dollars). It is not a representation of what is expected to happen to poor countries as they grow richer. Nonetheless, Figure 11.5 makes it clear why child labor is so often closely connected to poverty.

Sixty-two per cent of the cross-country variation in economic activity rates can be explained by GDP per capita alone.[4] If one were willing to impose a constant elasticity to the estimate, Figure 11.5 implies an elasticity of economic activity with respect to GDP per capita of –0.72. That is, the data pictured in Figure 11.5 imply that a doubling of GDP per capita should be associated with a 72 per cent reduction in the economic activity rate of children.

The schooling–GDP per capita relationship is similarly strong. Figure 11.6 plots net primary school enrollments (see Figure 11.3) against GDP per capita. The picture is not the reciprocal of Figure 11.5, because it is a different data source, and also because many out-of-school children are not economically active. In fact, Edmonds and Pavcnik (2005a) document that in the 36 countries they study most 'out of school' children are not economically active, although they are involved in domestic work.

Figures 11.5 and 11.6 do not depict a causal relationship, but the causal evidence generally depicts a similarly strong relationship.[5] For example, Edmonds and Schady (2009) consider the response of poor Ecuador families to the receipt of a lottery award that was delivered in the context of a social marketing campaign promoting human capital investment (although a lottery receipt was not

[4] Edmonds and Pavcnik (2005) present a similar picture using different data. Using the ILOSTAT estimates of economic activity rates of children aged 10 to 14 from 2000, and GDP per capita from the Penn World Tables, they find that that three-quarters of the cross-country variation in economic activity rates of children aged 10 to 14 can be explained by GDP per capita

[5] Several studies that explore the correlation between household asset wealth and child time allocation do not find a strong link. Basu et al. (2009) point out that this probably owes to the fact of the mixed role that assets play when labor markets are incomplete. Assets proxy for wealth, but they also raise the shadow value of child time by making child time in the household more productive.

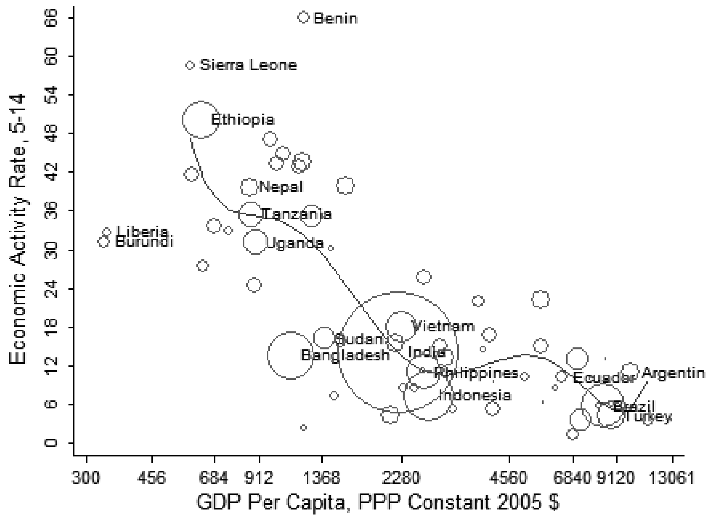

Figure 11.5: *Economic Activity and Gross Domestic Product*

GDP per capita is in constant PPP 2005 international dollars. It is taken from the World Development Indicators (series NY.GDP.PCAP.PP.KD) online in May 2009. See notes to Figure 11.1 for list of countries and the definition of the economic activity rate.

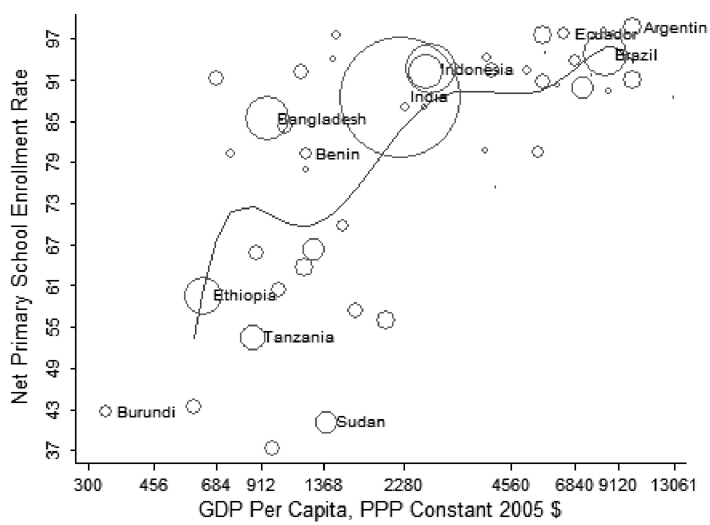

Figure 11.6: *Net Primary School Enrollment and Gross Domestic Product*

See notes to Figure 11.3 and Figure 11.4 for definitions of variables and country information.

conditional on taking any particular actions). They find that the lottery appears to be spent largely on education when families have children at the age of the transition from (free) primary school to (costly) secondary school. The result is an increase in education with the lottery, and a decline in child participation in paid employment such that total household income declines with the lottery award because of the decline in child labor.

There are many plausible reasons why poverty and child time allocation are so closely linked. One class of causal channels is associated with preferences. Child labor may be a bad, and schooling a good, in preferences. Poverty may lead households to choose to consume low-quality household-produced items over market-produced goods. Another class of reasons for a poverty-time allocation connection owes to liquidity constraints. Liquidity constraints may force families to underinvest in schooling because of schooling costs or a high marginal utility of current income. They may also cause households to underinvest in labor-saving technologies or other types of investment in human capital, such as nutrition, which in turn lowers the relative return to schooling.

Edmonds, Pavcnik, and Topalova (2007) shed some light on the question of why children work by examining how children in rural India were impacted by India's tariff reforms in the early 1990s. Since the 1950s, India has imposed large, distortionary tariffs on imported goods. These tariffs protected some jobs and employment opportunities at the expense of other workers and higher prices. Concurrent with the phased-in reduction in tariffs and other reforms that started in 1991, India's economy boomed. While much of India grew, rural areas with concentrations of pre-reform employment in industries that lost protection experienced smaller declines in poverty than the rest of India. Children living in these areas did not experience as large an increase in school attendance, or decline in work without school, as children residing in areas with lower pre-reform employment in heavily protected industries.

The attenuation in schooling improvements and child labour declines in these rural areas appears attributable to smaller reductions in poverty than elsewhere in India. There is little evidence that trade liberalization affected child labor demand or returns to education in rural areas, although this does not appear to be the case for urban India (Edmonds, Pavcnik, and Topalova 2009). The study looks at how children work in order to understand why there is such a close trade–poverty–child labor and schooling connection. Children are working more in areas that have lost protection relative to the national trend, but most of this work involves girls working around their family. Moreover, relatively more children are neither working as a principal activity nor attending school. For these children, their primary economic contribution to their family appears to be the avoidance of schooling costs, which can be considerable for a poor family. In fact, the study finds that the attenuation of schooling improvements associated with the loss of protection is smaller in parts of rural India where schooling is less costly.

These findings from India mirror the findings from another recent study by Edmonds and Pavcnik (2005b) in Vietnam. For a number of years, Vietnam used an

export quota to suppress rice exports out of a concern for domestic food security. In the 1990s, Vietnam liberalized its rice trade and allowed rice farmers to take advantage of higher international prices. The rice sector boomed and living standards of rice-producing households improved substantively. Edmonds and Pavcnik (ibid.) document that despite greater employment opportunities, children in households that benefited from higher rice prices became much less likely to work. Altogether, it appears that roughly one million fewer children worked as a result of rising rice prices in Vietnam, despite potentially more lucrative employment opportunities.

The primacy of living standards in the child labor–trade relationship has also been documented by other authors. Dammert (2008) shows that child labor increased in Peru in coca growing areas when eradication efforts and other attempts to limit coca trade lowered family incomes. Cogneau and Jedwab (2008) also find support for the primacy of income in the child labor–trade relationship. A permanent reduction in cocoa prices appears to have lowered family incomes in Cote d'Ivoire. As a result, children in cocoa growing households are attending school less and working more, relative to households without suitable land for cocoa cultivation in the same areas.

Interestingly, both the Vietnam and Cote d'Ivoire studies are ones where there is considerable possibility of a direct effect of trade on child labor working through labor demand that would be identifiable in the data. However, in both cases, despite the changes in the value of crops produced by the child labor, the main channel for influencing time allocation appears to be child income. At this point in the literature, it seems reasonable to suppose that the dominant channel through which trade will affect child labor depends on how trade changes the living standards of the poor, even in cases where children work in the sector where trade is changing.

4. PRIORITIES FOR FUTURE RESEARCH

Overall, the literature on trade and child time allocation is relatively small. From the perspective of the trade literature, part of the reason for this is that the relationship between trade and schooling on child labor is typically a collateral effect of trade's impact on some other part of the economy (like adult income). Still, anything that alters educational decisions or occupational choice has the potential to have long-term consequences in settings where poverty traps are possible. For example, a transitory shock on the income of an adult associated with the loss of protection may be much less substantive for the adult than for the affected child whose education is permanently disrupted. Better understanding of the nature of the collateral effects of trade adjustment, especially in agriculture, and how to ameliorate their effects on children merits further research.

There is considerable scope for expanding our understanding of the circumstances under which trade can have a direct effect on child time allocation by altering the demand for child labor. Why do children participate in some jobs and not others? Is child participation in a trade-related job simply an alternative to some other job

that is revealed to be worse through the child's job choice? What is the child's own wage elasticity of labor supply? When do children work no matter how low the wage? When might labor demand (relative price) effects be more important than income effects? These basic questions have not been answered conclusively in the child time allocation literature and are critical for improving our understanding of the relationship between trade, schooling, and child labor.

In fact, one of the appeals of studying issues related to trade, schooling, and child labor is that one can potentially observe all of these dynamics simultaneously. This creates interesting reduced forms where researchers can attempt to disentangle the underlying model driving the reduced forms. When this has been done (Edmonds and Pavcnik 2005b; Edmonds, Pavcnik, and Topalova 2007; Dammert 2008; Cogneau and Jedweb 2008), researchers have found changes in living standards to be the driving force between the change in trade or prices and time allocation, even in circumstances like Vietnam's rice sector or Cote d'Ivoire's cocoa sector where it would be reasonable to expect substantive changes in child wages. The next step in this line of research is to begin to impose some economic structure on this research in order to recover parameters that have a clearer policy interpretation and that can be compared across environments. Variation owing to a trade reform, while appealing in the reduced form, is somewhat problematic for a more structured approach because of the challenge of having one source of variation for so many different possible parameters.

The main question of interest to consumers in high-income countries with regard to trade and child labor is: 'To what extent does my consumption of goods from poor countries perpetuate child labor and low rates of schooling attendance in poor countries?' The literature to date can largely say that anything that raises income in poor countries will likely reduce child labor and increase schooling there. However, this sort of generality is likely to be unsatisfactory to a consumer asking this about a specific product. Future research that looks more at specific sectors, combined with an effort to develop estimates of parameters of clear economic interpretation, can help researchers provide a more precise answer to the consumer's question in the future.

Eric V. Edmonds is an Associate Professor of Economics at Dartmouth College.

BIBLIOGRAPHY

Atkin, David (2009). *Endogenous Skill Acquisition and Export Manufacturing in Mexico.* Unpublished Manuscript (Princeton, NJ: Department of Economics)

Bangladesh Bureau of Statistics (2003). Report on National Child Labour Survey 2002–2003 Government of the People's Republic of Bangladesh, Parishankhan Bhaban, Dhaka, Bangladesh

Basu, K, Das, S and B Dutta (2007). *Child Labor and Household Wealth: Theory and empirical evidence of an inverted-U.* Working Paper no. 139 Bureau for Research and Economic Analysis of Development Working Paper, Cambridge MA

BBC Two (2007). *Cotton Picked by Children Appears in UK High Street Clothes*, October 30, 2007 Broadcast http://www.bbc.co.uk/pressoffice/pressreleases/stories/2007/10_october/30/newsnight.shtml

Cogneau, Denis and Remi Jedwab (2008). *Family Income and Child Outcomes: The 1990 cocoa price shock in Cote d'Ivoire.* Unpublished manuscript, Paris France: Paris School of Economics

Dammert, Ana (2008). Child Labor and Schooling Responses to Changes in Coca Production in Rural Peru. *Journal of Development Economics* 86(1), April 2008, 164–80.

Dorosh, Peter A (2009). Price Stabilization, International Trade, and National Cereal Stocks: World price shocks and policy response in South Asia. *Food Security*, vol. 1, no. 2, June: 137–40, DOI 10.1007/s12571-009-0013-3

Edmonds, Eric V (2008). Child Labor, in T P Shultz and J Strauss, (Eds.) *Handbook of Development Economics,* vol. 4, Ch. 57, 3608-3709.

Edmonds, Eric V and Nina Pavcnik (2005a). Child Labor in the Global Economy. *Journal of Economic Perspectives*, vol. 18, no. 1, Winter, 199–220

Edmonds, Eric and Nina Pavcnik (2005b). The Effect of Trade Liberalization on Child Labor: Evidence from Vietnam. *Journal of International Economics* 65, 401–41

Edmonds, Eric and Nina Pavcnik (2006a). International Trade and Child Labor: Cross-country Evidence. *Journal of International Economics* 68 (1), January 2006, 115–40

Edmonds, Eric and Nina Pavcnik (2006b). Trade Liberalization and the Allocation of Labor between Households and Markets in a Poor Country. *Journal of International Economics* 69 (2), July 2006, 272–95

Edmonds, Eric, Pavcnik, Nina and Petia Topalova (2007). Trade Adjustment and Human Capital Investments: Evidence from Indian tariff reform. *American Economic Journal: Applied Economics,* forthcoming

Edmonds, Eric, Nina Pavcnik, and Petia Topalova (2009). Child Labor and Schooling in a Globalizing World: Some evidence from urban India. *Journal of the European Economics Association*, vol. 7, issue 2–3, 498–507

Edmonds, Eric and Norbert Schady (2009). *Poverty Alleviation and Child Labor.* NBER Working paper 15345

Fafchamps, Marcel and Jacqueline Wahba (2006). Child Labor, Urban Proximity, and Household Composition. *Journal of Development Economics* 79: 374–97

Feenstra, Robert and Gordon Hanson (1996). Foreign Investment, Outsourcing, and Relative Wages. In *Political Economy of Trade Policy: Essays in honor of Jadish Bhagwati.* R Feenstra and G Grossman (Eds), Cambridge MA: MIT Press

Kenyan Central Bureau of Statistics (2001). *The 1998/99 Child Labour Report.* Nairobi: Central Bureau of Statistics, Ministry of Finance and Planning, Republic of Kenya

Kruger, Diana. (2007). Coffee Production Effects on Child Labor and Schooling in Rural Brazil. *Journal of Development Economics* 82(2), March 2007, 448–63

Salaam-Blyther, Tiaji, Charles Hanrahan, and Nicolas Cook (2005). *Child Labor in West African Cocoa Production: Issues and US Policy*. Washington, D.C.: CRS Report for Congress Order Code RL32990, July 13, 2005

Shastry, Kartini (2008). *Human Capital Response to Globalization: Education and information technology in India*. Unpublished manuscript, Harvard University

Verhoogen, Eric (2008). Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector, *Quarterly Journal of Economics*, vol. 123, no. 2, 489-530, May

Wagner, Joachim (2007). Exports and Productivity: A Survey of the Evidence from Firm-Level Data, *The World Economy*, vol. 30, no. 1, January, 60–82

12

Adjustment to Internal Migration

GORDON H HANSON

1. INTRODUCTION

In this note I discuss recent empirical work on the consequences of global labor mobility. I examine how international migration affects the incomes of individuals in sending and receiving countries and of migrants themselves. In considering the effects of labor mobility, I give equal weight to sending and receiving economies, meaning neither receives in-depth treatment. For more detailed discussions of the literature on how immigration affects receiving countries, see Borjas (1999a), and Card (2005); and on research into how emigration affects sending countries, see Docquier and Rapoport (2008), and Hanson (2007).

In Section 2 I summarize facts about international migration that emerged from recently available data. In Section 3, I discuss empirical research on the consequences of labor flows for incomes in sending and receiving countries and for migrants and their family members. And in Section 4, I offer concluding remarks.

2. DATA ON INTERNATIONAL MIGRATION

Despite large differences in income between countries, international migration is uncommon. Figure 12.1, using UN data, shows that in 2005 individuals residing outside their country of birth comprised just 3 per cent of the world's population. During the last two decades, the stock of international migrants has grown modestly, from 2.2 per cent of the world population in 1980 to 2.9 per cent in 1990 and it grew marginally after that.

Table 12.1 shows the share of the population that is foreign born in selected OECD countries. The countries with the largest immigrant presence in 2005 are Australia (24 per cent), Switzerland (24 per cent), New Zealand (19 per cent), and Canada (19 per cent). Australia, New Zealand, and Canada use a point system to govern applications for admission, in which individuals with higher levels of skill are favored for entry. Next in line are the large economies of Germany (13 per cent), the United States (13 per cent), France (10 per cent), and the United Kingdom (10 per cent). The United States alone hosts 40 per cent of immigrants living in OECD countries making it the world's largest receiving country. The United States uses a quota system to govern legal immigration, with two-thirds of visas reserved for family members of US citizens or residents. European countries tend

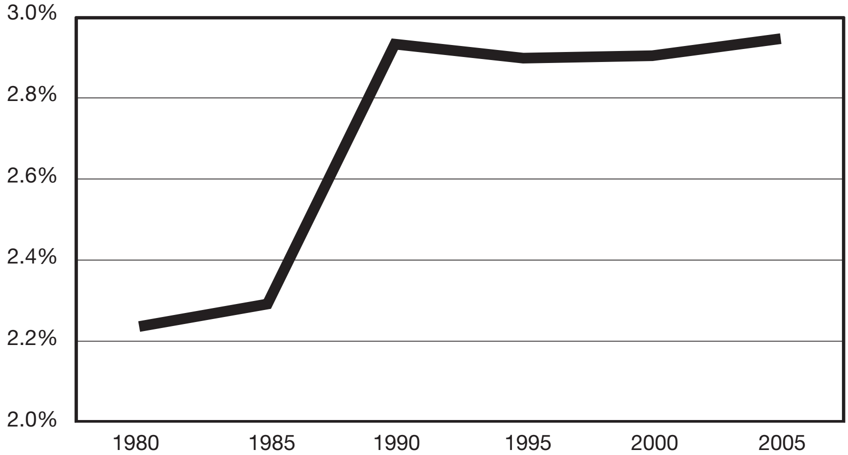

Figure 12.1: *Percentage of World Population Comprised of International Migrants*

Source: United Nations (2006)

to place more emphasis on an individual's refugee or asylum-seeking status in making immigrant admission decisions (Hatton and Williamson, 2004).

Inflows of illegal immigrants account for a substantial share of total immigration. In the United States, Passel and Cohn (2009) estimate that in 2008 there were 12 million illegal immigrants, which accounted for 35 per cent of the US foreign-born population, up from 28 per cent in 2000, and 19 per cent in 1996. In Europe, Jandl (2003) estimates that in 2003 there were 4 million illegal immigrants in the EU-15 countries, with the largest stocks in Germany, the United Kingdom, Italy, and France. Greece, Italy, Portugal, and Spain have engaged in repeated recent legalizations of illegal immigrants, meaning that the current stock of illegal immigrants in these countries understates the number of immigrants who first entered the country illegally.

Table 12.2, based on data from Beine, Docquier and Marfouk (2008), shows the share of the immigrant population in OECD countries by sending-country region. In 2000, 67 per cent of immigrants in the OECD were from a developing country, up from 54 per cent in 1990. Among developing-country sending regions, Mexico, Central America, and the Caribbean are the most important accounting for 20 per cent of OECD immigrants in 2000, up from 15 per cent in 1990. Half of this region's migrants come from Mexico, which in 2000 was the source of 11 per cent of OECD immigrants, making it the world's largest supplier of international migrants. The next most important developing source countries for OECD immigrants are Turkey (3.5 per cent of OECD immigrants); China, India, and the Philippines (each with 3 per cent); Vietnam, Korea, Poland, Morocco, and Cuba (each with 2 per cent); and Ukraine, Serbia, Jamaica, and El Salvador (each with 1 per cent).

Table 12.1: *Percent of Foreign-born Population in Total Population*

	1995	2000	2005	Change 1995–2005
Australia	23.0	23.0	23.8	0.8
Austria		10.5	13.5	
Belgium	9.7	10.3	12.1	2.4
Canada	16.6	17.4	19.1	2.5
Czech Republic		4.2	5.1	0.9
Denmark	4.8	5.8	6.5	1.7
Finland	2.0	2.6	3.4	1.4
France (a)		10.0	8.1	
Germany (b)	11.5	12.5		
Greece (c)		10.3		
Hungary	2.8	2.9	3.3	0.5
Ireland (d)	6.9	8.7	11.0	4.1
Italy (c)		2.5		
Mexico	0.4	0.5	0.4	0.0
Netherlands	9.1	10.1	10.6	1.5
New Zealand (d)	16.2	17.2	19.4	3.2
Norway	5.5	6.8	8.2	2.7
Poland				
Portugal	5.4	5.1	6.3	0.9
Slovak Republic (c)		2.5	3.9	
Spain (c)		5.3		
Sweden	10.5	11.3	12.4	1.9
Switzerland	21.4	21.9	23.8	2.4
Turkey		1.9		
United Kingdom	6.9	7.9	9.7	2.8
United States	9.3	11.0	12.9	3.6

Notes: (a) 2000 value is from 1999; (b) 2004 value is from 2003; (c) 2000 value is from 2001; (d) 1995 value is from 1996. Source: *International Migration Outlook*, OECD, 2006 (1995 data) and 2007.

Within sending countries, emigrants tend not to be drawn randomly from the population. Figure 12.2, taken from Grogger and Hanson (2008), plots the log odds of emigration for individuals with tertiary education (13 or more years) against the log odds of emigration for individuals with primary education (0 to 8 years). Nearly all points lie above the 45 degree line, indicating that in most countries individuals with more education are much more likely to leave. Migrants thus appear to be strongly positively selected in terms of schooling. It is high emigration rates for the more educated that have raised concerns about a brain drain from developing countries.

Table 12.3 compares education levels for adult immigrants and adult residents in Europe, North America, and Australia–New Zealand. In Europe and North America, immigrants are much more likely than residents to have less than a secondary education. In Australia and New Zealand, immigrants and residents have more similar education levels. These patterns matter for gauging the labor market impacts of immigration, for they mean that foreign labor inflows tend to increase the relative supply of low-skilled labor in receiving countries.

Table 12.2: *Share of OECD Immigrants by Sending Region, 2000*

	Share of Immigrants by OECD Receiving Region				Change in OECD Share
Low-income Sending Region	All OECD	North America	Europe	Asia, Oceania	1990 to 2000
Mexico, Cen. Am., Caribbean	0.202	0.374	0.025	0.002	0.053
Southeast Asia	0.102	0.137	0.039	0.160	0.016
Eastern Europe	0.099	0.049	0.161	0.116	0.042
Middle East	0.063	0.032	0.113	0.029	0.001
South Asia	0.052	0.052	0.055	0.036	0.011
North Africa	0.044	0.009	0.098	0.018	-0.006
South America	0.041	0.050	0.031	0.035	0.010
Cen., South Africa	0.036	0.021	0.061	0.021	0.007
Former Soviet Union	0.029	0.023	0.042	0.010	-0.002
Pacific Islands	0.004	0.003	0.001	0.027	0.000
Total	0.672	0.750	0.626	0.454	0.132
High-income Sending Region					
Western Europe	0.244	0.152	0.336	0.368	-0.111
Asia, Oceania	0.055	0.062	0.018	0.156	-0.010
North America	0.029	0.037	0.020	0.023	-0.011
Total	0.328	0.251	0.374	0.547	-0.132

Notes: This table shows data for 2000 on the share of different sending regions in the adult immigrant population of the entire OECD and of three OECD sub regions. 'High-income North America includes Canada and the United States and High-income Asia and Oceania includes Australia, Hong Kong, Japan, Korea, New Zealand, Singapore, and Taiwan. Source: author's calculations using data from Beine et al. (2007).

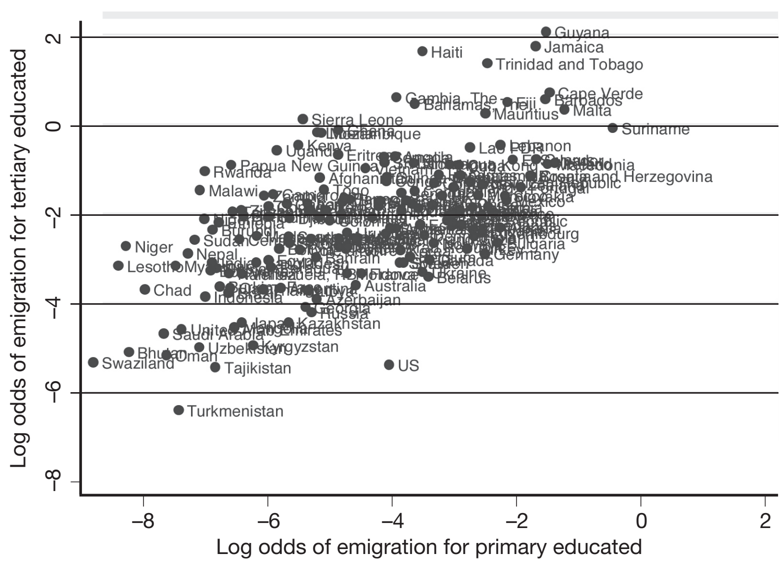

Figure 12.2: *Positive Selection of Emigrants—2000*

Source: Grogger and Hanson (2008)

Table 12.3: *Schooling of Residents and Immigrants by Destination Region*

		Share in adult immigrant population			Share in adult resident population		
		primary education	secondary education	tertiary education	primary education	secondary education	tertiary education
1990	Europe	0.616	0.210	0.174	0.332	0.485	0.183
	Canada, US	0.388	0.180	0.433	0.118	0.486	0.397
	Australia, N. Zealand	0.303	0.322	0.375	0.311	0.391	0.298
2000	Europe	0.510	0.240	0.250	0.233	0.541	0.226
	Canada, US	0.367	0.181	0.453	0.060	0.427	0.513
	Australia, N. Zealand	0.285	0.267	0.448	0.248	0.425	0.327

Note: This table shows the share of adult (25 years and older) immigrants or residents by education group: primary, secondary, and tertiary. Data for immigrants are from Beine et al. (2007); data for residents are from Docquier and Marfouk (2006). Europe consists of Austria, Belgium, Denmark, Finland, Germany, the Netherlands, Sweden, and the United Kingdom.

3. EMPIRICAL RESEARCH ON INTERNATIONAL MIGRATION

What does empirical research have to say about how international migration affects the incomes of individuals in sending and receiving countries? I begin the discussion by considering the gain in income to migrants, and evidence on the extent to which migrants share these gains with their family members in the sending country. I then consider the impact of global labor flows on labor market earnings, net tax burdens, and skill acquisition in sending and receiving countries.

3.1. Income Gains to Migrants

Combining household survey data in developing countries with data from the US Census, Clemons, et al. (2008) estimate the gains to international migration for individuals from a sample of 42 developing countries. For a young male with some secondary education, they find that the median annual gain from migrating to the United States is \$11,200 (after applying a correction for the self-selection of individuals into migration). Rosenzweig (2006) uses data from the New Immigrant Survey to examine the change in income for a random sample of new US permanent legal immigrants in 2003. He estimates that the annual gain from legal migration to the United States is \$10,600.

The income gain from migration captures the gross return from moving to another country. If migration costs are large, the net gain may be smaller. While there is research on migration networks (Hanson, 2007), there is little work that estimates the actual cost of migration. These costs include transport expenses in moving abroad, time lost in changing labor markets, administrative fees for legal migration, border crossing costs in illegal migration, the psychic costs of leaving home, and any perceived increase in uncertainty from living and working in another country.

Who benefits from the increase in income that migrants enjoy? Through remittances, migrants share a portion of their extra income with family members at home. Table 12.4 shows workers' remittances received from abroad as a share of GDP by geographic region. Remittances have increased markedly in East Asia and the Pacific, Latin America and the Caribbean, South Asia, and Sub-Saharan Africa. As of 2005, remittances exceeded official development assistance in all regions except Sub-Saharan Africa and were greater than 65 per cent of foreign direct investment inflows in all regions except Europe and Central Asia. Among the smaller countries of Central America, the Caribbean, and the South Pacific, remittances account for a large share of national income, ranging from 10 per cent to 17 per cent of GDP in the Dominican Republic, Guatemala, El Salvador, Honduras, Jamaica, and Nicaragua, and represent an astounding 53 per cent of GDP in Haiti (Acosta, et al., 2007). Remittances appear to have fallen sharply with the recent global economic downturn.

Having migrants abroad provides insurance to households, helping them smooth consumption in response to income shocks. Yang (2007) examines changes in remittances to households in the Philippines before and after the Asian financial crisis. As of 1997, 6 per cent of Philippine households had a member who had migrated abroad. Some had gone to countries in the Middle East, whose currencies appreciated sharply against the Philippine peso in 1997–98, while others had gone to East Asia, where currencies appreciated less sharply or even depreciated. Consistent with consumption-smoothing, remittances increased more for households whose migrants resided in countries that experienced stronger currency appreciation against the peso. Since income shocks associated with movements in exchange rates are largely transitory in nature, the response of remittances reveals the extent to which migrants share transitory income gains with family members at home. A 10 per cent depreciation of the Philippine peso is associated with a 6 per cent increase in remittances.

There is some evidence that increases in remittances are associated with increased expenditure on education and health. Yang (2007) also examines changes in household expenditure and labor supply in the Philippines. Households with migrants in countries experiencing stronger currency appreciation vis-à-vis the peso had larger increases in spending on child education, spending on durable goods (televisions and motor vehicles), children's school attendance, and entre-

Table 12.4: *Workers' Remittances and Compensation of Employees, % of GDP*

Region	1992	1996	2000	2006
East Asia and Pacific	0.55	0.69	0.97	1.46
Europe and Central Asia	0.32	1.04	1.45	1.42
Latin America and Caribbean	0.66	0.71	0.99	1.92
Middle East and North Africa	5.67	3.32	2.87	3.64
South Asia	1.74	2.39	2.83	3.47
Sub-Saharan Africa	0.68	0.94	1.35	1.45

Source: World Development Indicators

preneurial investments. In these households, the labor supply of 10 to 17 year old children fell by more, particularly for boys. Using cross-section data on Mexican states, Woodruff and Zenteno (2007) find a positive correlation between emigration and business formation. These results suggest that migration may help households to overcome credit constraints imposed by the sending-country financial markets.

3.2. Labor Market Consequences

The labor market consequences of international migration have inspired intense debate among scholars. Most research focuses on the impact of labor inflows on the US wage structure. Only recently has the literature begun to examine other receiving countries or the effects on sending economies. The US literature has been extensively reviewed elsewhere (see Borjas, 1999a; Card, 2005). I summarize the current state of the debate and identify questions that are central to resolving it.

Research using data on the national US labor market suggests that immigration depresses wages for US workers. Borjas (2003) defines labor markets at the national level according to a worker's education and labor market experience. Over the period 1960 to 2000, education–experience cells in which immigrant labor supply growth has been larger—such as for young high school dropouts—have had slower wage growth, even after controlling for education or experience-specific wage shocks. The evidence is consistent with immigration having depressed wages for low-skilled US workers (as well as for some high-skilled workers). The concern about this approach is that it might confound immigration with other labor market shocks that have hurt low-skilled workers, such as skill-biased technological change. Absent controls for these other shocks, one cannot be sure that the attributed wage changes are really due to immigration.

Applying a similar nationallevel approach to Canada, Aydemir and Borjas (1997) find comparable evidence of the wage effects of migration. In Canada, where immigration is dominated by workers at the top end of the skill distribution, immigration is negatively correlated with wages across education–experience cells, with more-educated workers being the ones who have suffered the largest wage effects. Since Canada is presumably subject to many of the same technology shocks as the United States, it would not appear that unobserved technology shocks could explain away the wage effects of immigration in both countries. Moreover, the national-level approach yields comparable results of the wage effects of migration in sending countries. Mishra (2007) finds a positive correlation between emigration and wages across education–experience cells in Mexico. In Mexico, emigrants come disproportionately from the middle of the skill distribution, meaning workers with close to average levels of education are those who have had the largest wage gains from labor outflows. Aydemir and Borjas (1997) obtain similar results and also find that the elasticity of wages with respect to labor supply is roughly similar in Canada, Mexico, and the United States. In all three countries, a 10 per cent change in labor supply due to migration is associated with a 4 to 6 per cent change in wages.

An older and larger literature has searched for immigration's impact by correlating the change in wages for low-skilled US natives with the change in the immigrant presence in local labor markets, typically at the level of US cities. These area studies tend to find that immigration has little if any impact on US wages (Borjas, 1999a). Card (2005) argues that if immigration has affected the US wage structure, one should see larger declines in the wages of native high school dropouts (relative to, say, native high school graduates) in US cities where the relative supply of high school dropouts has expanded by more. In fact, the correlation between the relative wage and the relative supply of US high school dropouts across US cities is close to zero.

One type of cross-sectional evidence is consistent with immigration having lowered wages. Cortes (2008) finds that in the 1980s and 1990s US cities with larger inflows of low-skilled immigrants experienced larger reductions in prices for housekeeping, gardening, child care, dry cleaning, and other labor-intensive services. A 10 per cent increase in the local immigrant population is associated with decreases in prices for labor-intensive services of 1.3 per cent percent. A mechanism through which immigration could have lowered prices is through its effects on wages.

The area study approach also has problems. Immigrants may tend to settle in US regions in which job growth is stronger, causing one to underestimate the wage impact of immigration when using city or state-level data. As a correction, many studies instrument for growth in local immigrant labor supply using lagged immigrant settlement patterns. But this strategy requires rather strong identifying assumptions. It would be invalid for instance, if the labor demand shocks that influence immigrant settlement patterns are persistent over time (Borjas, Freeman, and Katz, 1997).

Research on other receiving countries tends to report negligible estimated impacts of immigration on wages. After the fall of the Soviet Union, there was sizeable migration of Russian Jews to Israel, which increased the Israeli population by 12 per cent. Over the course of the Russian influx, Friedberg (2001) finds that occupations that employed more immigrants had slower wage growth, but the correlation is zero once she instruments for immigrants' occupational choice.[1] In applications of the area studies approach outside of the United States, findings of little or no impact of immigration on regional wages include Addison and Worswick (2002) for Australia; Pischke and Velling (1997) for Germany; Zorlu and Hartog (2005) for the Netherlands and Norway (2005); Carrasco, Jimeno, and Ortega (2008) for Spain; and Dustmann et al. (2005) in the United Kingom.[2]

The impact of immigration on the migration of native labor is another issue about which there is disagreement. Card (2001; 2005) finds that across US cities, higher presence of low-skilled immigrants is associated with higher levels of em-

[1] Hunt (1992) and Carrington and de Lima (1996) find evidence of minimal labor market effects from the forced return of expatriates in France and Portugal, following the end of colonialism.

[2] Negative wage effects of immigration have been found in Germany (De New and Zimmerman, 1994) and Austria (Hofer and Huber, 2003).

ployment of low-skilled labor, with one new immigrant net adding about one new net worker to a labor market, suggesting that native outmigration does not offset the labor supply effects of arriving immigrant workers. Pischke and Velling (1997) find a similar absence of native displacement effects in Germany. Borjas (2006), using the regional counterpart to the national level education–experience cells as in Borjas (2003), comes to the opposite conclusion. He finds that the growth in the native workforce is smaller in regional education–experience cells in which the growth in immigrant presence has been larger. Moreover, he shows that not accounting for the internal migration of natives causes area studies' regressions to understate the wage effects of immigration by about half. Hatton and Taini (2005), using data on regional labor markets in the United Kingdom, also find evidence that the arrival of immigrant workers displaces local native workers. Again, we have an instance in which national level and regional level approaches yield different results.

Does immigration induce firms to raise investment and increase innovation, while partially or fully offsetting the wage impacts of labor inflows? While the idea is plausible, there is relatively little empirical research on the impact of immigration on investment or innovation at the regional or national level. There is evidence that immigration is associated with changes in production techniques. Lewis (2005) finds that regions absorb immigrants through their industries becoming more intensive in the use of immigrant labor. Industries in US cities that have received larger inflows of low-skilled immigrant labor have increased their relative labor intensity by more than cities receiving lower inflows. These industries have also been slower to adopt new technologies, suggesting that changes in labor supply affect incentives for technology adoption, as in Acemoglu (1998). Lewis's (2005) results rule out changes in sectoral mix accounting for regional absorption of immigrant labor, as could occur in a simple Heckscher–Ohlin model. He finds little evidence that regions have absorbed incoming immigrants by shifting employment towards sectors that are more intensive in low-skilled labor.

In initial work, Ottaviano and Peri (2007) found evidence that immigrant and native labor were imperfect substitutes.[3] They estimated a negative and significant correlation between immigrant–native relative wages and immigrant–native relative employment, across Borjas's (2003) education–experience cells. However, their results are sensitive to how one defines skill groups. Dropping high school students from the sample, the finding of imperfect substitutability between immigrants and natives disappears. Borjas, Grogger, and Hanson (2008) show that for many specifications and factor-supply definitions one cannot reject the hypothesis that comparably skilled immigrants and natives are perfect substitutes in employment, in line with Jaeger (1997). Whatever one thinks about the wage effects of immigration, low-skilled immigrant and native workers appear to be in the same labor market, at least in the United States.

[3] Galosto, Venturini, and Villosio (1999) find evidence of imperfect substitutability between immigrants and natives in Italy.

To date, the literature offers two approaches for estimating the wage effects of migration, which yield quite different results. The national approach is subject to concerns about how one controls for changes in technology, though these should be at least partly allayed by the fact that countries with very different types of migration shocks exhibit similar migration wage elasticities. The area studies approach is subject to concerns about the endogeneity of immigrant settlement patterns, with it being difficult to assess the validity of proposed solutions to this problem. An issue often overlooked is that in an economy without distortions, even if all workers lose from immigration, the income gain to capital owners will be sufficient to ensure that national income increases. Indeed, it is unlikely that an economy could experience a gain in national income from immigration without some workers losing out.

3.3. Fiscal Consequences

By changing absolute and relative labor supplies, international migration may have consequences for a country's fiscal accounts. With emigrants being positively selected in terms of schooling, it is individuals with relatively high skill levels who leave sending countries, depriving them of higher-income taxpayers. To the extent that education and health care are public, sending countries may have made substantial investments in these individuals while they were young, only to have receiving countries reap the returns. The potentially adverse effects of brain drain on economic growth may be compounded by net tax losses from high-skilled emigration.

While there is a large body of theoretical literature on the taxation of skilled emigration (see Docquier and Rapoport, 2007), empirical research on the subject is sparse. In a recent contribution, Desai et al. (2008) examine the fiscal effects of brain drain from India. In 2000, individuals with tertiary education accounted for 61 per cent of Indian emigrants but just 5 per cent of India's total population. Between 1990 and 2000, the emigration rate for the tertiary educated rose from 2.8 per cent to 4.3 per cent, but from just 0.3 per cent to 0.4 per cent for the population as a whole. Desai et al. (ibid.) examine Indian emigration to the United States, which in 2000 was host to 65 per cent of India's skilled emigrants. First, they use Mincer wage regressions to produce a counterfactual income series that gives emigrants the income they would have earned in India based on their observed characteristics and the returns to these characteristics in India. On the tax side, they calculate income tax losses by running the counterfactual income series through the Indian income tax schedule. They also calculate indirect tax losses using estimates of indirect tax payments per unit of gross national income. On the spending side, they calculate expenditure savings by taking the categories for which savings would exist—which are most government accounts except interest payments and national defense—then estimating savings per individual. Their results suggest Indian emigration to the United States cost India net tax contributions of 0.24 per cent of GDP in 2000. Remittances by skilled emigrants generated a tax gain of 0.1 per cent of GDP, partially offsetting these losses. For

India, the tax consequences of skilled emigration appear to be small, though small countries with high emigration rates may face larger impacts.

In receiving countries, immigration may exacerbate inefficiencies associated with a country's system of public finance. Where immigrants pay more in taxes than they receive in government benefits, immigration reduces the net tax burden on native taxpayers. The total impact of immigration on native residents—the sum of the immigration surplus (the pretax income gain) and the net fiscal transfer from immigrants—would be positive. With progressive taxes and means-tested entitlements in many receiving countries, positive fiscal consequences from immigration would be more likely, the more skilled is the labor inflow. In contexts where immigrants pay less in taxes than they receive in government benefits, immigration increases the net tax burden on natives, necessitating an increase in taxes on natives, a reduction in government benefits to natives, or increased borrowing from future generations.

There are dynamic fiscal effects from immigration (Auerbach and Oreopoulos, 1999). If the net tax burden on residents of a country is expected to increase in the future, immigration increases the tax base over which the burden is spread and reduces the increase that natives have to bear (Collado and Valera, 2004). But this is only true if the descendents of immigrants make positive net tax contributions. If the children of immigrants have low educational attainment, high levels of immigration today could instead increase the future tax burden on the native population.

In the United States, immigrant households have historically made greater use of subsidized health care, income support to poor families, food stamps, and other types of public assistance (Borjas and Hinton, 1996). The reason for native–immigrant differences in the uptake of welfare programs is simple. Immigrant households tend to be larger than native households; they have more children, and have very low incomes, making them eligible for more types of benefits. In the last decade, however, the difference between immigrant and native use of welfare programs in the United States has fallen because of reforms to welfare policy, which restricted non-citizens from having access to many federally funded benefit programs. While immigrant households still make greater use of public healthcare than native households do, their use of other types of public assistance has fallen (Borjas, 2003; Capps et al., 2005). In the European Union, enlargement to include lower income countries in Central and Eastern Europe has lead to low-skilled migration to higher income countries, possibly increasing welfare usage (Sinn, 2002).

Calculating the total fiscal consequences of immigration, while straightforward conceptually, is difficult in practice. To estimate them correctly, one needs to know many details about the income, spending, and employment behavior of the population of immigrants. As a result, there are few comprehensive national level analyses of the fiscal impact of immigration. In one of the few such studies, Smith and Edmonston (1997) estimate that in 1996 immigration imposed a short-run fiscal burden on the average US native household of \$200, or 0.2 percent of US GDP. In that year, a back of the envelope calculation suggests that the immigra-

tion surplus was about 0.1 percent of GDP (Borjas, 1999b), meaning that the short-run immigration in the mid-1990s reduced the annual income of US residents by about 0.1 percent of GDP. Given the uncertainties involved in making this calculation, this estimate is unlikely to be statistically indistinguishable from zero. While we cannot say with much conviction whether the aggregate fiscal impact of immigration on the US economy is positive or negative, it does appear that the total impact is small.[4]

Tax and transfer policies create a motivation for a government to restrict immigration, even where the level of immigration is set by a social planner. If immigrants are primarily individuals with low incomes relative to natives, increased labor inflows may exacerbate distortions created by social-insurance programs or means-tested entitlement programs, making a departure from free immigration the constrained social optimum (Wellisch and Walz, 1998).[5] Pay-as-you-go pension systems create a further incentive for politicians to manipulate the timing and level of immigration (Scholten and Thum, 1996; Razin and Sadka, 1999). Given its graying population and unfunded pension liabilities, one might expect Western Europe to be opening itself more aggressively to foreign labor inflows. However, concerns over possible increases in expenditure on social insurance programs may temper the region's enthusiasm for using immigration to solve its pension problems (Boeri and Brücker, 2005).

3.4. Human Capital Accumulation

International migration has the potential to affect the accumulation of human capital in both sending and receiving countries. In receiving countries, migration may increase the relative supply of high-skilled labor (for example, Canada), low-skilled labor (for example, Spain), or both high and low-skilled labor (for example, the United States). To the extent that wages fall for the skill group whose relative supply increases, native workers have an incentive to select out of that skill group. Alternatively, immigration may affect native schooling decisions by increasing competition for scarce educational resources.

Using data on the United States, Borjas (2004) estimates a negative correlation between the number of foreign students and the number of native-born students in university graduate programs, suggesting that foreign students may crowd out natives. Even with crowding out, the arrival of foreign students may still lead to an increase in the net supply of skilled labor in the United States. Stuen et al. (2006) find that university departments with more foreign graduate students have more publications in scientific journals, suggesting inflows of foreign students may spur knowledge creation.

Betts and Lofstrom (2000) and Hoxby (1998) present evidence that immigration reduces college attendance for US natives, particularly for minority students; and

[4] This estimate is based on short-run considerations. Going from the short run to the long run can change the results dramatically.

[5] In the long run, immigrants may affect voting outcomes directly through their participation in the political process (Razin et al., 2002; Ortega, 2004).

Betts (1999) finds that increases in the number of student-age immigrants in a US locality are associated with decreases in the likelihood that local minority students complete a high school degree.[6] For Israel, Paserman and Gould (2008) find that having more immigrants in one's grade school class is associated with a lower likelihood that a student will subsequently matriculate in or graduate from high school (even controlling for the overall immigrant presence in one's grade school). While the precise mechanisms behind these relationships are unclear, it does appear that the performance of native students deteriorates following a local influx of immigrant students.

In sending economies, the focus of research has been less on how migration affects competition for schooling and more on how opportunities for emigration affect the incentive to acquire skill. In poor countries, the income gain from emigration is often substantial, promising to raise real earnings by two to four times (Clemons et al. 2008). Moreover, the gain to migration is larger for individuals with higher education levels (Rosenzweig, 2007; Grogger and Hanson, 2008). An increase in the probability that individuals from a poor sending country will be allowed to emigrate to the United States or Europe may thus increase the incentive to obtain higher levels of education. The quantitative impact of this brain gain effect depends on the elasticity of the sending-country supply of educational services, and the perceived probability of migrating successfully. Where seats in colleges and universities are in limited supply, increases in the demand for higher education may have little effect on the local number of educated workers.[7] Related to this, where receiving countries allocate immigration visas in a non-random manner (say, by reserving entry slots for family members of existing US residents), many sending-country residents may have little hope of moving abroad, leaving their incentive to acquire skill unaffected by emigration opportunities.

Only a handful of empirical papers have examined the relationship between emigration and human-capital accumulation. For a cross-section of countries, Beine, et al. (2006) report a positive correlation between emigration to rich countries (measured by the fraction of the tertiary-educated population living in OECD countries in 1990) and the increase in the stock of human capital (measured as the 1990 to 2000 change in the fraction of adults who have tertiary education). This finding is consistent with emigration increasing the incentive to acquire education. However, it is not clear that one can make inferences about the causal impact of brain drain on educational attainment from the cross-section correlation between emigration and schooling. Individuals are likely to treat education and migration as joint decisions, making the two outcomes simultaneously determined.

[6] In related work, Betts and Farlie (2003) find that immigration induces natives to select out of public schools and into private schools.

[7] This is unless, of course, individuals are able to migrate abroad for their education. See Rosenzweig (2006).

Some evidence suggests that international migration may increase the flow of ideas between countries. In China, India, and Taiwan, the migration of skilled labor to Silicon Valley–where Indian and Chinese immigrants account for one-third of the engineering labor force–has been followed by increased trade with and investment from the United States, helping foster the creation of local high-technology industries (Saxenian, 2002). When individuals live and work in another country they are exposed to new political ideologies and alternative systems of government. Spilimbergo (2008) suggests there is an association between a country's democratic tendencies and the political systems of the countries under which its students did their university training. He finds a positive correlation between the democracy index in a sending country and the average democracy index in the countries in which a country's emigrant students have studied. Migration flows may also help to erode barriers to trade. Successive waves of emigration from China have created communities of ethnic Chinese throughout Southeast Asia, as well as in South Asia, and on the east coast of Africa. Rauch and Trindade (2002) find that bilateral trade is positively correlated with the interaction between the two countries' Chinese populations, consistent with ethnic business networks facilitating trade.

4. SUMMARY

There is ample evidence that international migration raises gross incomes for migrants, while it redistributes incomes within sending and receiving countries. Because the net impact of immigration on receiving countries appears to be small and the gain to migration appears to be so large, it is natural to presume international migration raises global income. Still, there remain many unknowns in evaluating migration's impact.

Economic theory suggests that international migration expands global output. Moving labor from low-productivity to high-productivity countries improves world allocative efficiency. No study suggests there are large negative consequences from global migration. In the United States, which is the largest receiving country for immigrants, the net impact of immigration, to a first approximation, appears to be awash (Borjas, 1999b). The global gains from migration are largely captured by migrants themselves, which they share with family members at home through remittances. Unless there are large unmeasured negative externalities from migration, or that migration exacerbates existing distortions in ways that have not yet been detected, it would be difficult to justify restrictive barriers to global labor flows. While the gross income gain to migration appears to be large, the net gain is unknown, given little evidence on migration costs.

The impact of immigration on receiving country labor markets is hotly disputed. The evidence would seem to favor the argument that wage effects from immigration do exist. Studies using national level data, while subject to concerns about their ability to control for all relevant labor market shocks, yield consistent qualitative results across sending and receiving countries (with Israel being

an exception). The results are also consistent with observed changes in native labor supply. Studies using local level data, whose results suggest immigration has little wage impact, are subject to concerns about the endogeneity of immigrant settlement patterns that have yet to be resolved fully. The literature has focused on the wage effects of immigration, while largely ignoring impacts on non-labor income. In theory, one would expect the gains in non-labor income (plus the gains to workers that complement foreign labor) to more than offset the losses to workers who compete with immigrant labor.

The net fiscal consequences of international migration are also poorly understood. In sending countries, there are only a handful of studies on emigration's fiscal impacts and these focus on the movement of high-skilled workers to high-income destinations. In receiving countries, the impact of immigration on the tax burden of natives is a central issue in political opposition to labor inflows. While there are many studies on how immigration affects government expenditure, there are few on how it affects government revenue, making it difficult to evaluate the net fiscal impact of labor inflows.

Another unknown is the effect of emigration on the incentive to acquire skill in sending countries. In the cross-section evidence, countries that have higher emigrant stocks abroad also have faster growth in the number of educated adults, but this association may not be informative about the consequences of brain drain. We still do not know how changes in the opportunity to emigrate affect human capital accumulation. Many individuals migrate abroad to complete their education, with many ultimately returning to their home countries. Circular migration is important for the accumulation of skill in developing countries, though migrants from the poorest countries are those most tempted to emigrate permanently. Even where migration is permanent, having emigrants abroad may help a country lower its barriers to trade, investment, and technology flows.

Gordon Hanson is the director of the Center on Pacific Economies and is a professor of economics at UC San Diego.

BIBLIOGRAPHY

Acosta, Pablo, Fajnzylber, Pablo and Humberto Lopez 2007. How Important are Remittances in America? In Pablo Fajnzylber and Humberto Lopez, (Eds.), *Close to Home*, Washington, DC: The World Bank, pp. 21-50.

Addison, Thomas, and Christopher Worswick 2002. The Impact of Immigration on the Earnings of Natives: Evidence from Australian Micro Data. *The Economic Record*, 78(1): 68–78

Auerbach, Alan J, and Philip Oreopoulos 1999. Analyzing the Fiscal Impact of US Immigration. *American Economic Review*. Papers and Proceedings, 89(2), 176–80

Aydemir, Abdurrahman and George J Borjas 2007. A Comparative Analysis of the Labor Market Impact of International Migration: Canada, Mexico, and the United States. *Journal of the European Economic Association*, forthcoming

Beine, Micahael, Docquier, Frédéric and Hillel Rapoport 2006. *Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers*. IRES Working Paper no. 2006–23

Beine, Michel, Docquier, Frédéric and Hillel Rapoport 2007. Measuring International Skilled Migration: New Estimates Accounting for Age of Entry. *World Bank Economic Review*, 21(June 2): 249–54

Betts, Julian 1999. *Have Inflows of Immigrants Diminished Natives' Educational Attainment?* Public Policy Institute of California

Betts, Julian, and Robert Farlie 2003. Does Immigration Induce Native Flight' from Public into Private Schools. *Journal of Public Economics*. 87(5–6): 987–1102

Betts, Julian, and Magnus Lofstrom 2000. The Educational Attainment of Immigrants: Trends and Implications. In George J. Borjas, (Ed.), *Issues in the Economics of Immigration*, Chicago: University of Chicago and NBER

Borjas, George J 1999a. The Economic Analysis of Immigration. In Orley C Ashenfelter and David Card, (Eds.), *Handbook of Labor Economics, Amsterdam*: North-Holland, 1697–1760.

Borjas, George J 1999b. *Heaven's Door: Immigration Policy and the American Economy*. Princeton, N.J.: Princeton University Press

Borjas, George J 2003. The Labor Demand Curve Is Downward Sloping: Reexamining the impact of immigration on the labor market. *Quarterly Journal of Economics* 118 (November): 1335–1374

Borjas, George J 2003. *Welfare Reform, Labor Supply, and Health Insurance in the Immigrant Population*. NBER Working Paper no. 9781

Borjas, George J 2004. *Do Foreign Students Crowd out Native Students from Graduate Programs?* NBER Working Paper no. 10349

Borjas, George J 2006 Native Internal Migration and the Labor Market Impact of Immigration. *Journal of Human Resources* 41 (Spring): 221–58

Borjas, George J, Freeman, Richard B and Lawrence F Katz 1997. How Much Do Immigration and Trade Affect Labor Market Outcomes? *Brookings Papers on Economic Activity* 1: 1–90

Borjas, George J, Grogger, Jeffrey and Gordon H Hanson 2008. *Imperfect Substitution between Immigrants and Natives: A Reappraisal*. NBER Working Paper no. 13887

Borjas, George J., and Lynette Hilton. 1996. Immigration and the Welfare State: Immigrant Participation in Means-tested Entitlement Programs. *Quarterly Journal of Economics*, 111(2): 575–604

Capps, Randall, Fix, Michael Henderson, Everett and Jane Reardon-Anderson 2005. *A Profile of Low-income Working Immigrant Families*. Urban Institute Working Paper B–67

Card, David 2001. Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration, *Journal of Labor Economics* (January): 22–64

Card, David 2005. Is the New Immigration Really so Bad? *Economic Journal*, 115(507): 300–23

Carrasco, Raquel, Jimeno, Juan F and A Carolina Ortega 2008. The Effects of Immigration on the Labor Market Performance of Native Born Workers: Some Evidence for Spain. *Journal of Population Economics*, 21: 627–48

Carrington, William, and Pedro de Lima 1996. The Impact of the 1970s Repatriates from Africa on the Portuguese Labor Market. *Industrial and Labor Relations Review*, 49(2): 330–47

Clemons, Michael, Montenegro, Claudio and Lant Pritchett 2008. *The Great Discrimination: Borders as a Labor Market Barrier.* Center on Global Development

Collado Inigo Iturbe-Ormaetxe, M Dolores, and Guadalupe Valera 2004. Quantifying the Impact of Immigration on the Spanish Welfare State. *International Tax and Public Finance*, 11: 335–53

Cortes, Patricia. 2008. The Effect of Low-skilled Immigration on US Prices: Evidence from CPI Data. *Journal of Political Economy*, vol. 116, no. 3], pp. 381-422.

De New, John P, and Klaus F Zimmerman 1994. Native Wage Impacts of Foreign Labor: A Random Effects Panel Analysis. *Journal of Population Economics*, 7: 177–92

Desai, Mihir, Devesh Kapur, and John McHale. 2008. The Fiscal Impact of High Skilled Emigration: Flows of Indians to the US. *Journal of Development Economics*, forthcoming

Docquier, Frederic and Hillel Rapoport. 2008. Skilled Migration: The Perspective of Developing Countries. In Jagdish Bhagwati and Gordon Hanson, eds., *Skilled Immigration Today: Problems, Prospects, and Policies*. Oxford: Oxford University Press, forthcoming

Dustmann, Christian, Fabbri, Francesca and Ian Preston 2005. The Impact of Immigration on the British Labour Market. *The Economic Journal*, 115: F324–F341.

Friedberg, Rachel 2001. The Impact of Mass Migration on the Israeli Labor Market. *Quarterly Journal of Economics*, 1373–408

Galosto, Andrea, Venturini, Alessandra and Claudia Villosio 1999. Do Immigrants Compete with Natives? *Labour*, 13(3): 603–22

Grogger, Jeffrey, and Gordon H. Hanson. 2008. Income Maximization and the Selection and Sorting of International Migrants, NBER Working Paper No. 13821, February 2008.

Hanson, Gordon H 2007. International Migration and the Developing World. Mimeo, UCSD

Hatton, Timothy J., and Massiliano Tani. 2005. Immigration and Inter-Regional Mobility in the UK, 1982–2000. *The Economic Journal*, 115: F342–F358

Hatton, Timothy J, and Jeffrey G Williamson 2004. *Refugees, Asylum Seekers and Policy in Europe*. NBER Working Paper no. 10680

Hofer, Helmut, and Peter Huber 2003. Wage and Mobility Effects of Trade and Migration on the Austrian Labour Market. *Empirica*, 30: 107–25

Hoxby, Caroline M 1998 Do Immigrants Crowd Disadvantaged American Natives our of Higher Education? In Daniel S. Hamermesh and Frank D. Bean, (Eds.), *Help or Hindrance? The Economic Implications of Immigration for African–Americans*. New York: Russell Sage Foundation, 282–321

Hunt, Jennifer 1992. The Impact of the 1962 Repatriates from Algeria on the French Labor Market. *Industrial and Labor Relations Review*, 45(3): 556–72

Jaeger, David A 1997. Reconciling the Old and New Census Education Questions: Recommendations for Researchers. *Journal of Business and Economic Statistics*, 15(3): 300-309

Jandl, Michael 2003. Estimates on the Number of Illegal and Smuggled Immigrants in Europe. International Centre for Migration Policy Development

Mishra, Prachi 2007. Emigration and Wages in Source Countries: Evidence from Mexico. *Journal of Development Economics*, 82(1): 180–99

Ottaviano, Gianmarco, and Giovanni Peri 2007. Rethinking the Effects of Immigration on Wages. Mimeo, UC Davis

Passel, Jeffrey S. and D'Vera Cohn. 2009. "A Portrait of Unauthorized Immigrants in the United States." Pew Hispanic Center, Washington, DC.

Pischke, Jorn-Steffen, and Johannes Velling 1997. Employment Effects of Immigration to Germany: An analysis based on local labor markets. *Quarterly Journal of Economics*, 594–604

Rauch, James E and Vitor Trindade 2002. Ethnic Chinese Networks in International Trade. *Review of Economics and Statistics*, 84(1), February, 116–30

Razin, Asaf, and Efraim Sadka 1999. Migration and Pensions with International Capital Mobility. *Journal of Public Economics*, 74: 141–50

Razin, Asaf, Sadka, and Phillip Swagel 2002. Tax Burden and Migration: Political Economy Theory and Evidence. *Journal of Public Economics* 85: 167–90

Rosenzweig, Mark 2006. Global Wage Differences and International Student Flows. *Brookings Trade Forum.*

Rosenzweig, Mark 2007. Education and Migration: A Global Perspective. Mimeo, Yale

Saxenian, AnnaLee 2002. Local and Global Networks of Immigrant Professionals in Silicon Valley. Monograph, Public Policy Institute of California

Scholten, Ulrich, and Marcel Thum 1996. Public Pensions and Immigration Policy in a Democracy. *Public Choice*, 66: 347–61

Sinn, Hans-Werner 2002. EU Enlargement and the Future of the Welfare State. *Scottish Journal of Political Economy*, 49(1): 1–4–115.

Smith, James P and Barry Edmonston 1997. (Eds), *The New Americans: Economic, Demographic, and Fiscal Effects of Immigration.* Washington, DC: National Academy Pres, 1997

Spilimbergo, Antonio 2009. Foreign Students and Democracy. American *Economic Review*, 99(1): 528–43.

Stuen, Eric T, Ahmed Mobarak, and Keith Maskus 2006. Foreign Graduate Students and Knowledge Creation at US Universities. Mimeo, University of Colorado

United Nations 2006. *World Migrant Stock: The 2005 Revision Population Database* http://esa.un.org/migration/

Woodruff, Christopher and Rene Zenteno 2007. Migration Networks and Microenterprises in Mexico. *Journal of Development Economic*s, 82(2): 509-528

Yang, Dean 2008. International Migration, Remittances, and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks. *Economic Journal*, 118: 591-630

Zorlu, Aslan, and Joop Hartog. 2005. The Effect of Immigration on Wages in Three European Countries. *Journal of Population Economics*, 18: 113–51

13

Exporter Adjustment to the End of Trade Preferences: Evidence from the abrupt end of textile and apparel quotas

JAMES HARRIGAN

1. INTRODUCTION

Trade policies are often discriminatory: tariffs and quotas apply unevenly to a country's trading partners. Economics teaches that such discrimination can be even more distortionary than a Most Favored Nation (MFN) trade policy, since discrimination may lead an importer to import inefficiently from high-cost suppliers. This inefficiency is of course costly for the importing country, but it also can distort the global pattern of specialization: countries with preferential access invest too much in providing the protected products, while low-cost suppliers are prevented from fully exploiting their competitive advantage.

These insights from economic theory are the basis of much policy advice, and in particular the prescription that if countries need to protect their domestic markets then they should do so using nondiscriminatory policies. This note provides empirical evidence to support the theory and policy advice from one of the largest and most abrupt trade policy changes in recent years: the end of the global regime of textile and apparel quotas on January 1 2005. This regime, best known as the 'Multi Fiber Arrangement' (MFA) but renamed the 'Agreement on Textiles and Clothing' (ATC) at the conclusion of the Uruguay Round in 1995, was long thought to distort global textile and apparel trade severely, and its demise has confirmed this view.[1]

This paper, based on my joint work with Geoffrey Barrows (Harrigan and Barrows, 2009), focuses on how exporters to the US market adjusted when the MFA ended[2]. As discussed in greater detail in Harrigan and Barrows (2009), the bulk of MFA liberalization was put off until the last minute, with the result that US imports of textiles and apparel changed dramatically in 2005 in ways that can be credibly attributed to the end of the MFA. I will show that

[1] For simplicity, I will continue to use the older term MFA in the remainder of this note.

[2] For background on the rise and especially fall of the MFA, see Harrigan and Barrows (2009); Evans and Harrigan (2005a); and Brambilla et al. (2007).

- China had been severely constrained by the MFA. Chinese export prices plunged, and volumes soared, when the MFA ended.
- Other low-wage MFA-constrained exporters also saw big increases in their exports to the United States in 2005.
- The 'East Asian Miracle' exporters saw big drops in export values, despite having large shares of filled quotas in 2004.
- Mexico was a big loser from the end of the MFA, as it lost its previously privileged access to the US market. Mexican export volumes fell substantially in the face of greater competition from China and other Asian exporters. However, by specializing in goods where Mexico's proximity to the United States gave it a competitive advantage, Mexico's losses from the end of the MFA were smaller than what some observers had feared.

All figures and calculations in this paper refer to Harrigan and Barrows (2009) unless otherwise noted.

2. THE END OF THE MFA IN THE UNITED STATES

Table 13.1 summarizes the US import market for apparel and textiles in the last year of the MFA, 2004, and in 2005. In 2004, 17 per cent of US imports came in under a binding quota.[3] China was the largest exporter, with 21 per cent of the market,

Table 13.1: *US Apparel and Textile Imports: Top 20 exporters*

	Quota coverage 2004	Market share	
		2004	2005
China	18.0	20.7	27.8
Mexico	0.0	9.7	8.4
India	36.0	4.5	5.3
Canada	0.0	4.1	3.6
Hong Kong	50.4	3.6	2.9
Korea	27.7	3.4	2.4
Honduras	0.0	3.1	2.8
Vietnam	29.1	3.1	3.0
Indonesia	64.2	2.9	3.2
Pakistan	42.3	2.9	3.1
Italy	0.0	2.8	2.5
Taiwan	18.5	2.6	2.0
Thailand	18.0	2.6	2.3
Bangladesh	34.9	2.3	2.6
Philippines	31.6	2.2	2.0
Turkey	2.3	2.0	1.7
El Salvador	0.0	2.0	1.7
Sri Lanka	28.0	1.8	1.8
Guatemala	0.3	1.7	1.5
Cambodia	44.4	1.6	1.8
Other		20.5	17.5
Total	16.7	100.0	100.0

[3] A quota is defined as binding if imports are at least 90 per cent of the quota level.

but at first glance was not highly constrained by the MFA: only 18 per cent of Chinese imports came into the United States under a binding quota, no different from the overall average. Mexico was the second largest exporter with 10 per cent of the market, and all of its exports arrived unimpeded by quotas. The big South Asian exporters (India, Pakistan, Bangladesh, and Sri Lanka) all had greater shares of quota-covered exports than did China, while the exporters with the highest shares of quota coverage were Hong Kong (50 per cent) and Indonesia (64 per cent).

2.1 Overall effects

The low overall quota coverage in the US market in 2004 might at first glance suggest that the MFA was not severely constraining US imports. Similarly, China's fairly modest share of quota-covered exports might be taken as a hint that it was not unduly restricted. The final column of Table 13.2 shows that such inferences are wrong. When the MFA expired China's export value exploded by 45 per cent. India (+28), Indonesia (+18), Pakistan (+14), Bangladesh (+18), and Cambodia (+20) also saw double-digit increases in exports. Mexico and Canada, which lost their preferential access to the US market, saw their exports fall by more than 5 per cent.

Table 13.2: *Quantity, Price, Quality and Value Change 2004–2005: US Apparel and textile imports, top 20 exporters*

	Quantity			Price			Quality			Value
	Total	Not bound	Bound 2004	Total	Not bound	Bound 2004	Total	Not bound	Bound 2004	Total
China	155.9	51.8	449.6	−10.2	−1.0	−37.8	−3.0	−0.3	−11.2	44.7
Mexico	−7.0	−7.0		4.0	4.0		0.6	0.6		−6.5
India	124.5	166.3	54.1	−1.9	2.3	−9.2	−1.2	−0.4	−2.7	27.6
Canada	−2.0	−2.0		1.7	1.7		−0.9	−0.9		−5.3
Hong Kong	21.8	27.5	16.9	−2.2	−1.6	−2.8	0.6	5.5	−4.0	−13.7
Korea	−11.3	−14.0	−3.7	3.9	7.1	−4.9	−2.3	−1.7	−3.8	−21.9
Honduras	1.8	1.8		−1.8	−1.8		−2.3	−2.3		−1.9
Vietnam	11.0	18.5	−9.4	3.0	2.2	5.0	0.4	0.6	0.0	5.9
Indonesia	41.7	27.8	49.3	−5.0	1.7	−8.4	−1.7	0.4	−2.7	18.0
Pakistan	54.7	−0.1	113.3	−8.6	−0.4	−17.6	0.4	0.5	0.2	14.4
Italy	−11.5	−11.5		12.4	12.4		0.3	0.3		−4.0
Taiwan	−11.7	−11.1	−14.6	3.5	3.4	4.1	0.4	0.7	−0.9	−19.3
Thailand	6.9	−1.4	39.6	0.1	1.8	−6.5	−3.1	−3.3	−2.3	−3.1
Bangladesh	40.0	25.0	64.6	−6.6	−4.9	−9.2	−1.6	−2.1	−0.7	18.8
Philippines	19.2	8.0	41.6	−3.4	−0.2	−9.7	−2.7	−2.2	−3.7	−1.1
Turkey	−7.6	−9.6	53.7	8.4	8.5	6.1	4.6	5.3	−16.6	−8.3
El Salvador	0.8	0.8		−2.7	−2.7		−4.1	−4.1		−6.3
Sri Lanka	19.2	9.6	41.5	−3.1	−0.6	−8.4	−2.4	−2.6	−2.1	5.5
Guatemala	2.3	0.8	202.8	−3.9	−4.0	0.2	−0.4	−0.3	−7.5	−6.1
Cambodia	41.2	8.9	75.5	−7.0	1.7	−16.1	−2.6	0.0	−5.3	19.9

Notes: All entries are percent changes between 2004 and 2005. Columns headed 'bound 2004' aggregate products subject to a binding quota in 2004, with all other products aggregated in the 'not bound' columns.

2.2 Price and quantity effects

The overall changes in export volumes obscure much of what happened in 2005. Using detailed data and index number theory, Harrigan and Barrows (2009) also compute price, quantity, and quality changes.[4] Figures 13.1a and 13.1b illustrate the first nine columns of Table 13.2: China's overall quantity of exports grew an eye-popping 150 per cent, an increase driven largely by an incredible 450 per cent increase in exports of previously constrained goods. To achieve this level of sales required big price drops: the price of Chinese exports fell by 10 per cent, almost entirely due to a 40 per cent drop in the prices of previously constrained goods.

Other big exporters saw similar patterns. Pakistan's exports of previously constrained goods grew by over 100 per cent, accompanied by price drops of 18 per cent. The experience of India, Bangladesh, Indonesia, and the Philippines was only slightly less dramatic, with quantities of previously constrained goods increasing by around 50 per cent and prices falling by about 9 per cent.

Interestingly, some of the exporters that had the largest shares of quota-constrained exports saw more modest changes in 2005. Hong Kong, which had half of its exports in constrained categories in 2005, saw sales of these goods rise just 17 per cent and prices fell 3 per cent. Korea's exports of previously constrained goods actually *fell* 4 per cent, despite prices that fell by 5 per cent. A similar thing happened to Taiwanese exports of previously constrained goods: they fell 15 per cent, as prices rose 4 per cent.

2.3 Efficient quality downgrading

A more subtle effect of quotas is to raise quality inefficiently. The reason is that with a quantity constraint, exporters will choose to set prices so that the quota rent per unit of quantity is the same. This has the effect of lowering the relative price of high-cost/high-quality goods, thus tilting sales toward these higher-end varieties.[5] An immediate corollary of this theory is that abandoning quotas should see efficiency-enhancing quality downgrading: the mix of exports should shift toward less expensive items. Table 13.2 shows some evidence of this phenomenon in the data. In particular, the quality of previously constrained goods from China fell by more than 10 per cent (other quality effects in the data are smaller, and Harrigan and Barrows (2009) show that except for China the hypothesis that quality downgrading was random can not be rejected).

3. COMPARATIVE ADVANTAGE AND THE END OF THE MFA

The changes detailed above can be summarized compactly. When the MFA ended:

- China and other low-wage exporters (South Asia, Indonesia, and the Philippines) had huge gains in sales, along with sharp drops in prices.

[4] For details on the calculations, see the Box 1.

[5] See Falvey (1979) and Rodriguez (1979) for the theory of quality upgrading under quotas, and Boorsten and Feenstra (1991) and Feenstra (1988) for applications.

COMPUTING PRICE, QUANTITY, AND QUALITY INDEXES

The trade data used by Harrigan and Barrows (2009) come from the Census Bureau, and are analyzed at the 10-digit Harmonized System (HS) level, the most disaggregated classification available. Each import observation includes information on date, source country, value in dollars, and physical quantity (such as number of shirts). The analytical challenge is to aggregate these data into a form that is easy to understand.

Harrigan and Barrows (2009) use modern index number theory to do the aggregation. The basic building block is the Feenstra (1994) index, which allows for changes in the set of goods exported over time. A simplified version of the Feenstra price index, which ignores the new goods correction, is

$$\ln F\left(\mathbf{p}_t, \mathbf{p}_{t-1}, \mathbf{x}_t, \mathbf{x}_{t-1}\right) = \sum_{i \in I} w_{it} \ln \frac{p_{it}}{p_{it-1}}$$

where p_{it} is the price of product i in period t, and

$$w_{it} = \frac{s_{it} + s_{it-1}}{2}$$

is the average share of good i in the aggregate in the two periods. In words, the Feenstra index is an expenditure-share weighted average of price changes of individual goods.

For any aggregate Feenstra price index, the corresponding quantity index is computed from the identity that value = price × quantity.

The computation of the quality index exploits the difference between a price index and a so-called unit value index,

$$UV\left(\mathbf{p}_t, \mathbf{x}_t\right) = \sum_{i \in I} \omega_{it} p_{it}, \qquad \omega_{it} = \frac{x_{it}}{\sum_{j \in I} x_{jt}}$$

where the weights ω_i are quantity weights: the number of units (count, kilos, square meters, and so on) of good i as a share of the total number of units in the aggregate. It is clear from the definition that UV can change over time, even if no individual prices change, just by changing the number of units of each good in the aggregate. Boorstein and Feenstra (1991) show that an indicator of quality change over time is given by the difference between unit value and price change. Expressed in logs, change in quality Q is measured as

$$\ln \frac{Q_t}{Q_{t-1}} = \ln \frac{UV_t}{UV_{t-1}} - \ln F\left(\mathbf{p}_t, \mathbf{p}_{t-1}, \mathbf{x}_t, \mathbf{x}_{t-1}\right)$$

The interpretation is that if the unit value index increases more than the price index, then the quality index rises, reflecting the fact that consumption has shifted towards more expensive goods within the category.

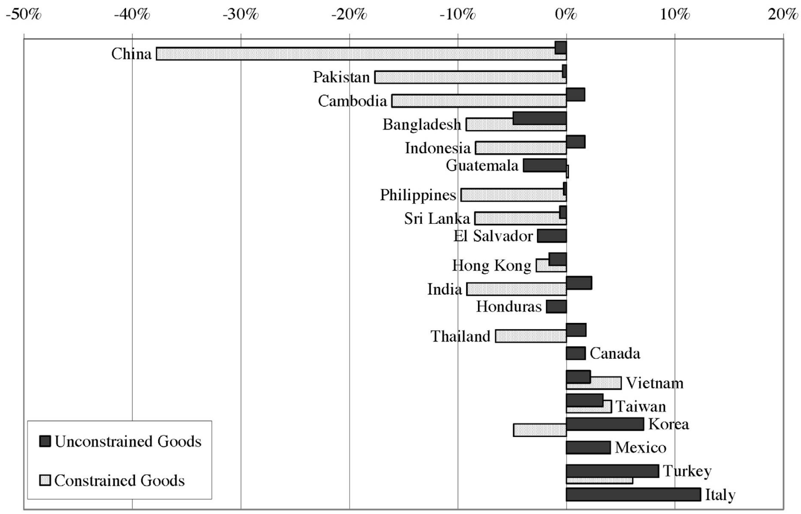

Figure 13.1a: *Price changes 2004–2005: Top 20 exporters, ordered by total price change*

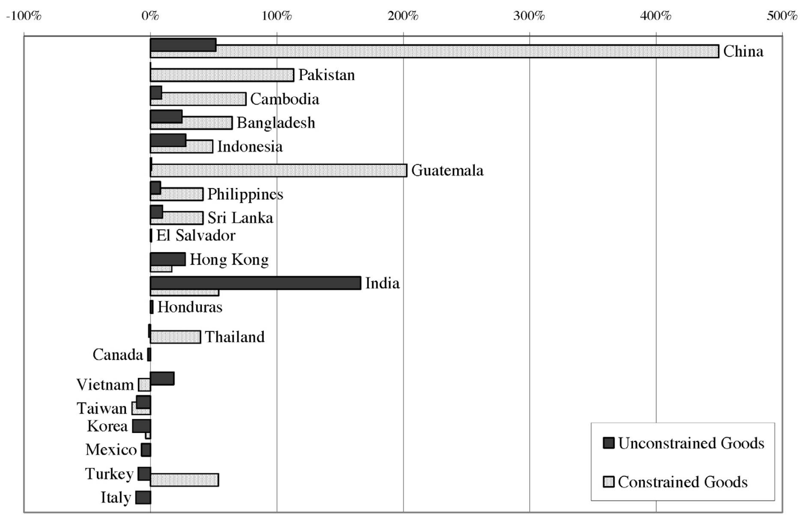

Figure 13.1b: *Quantity changes 2004-2005: Top 20 exporters, ordered by total price change*

- The NAFTA exporters saw sales fall modestly, and prices rose somewhat.
- The relatively high-wage 'East Asian Miracle' exporters (Hong Kong, Korea, Taiwan) saw the value of their sales fall precipitously, despite having large shares of their exports in constrained categories in 2004.

This pattern of results is very consistent with a view that competitive advantage in the apparel and textile industry was being driven primarily by low wages. Once China and the other low-wage exporters were no longer constrained, they elbowed aside exporters from relatively high-wage East Asian exporters that, for historical reasons, had large quota allocations under the MFA.

An interesting feature of the results is that Mexico was not hurt as much as some had feared: despite losing preferential access, and in the face of a flood of low-cost imports from the low-wage countries, Mexico's market share fell only slightly more than 1 percentage point. Canada's market share fell by only half of a percentage point. What Canada and Mexico have in common, of course, is a competitive advantage that can not be eroded by shifts in trade policy: proximity to the United States. In Evans and Harrigan (2005b), the authors show that Mexico used its preferential access in apparel and textiles under NAFTA to expand sales disproportionately in goods where timely delivery was valuable to US retailers. My conjecture is that this durable market niche insulated Mexico from the competitive pressures from low-cost, but faraway, suppliers in Asia.

3.1 Winners and losers from the end of the MFA

Changes in economic policy, however salutary in the aggregate, always create winners and losers. A full accounting of the welfare effects of the end of the MFA is beyond the scope of this paper, but Harrigan and Barrows (2009) show that the MFA's demise saved US consumers about $7 billion, mainly due to lower prices on Chinese imports.

The gains to US consumers came in large part at the expense of quota license holders in the exporting countries. This raises the possibility that the end of the MFA may have actually made exporters worse off. China is a particularly provocative case: while China increased its market share from 21 per cent to 28 per cent, it accomplished this by lowering prices by 10 per cent overall, and by 38 per cent in previously constrained categories. The partial effect of the terms of trade deterioration is negative, although the more effective exploitation of China's comparative advantage in labor-intensive manufactured goods might outweigh the terms of trade deterioration. Similar considerations apply to the other big low-wage exporters (South Asia, Indonesia, and the Philippines).

The verdict for the 'East Asian Miracle' exporters (Hong Kong, Korea, Taiwan) is unambiguous: their welfare fell as a result of the end of the MFA. This is shown by the drop in overall export value, driven by falling prices, with only Hong Kong seeing any increase in real sales of previously constrained goods. These countries' textile and apparel exports had apparently been driven by quota rent-seeking, and when these rents evaporated the rationale for large-scale exports of labor-intensive apparel and textiles evaporated as well.

4. CONCLUSION

The end of the MFA in 2004 led to big changes in the pattern of US imports of textiles and apparel. The new pattern seems much more reflective of comparative advantage: the big low-wage countries, especially China, expanded their sales at the expense of higher-wage exporters that had preferential access to the US market under the MFA. The NAFTA exporters, whose competitive advantage in proximity to the US was not eroded by the end of the MFA, suffered less than higher-wage East Asian exporters who owed their share of the US market to their historically commodious quota allocations.

James Harrigan is a Professor of Economics at the University of Virginia

BIBLIOGRAPHY

Brambilla, Irene, Khandelwal, Amit and Peter Schott, 2007, China's Experience Under the Multi Fiber Arrangement (MFA) and the Agreement on Textiles and Clothing (ATC), NBER Working Paper no. 13346 (August)

Boorstein, Randi, and Robert C Feenstra, 1991, Quality Upgrading and Its Welfare Cost in US Steel Imports, 1969–74, in Elhanan Helpman and Assaf Razin, Editors, *International Trade and Trade Policy*, Cambridge, MA: MIT Press

Evans, Carolyn L, and James Harrigan, 2005a, Tight Clothing: How the MFA Affects Asian Apparel Exports, *East Asian Seminar on Economics 14: International Trade*, Chapter 11, 2005, Takatoshi Ito and Andrew Rose, Eds., University of Chicago Press

—2005b, Distance, Time, and Specialization: Lean Retailing in General Equilibrium, *American Economic Review* 95 no. 1 (March): 292–313

Falvey, Rodney E, 1979, The composition of trade within import-restricted categories, *Journal of Political Economy 87* (5): 1105–14

Feenstra, Robert C, 1988, Quality Change under Trade Restraints in Japanese Autos, *Quarterly Journal of Economics* 103:131–46

Feenstra, Robert C, 1994, New Product Varieties and the Measurement of International Prices, *American Economic Review* 84 (1): 157–77 (March)

Harrigan, James, and Geoffrey Barrows, 2009, Testing the theory of trade policy: Evidence from the abrupt end of the multifiber arrangement, *Review of Economics and Statistics* 91 (2): 282–94 (May)

Rodriguez, Carlos Alfredo, 1979, The quality of imports and the differential welfare effects of tariffs, quotas, and quality controls as protective devices, *Canadian Journal of Economics 12* (3): 439–49

14

New Kids on the Block: Adjustment of Indigenous Producers to FDI Inflows

BEATA S JAVORCIK

1. INTRODUCTION

The past several decades have witnessed a spectacular increase in international trade and foreign direct investment (FDI) in both developed and developing countries. While a large literature has investigated the adjustment process taking place in the aftermath of trade liberalization, there is much less systematic evidence on how countries adjust after receiving large inflows of FDI.

The purpose of this chapter is to summarize the existing empirical evidence on how indigenous producers are affected by the presence of foreign investors and how they respond to new opportunities and challenges created by FDI inflows. As the existing knowledge on this subject is still limited, the note points to potentially fruitful areas for future research.

The note starts with a brief review of the arguments for why indigenous producers should be affected by FDI. There is a consensus in the literature that multinational corporations (MNCs) are characterized by large endowments of intangible assets which translate into superior performance. This implies that MNCs present formidable competition for indigenous producers in any host country, while at the same time being a potential source of knowledge spillovers (Section 2). To substantiate this view, the note presents empirical evidence from Indonesia indicating that new foreign entrants taking the form of greenfield projects exhibit higher productivity than domestic entrants or mature domestic producers. Further, the note reviews evidence on foreign acquisitions of Indonesian plants which suggests that such acquisitions lead to large and rapid productivity improvements taking place through deep restructuring of the acquisition targets (Section 3).

Next, the note discusses findings of enterprise surveys and econometric firm-level studies which suggest that FDI inflows increase competitive pressures in their industry of operation (Section 4) and lead to knowledge spillovers within and across industries (Section 5). In Section 6, the implications of inflows of FDI into service sectors are reviewed. In particular, it is argued that the presence of foreign service providers may increase the quality, range, and availability of serv-

ices, thus benefiting downstream users in manufacturing industries and boosting their performance. The focus then shifts to global retail chains and their impact on the level of competition in the supplying sectors. The last section concludes with suggestions for future research.

2. WHY SHOULD WE EXPECT INDIGENOUS PRODUCERS TO BE AFFECTED BY FDI?

A basic tenet of the theory of the multinational firm is that such firms rely heavily on intangible assets to compete in distant and unfamiliar markets successfully. These assets, named 'ownership advantages' by Dunning (1983), may take the form of new technologies and well-established brand names, know-how, management techniques, etc. The theory further postulates that intangible assets, developed in headquarters, can be easily transferred to foreign subsidiaries and their productivity is independent of the number of facilities in which they are employed. The multinational thus offers the world increased technical efficiency by eliminating the duplication of the joint input that would occur with independent national firms (Markusen 2002). Similarly, recent theoretical work focusing on heterogeneous firms predicts that only the most productive firms can afford the extra cost of setting up production facilities in a foreign country, and thus multinationals come from the upper part of the productivity distribution of firms in their country of origin (Helpman et al. 2004).

The data confirm that multinationals are responsible for most of the world's research and development (R&D) activities. In 2003, 700 firms, 98 percent of which are multinational corporations, accounted for 46 percent of the world's total R&D expenditure and 69 percent of the world's business R&D. Considering that there are about 70,000 multinational corporations in the world, this is a conservative estimate. In 2003, the gross domestic expenditure on R&D by the eight new members of the European Union at 3.84 billion dollars[1] was equal to about half of the R&D expenditure of the Ford Motor Company (6.84 billion), Pfizer (6.5 billion), DaimlerChrysler (6.4 billion), and Siemens (6.3 billion) during the same year. It was comparable to the R&D budget of Intel (3.98 billion), Sony (3.77 billion), and Honda and Ericsson (3.72 billion each) (see UNCTAD 2005). Aggregate data reveal a similar pattern–more than 80 percent of global royalty payments for international transfers of technology in 1995 were made from subsidiaries to their parent firms (UNCTAD 1997).

Even though most of the R&D activity undertaken by multinational corporations remains in their home country, recent years have witnessed a growing internationalization of R&D efforts. According to the data collected by UNCTAD (2005) in their 2004–5 survey of the world's largest R&D investors, the average respondent spent 28 percent of its R&D budget abroad in 2003, including in-

[1] The group includes the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia, and Slovenia. As the 2003 figures were not available for Lithuania and Slovenia, the 2002 data were used for these countries.

house expenditure by foreign affiliates and extramural spending on R&D contracted to other countries. Consider that 62.5 percent of business R&D conducted in Hungary was undertaken by foreign affiliates; the corresponding figure for the Czech Republic was 46.6 percent, while in Poland and Slovakia foreign affiliates accounted for 19 percent of business R&D.

It has also been demonstrated that multinational companies tend to invest more in labor training than local firms in host countries do.[2] A significant portion of outlays on employee training is associated with technology transfer from the parent company to its foreign subsidiaries. It is not uncommon for staff from headquarters to conduct training in subsidiaries, or for subsidiary staff to be trained at headquarters (Ramachandaram 1993).

The combination of large endowments of intangible assets and high investment in staff training has three implications. First, foreign affiliates should exhibit superior performance relative to indigenous producers and thus should directly contribute to increasing the productivity level of the host country. Second, inflows of FDI should lead to increased competitive pressures in their sector of operation in the host country. And third, the presence of FDI is likely to benefit indigenous producers through knowledge spillovers.

3. ARE FOREIGN AFFILIATES DIFFERENT FROM INDIGENOUS PRODUCERS?

This section aims to substantiate the claims about superior performance of foreign affiliates by drawing on the empirical evidence from Indonesia. First, we compare the characteristics of new FDI greenfield projects to those of new Indonesian entrants and mature Indonesian producers. Then, we present evidence on how foreign ownership affects the performance of acquired Indonesian plants.

Our exercise is based on the plant-level information from the Indonesian Census of Manufacturing. The census surveys all registered manufacturing plants with more than 20 employees. The sample covers the period 1983–2001 and contains more than 308,439 plant observations, of which about 5.5 percent belong to foreign-owned plants. The average spell a plant remains in the sample is about 11 years.

In the first part of the exercise, which draws on Arnold and Javorcik (2009a), the following empirical specification is estimated:

$$Y_{it} = \alpha + \beta_1 \text{ Foreign greenfield entrant}_{it} + \beta_2 \text{ Domestic entrant}_{it} + \beta_3 \text{ Other foreign affiliate}_{it} + \gamma_j + \delta_r + \varphi_t + \varepsilon_{it} \quad (1)$$

where Y_{it} is one of a series of outcome variables pertaining to plant i observed at time t. ‘Foreign greenfield entrant’ is an indicator variable taking the value of one

[2] For instance, according to the survey described by Kertesi and Köllö (2001), foreign-owned firms in Hungary spent 14.2 percent of their investment on training, as compared to 2.4 percent in the case of domestic firms. Similarly, a World Bank study focusing on Malaysia also showed that foreign-owned firms provide more training to their workers than domestic enterprises do (World Bank 1997).

for plants which are no more than three years old and which at the moment of establishment had a foreign ownership share of at least 20 percent of total equity. Such plants are identified on the basis of foreign ownership and age.[3] The variable is equal to zero in all other cases. 'Domestic entrants' are defined as domestic plants in the first three years of their operations. The category 'other foreign affiliate' encompasses all establishments with a foreign ownership share of at least 20 percent of total equity, which are not foreign greenfield entrants. Thus, the comparison group in this exercise is the mature indigenous producers. To capture differences between industries, regions, and time periods the specification also includes 4-digit industry fixed effects (γ_j), 27 province fixed effects (δ_r), and year fixed effects (φ_t).

If foreign ownership is indeed associated with superior performance, this pattern should already be observable among new greenfield entrants who should exhibit different characteristics from new domestic establishments. As evident from Table 14.1, this is indeed the case. New greenfield projects exhibit higher total factor productivity (TFP) and labor productivity levels as well as a higher TFP growth than both new and mature Indonesian producers. They are also larger in terms of output and employment. They pay higher wages and employ a larger proportion of skilled workers. They are more capital intensive and they invest more in general, as well as in machinery. They export a larger share of their output and are more reliant on imported inputs. In all cases, the difference between domestic and foreign entrants is statistically significant.

The performance of new foreign entrants is, however, dwarfed by the TFP and labor productivity of mature foreign affiliates. The difference between the two groups is statistically significant. In contrast, when compared to mature foreign affiliates, new greenfield projects appear to have higher investment outlays in general, as well as higher investment outlays on machinery. They have a higher capital-labor ratio and experience faster productivity growth. They also appear to be more connected to international production networks, as evidenced by a higher reliance on export markets and imported inputs.

The magnitudes of the estimated coefficients are economically meaningful. For instance, while new domestic entrants are on average 7 percent less productive than mature Indonesian producers, new greenfield operations exhibit on average a 4 percent higher TFP level, and for mature foreign affiliates the premium reaches 36.6 percent. The share of output exported by new indigenous producers is 3 percentage points higher when compared to mature Indonesian plants. This figure is an impressive 35 percentage points for the new foreign entrants and 21 percentage points for mature foreign affiliates.[4]

[3] The information on foreign ownership and age, needed to identify greenfield projects, is available starting in 1975.

[4] If entrants (domestic and foreign) were defined as plants in the first two years of their operation, we would find that domestic entrants are characterized by lower TFP and labor productivity than foreign entrants and mature Indonesian plants are. Foreign entrants would be found to outperform mature Indonesian plants in terms of labor productivity, but would not be significantly different in terms of TFP. If entrants were defined as plants in the first year of their operation, the conclusions would be the same as those just stated, except for foreign entrants exhibiting lower TFP than mature domestic producers.

Table 14.1. *Comparison of foreign and domestic entrants*

	TFP	Labor productivity	Output	Employment	Average wage	Investment	Investment in machinery	Export share	Imported input share	Capital intensity	Skilled labor share	TFP growth
Foreign affiliate (> 3 yrs old)	0.366***	1.290***	2.687***	1.375***	0.766***	2.307***	2.217***	21.234***	0.286***	0.875***	0.045***	0.009*
	(0.008)	(0.012)	(0.020)	(0.012)	(0.008)	(0.035)	(0.029)	(0.330)	(0.003)	(0.017)	(0.002)	(0.005)
New foreign entrant (yrs 1–3)	0.040**	0.927***	1.614***	0.663***	0.505***	3.090***	2.892***	34.754***	0.342***	1.455***	0.020***	0.070***
	(0.017)	(0.025)	(0.043)	(0.026)	(0.017)	(0.073)	(0.061)	(0.640)	(0.005)	(0.037)	(0.003)	(0.014)
New domestic entrant (yrs 1–3)	−0.075***	−0.046***	−0.279***	−0.213***	−0.077***	0.433***	0.339***	2.835***	−0.003**	0.341***	0.001	0.037***
	(0.005)	(0.006)	(0.010)	(0.006)	(0.004)	(0.018)	(0.015)	(0.178)	(0.001)	(0.011)	(0.001)	(0.004)
R2	0.45	0.32	0.28	0.17	0.64	0.10	0.10	0.21	0.23	0.20	0.19	0.14
No. of obs.	199,479	308,358	308,439	308,441	308,436	304,940	283,773	212,728	295,795	203,266	252,448	164,130

*, **, and *** indicate statistical significance at the 10, 5 and 1 percent levels, respectively.
Source: Arnold and Javorcik (2009a).

As the majority of global FDI flows take the form of acquisitions rather than greenfield projects, the next question one would like to ask is whether foreign ownership leads to an improved performance in the acquired establishments. This view is confirmed by the empirical analysis of Arnold and Javorcik (2009b) who use the data mentioned above, and control for the selection of acquisition targets by combining propensity score-matching with a difference-in-differences approach. They show that foreign ownership leads to significant productivity improvements in the acquired plants. The improvements become visible in the acquisition year and continue in subsequent periods. After three years, the acquired plants exhibit a 13.5 percent higher productivity than the control group. The rise in productivity is a result of restructuring, as acquired plants increase investment outlays, employment, and wages. Foreign ownership also appears to enhance the integration of plants into the global economy through increased exports and imports. Similar productivity improvements and evidence of restructuring are also found in the context of foreign privatizations.

The profound changes taking place in the acquired plants, documented by Arnold and Javorcik (2009b), do not extend to all aspects of plant operations. FDI does not appear to induce increases in the skill intensity of the labor force (defined as the share of white collar workers in total employment) or the capital-labor ratio.

How can we reconcile an increase in TFP, labor productivity and wages with no evidence of changes to skill composition or the capital-labor ratio? One possibility is that new foreign owners introduce organizational and managerial changes that make the production process more efficient by reducing waste, lowering the percentage of faulty product, and using labor more effectively.[5] Another possibility is that while foreign owners do not alter the skill composition of labor, they are able to attract more experienced and motivated workers.[6] They may also substitute expatriate staff for local managers and introduce pay scales linked to performance to motivate their staff.[7] This possibility is in line with the earlier observation that acquired plants hire a large number of new employees and raise the average wage. Further, foreign owners may invest more in staff training, which is consistent with the international experience mentioned earlier. Yet another possibility is that the use of higher quality inputs or more suitable parts and components translates into higher productivity.[8] This possibility is supported by the observation of FDI leading to a greater reliance on imported inputs.

[5] A relevant example of organizational changes introduced by a foreign investor in its Chinese affiliate is presented in Sutton (2005). According to the interviewed engineer, what mattered was not the obvious alteration to the physical plant, but rather, inducing a shift in work practices. This shift involved a move away from traditional notions of inspection at the end of the production line to a system in which each operator along the line searched for defects in each item as it arrived and as it departed. The idea of such constant monitoring was in part to avoid adding value to defective units. More importantly, this system allowed for a quick identification and rectification of sources of defects.

[6] About 10 percent of Czech firms surveyed by the World Bank in 2003 reported that they lost employees as a result of FDI entry into their sector (Javorcik and Spatareanu 2005).

[7] Lipsey and Sjöholm (2004) found that foreign affiliates in Indonesia paid higher wages to workers with a given educational level relative to domestic producers.

[8] For instance, a lower percentage of faulty inputs translates into fewer final products that must be rejected at the quality control stage.

4. DOES FDI INCREASE COMPETITIVE PRESSURES IN THE HOST COUNTRY?

The superior performance of foreign affiliates documented in the previous section suggests that inflows of FDI are likely to increase competitive pressures in the host country, provided that at least part of their output is destined for the host country market. This view is supported by two types of evidence.

First, the most direct (though subjective) evidence comes from enterprise surveys where managers are directly asked about the implications of FDI inflows into their sectors. As reported by Javorcik and Spatareanu (2005), 48 percent of Czech firms interviewed believed that the presence of multinationals increased the level of competition in their sector. The same was true of two-fifths of Latvian enterprises. Almost thirty percent of firms in each country reported losing market share as a result of FDI inflows.

Increased competitive pressures resulting from FDI inflows are likely to lead to adjustments similar to those documented in the literature on tariff liberalization (for example, Pavcnik 2002): exit of the least-productive indigenous firms and expansion of better performers. While no direct evidence of such adjustment is available for episodes of large FDI inflows, it is interesting to note that Czech firms reporting (in a 2003 survey) rising competitive pressures, as a result of foreign entry, experienced faster productivity growth and a larger increase in employment in 1997–2000 than other firms did (Javorcik and Spatareanu 2005). This pattern is consistent with the idea that only firms able to make improvements were able to withstand increased competition and survive. In contrast, Czech firms reporting loss of market share, which they attributed to foreign presence in their sector, experienced a much larger decline in employment and a slower TFP growth than other firms, which supports the idea that weaker performers decline in the face of increased competition.

The second piece of evidence comes from firm-level panel studies, some of which have documented a negative relationship between the presence of foreign affiliates in the sector and the performance of indigenous producers. Such a pattern was, for instance, found by Aitken and Harrison (1999) in Venezuela. The authors' interpretation of this finding was that the expansion of foreign affiliates took part of the market share away from local producers, forcing them to spread their fixed cost over a smaller volume of production, resulting in a lower observed TFP. As pointed out by Moran (2007), during the time period considered in the study, Venezuela was pursuing an import-substitution strategy, thus indigenous producers were not exposed to significant competition from abroad. It is not surprising therefore that FDI inflows could have had a large negative effect on market shares of indigenous producers.

Though this issue has not been formally investigated, the magnitude of the increase in competitive pressures resulting from FDI inflows is likely to depend on host country characteristics. It will be limited in countries with liberal trade regimes, and quite large in countries with restrictive trade policies.

5. DOES FDI LEAD TO KNOWLEDGE SPILLOVERS?

The combination of large endowments of intangible assets and high investment in staff training, both of which characterize multinational companies, suggests that FDI can potentially lead to knowledge spillovers in a host country. The existence of such spillovers is supported by several types of evidence.

First, evidence of knowledge spillovers appears in enterprise surveys. For instance, according to Javorcik and Spatareanu (2005), 24 percent of Czech firms and 15 percent of Latvian firms reported learning about new technologies from multinationals operating in their sector. The difference in the ability to learn about marketing techniques was much less pronounced (about 12 percent of respondents in each country). Whether these differences stem from differences in the composition of FDI inflows or differences in local firms' ability to absorb knowledge spillovers, the key message is that host country conditions affect the extent of knowledge spillovers.

The second type of evidence on spillovers from FDI comes from studies asking whether the movement of employees from MNCs to local establishments benefits the productivity of the latter. This is a very promising area for future research, though due to high data requirements, there exist only a few studies on this topic. Görg and Strobl (2005) use Ghanaian data on whether or not the owner of a domestic firm had previous experience in a multinational, which they relate to firm-level productivity. Their results suggest that firms run by owners who worked for multinationals in the same industry immediately before opening their own firm are more productive than other domestic firms are. Using matched employer-employee data from Norway (though Norway is not a developing country, the study is worth mentioning here), Balsvik (2009) finds that hiring workers with MNC experience boosts the productivity of domestic firms (controlling for other factors, including the total number of new employees). In a related study, Poole (2009) argues that if movement of labor is a spillover channel, we should observe that workers in domestic establishments with a greater share of employees with MNC experience should enjoy higher wages. Using a matched employer-employee data set from Brazil, she finds results consistent with the existence of such spillovers. Namely, *ex-ante* identical workers in establishments with a higher proportion of workers with some experience at a multinational firm earn higher wages, though this effect is statistically significant in only some industries.

The third type of evidence on spillovers from FDI comes from firm-level panel studies. While the identification of knowledge spillovers within an industry is complicated by the existence of the competition effect mentioned in the previous section, spillovers through linkages to the supplying sectors seem (at least *a priori*) to be easier to capture. Evidence consistent with productivity spillovers benefiting upstream producers has been found in Lithuania by Javorcik (2004) and in Indonesia by Blalock and Gertler (2008). Such evidence, though convincing, relies on input-output matrices to capture linkages between MNCs and their suppliers, rather than on information on actual relationships between indigenous producers and multinationals. Ideally, the literature should move towards iden-

tifying MNC suppliers and analyzing the causal relationship between doing business with MNCs and supplier performance.

The limited information available suggests that MNC suppliers are different from other indigenous producers. Controlling for industry affiliation and year fixed effects, Javorcik and Spatareanu (2009a) find that Czech firms supplying MNCs tend to be 13 percent larger in terms of employment and 18 percent larger in terms of sales value, though they do not experience faster sales growth. They tend to have higher TFP levels (14 percent premium), and higher labor productivity measured as value added per worker (23 percent premium). They also appear to be more capital intensive (17 percent) and pay higher wages (12 percent). Controlling for firm size does not change these conclusions (see Table 14.2).

Table 14.2. *Supplier Premium*

	(a) (%)	(b) with controls for firm size
Total employment	12.8	–
Sales	17.7	11.1
Sales growth	ns	ns
Capital per worker	16.6	18.6
TFP	14.1	11.6
Value added per worker	23.2	12.2
Wages per worker	11.7	14.4

The(a) The premium is based on coefficients of the Supplier dummy in the following regressions:
$ln\ X_{it} = \alpha + \beta\ Supplier_{it} + \mu_j + \mu_t + \varepsilon_{it}$
where μ_j stands for two-digit industry and μ_t for year fixed effects.
The(b) The premium is based on the following regression:
$ln\ X_{it} = \alpha + \beta\ Supplier_{it} + \delta\ ln\ Employment_{it} + \mu_j + \mu_t + \varepsilon_{it}$
ns denotes a coefficient not statistically significant at conventional levels.
Source: Javorcik and Spatareanu (2009).

Further, Javorcik and Spatareanu (2009a) find that while more productive firms self-select into supplying relationships with multinationals, the results from the instrumental variable approach are suggestive of learning from the relationships with MNCs. However, as these conclusions are based on a small sample, they should be treated with caution and tested using a larger data set.

From the perspective of policy makers, more interesting is the observation that Czech firms supplying MNCs are less credit-constrained than non-suppliers are. A closer inspection of the timing of the effect suggests that this result is due to less-constrained firms self-selecting into becoming MNC suppliers, rather than from the benefits derived from the supplying relationship (Javorcik and Spatareanu 2009b). This result is not surprising, as survey evidence suggests that supplying MNCs often requires significant investment outlays and obtaining costly quality certifications (for example, ISO 9000). For instance in a survey of Czech enterprises, 40 percent of all respondents who reported having such an ISO certification obtained the certification to become suppliers to multinationals.

The above evidence is suggestive of well functioning credit markets being important in facilitating business relationships between local firms and MNCs,

though they *do not* suggest that a well developed financial market is a *sufficient* condition for such relationships to take place. Other factors, such as a certain level of sophistication of the local manufacturing sector, may be needed for these relationships to materialize.

6. FDI IN SERVICES

Most of the barriers to FDI today are not in goods but in services (UNCTAD, 2004), reflecting the unwillingness of governments, particularly in the developing world, to allow unrestricted foreign presence in what they believe are ‘strategic’ sectors. For instance, even though economies in South East Asia, such as Malaysia and Thailand, which have reaped huge benefits from the liberalization of trade and investment in goods, continue to maintain restrictions on foreign ownership in services ranging from transport to telecommunications. India, which is emerging as a highly competitive supplier of a range of skilled labor-intensive services, still restricts foreign ownership in banking, insurance, telecommunications, and retail distribution.

Yet FDI in services presents a large source of potential gains for the host country. The nature of many service industries and the existing barriers to trade in services mean that the scope for using cross-border trade to substitute for domestically produced service inputs is limited. Therefore, the competitiveness of manufacturing sectors is tied more directly to the quality and availability of services supplied domestically than is the case for physical intermediate inputs. As virtually all enterprises use basic services, such as telecommunications and banking, improvements in these services are likely to affect all industries.

Starting with the theoretical contribution of Ethier (1982), researchers have argued that access to a greater variety of inputs raises the productivity of downstream industries. Access to a larger range or higher-quality inputs is one of the oft-cited arguments in favor of trade liberalization. A similar argument could be made about FDI inflows, especially into service industries.

Foreign entry into the service industry may improve and expand the set of available producer services and introduce international best practices. It may also induce domestic competitors to make similar improvements. In Mexico, for example, Wal-Mart introduced cutting-edge retail practices (central warehousing, an appointment system, use of pallets), which significantly cut distribution costs. These practices were quickly adopted by other domestic retail chains competing with Wal-Mart (Javorcik, Keller, and Tybout 2008).

Survey data from the Czech Republic reveal that local entrepreneurs have positive perceptions of opening the service sector to foreign entry. A vast majority of respondents reported that liberalization contributed to improvements in the quality, range, and availability of services inputs. The positive perceptions ranged from 55 percent of respondents asked about the quality of accounting and auditing services to 82 percent for telecommunications. With regard to the variety of products offered, the positive views of liberalization ranged from 56 percent of respondents evaluating accounting and auditing services, to 87 percent of re-

spondents who were asked about telecommunications. The corresponding figures for the effect on service availability ranged from 47 percent in accounting and auditing to 80 percent in telecommunications (Arnold et al. 2007).

Arnold et al. (2007) formally examined the link between FDI in services and the performance of domestic firms in downstream manufacturing. Using firm-level data from the Czech Republic for 1998–2003, they measure the presence of FDI in services by the share of services output provided by foreign affiliates. The manufacturing–services linkage is captured using information on the degree to which manufacturing firms rely on intermediate inputs from service industries. The econometric results indicate that opening services to foreign providers leads to improved performance of downstream manufacturing sectors. This finding is robust to several econometric specifications, including controlling for unobservable firm heterogeneity and other aspects of openness, and instrumenting for the extent of foreign presence in service industries. The magnitude of the effect is economically meaningful: a one standard deviation increase in foreign presence in service industries is associated with a 3.8 percent increase in the productivity of manufacturing firms relying on service inputs.

FDI inflows into services, and more specifically into the wholesale and retail sector, may also lead to increased competition in the manufacturing industries of a host country. Global retail chains with their extensive supplier networks spanning multiple countries, if not continents, are much better positioned than smaller national chains to put pressure on indigenous suppliers. This view is supported by the results of a case study of the soap, detergent, and surfactant (SDS) producers in Mexico. According to Javorcik et al. (2008), entry of Wal-Mart into Mexico changed the way that SDS producers and other suppliers of consumer goods interacted with retailers. By exercising its bargaining power, Wal-Mart squeezed profit margins among the major brands, offering them higher volumes in return. It also engaged the most efficient small-scale local producers as suppliers of store brands, thereby creating for itself a residual source of SDS products that could be used in bargaining with the major (multinational) branded suppliers. Those local firms that were not efficient enough to meet Wal-Mart's terms lost market share, and many failed. At the same time, the limited set of producers that survived grew, and with prodding from Wal-Mart they became more efficient and innovative, adopting innovations first introduced to the market by their multinational competitors.

This view finds further support in the results of Javorcik and Li (2009) who examine how the presence of global retail chains affects firms in the supplying industries in Romania. Applying a difference-in-differences method and the instrumental variable approach, the authors conclude that expansion of global retail chains leads to a significant increase in the TFP in the supplying industries. Presence of global retail chains in a Romanian region increases the TFP of firms in the supplying industries by 3.8 to 4.7 percent and doubling the number of chains leads to a 3.3 to 3.7 percent increase in total factor productivity. However, the expansion benefits larger firms the most and has a much smaller impact on small enterprises.

7. FUTURE RESEARCH

The evidence presented in this note has several implications for the direction of future research on the adjustment process taking place in the aftermath of FDI inflows. First, it suggests that the focus of the debate should shift from attempting to generalize about whether or not FDI spillovers exist, to determining the conditions under which they are likely to be present, and investigating under what conditions their positive effect on indigenous firms' performances will be dwarfed by the increased competition resulting from foreign presence.[9] Examining the impact of FDI in the context of one country at a time is unlikely to be very productive. What is needed is a multi-country study based on comparable high-quality, firm-level panel data that could take into account host country characteristics. Conducting a meta-study focusing on the host country business environment and level of development could be another promising avenue for future research.

Second, more effort should be directed at understanding the exact mechanisms behind the observed patterns. Rather than correlating the performance of host country firms with the presence of multinationals in their or other sectors, researchers should look at the flows of workers between the two types of firms, identify domestic suppliers of foreign customers, consider the effect of foreign presence on the entry of new firms and their characteristics, and ask firms detailed questions about the sources of their innovation. Some researchers have already pursued this line of study, but more work is needed. While it creates new challenges in terms of finding appropriate econometric strategies, collecting data, and overcoming the fear of relying on surveys, this area of research probably has the greatest potential.

Third, the scope of investigations should be extended to encompass the service sector. Anecdotal evidence suggests that the movement of service industry professionals to executive positions in other firms may also constitute an important spillover channel to other service firms and to manufacturing industry. For instance McKendrick (1994) reports that local banks and financial institutions in Latin America and South Asia are filled with the 'alumni' of Citibank and BNP. Moreover, because the nature of the sector and trade barriers limit cross-border trade in services, opening service industries to foreign providers may confer large benefits on downstream manufacturing. Of course, allowing entry of foreign services providers without undertaking complementary reforms (competition, regulation) is unlikely to be productive. More research is certainly needed to assess the conditions under which a host country can maximize the benefits from FDI in services.

Beata Javorcik is a reader in economics at the University of Oxford and a CEPR Research Affiliate.

[9] This is not to say that an increase in competition is not a desirable effect.

BIBLIOGRAPHY

Aitken, Brian and Ann Harrison (1999). Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela. *American Economic Review* 89(3): 605–18

Arnold, Jens Matthias and Beata S Javorcik (2009a). *Gifted Kids or Pushy Parents? Foreign Direct Investment and Plant Productivity in Indonesia*, University of Oxford, Department of Economics Discussion Paper No. 434

Arnold, Jens Matthias and Beata S Javorcik (2009b). Gifted Kids or Pushy Parents? Foreign Direct Investment and Plant Productivity in Indonesia, *Journal of International Economics*, 79(1): 42–53

Arnold, Jens Matthias, Beata S Javorcik and Aaditya Mattoo (2007). Does Services Liberalization Benefit Manufacturing Firms? Evidence from the Czech Republic, *World Bank Working Paper* No. 4109

Balsvik, Ragnhild (2009) Is Labor Mobility a Channel for Spillovers from Multinationals? Evidence from Norwegian Manufacturing. *Review of Economics and Statistics*, forthcoming

Blalock, Garrick and Paul J Gertler (2008). Welfare Gains from Foreign Direct Investment through Technology Transfer to Local Suppliers. *Journal of International Economics*, 74(2): 402–21

Dunning, John H (1993). *Multinational Enterprises and the Global Economy*. Wokingham, England: Addison-Wesley Publishing Company

Ethier, Wilfred (1982). National and International Returns to Scale in the Modern Theory of International Trade. *American Economic Review* 72: 389–405

Görg, Holger and Eric Strobl (2005). Spillovers from Foreign Firms through Worker Mobility: An Empirical Investigation. *Scandinavian Journal of Economics* 107(4): 693–709

Helpman, Elhanan, Marc J Melitz and Stephen R Yeaple (2004). Export Versus FDI with Heterogeneous Firms, *American Economic Review* 94(1): 300–16

Javorcik, Beata S (2004). Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages, *American Economic Review* 93(3): 605–27

Javorcik, Beata S (2008). Can Survey Evidence Shed Light on Spillovers from Foreign Direct Investment? *World Bank Research Observer*, 23(2): 139–159

Javorcik, Beata S, Wolfgang Keller and Jame Tybout (2008). Openness and Industrial Response in a Wal-Mart World: A Case Study of Mexican Soaps, Detergents and Surfactant Producers, *The World Economy* 31(12): 1558–1580

Javorcik, Beata S and Yue Li (2009). Do the Biggest Aisles Serve a Brighter Future? Global Retail Chains and Their Implications for Romania, University of Oxford, mimeo

Javorcik, Beata S and Mariana Spatareanu (2005). Disentangling FDI Spillover Effects: What Do Firm Perceptions Tell Us? in *Does Foreign Direct Investment Promote Development?* T Moran, E Graham and M Blomstrom, eds., Institute for International Economics

Javorcik, Beata S and Mariana Spatareanu (2009a). Tough Love: Do Czech Suppliers Learn from Their Relationships with Multinationals? *Scandinavian Journal of Economics*, 111(4): 811–833

Javorcik, Beata S and Mariana Spatareanu (2009b). Liquidity Constraints and Linkages with Multinationals, *World Bank Economic Review*, 23(2): 323-346

Kertesi, G and J Köllö (2001). A gazdasági átalakulás két szakasza és az emberi töke átértékelödése, (Two phases of economic transformation and the revaluation of human capital). *Közgazdasági Szemle* 47: 897–919

Lipsey, Robert E and Fredrik Sjöholm (2004). Foreign direct investment, education and wages in Indonesian manufacturing, *Journal of Development Economics*, 73(1): 415–22

Markusen, James (2002). *Multinational Firms and the Theory of International Trade*, Cambridge: MIT Press.

Moran, Theodore (2007). *How to Investigate the Impact of Foreign Direct Investment on Development and Use the Results to Guide Policy*. Georgetown University mimeo

Poole, Jennifer (2009). *Knowledge Transfers from Multinational to Domestic Firms: Evidence from Worker Mobility*, University of California Santa-Cruz, mimeo

Ramachandaram, Vijaya (1993). Technology transfer, Firm Ownership, and Investment in Human Capital. *Review of Economics and Statistics* 75(4): 664–70

Sutton, John (2005). The Globalization Process: Auto-Component Supply Chain in China and India, in Bourguignon, F, Pleskovic B and Andre Sapir, eds. Annual World Bank Conference on Development Economics – *Europe. Are We on Track to Achieve the Millennium Development Goals?*, Washington DC: World Bank and Oxford University Press

UNCTAD (United Nations Commission on Trade and Development) (1997). *World Investment Report: Transnational Corporations, Market Structure, and Competition Policy*. New York: United Nations

UNCTAD (2004). *World Investment Report. The Shift Towards Services*. New York and Geneva: United Nations

UNCTAD (2005). *World Investment Report: Transnational Corporations and the Internalization of R&D*. New York and Geneva: United Nations

World Bank (1997). *Malaysia: Enterprise Training, Technology, and Productivity*. World Bank. Washington, DC

15

Adjustment to Foreign Changes in Trade Policy Under the WTO System

CHAD P BOWN

1. INTRODUCTION

How do domestic economies adjust when other countries change their trade policies? This question is increasingly important for at least two reasons associated with the current state of the global economy and its rules-based trading system. First, the global economy is highly integrated through foreign direct investment, supply chains, and international trade flows. These are the result of decades of multilateral negotiations under the GATT/WTO system, which have led the major nations to impose and legally 'bind' their import tariffs at historically low average levels. But the same rules-based system that has led to low average tariffs also allows its members to access a number of liberal trade 'exceptions' (such as safeguards and antidumping), which permit WTO members to change their trade icy in response to political–economic shocks. Given the current WTO framework, which results in both openness to foreign shocks because of liberal trade, and yet the possibility of significant changes in countries' trade policies over time, there is an important need for research to improve our understanding of how trade flows, industries, firms, and factors of production adjust to foreign changes in trade policy.

The lack of understanding of how domestic economies adjust to changes in foreign market access and trade policy abroad is not because theorists have failed to motivate the importance of the issue. Terms of trade theory has long suggested that whenever a policy-changing country is 'large' and thus able to affect international prices, a policy change is expected to feed back into the domestic economy of trading partners, thus imposing an adjustment process on their economies as well. Indeed, a now dominant strand of the theoretical literature on trade agreements (Bagwell and Staiger, 1999; 2002) identifies a fundamental *raison d'être* of the WTO as the necessity to confront the international cost-shifting motive of large country members whose policy changes affect other economies via their impact on the terms of trade (exporter-received prices), or equivalently, the terms of export market access (export sales volumes), holding all else constant.

While the international externality implications of trade policy changes have a long history in the theoretical literature, the first round of empirical support for

these theories has only recently emerged. In particular, Broda, Limão, and Weinstein (2008) is the first evidence consistent with the theory that, when unconstrained by WTO rules, countries set import tariffs with terms of trade considerations (and hence international cost-shifting motives) in mind.[1] From our perspective, this line of research motivates the need for additional work to examine the process by which domestic economies adjust to frequent changes in trade policies abroad. The Broda, Limão, and Weinstein (2008) evidence suggests that such trade policy changes are likely to have important international externality implications.

Any new literature on domestic adjustment to changes in export market access can be expected to draw substantially on the well-established parallel research literature on the process of adjustment to new import competition. Indeed, most of the relevant empirical research examining how trade policy affects the adjustment process has focused largely on the 'own' environment—that is, how a country's change in its own trade policy affects its own imports, domestic industries, domestic firms, and domestic factors of production. This research typically investigates the adjustment experience of a trade liberalization episode that increased the domestic economy's own openness to imports. [2] An extensive and impressive literature has examined various aspects of the domestic adjustment experiences associated with trade liberalization shocks across a diverse set of countries and time periods.[3]

Why has the research on changes in export-market access that would be the natural parallel to the literature on import-market access liberalization not yet materialized? The paucity of research on the response to changes in export-market access can be explained at least partially by the challenges that confront researchers attempting to estimate the adjustment impact of other countries' changes to trade policy. The first challenge is to create the same sort of 'natural experiment' testing environments analogous to import-market access shocks (exogenous, unilateral trade liberalizations) on the export side of the market, which is necessary to identify the causal link between changing conditions of export-market access and the process of adjustment. Partly because of a lack of sufficiently detailed data needed to control for other factors, researchers have so far not used most of the same exogenous and large-scale import-market access trade liberalization episodes of the parallel literature to examine the adjustment process

[1] See also the empirical evidence of Bagwell and Staiger (forthcoming).

[2] In one of the few research papers examining the domestic effects of a country's liberalization of exports, Edmonds and Pavcnik (2006) examine the micro-level impacts of Vietnam's removal of restrictions on rice exports in the 1990s. The research that we motivate and describe in more detail below also examines the adjustment process of those associated with exporting, but through the alternative channel of foreign changes in trade policy, not the change in the exporting country's own policy.

[3] Examples of countries and trade liberalization environments frequently studied in this context include Brazil, Canada, Chile, Cote d'Ivoire, India, Mexico, and Turkey. For recent surveys, Tybout (2000) and Erdem and Tybout (2004) examine the responses of domestic import-competing firms and industries to these types of shocks, while Goldberg and Pavcnik (2005; 2007) respectively examine the literature on the effects on poverty rates and income inequality of import market liberalization.

from the perspective of the exporters abroad.[4] Improvements in data availability are making these sorts of approaches more plausible, and in Section 3 we describe a number of research approaches that adapt the identification strategy to exploit a smaller-scale approach. [5] In particular, several papers take advantage of product-specific or industry-specific exogenous changes in foreign market access as part of an identification strategy to estimate the impact of policy changes abroad on the domestic adjustment process.

Before turning to these specific examples of research, the next section briefly describes the most relevant features of the WTO: the foundation of the current rules-based trading system. Section 2 therefore describes the key elements of the WTO that establish the exceptions and procedures, that is, the WTO features that national governments use and those which create the identification opportunities that the research described in Section 3 exploits. More than 60 years of GATT/WTO negotiations have resulted in a WTO agreement that is largely responsible both for today's liberal trading environment and the rules under which certain forms of trade policy changes occur. Given the lack of major reform proposals in the ongoing Doha Round of WTO negotiations, these rules and procedures governing how the current system accommodates national changes in trade policy at the industry or product level are likely to become even more relevant in the future.[6] Especially as more developing countries increase their openness to trade and are encouraged to adopt the WTO system's approach to accommodating national changes in trade policy through 'exceptions' such as safeguards and antidumping, research in this area is increasingly important and relevant for policy.

2. INSTITUTIONAL BACKGROUND: USING THE WTO SYSTEM FOR IDENTIFICATION

In this section we briefly describe two elements of the current WTO system that may provide fertile testing environments for research on how foreign trade pol-

[4] To see one important part of the problem, consider the case of an exporting firm that serves two or more foreign markets. If all of the foreign markets don't make their detailed trade policy data easy to observe (and collect), the data problem can become insurmountable, as it is impossible to control for other foreign countries' trade policy changes that may equally affect the exporting firm's adjustment process.

[5] Papers such as Trefler (2004) and Lileeva and Trefler (forthcoming) for the Canada–US Free Trade Agreement, and Bustos (forthcoming) for MERCOSUR do exploit the fact that certain countries' exports may be highly concentrated toward one foreign market, and thus when that foreign market undertakes additional (and preferential) import liberalization, the concern of not having access to data on trade policy changes in other foreign markets is less problematic. However, a secondary concern for even these types of studies could be that the export market access changes embodied in these trade agreements may not have been exogenous or unanticipated, which may lead to additional challenges for identification.

[6] This assumes there is no large-scale protectionist retreat associated with the global financial crisis. While the severity of the global recession caused by the crisis remains uncertain, as is the extent of an associated protectionist response, early evidence from policy changes during the crisis indicates that countries may be refraining from large-scale protectionism. Bown (2009a) presents some evidence of a moderate increase in the use of new import restrictions in the form of antidumping and safeguards, at least through the first quarter of 2009, associated with the crisis. On more general trends in protectionism during the crisis, see the other contributions in Evenett and Hoekman (2009).

icy changes affect the domestic adjustment process.[7] The first is how WTO exceptions such as antidumping and safeguards allow countries to change the conditions of trading partners' export market access via imposition of new trade restrictions. The second is how WTO dispute-settlement provisions facilitate changes in the conditions of export market access via removal of partners' trade-distorting policies.

Before turning to this discussion, it is important to note that other WTO rules are likely to affect each of these areas in ways that may ultimately influence the identification strategies that researchers use in econometric applications. One important WTO principle is most-favored-nation (MFN) treatment, by which the WTO system requires its members to apply nondiscriminatory treatment across trading partners.

2.1 Imposition of new trade barriers under WTO exceptions

Under the GATT/WTO system, many of the major economies have relatively low average tariffs as well as applied tariff rates that are quite close to their bound rates. Table 15.1 documents this for economies such as the United States, the European Union, and Japan, which have legally bound virtually all of their import tariff lines under the WTO and have applied and bound rates on manufactures imports, if not necessarily agriculture, in the range of only 2 to 4 per cent on average. Even China, despite being a developing economy, has average tariffs that are relatively low and applied tariff rates that are close to the binding levels of its tariffs, especially compared to other emerging economies such as Brazil and India.

Under the rules of the WTO system, if such economies feel pressure to raise trade barriers because of either domestic political–economic or foreign supply-induced shocks, policymakers in these economies have relatively limited options. If their applied tariff rates are close to their binding levels, they cannot simply raise tariff rates (or impose new quantitative restrictions), as this would be in blatant violation of WTO rules. Instead, these economies can use the 'exceptions' to liberal trade that are embedded in the WTO agreements in the form of policies such as antidumping and safeguards. As the last column of Table 15.1 indicates, for example, each of these economies except Japan is also among the WTO membership's most frequent users of antidumping to impose new product-specific import-restricting trade policies. Even focusing on only the WTO member countries listed in Table 15.1, and solely on their use of antidumping, the table shows that there have been hundreds of instances in which these countries imposed new product-level import restrictions, typically for at least five years.[8] While not shown in the table, many of these economies are also frequent users of the other major WTO-permitted exception—the global safeguard, which countries typically impose for three or four years. [9]

[7] For an extensive introduction to and discussion of the WTO, see Hoekman and Kostecki (2009).

[8] While only 50 to 60 per cent of the US or European Union investigations result in the imposition of new antidumping measures, the figure is over 80 per cent for India.

Table 15.1: *Selected WTO Members' applied tariffs and tariff bindings in 2007 and cumulative use of antidumping since China's 2001 accession*

Country/ Territory	Product Category	Binding coverage (%)	Average bound tariff (%)	Average applied tariff (%)	Number of antidumping initiations, 2002–2008 (WTO rank)
US	All	100	3.5	3.5	162 (2)
	Agriculture	na	5.0	5.5	
	Non-agriculture	100	3.3	3.2	
EU	All	100	5.4	5.2	145 (3)
	Agriculture	na	15.1	15.0	
	Non-agriculture	100	3.9	3.8	
Japan	All	99.6	5.1	5.1	4 (26)
	Agriculture	na	22.7	21.8	
	Non-agriculture	99.6	2.4	2.6	
China	All	100	10.0	9.9	131 (4)
	Agriculture	na	15.8	15.8	
	Non-agriculture	100	9.1	9.0	
Brazil	All	100	31.4	12.2	74 (7)
	Agriculture	na	35.5	10.3	
	Non-agriculture	100	30.8	12.5	
India	All	73.8	50.2	14.5	312 (1)
	Agriculture	na	114.2	34.4	
	Non-agriculture	69.8	38.2	11.5	

Source: compiled by the author from WTO's *World Tariff Profiles 2008*. The entry 'na' indicates not available. Binding coverage is defined as share of HS six-digit subheadings containing at least one bound tariff line. Simple averages are of the *ad valorem* (*ad valorem* equivalent) six-digit HS duty averages.

WTO rules require tariffs to be applied on an MFN basis, that is, in a way that does not discriminate across foreign export sources; nevertheless, countries frequently impose new trade barriers such as antidumping and global safeguards in a discriminatory way. While global safeguards especially are supposed to be imposed on an MFN basis, countries sometimes apply the policy so as to exempt certain exporters. Some of the research described in the next section exploits this discriminatory behavior. And while antidumping investigations are carried out at the level of a single foreign exporting country, the new trade restrictions can actually be applied on an exporting firm-specific basis, allowing policy-imposing countries to discriminate even across firms if they determine that different firms had different dumping (or pricing at 'less than fair value') margins.

In cases in which these policies change in a discriminatory manner, the adjustment is then not necessarily limited to the two countries directly impacted—that

[9] Bown (2009b) provides detailed data on WTO member use of antidumping, global safeguards, countervailing duties, and China-specific safeguards. For policies such as antidumping, for some countries this database also includes information on the size of the firm-specific new trade barriers.

is, the policy-imposing country's import-competing firms and the firms in the exporting country targeted by the trade restriction—but it likely affects firms in other exporting countries as well. The discriminatory imposition of a new trade barrier against one foreign export source but not another creates an implicit preference for exporters not subject to the policy. From the perspective of the econometrician looking for identification, this group of non-targeted exporters may be an especially interesting cohort to examine. If it turns out that they are not a cause of the 'problem,' a policy change affecting other foreign suppliers to the same export market may be viewed as an exogenous event from their perspective.[10]

2.2 Removal of trade barriers under WTO dispute settlement

A second important institutional aspect of the WTO that results in members changing their trade policy in ways likely to affect the adjustment process in foreign countries is through formal dispute settlement.[11] The chain of events associated with a typical WTO dispute is the following. One WTO member country—ultimately the defendant—imposes a WTO-illegal policy or refuses to live to up to a commitment negotiated in an earlier negotiating round that infringes on the export market access for which another WTO member country negotiated. The infringement may be the result of a newly imposed, but WTO-inconsistent, antidumping or safeguard policy such as those described in the last section, or it may be an illegal subsidy, a standards barrier, or some other non-tariff barrier to trade. The WTO dispute settlement process typically results in the defendant country removing its WTO-inconsistent policy. There have been over 400 disputes of this type since the WTO's inception in 1995.

The identifiable and discrete nature of the removal of the WTO-inconsistent policy may provide a useful environment in which to study the adjustment process. Analogous to the cases described in the previous section, because of the WTO's MFN principle, the restoration of market access available to exporters in the complaining country in the dispute is also likely to impact exporters (and hence the adjustment process) in other WTO member countries that export the same product as the one under dispute. If the initial WTO violation being eliminated was itself applied on an MFN basis, the exogenous (from the perspective of third countries) removal of the policy would be expected to have a positive market-access impact and thus positive implications for adjustment by the supplying sector in such third countries. But if the initial WTO violation was applied on a discriminatory basis (and thus afforded implicit preferential access to the third countries), the exogenous removal of the policy would be expected to have

[10] By not part of the problem, we mean did they not contribute to a surge in imports that may have been the shock triggering the new demand for import protection resulting in a change in foreign market access.

[11] Bown (2009c) provides an analysis of WTO dispute settlement from the perspective of developing countries in particular. Bown (2009d) presents a taxonomic approach that identifies the ways in which various types of and resolutions to WTO disputes can be expected to cause third country trade flows to adjust.

a negative market-access impact and implications for adjustment within such third countries.

While the adjustment process in such a trade dispute context has been subject to only limited empirical analysis, micro-level studies in this area, in the flavor of those that we review in Section 3.2, would seem to be an important component of some high-profile WTO disputes.[12] These include cases challenging European Community policies over imported bananas and sugar, policies which afforded large initial (but WTO-violating) discriminatory preferences for many developing countries. These affected countries would have then been forced to adjust when the European Union removed the violations and restored basic MFN treatment.

3. RESEARCH ON ECONOMIC ADJUSTMENT TO FOREIGN CHANGES IN TRADE POLICY

The previous section identified how WTO rules and exceptions establish a number of testing environments that researchers may find create differential treatment of the kind useful for identification. First is the application of new barriers under the agreement's exceptions (such as safeguards and antidumping) that eliminate foreign market access for some countries and, in the case of discriminatory application of the new policy, potentially create implicit preferential market access for other (non-targeted) exporters. Second is the removal of these and other similar barriers through the WTO's formal dispute settlement procedures, which can create (or at least restore) foreign market access for some exporters and, along the same lines, may also eliminate (implicit) preferential market access that other exporters had enjoyed due to violations of the rules on nondiscrimination.

In the next two subsections we explore examples of how in practice researchers are exploiting these sorts of testing environments to assess the impact of foreign changes in trade policy on the adjustment process. First we examine the more direct impact through the effect on trade flows, and we then describe how research is looking beyond trade flows to adjustments taking place at the micro level of individuals, households, or firms.

3.1 Trade-flow adjustment to foreign changes in trade policy

Bown and Crowley (2007) empirically examine whether a country's use of an import-restricting trade policy distorts a second country's exports to third markets. As in Figure 15.1, they first develop a theoretical model of trade between three countries (A, B, C), in which the imposition of tariffs by one country (A) causes

[12] The only empirical study of which we are aware that examines even the third country trade flow impact of WTO dispute settlement decisions and outcomes is Bown (2004a). That paper use a sample of GATT/WTO disputes over the 1991–98 period to assess the extent to which product-specific trade liberalization that the defendant country extends to the complaining country after an economically successful trade dispute is also extended to (non-complaining) third country exporters of the same product under the MFN rule.

significant distortions in 'world' trade flows. For example, when country A imposes a discriminatory tariff on imports from country B, the first-order impact is a simple 'destruction' of A's imports from B and an increase in A's imports from the non-targeted exporter C through the traditional channel of 'trade diversion' (Viner, 1950). The novel element of the paper is to focus on the tariff's additional impacts on trade with third markets. Specifically, A's import tariff on B leads B to 'deflect' some of its exports to country C; A's tariff on B also leads to increased

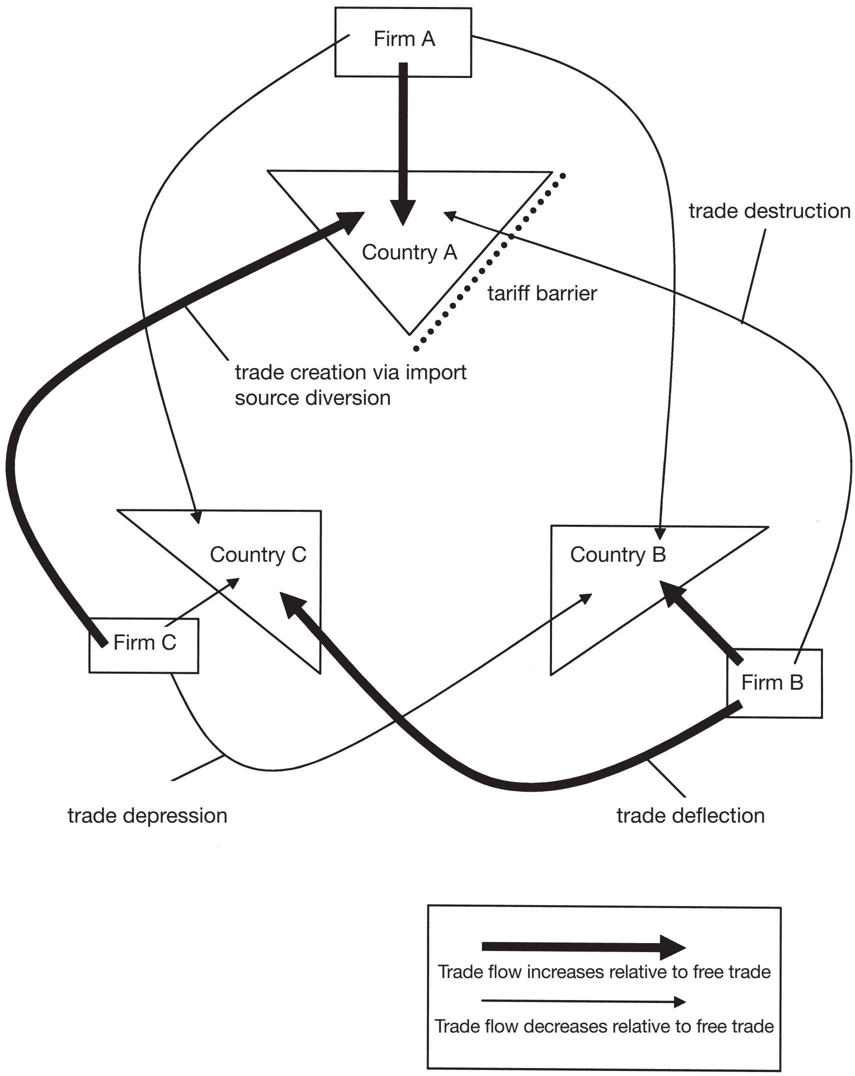

Figure 15.1: *Trade flow response to a discriminatory import duty in a three country model*

Source: Figure 2 of Bown and Crowley (2007, 181).

domestic consumption of the affected good in country B, which then crowds out B's imports from C, a phenomenon termed 'trade depression.'

The main contribution of Bown and Crowley (2007) is to provide a first empirical test of these third market effects on trade flows of trade deflection and trade depression. The paper investigates the effect of US antidumping and safeguards import restrictions on Japanese exports of nearly 5000 6-digit Harmonized System (HS) products to 37 countries between 1992 and 2001. Their evidence suggests that application of new US import restrictions both deflected and depressed Japanese trade with third countries during the period. Imposition of a US antidumping measure against Japan deflected trade; the average antidumping duty on Japanese exports to the United States led to a 5 to 7 per cent increase in Japanese exports of the same product to the average third-country market. The imposition of a US antidumping measure against a third country depressed Japan's trade with that country; the average US duty imposed on a third country led to a 5 to 19 per cent decrease in Japanese exports of that same product to the average third-country market.

Bown and Crowley (2006) present an extension to the study that looks in depth at the international externalities associated with US use of antidumping (AD) against Japanese exports to the United States and the European Union over the 1992–2001 period.[13] Following Prusa (1997; 2001), this paper first examines the trade destruction and trade diversion associated with the US AD duties on Japanese exports to the US market, and then documents sizeable trade deflection and trade depression in the European Union market resulting from the new US import restrictions. Model estimates indicate that, on average, roughly one quarter to one third of the value of Japanese exports to the United States apparently destroyed by US antidumping was actually deflected to the European Union in the form of a contemporaneous increase in exports. The paper also presents new evidence that US antidumping causes terms of trade externalities in non-targeted markets. New US tariffs on Japanese exports are associated with a substantial reduction of Japanese prices of these exports in the European Union market.

In a third paper in this line of research, Bown and Crowley (forthcoming) look for evidence of trade deflection in the context of China's historical exports. This particular empirical application was motivated by China's 2001 accession to the WTO, which allowed for current members to deviate from core WTO principles of reciprocity and MFN treatment by introducing access to a discriminatory, import-restricting 'China safeguard' that could be triggered by the mere threat of trade deflection. An *ex post* assessment on use of this safeguard indicates that between 2002 and 2009, industries in at least 10 different WTO members sought access to this particular import-restricting policy on more than 25 different occasions (Bown, 2009b).

Bown and Crowley (forthcoming) examine whether there is historical evidence that imposing discriminatory trade restrictions against China during its pre-WTO

[13] In related work, Staiger and Wolak (1994) study the 'own' impact of US policy on micro-level activity; part of their examination focuses on the US industry-level activity associated with US use of antidumping.

accession period led to Chinese exports surging to alternative markets. They construct a data set of product-level, discriminatory trade policy actions imposed on Chinese exports to two of its largest destination markets during 1992–2001. Perhaps surprisingly, they find no systematic evidence that either US or EU imposition of such import restrictions during this period deflected Chinese exports to alternative export destinations. To the contrary, there is evidence that such import restrictions may have had a chilling effect on China's exports of these products to alternative markets. The conditional mean US antidumping duty on China is associated with a 20 percentage point reduction in the relative growth rate of China's targeted exports to alternative markets during this period.

The question of the third-country effects of discriminatory use of trade policies under permitted WTO exceptions has also been the subject of a number of other papers that focus on global trade in particular products or industries. For example, Durling and Prusa (2006) focus exclusively on the global hot rolled steel market. They examine the impact in this particular product market of the use of such import barriers during the 'antidumping epidemic' of new trade restrictions during 1996–2001. Similarly, Debaere (2010) focuses on the global market for shrimp, which he uses to examines how the EU's discriminatory trade policy change (revocation of preferential tariff treatment for Thai exporters under the Generalized System of Preferences) affects the trade volumes and prices of traded shrimp in a third-country import market like the United States.

Table 15.2 summarizes the research described in this section. The literature provides evidence that the rules (and exceptions) of the WTO-based trading system lead countries to make trade-policy changes with economically significant impacts on the resulting exports and trade flows to non-targeted third country markets.[14] Using the terminology of Bown and Crowley (2007), sometimes affected exporters are able to deflect trade and sometimes they are not; sometimes trade is depressed, while sometimes it is not. The impact on third-country trade flows implies a likely need for the industries, firms, and factors of production that underlie these trade flows to adjust as well—a level of analysis that is just beginning to be taken up by formal empirical research. Research movements in this direction are described in Section 3.2.

[14] Other approaches to the third-party effects of discriminatory trade policy focus on the WTO exception to MFN found under the GATT's Article XXIV allowance that members be permitted to pursue preferential trade arrangements covering 'substantially all trade' with particular trading partners. Chang and Winters (2002) examine the effects of MERCOSUR on the export prices of nonmember countries to Brazil. They find that Brazil's tariff preferences to Argentina, Paraguay, and Uruguay result in competitive pressure in which exporters in other countries significantly reduce their prices and worsen their terms of trade. Similarly, Romalis (2007) examines the impact of the Canada–US Free Trade Area (CUSFTA) and the subsequent addition of Mexico (NAFTA). This study also finds that the implicitly discriminatory treatment to non-CUSFTA/NAFTA exporters results in a substantial impact on international trade volumes through a reduction in imports from nonmember countries.

Table 15.2: *Examples of research examining the adjustment response to foreign changes in trade policy*

Foreign change in trade policy	Testing environment to examine adjustment	Paper
US imposition of antidumping and safeguards	Japan's export volumes and term of trade to third (non-US) markets	Bown and Crowley (2006, 2007)
US and EU imposition of antidumping and safeguards	China's exports to third (non-US, non-EC) markets	Bown and Crowley (forthcoming)
Global use of antidumping on hot-rolled steel trade	Multiple countries' exports to third markets in hot-rolled steel	Durling and Prusa (2006)
EU withdrawal of GSP preferences for Thai shrimp	Thailand's shrimp exports to third (US) market	Debaere (2010)
US imposition of antidumping on Vietnamese catfish	Vietnam's household decisions and domestic labor markets	Brambilla, Porto and Tarozzi (2008)
US, EU, and China's imposition of steel safeguard import restriction on non-Indian exporters	Indian steel firms' input, output, product-switching, and export responses to the unexpected implicit preference	Bown and Porto (2009)

3.2 Adjustments at the 'micro level' to foreign changes in trade policy

The first paper of which we are aware to use a testing environment created by a foreign change in trade policy (permitted under a WTO exception) to examine the domestic, micro-level adjustment process is Brambilla et al. (2008). Their study examines the Vietnamese response to the US imposition in 2003 of antidumping tariffs on imports of catfish from Vietnam. As expected, this resulted in trade destruction through a major decline of Vietnamese exports of catfish to the US market. Using panel data on Vietnamese households, the paper examines the responses of catfish producers in the Mekong Delta between 2002 and 2004 and finds income growth for households relatively more involved in catfish farming in 2002 was significantly lower than for other comparable households. They also document how the US antidumping shock triggered significant Vietnamese exit from catfish farming. The paper traces how Vietnamese households adjusted by moving into wage labor markets and agriculture, but not into other areas of aquaculture. Thus, it would appear that not only were Vietnamese households unable to deflect their catfish exports to new markets in response to the new US important restrictions, but the technology with which they had farmed catfish was not readily transferable to other forms of aquaculture (for example, shrimp) in which Vietnamese exporters may have had access to relatively open international markets.[15]

Bown and Porto (2009) analyze a related micro-level adjustment to a foreign market access shock. They study the micro-level response in India to US imposition of significant new 'safeguard' import restrictions on steel products in

[15] Of course it would ultimately turn out that even the US import market for shrimp would not be open for much longer. In 2004, the United States initiated an antidumping investigation on shrimp imports from Vietnam and five other exporting countries, and this resulted in the imposition of new duties on Vietnamese shrimp in 2005, though mostly at a low level (4.57 per cent).

2002–03.[16] However, this paper is fundamentally different from Brambilla et al. (2008) along at least two dimensions. First is the underlying nature of the shock. Unlike the negative foreign market access shock associated with the US antidumping duties on Vietnamese catfish, Bown and Porto (2009) examine a potentially positive foreign market access shock that arose because India was granted an implicit tariff preference to the US steel market, based on the way in which the United States constructed its policy.[17] Second, instead of focusing on households (the unit of observation in the catfish study), the steel study focuses on the Indian firm-level response to the changing terms of market access associated with the foreign change in trade policy.

The paper provides evidence that Indian firms with historic export ties to the preference market responded more quickly to the changing market conditions in order to increase sales, exports, profits, and also to make adjustments to their use of inputs.[18] The paper also explores the source of firm-level entry into these new products (product-switching) and finds evidence that it was predominantly undertaken by larger firms that had previous experience exporting other types of steel products.

These studies are only two examples of research that uses foreign changes in trade policy to establish 'natural experiment' type environments that are useful for empirical testing. In the latter example, the idea is that third country changes in policy are reasonably unanticipated and exogenous events and thus can be used to identify the effects of changes in policy on adjustment-related decisions and outcomes.

4. CONCLUSIONS

This paper highlights a promising new area of research that empirically assesses some of the domestic adjustment to foreign changes in trade policies. While we have identified a number of important issues that this research is examining, we highlight two important caveats before concluding with one final policy motivation for the importance of this line of research.

First, the applicability of research findings on the adjustment associated with foreign changes in market access is expected to have its limits. In particular, since

[16] While we describe the policy as a US-imposed safeguard, in reality the testing environment in the paper is exploitable because the United States, the European Union, and China simultaneously imposed similar policies, thus granting preferential access to their markets to Indian firms over a similar set of steel products. Bown (2004b) provides evidence that the discriminatory application of the US steel safeguard in 2002–03 led to substantial trade diversion in the form of increased US imports from India and a number of other developing countries in the product categories targeted by the policy.

[17] From the perspective of our motivating discussion in Section 2, India is really the third country. The United States imposed discriminatory import restrictions targeting a number of other exporting countries but not India.

[18] While there is a substantial and growing empirical literature examining exporting firms and the process by which they adjust, this literature has not typically focused on the sort of environment created by an exogenous foreign market access shock studied in Bown and Porto (2009). For a recent survey of the exporting firm literature, see Bernard et al. (2007).

many of the foreign changes in trade policy being used for identification in the studies we have highlighted are product- or industry-specific, the research contributes less to our understanding of the broader general equilibrium types of issues than some of the parallel literature on the adjustment to new import competition.

Second, coming up with sufficiently 'clean' environments, such as those that the research described in Section 3 uses as identification, is not trivial and may become increasingly difficult for reasons we have not yet mentioned. Anecdotal evidence suggests that the trade policy changes that would be used for identification may be inter-related across countries, even at the product level.[19] A separate line of research examines such cross-country linkages and identifies a number of possible mechanisms through which this may occur. When it comes to policies such as safeguards or antidumping, some of it may be associated with retaliation (Blonigen and Bown, 2003), a reaction to the prospect of trade deflection (Bown and Crowley, 2007), or 'cascading protection' in which new trade barriers on inputs feed into downstream demands for new protection for domestic producers that use those more costly imported inputs (Hoekman and Leidy, 1992). While the data on how countries are changing their trade policies is increasingly available, it will be important for these studies to control adequately for the possibility that multiple jurisdictions may be changing their trade policies over identical or related products almost simultaneously.

Despite these caveats, there are policy-based reasons to motivate the importance of continued research in this area. WTO dispute-settlement rulings in particular lead one country to change its policies to comply with obligations and market access interests that other complaining WTO members have brought forward. These changes affect its economic environment as well as that of the complaining countries. However, mostly overlooked is the fact that in many instances these same WTO disputes also change the competitiveness conditions and foreign market access available to third countries that were not the original complainers but that will also have the need to adjust.

Thus far very little research or policy attention has focused on the parties in these third countries, and the benefits and costs associated with their adjustment practices. On the positive side, the MFN rule implies that many of the benefits that complaining countries achieve by winning WTO disputes and getting respondent countries to remove (non-MFN-violating) trade barriers spill over to benefit other countries by improving their terms of market access as well. On the negative side, some important and high profile WTO disputes involve the elimination of WTO-inconsistent policies that may have provided implicit preferential treatment to developing countries in politically sensitive products (for example, sugar and bananas). The elimination of such preferences is thus expected to result in a nega-

[19] For example, during the recent crisis period alone, Bown (2009a, Table 5) identifies more than 70 distinct 6-digit Harmonized System (HS) product codes with at least two different countries newly initiating trade remedy (antidumping, global safeguard, countervailing duty, or China-specific safeguard) investigations over the same code between 2007 and the first quarter of 2009.

tive adjustment impact on many vulnerable economies. Without a full accounting of the third-country adjustment implications of removal of trade barriers resulting from the WTO dispute settlement process, there is much that we do not know about the size of the true costs and benefits of the WTO system, under which such changes to national trade policies frequently occur.

Chad P. Bown is Senior Economist in the World Bank's Development Research Group, Trade and International Integration (DECTI) in Washington, DC.

BIBLIOGRAPHY

Bagwell, Kyle and Robert W Staiger (forthcoming). What Do Trade Negotiators Negotiate About? Empirical Evidence from the World Trade Organization, *American Economic Review*

—(2002). *The Economics of the World Trading System*. Cambridge, MA: The MIT Press

—(1999). An Economic Theory of GATT, *American Economic Review* 89(1): 215–48

Bernard, Andrew B, Jensen, J Bradford, Redding, Stephen and Peter K Schott (2007). Firms in International Trade, *Journal of Economic Perspectives* 21(3): 105–30

Bown, Chad P. (2009a). The Global Resort to Antidumping, Safeguards, and other Trade Remedies Amidst the Economic Crisis, in Simon J. Evenett, Bernard M. Hoekman and Olivier Cattaneo (Eds.), *Effective Crisis Response and Openness: Implications for the Trading System*, London, UK: CEPR and World Bank

—(2009b). Global Antidumping Database, [Version 5.0, July], available at www.brandeis.edu/~cbown/global_ad/

—(2009c). *Self-Enforcing Trade: Developing Countries and WTO Dispute Settlement*. Washington, DC: Brookings Institution Press

—(2009d). MFN and the Third-Party Economic Interests of Developing Countries in GATT/WTO Dispute Settlement, in Chantal Thomas and Joel P. Trachtman (Eds.), *Transcending the Ostensible: Developing countries in the WTO legal system*. Oxford, UK: Oxford University Press

—(2004a). Trade Policy under the GATT/WTO: Empirical Evidence of the Equal Treatment Rule, *Canadian Journal of Economics* 37(3): 678–720

—(2004b). How Different Are Safeguards from Antidumping? Evidence from US Trade Policies toward Steel, Brandeis University manuscript, July

Bown, Chad P and Meredith A Crowley (forthcoming). China's Export Growth and the China Safeguard: Threats to the World Trading System? *Canadian Journal of Economics*

—(2007). Trade Deflection and Trade Depression, *Journal of International Economics* 72(1): 176-201

—(2006). Policy Externalities: How US Antidumping Affects Japanese Exports to the EU, *European Journal of Political Economy* 22(3): 696–714

Bown, Chad P and Guido Porto (2009) Exporters in Developing Countries: Adjustment to Foreign Market Access after a Trade Policy Shock, Brandeis University manuscript, January

Brambilla, Irene, Guido Porto, and Alessandro Tarozzi (2008). Adjusting to Trade Policy: Evidence from US Antidumping Duties on Vietnamese Catfish, *NBER Working Paper no.* 14495, November

Broda, Christian, Nuno Limão, and David E Weinstein (2008). Optimal Tariffs and Market Power: The Evidence, *American Economic Review* 98(5): 2032–65

Bustos, Paula (forthcoming). Trade Liberalization, Exports and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinean Firms, *American Economic Review*

Chang, Won and L Alan Winters (2002). How Regional Blocs Affect Excluded Countries: The Price Effects of MERCOSUR, *American Economic Review* 92(4): 889–904

Debaere, Peter (2010). Small Fish—Big Issues: The Effect of Trade Policy on the Global Shrimp Market, *World Trade Review* 9(2): 353–374.

Durling, James P.. and Thomas J.. Prusa (2006). The Trade Effects Associated with an Antidumping Epidemic: The Hot-Rolled Steel Market, 1996-2001, . *European Journal of Political Economy* 22(3): 675–95

Edmonds, Eric and Nina Pavcnik (2006). Trade Liberalization and the Allocation of Labor between Households and Markets in a Poor Country, *Journal of International Economics* 69(2): 257–95

Erdem, Erkan and James Tybout (2004). Trade Policy and Industrial Sector Responses: Using Evolutionary Models to Interpret the Evidence, in Susan Collins and Dani Rodrik (Eds.), *Brookings Trade Forum 2003*. Washington, D.C.: The Brookings Institution

Evenett, Simon and Bernard Hoekman, Eds. (2009). *Trade Implications of Policy Responses to the Crisis*. VoxEU.org e-book

Goldberg, Pinelopi K and Nina Pavcnik (2007). Distributional Effects of Globalization in Developing Countries, *Journal of Economic Literature* 45(1): 39–82

—(2005). Trade, Inequality, and Poverty: What Do We Know? Evidence from Recent Trade Liberalization Episodes in Developing Countries, *Brookings Trade Forum 2004*. Washington, D.C.: The Brookings Institution

Hoekman, Bernard M and Michael P Leidy (1992). Cascading Contingent Protection, *European Economic Review* 36(4): 883–92

Hoekman, Bernard M and Michael M Kostecki (2009). *The Political Economy of the World Trading System: The WTO and Beyond* (3rd edition.) Oxford, UK: Oxford University Press

Lileeva, Alla and Daniel Trefler (forthcoming). Improved Access to Foreign Markets Raises Plant-Level Productivity for Some Plants, *Quarterly Journal of Economics*

Prusa, Thomas J (1997). The Trade Effects of US Antidumping Actions, in Robert C. Feenstra (ed.), *The Effects of US Trade Protection and Promotion Policies*. Chicago, IL: University of Chicago Press for NBER

—(2001). On the Spread and Impact of Anti-dumping, *Canadian Journal of Economics* 34(3): 591–611

Romalis, John (2007). NAFTA's and CUSFTA's Impact on International Trade, *The Review of Economics and Statistics* 89(3): 416–35

Staiger, Robert W and Frank A Wolak (1994). Measuring Industry-Specific Protection: Antidumping in the United States, *Brookings Papers on Economic Activity: Microeconomics* 51–118

Trefler, Daniel (2004). The Long And Short Of The Canada-U.S. Free Trade Agreement, *American Economic Review* 94(4): 870–95

Tybout, James (2000). Manufacturing Firms in Developing Countries: How Well do They do and Why? *Journal of Economic Literature* 38(1): 11–44

Viner, Jacob (1950). *The Customs Union Issue*. New York: Carnegie Endowment for International Peace

PART C

FACTORS THAT AFFECT TRADE

16

Transportation Costs and Adjustments to Trade

DAVID HUMMELS

This note discusses the interaction between transportation costs and trade shocks. The interactions in question are two-way, including how shocks affect transportation and how transportation affects the shocks or shapes their impact on national economies. Trade shocks in this context will be taken to mean any changes in the volume or composition of trade. Some of these changes will be high frequency in nature, while others will reflect long-run trends.

Transportation has traditionally either been ignored by international trade scholars, or treated reductively as an 'iceberg cost'. More recently, researchers have begun providing richer modeling of transportation in order to understand its interactions with trade better. These effects can be grouped into three broad categories of interactions: transport as a non-iceberg cost; transport as a produced service; and 'transport costs' more broadly defined to include a range of non-tariff barriers to trade. I discuss each of these in turn.

1. NON-ICEBERG COSTS

In the traditional iceberg formulation, transport is treated as an exogenous friction τ that is fixed and proportional to the value shipped, with the value added of transportation services treated as pure waste, or 'melt'. Given this treatment, transport costs shift relative prices so that the delivered price equals the origin price multiplied by the iceberg factor, $p^{*}=p(1+\tau)$, or as a ratio,

(0.1) $p^{*}/p=p(1+\tau)$.

Costs specified in this way simply introduce a wedge between origin and destination prices. When combined with CES preferences, the most common formulation, they are used to explain the distribution of purchases from domestic versus foreign sources, or the distribution across foreign sources depending on proximity.[1] In combination with scale economies in production, as in the New Geography or Home Market Effect literatures, iceberg costs create interesting feedback loops, as better (lower τ) access to foreign markets becomes a source of compar-

[1] Simple extensions allow the constant of proportionality to co-vary with simple geographical determinants such as distance between markets i and j, e.g. $\tau_{ij}=\alpha(DIST_{ij})^{\delta}$

ative advantage for firms. Beyond these two main effects, iceberg transportation costs do not interact in especially interesting ways with trade or trade shocks. Indeed, from a modeling perspective, the whole point of this specification is to introduce frictions in a manner exactly like *ad valorem* tariffs and to proceed with the rest of the analysis unimpeded.

However, a slightly more general formulation better fits facts about the shape of transportation costs and yields a host of interesting interactions. Denote the price per kilogram of a good as p (so that $1/p = weight/value$), and the shipping charge per kilogram shipped as f. If the shipping charge is independent of the goods price, the ratio of destination to origin prices is $p^{*}=p+f$, or as a ratio, $p^{*}/p=1+f/p$ Of course, the shipping charge f may be increasing in p because higher-value goods require more careful handling and a larger insurance premium. We can then write the per kilogram shipping charge as $f=p^{\beta}X$, where X represents other costs shifters such as distance, port quality, and so on. In this case we have $p^{*}=p+p^{\beta}X$, or in ratios

$$(0.2) \quad p^{*}/p = 1+(\frac{weight}{value})^{1-\beta} X$$

Unless $\beta=1$, the weight/value ratio of a product will be an important determinant of the transportation expenses incurred when trading that product. Hummels and Skiba (2004b) estimate that a 10 per cent increase in product weight/value leads to a 4 to 6 per cent increase in shipping costs measured *ad valorem*, that is, relative to the value of the good shipped.

Consider four implications for trade and trade shocks. First, compared to expression (0.1), transportation is no longer an exogenous constant but instead depends on the composition of what is shipped. For some goods like scrap metal, the price per kilogram is low (weight/value is high), and the ratio p*/p is high. That is, shipping charges drive a large wedge between the prices at the origin and destination. For computer microchips, p is very high (weight/value is very low), the ratio p*/p is close to 1, and shipping charges drive only a small wedge between prices at the origin and destination.

Second, product weight/value, which varies widely across goods, explains far more variation in *ad valorem* transportation costs than do other observables including: the distance goods are shipped, the technology with which they are shipped, the quality of port infrastructure, or the intensity of competition between carriers on a trade route. Differences across countries in the product composition (weight/value ratio) of their trade largely explain why developing countries pay nearly twice as much as developed countries for transporting goods internationally.[2]

Third, because consumers are sensitive to delivered prices, non-iceberg costs change relative demands for products. In particular, the existence of per unit transport charges raises the relative demand for high-quality goods. This is known as the Alchian–Allen effect and can be simply illustrated. Suppose I have two

[2] Hummels, Lugovskyy, and Skiba 2009

bottles of wine, high and low quality, with factory gate prices of \$20 and \$10. The relative price of the high-quality bottle at the factory gate is 2/1. When we include international shipping at \$5 a bottle, the relative delivered price falls to 5/3, that is, the price premium for the better bottle of wine falls from 100 per cent to 66 per cent. Raising international shipping costs to \$10 a bottle, pushes the relative price at the point of delivery down to 3/2, or a premium of 50 per cent. This effect significantly alters the pattern of international trade. Even within narrowly defined product categories, exporters shift the mix of goods sold toward higher-price varieties when selling to destinations for which transport costs are high. The strength of this effect is greater the larger is X in equation (0.2). It is stronger for more distant markets, for countries with poor transport infrastructure, and in periods of high oil prices.[3]

Fourth, suppose the price of the same good changes over time, due perhaps to quality upgrading or the general equilibrium effects of trade liberalization on production costs. Holding shipping charges per unit, f, fixed, product price increases lower the *ad valorem* cost imposed by transportation, while product price decreases raise the *ad valorem* cost of transport. The same is true of high frequency movements in product prices. In essence, the non-iceberg nature of transport costs acts as a kind of shock absorber, dampening the transmission of product price shocks to delivered prices.

2. TRANSPORT AS A PRODUCED SERVICE

I turn next to discussing the transportation charge, f, not as a trade friction but rather as the price of a produced service, with a somewhat narrow focus on the interactions between the production of transport services and trade. I examine input costs, economies and diseconomies of scale, and the liberalization of cargo services.

Shocks to the global demand for and supply of oil affect the price of transportation fuels which are an important component of costs. The effect on transport has been especially pronounced in the last two decades, with oil prices falling throughout much of the 1990s and then rising sharply since 2002. Data from the Air Transport Association show that airline operating costs have risen 89 per cent since 2000 and much of that increase can be ascribed to fuel cost increases (between 2002Q1 and 2008Q1 jet fuel rose from 9.9 per cent to 29.4 per cent of airline operating expenses). A simple back of the envelope calculation using the ATA data suggests that doubling fuel prices leads to a nearly 50 per cent increase in aviation costs.

Fuel price shocks also change relative prices of internationally transported goods. Recalling the Alchian–Allen effect above, fuel prices enter the X term in equation (0.2), which means that rising fuel prices shift demand toward high-value goods. In addition, different transportation modes use fuel with different intensity (in order: planes, trucks, trains, boats), which means that rising fuel

[3] Hummels and Skiba 2004b

prices shift demand toward goods using fuel efficient modes. As an example, time-sensitive products can be sourced locally and shipped via trucks or sourced globally and shipped via airplane in roughly the same time frame.[4] However, the fuel intensity of the plane is many times greater than the truck, and so its use will be much more sensitive to rising fuel prices.[5]

An important way that trade shocks interact with transport costs is through economies or diseconomies of scale. In periods of rapidly rising demand, shipping capacity becomes scarce, ports become congested, and spot shipping prices rise quickly. The reverse is true in periods of rapidly falling demand. As an example, data from Containerisation International show that the cost of shipping a standard container from East Asia to the United States fell 33 per cent from the height of the US business cycle in 1999–2000 to the low point of the US recession in 2001–02. Similarly, large bilateral trade imbalances cause ships to run fully loaded in one direction but at a fraction of capacity on the return voyage. This differential is congestion priced which leads to large differences in rates on the same trade route, depending on the direction of the flow. This effect has been especially pronounced for the United States, whose bilateral imbalances with Asia are large. Since 2000, the cost of shipping containers eastbound from Asia to the United States has been consistently two to three times higher than the cost of shipping containers westbound on the same route.

Over longer periods however, rising demand for shipping may actually lower shipping prices, especially in smaller countries with initially low trade volumes. This suggests an important secondary benefit to tariff liberalization—tariff reductions that boost trade volumes may spur ancillary reductions in shipping costs. To explain, the capacity of a modern ocean-going vessel is large relative to the quantities shipped by smaller exporting nations. As a consequence, vessels may stop in a dozen ports and in different countries to reach capacity. As trade quantities increase, it is possible to realize gains more effectively from several sources.

First, a densely traded route allows for effective use of hub and spoke shipping economies—small container vessels move quantities into a hub where containers are aggregated into much larger and faster containerships for longer hauls. Second, the frequency of port calls rises, which is highly beneficial for time-sensitive products[6]. Third, larger trade volumes allow for the introduction of specialized vessels (reefers, 'ro-ros') that are adapted to specific cargos, and larger ships that enjoy substantial cost savings relative to older, smaller models still in use. Fourth, larger trade volumes induce investment in port infrastructure and better port infrastructure is highly correlated with lower shipping costs[7]. Fifth, ris-

[4] See Evans and Harrigan, 2005; Harrigan and Venables, 2006; and Hummels and Schaur 2010 for a discussion of timeliness in trade and the substitution possibilities available to firms interested in timely delivery.

[5] See Hummels et al 2009 for a calculation of fuel-use intensities for different transport modes described in tonnes per kilometer carried. Shifting from trucks to planes raises the fuel use per tonnes per kilometer by a factor of 10 to 20, while shifting from local to global sources could raise the kilometers shipped enormously.

[6] Hummels, Skiba (2004a)

[7] Limao and Venables (2001); Clark et al. (2004); Fink et al. (2002); and Haveman et al. 2008

ing trade volumes promote entry into shipping markets and pro-competitive effects on shipping markups, which can substantially lower shipping prices.

On this last point, Hummels, Lugovskyy and Skiba (2009) systematically examine the effect of market power in shipping. They report that in 2006 one in six importer–exporter pairs was served by a single liner service, and over half were served by three or fewer. In general, large countries with higher trade volumes enjoy a greater number of shipping firms competing for their trade. To explain these facts, Hummels, Lugovskyy and Skiba (2009) model the shipping industry as a Cournot oligopoly and determine optimal shipping markups as a function of the number of carriers and the elasticity of transportation demand faced by carriers. A key insight of the model is that transportation is not consumed directly; instead, carriers face transportation demand derived indirectly from import demand. This implies that the impact of an increased shipping markup on the demand for transportation depends on the share of transportation costs in the delivered price of the good, and elasticity of import demand. The first effect is closely related to the wine bottle example above. For expensive goods, the marginal cost of shipping represents a smaller fraction of the delivered product price, which enables cargo carriers to charge larger markups without a large demand response.

The second effect relates to the responsiveness of trade volumes to increased prices. Suppose we have two goods for which shipping prices, including markups, will yield an equal 5 per cent increase in the delivered price of the good. The first good is a differentiated product with import demand elasticity equal to 1.1. Here, a markup that yields a 5 per cent increase in delivered price reduces traded quantities, and therefore demand for transportation services, by only 5.5 per cent. The second good is a highly substitutable commodity and faces an import demand elasticity of 10. Here, the markup raises prices by 5 per cent but lowers quantities traded and demand for transportation services by 50 per cent! In the latter case the identical markup reduces import (and therefore transportation) demand to a much greater degree, limiting the carrier's optimal markup.

The implication is that the market power of shipping firms is extremely high when they are moving goods with inelastic import demand, and when marginal costs of shipping comprise a small fraction of the overall delivered price. In these cases, it is easy to generate examples where optimal markups could be an order of magnitude higher than the marginal cost of shipping. However, entry by rival liner companies can very quickly erode this pricing power. This suggests an especially important role for government policy in promoting competition in the transportation industries. Not only is transport an input into merchandise trade, but trade in transportation services could yield substantial gains for the countries involved. Such policy might take two forms. The first is regulating monopoly in an industry that may be ripe for collusive behavior. As an example, the European Union Competitiveness Council recently concluded that cartelization in maritime shipping had led to a less competitive shipping market and higher shipping prices. The Council repealed a long-standing block exemption to its competition laws that had allowed carriers serving the European Union to collude in setting prices and market shares.

The second form of policy is simply allowing entry into national markets by international firms. Transportation services worldwide are tightly regulated both in terms of which firms can enter, and the service quality and (or) safety they must provide. Recent efforts to liberalize air cargo services via 'Open Skies Agreements' that allow open competition to foreign carriers have yielded substantial reductions in air freight rates.[8] However, there is scope for significant further liberalization, primarily because there is no internationally integrated market for land-transport services. Roughly one-quarter of international trade occurs between countries sharing a land border, and these flows are dominated by rail and truck services that are generally national, or at best regional, in scope. And when ocean or air carriers are employed to leap oceans or long distances, the door-to-door transport chain involves land-based modes at both ends and these costs likely represent the majority of the total transport bill. Efforts to integrate these markets have been halting at best, even in cases where liberalization has already been agreed to (compare with the North American trucking industry for instance).

3. 'TRANSPORTATION COSTS' BROADLY DEFINED

Many scholars use the phrase 'transportation costs' more broadly than I have discussed here, including in the phrase any non-tariff cost of trade. For example, information about foreign markets is almost certainly an important cost of trade, especially trade in complex and differentiated goods. Similarly, marketing and distribution costs, and product adaptation to local tastes and regulatory requirements are important costs that segment markets for all but the simplest of commodities. Space constraints preclude a systematic treatment of this broader meaning, but even focusing narrowly on transportation itself there are aspects of transport services that are qualitatively richer than simple expenditures on freight.

One important example is timeliness. A standard contract for shipping services specifies from where and to where cargos will be transported, as well as a delivery schedule. More rapid transport can be purchased at a premium—priority handling in ports, faster ships or more direct routing, or upgrading to the use of air cargo. With speed comes flexibility as well—the ability to reach foreign markets in a day means that purchasing decisions can be put off until after uncertainty is resolved. This allows firms to adjust to shocks in real time.

How quantitatively important is speed and flexibility? Worldwide, air cargo has grown 8.3 per cent per year since 1975, much faster than trade as a whole.[9] More directly, Hummels and Schaur (2010) combine data on the premium paid for air shipment along with the greater transit time needed for ocean shipment estimates, the implicit value that firms attach to timeliness. They find that firms are willing to pay just under 1 per cent of the value of the good for each day saved in transit.

[8] Micco and Serebrisky 2006
[9] Hummels 2007

But what exactly is driving the rapid growth in air cargo? A few factors closely related to the changing nature of trade and trade shocks seem especially relevant. These are: a fall in the weight of trade, rising incomes, vertical specialization and (or) fragmentation, testing new markets, and trade between geographically remote locations.

Above I discussed how high-value goods enjoy a lower *ad valorem* transport costs. The same logic can be employed to show that the premium charged for air-shipping, measured as a proportion of the final price, will be lower when goods have higher prices. This means that a compositional shift in trade toward high-value manufactures makes air transport feasible for a larger set of goods.

Rising incomes worldwide affect demand for air transport in three ways. First, high-income households buy higher-quality goods which have higher prices and therefore a lower *ad valorem* transportation cost. Second, as consumers grow richer, so does their willingness to pay for precise product characteristics. That in turn puts pressure on manufactures to produce to those specifications, and to be rapidly adaptable to changing tastes. Third, as evidenced by the success of online retailers like Amazon, delivery speed is itself an important characteristic of product quality and will be in greater demand as income grows.

Much of recent trade growth has occurred through the fragmentation of international production processes, also known as vertical specialization.[10] Multistage production may be especially sensitive to lags and variability in timely delivery, and both are reduced by using airplanes. Of course, airplanes move people in addition to cargo. Multinational firms with foreign production plants rely heavily on the ability to fly executives and engineers for consultations with their foreign counterparts.[11]

Airplanes are especially useful for firms who are expanding trade by selling new goods for the first time. Consider a stylized description of export expansion. Firms begin producing for the local market, slowly expand sales within their own country, and some fraction of firms gradually expand sales abroad. When serving new markets, firms face uncertainty about demand, quantities sold are likely to be very low initially, and most trading relationships fail in a few years. All of these characteristics, initially small quantities of uncertain demand in distant markets, are precisely the characteristics that make air shipping particularly attractive.[12]

4. CONCLUSION

There is a central idea that runs through all the preceding analysis. Transportation costs are not an exogenous friction or drag on trade. Rather, they are en-

[10] Hummels, Ishii, and Yi (2001)

[11] Cristea (2009) and Poole (2010) provide evidence showing a linkage between business travel and the ability to export.

[12] Aizenman (2003) and Hummels and Schaur (2009) examine the use of airplanes in hedging demand volatility. Evans and Harrigan (2005) and Harrigan and Venables (2006) discuss the importance of demand volatility in determining comparative advantage and industrial agglomerations.

dogenous to how production is organized and what is traded. Changes in product composition and product prices alter the *ad valorem* impact of transport. Changes in trade volumes and trade policies affect the organization of transportation as a produced service. And changes in the organization of production alter the qualitative characteristics of transportation services demanded.

David Hummels is Professor of Economics at the Krannert School of Management, Purdue University.

BIBLIOGRAPHY

Aizenman, Joshua (2004). Endogenous Pricing to Market and Financing Cost. *Journal of Monetary Economics* 51(4), 691–712

Clark, X, Dollar, D, and A Micco (2004) Port Efficiency, Maritime Transport Costs and Bilateral Trade. Journal of Development Economics vol 75 (2) pp 417–450.

Cristea, Anca (2008). Information as an Input into International Trade mimeo, Purdue University

Evans, Carolyn and Harrigan James (2005). Distance, Time, and Specialization *American Economic Review,* 95(1), 292–313

Fink, A Matoo and I C Neagu, 2002. Trade in international maritime services: how much does policy matter?, *World Bank Economic Review* 16 (1), 81–108

Harrigan, James and Anthony Venables (2006). Timeliness and Agglomeration *Journal of Urban Economics* 59(2), 300–316

Haveman, Jon, Ardelean, Adina, and Christopher Thornberg (2008). Trade Infrastructure and Trade Costs: A study of select Asian Ports, in Douglas H Brooks and David Hummels, (Eds.) *Infrastructure's Role in Lowering Asia's Trade Costs: Building for Trade*. Edward Elgar Publishers

Hummels, David (2007). Transportation Costs and International Trade In the Second Era of Globalization, *Journal of Economic Perspectives* 21 pp. 131–54

Hummels, D. Avetisyan, M, Cristea, A and Puzzello, L) (2009). How Further Trade Liberalization would Change Greenhouse Gas Emissions from International Freight Transport, mimeo, Purdue University.

Hummels, David, Ishii, Jun, and Kei-Mu Yi, (2001). The Nature and Growth of Vertical Specialization in World Trade, *Journal of International Economics*, 54 75–96

Hummels, David, Volodymyr Lugovskyy, Alexandre Skiba (2009). The Trade Reducing Effects of Market Power in International Shipping*, Journal of Development Economics* 89(1), 84–97.

Hummels, David and Georg Schaur (2009), Hedging Price Volatility Using Fast Transport, NBER Working Paper 15154

Hummels, David and Georg Schaur (2010). Time as a Trade Barrier, mimeo, Purdue University

Hummels David, and Alexandre Skiba (2004a). A virtuous circle? Regional tariff liberalization and scale economies in transport, in: Antoni Estevadeordal, Dani Rodrik, Alan M. Taylor, Andrés Velasco (Eds.), *FTAA and Beyond: Prospects for Integration in the Americas*. Harvard University Press.

Hummels, David and Alexandre Skiba (2004b) Shipping the Good Apples Out: An empirical confirmation of the Alchian-Allen conjecture. *Journal of Political Economy* 112 (2004) 1384–1402

Limão, Nuno and Venables, Anthony 2001 Infrastructure, geographical disadvantage, transport costs, and trade, *World Bank Economic Review.* 15 (2001) (3), 451–479.

Micco, Alejandro and Tomas Serebrisky (2006). Competition Regimes and Air Transport Costs: The effects of open skies agreements. *Journal of International Ec*onomics 70 (2006) 25–51

Poole, Jennifer, (2010) "Business Travel as an Input into International Trade" mimeo, University of California Santa Cruz.

17

The Duration of Trade Relationships

TIBOR BESEDEŠ AND THOMAS J. PRUSA

1. INTRODUCTION

When countries trade, how long do their trade relationships last? Are they exchanging products over long or short periods of time? To answer these questions, Besedeš and Prusa (2006a; 2006b) and Besedeš (2008) have studied the duration of trade relationships. These studies have found that international trade relationships are far more fragile than previously thought. The median duration of exporting a product to the United States is very short, on the order of two to four years. More recently, Besedeš and Prusa have also shown that brief trade duration holds for developed and developing countries (Besedeš and Prusa, 2007).

The finding that most trade relationships are brief is important given that trade theories suggest that most relationships will be long-lived. Under the factor proportions theory, trade is based on factor endowment differences, and since endowments change gradually trade patterns are likewise expected to evolve slowly. Once a country develops a comparative advantage in a particular product, that advantage should last. This suggests that once two countries begin to trade a particular product, the relationship is likely to persist, as endowments are rarely subject to large shocks. Similarly, the continuum of goods Ricardian trade model suggests that most of the product dynamics will be confined to borderline or marginal goods. Most products are located far away from the margin and hence will be regularly exported (or imported). Only the goods near the margin should display fragility.

Models of trade dynamics such as Vernon's (1966) seminal product cycle theory also suggest that trade relationships will be long-lived. Technological leaders develop and export a product until others learn how to manufacture it and enter the market. As technology becomes more standardized, other countries will begin to produce and export the product. If follower countries have relatively low labor costs, they will eventually take over the market and push out the leaders. Product cycle models imply a fairly predictable pattern of trade dynamics where dynamics evolve either slowly or in a logical progression from developed to developing countries. Melitz's (2003) seminal paper also suggests that relationships will be relatively long-lived; once a firm makes its sunk cost investment to export, the ongoing cost of servicing a foreign market is modest; said differently, in the Melitz formulation once relationships are established they will tend to be robust.

Our results indicate that more is happening at the micro-level than suggested by the dominant theories of trade. Not only is there a remarkable amount of entry and exit in the US import market, but the period of time a country is 'in' a given product market is often fleeting. The vast majority of exporting relationships are only active for a short period of time. More than half of all trade relationships are observed for a single year and approximately 80 per cent are observed for less than five years. Relatively few relationships survive 10 years, but those that do account for a disproportionate amount of trade. The results are remarkably robust to a large number of alternative empirical specifications and data sets.

The results suggest that entering an export market—growth along the extensive margin—is no guarantee of long-term export presence. Most products sold by most countries to most destination markets will not be exported within a few years. These findings should spur trade economists to re-examine export incentives and should serve as a warning to exporters that breaking into a market is no guarantee that they will be servicing the market for more than a fleeting moment.

The remainder of the paper is organized as follows. In the next section we will discuss the data used and discuss how we define trade relationships, spells of service, duration, and so on. In Section 3 we present our main findings. In Section 4 we discuss the extent to which a matching model of trade can explain the duration phenomenon. In Section 5 we briefly mention recent work that applies Besedeš and Prusa's (date) approach to other datasets and which confirms our findings.

2. TRADE RELATIONSHIPS AND SPELLS OF SERVICE

Firm-level export and import transaction data are not widely available, so instead the empirical analysis is based on bilateral 'tariff line' import statistics. Using highly disaggregated data for duration analysis is imperative. The more aggregated the data, the more the analysis identifies industry trends rather than competitive dynamics at the product level. That country *c* in industry *j* has a long duration may tell us little about duration of commodity trade or underlying trade dynamics.

Line item data are the most disaggregated trade data widely available.[1] For most of our discussion we will focus on the United States as the destination market of interest. One limitation with using highly disaggregated trade data is that the product classification system was changed in 1989, so our product level analysis often appears into two distinct subsets: pre- and post-1989.[2] From 1972 through 1988 import products were classified according to the 7-digit Tariff Schedule of the United States (TS). Since 1989 imports have been classified ac-

[1] US line item data has been compiled by Feenstra (1996) and was later augmented by Feenstra, Romalis, and Schott (2002). For other countries the UN's Comtrade database provides line-item bilateral trade data—http://comtrade.un.org/db/.

[2] Prior to 1989, countries had their own line-item classification systems.

cording to the 10-digit Harmonized System (HS).[3] The United States is the only country we know of where line-item trade data are consistently available prior to 1989. We later use the more aggregated SITC bilateral trade data to create longer time horizons and also to check robustness of our disaggregated results.

Table 17.1: *Summary Statistics*[4]

	Observed Spell Length (years)		Estimated KM Survival Rate			Total Number	Total Number
1972–1988	Mean	Median	1 year	4 year	12 year	of Spells	of Product Codes
Benchmark data							
TS7	2.7	1	0.67	0.49	0.42	693,963	22,950
Industry-level aggregation							
SITC5	3.9	1	0.58	0.37	0.31	157,441	1,682
SITC4	4.2	2	0.58	0.38	0.33	98,035	827
SITC3	4.7	2	0.60	0.40	0.35	43,480	253
SITC2	5.5	2	0.65	0.46	0.41	15,257	69
SITC1	8.4	5	0.78	0.64	0.60	2,445	10
TS7 Alternatives							
Modified Censoring	2.7	1	0.55	0.28	0.19	693,963	22,950
First Spell	2.9	1	0.70	0.57	0.53	495,763	22,950
One Spell Only	3.2	1	0.74	0.65	0.63	365,491	22,950
Gap Adjusted	3.3	2	0.73	0.58	0.49	593,450	22,950
1989–2001							
Benchmark data							
HS10	3.1	1	0.66	0.48	0.43	918,236	22,782
Industry-level aggregation							
SITC5	4.1	2	0.64	0.46	0.42	156,110	1,664
SITC4	4.4	2	0.65	0.48	0.45	92,566	768
SITC3	4.7	2	0.67	0.51	0.48	40,068	237
SITC2	5.3	2	0.71	0.56	0.53	14,373	68
SITC1	7.8	10	0.85	0.73	0.72	2,292	10
HS10 Alternatives							
Modified Censoring	3.1	1	0.61	0.39	0.33	918,236	22,782
First Spell	3.5	1	0.69	0.55	0.52	620,177	22,782
One Spell Only	4.2	2	0.72	0.63	0.61	415,851	22,782
Gap Adjusted	3.9	1	0.72	0.58	0.51	768,048	22,782

[3] Many TS products are mapped into multiple HS codes and vice versa making it impossible to create a long product-level panel. As a robustness check, we also aggregate from the product level to the industry level data where we use the Standard International Trade Classification (SITC) industry codes to define trade relationships.

[4] Extract from Table 2 in Besedeš and Prusa (2006a).

We define a trading relationship x_{ei} as country e exporting a good x to a particular destination market, i for a continuous period of time. Our interest is to study the length of time until the relationship ceases to be active, an event we will refer to as a 'failure.' Calendar time is not as important as analysis time, which measures the length of time of continuous exporting. For each product and country pair we use the annual data to create spell data. For example, if country e exports good x to destination country i from 1976 to 1980 then relationship x_{ei} has a spell length of five years.

For each product we create a panel of countries which export the product to the United States As shown in Table 17.1 there are 693,963 (918,236) observed spells of service at the TS (HS) level. Both TS and HS datasets have a median spell length of one year and a mean spell length of about three years. We note that Table 17.1 also includes summary statistics for trade relationships created from the more aggregated SITC industry level data. In the top panel of the table we see that in the 1972–1988 period there are 157,441 observations at the 5-digit SITC level, 43,480 observations at the 3-digit level and just 2,445 observations at the 1-digit level. As expected, aggregating the data diminishes the ability to observe entry and exit—for the 1972–1988 period the mean spell length increases from 2.7 years in the 7-digit TS data to 3.9 years in the 5-digit SITC data to 8.4 years in the 1-digit SITC data. (Table 17.1 also includes some robustness results which we will discuss below.)

One complicating factor is that some trade relationships reoccur, exhibiting what we will refer to as multiple spells of service. A country will service the market, exit, then re-enter the market, and then almost always exit again. Approximately 30 per cent of relationships experience multiple spells of service in the disaggregated product level data. About two-thirds of relationships with multiple spells experience just two spells; less than 10 per cent have more than three spells. We begin by treating multiple spells as independent. While the assumption is made primarily in the interest of simplicity, we spend a great deal of time on alternative approaches to handling the issue and find the results are consistent across all methods.[5]

Once we begin to think of data in terms of spells it becomes apparent that we need to account for censoring, by which we mean losses from the sample before the final outcome is observed. Censoring is a result of two phenomena. First, we do not have information on trade relationships for the years before the beginning and after the end of the sample. All relationships active in either the first year or last year of our data will be classified as censored, as we are not certain how long they truly were active. Consider, for example, a relationship that starts in 2000 and is also observed in 2001 (the last year of our data). That relationship may indeed fail after two years but we simply are not sure. By classifying it as censored we interpret the duration of at least two years. Second, product definitions for the tariff codes are revised on an ongoing basis. Sometimes a single

[5] In Besedeš and Prusa (2006a) we discuss why the independence assumption is a reasonable starting place.

code is split into multiple codes and at other times multiple codes are refined into fewer codes. Unfortunately, there is no information to allow us to map old product codes systematically into new ones. We can observe when a code is changed but we cannot observe if the trade associated with that code ceased (a true failure) or is now classified under a new code. For the bulk of our analysis we choose to be cautious and classify all such changes as censored. Reclassified relationships are interpreted as having duration of at least x years (where x is the number of years when trade in the original code was observed). We are sure that some spells are classified as being censored when they were truly associated with either entry or exit. As a result, our benchmark results will overstate the true duration of a typical trading relationship.[6] We will also report alternative censoring approaches.

Once we have expressed the annual trade data into spell data and applied the censoring definitions, we can use well-developed survival analysis statistical methods to analyze the duration of trade. The Kaplan–Meier product limit estimator will be used to non-parametrically characterize the survivor function. Loosely speaking, the Kaplan–Meier estimator gives the fraction of spells that will survive at least t years. An important advantage of the Kaplan–Meier curve is that it is robust to censoring and uses information from both censored and non-censored observations.

We also use the Cox proportional hazards model to derive semi-parametric estimates for the factors determining survival. The Cox model is computationally convenient and allows us to consider survival as consisting of two parts: the underlying hazard function, describing how hazard changes over time, and the effect parameters, describing how hazard relates to other factors. A particular advantage of the Cox model is that the baseline hazard is left unspecified and is not estimated.

3. EMPIRICAL RESULTS

We begin by examining the benchmark 7-digit TS data and report our findings in Table 17.1. We report the 1-, 4-, and 12-year survival rates for benchmark product level data and also for the industry data. The table conveys several important lessons about duration of trade.

First and foremost, a very large fraction of relationships fail after only a year or two. For the benchmark TS data, only 67 per cent of relationships survive one year; 49 per cent survive four years; and 42 per cent survive 12 years. An almost identical survival experience is found in HS data. In fact, as we discuss below, a qualitatively similar experience is seen across all estimates. The evidence is quite clear: the typical US trade relationship is very short-lived.

The second important finding is the sharp decline of the risk of failure. It is quite high in the early years, but then rapidly falls once a trade relationship survives a threshold duration. As shown, a large number of relationships fail dur-

[6] Only the first type of censoring is present for SITC data.

ing the first four years, especially in the first year when the hazard rate is 33 per cent for TS data. However, after about four or five years failure becomes a lot less common. The hazard rate between year one and year five is an additional 30 per cent and just 12 per cent for the remaining twelve years.[7] The results indicate negative duration dependence—the conditional probability of failure decreases as duration increases. There is a type of a threshold effect. Once a relationship is established and has survived the first few years it is likely to survive a long time.

A picture of the estimated overall survival function is given in Figure 17.1—the upper figure is based on TS data and the lower figure is based on HS data. In both cases the survival function is downward sloping with a decreasing slope. We note that the Kaplan–Meier estimated probability of exporting a product for more than 17 years is 41 per cent. Almost exactly the same long-run survival rate is found in HS data. Said differently, taking into account both types of censoring, about 40 per cent of relationships will survive more than 17 years. This is noteworthy for at least two reasons. First, as discussed above, a remarkably large number of relationships fail within the first few years of service; only about half of all relationships will survive the four years. But, after the initial 'shake-out' the hazard rate falls dramatically. Second, a simple look at data reveals that less than two per cent of all trade relationships span the entire sample; that is, less than two per cent are present every year from 1972 to 1988 (TS data) or from 1998 to 2001 (HS data). From this vantage point, a 40 per cent long-run survival is superb. The explanation for the seemingly inconsistent results is the prevalence of censoring at the product level. Many relationships observed to the end are censored and are not classified as failures in benchmark results.

The impact of censoring due to product code changes can be identified if we estimate the Kaplan–Meier survival function using a modified censoring approach where we interpret all changes and reclassifications in TS codes as starts and failures (that is, we ignore the second type of censoring). This alternative approach leads to much more entry and exit, and as a result duration is significantly shorter than in the benchmark case: the median duration falls to just two years as compared with four years, while the 75th percentile is just six years. The probability of exporting a product for more than 17 years under modified censoring is only 18 per cent—less than half the benchmark—but still considerably higher than the observed two per cent of trade relationships that span the entire sample. The first type of censoring accounts for the difference.

Besedeš and Prusa (2006a) also find that duration varies by source country and region with short-lived relationships characterizing trade by most countries. Short relationships are prevalent for both OECD and non-OECD countries although OECD trade relationships exhibit systematically longer survival.

There is compelling evidence that the results are robust to aggregation. We calculated spells of service using SITC industry (revision 2) definitions ranging from the 5-digit to the 1-digit level. Aggregation dramatically decreases the number

[7] In TS data more than 50 per cent of observed spells of service fail within the first four years, but over the next thirteen years about 7 per cent of spells fail.

of observed relationships, but the impact on duration is much more modest. The median survival time for SITC data is only two to three years until we aggregate to the 1– or 2–digit level. Within the SITC classification, higher levels of aggregation are associated with longer survival times (Table 17.1), but the impact on median survival time is modest until data are highly aggregated.

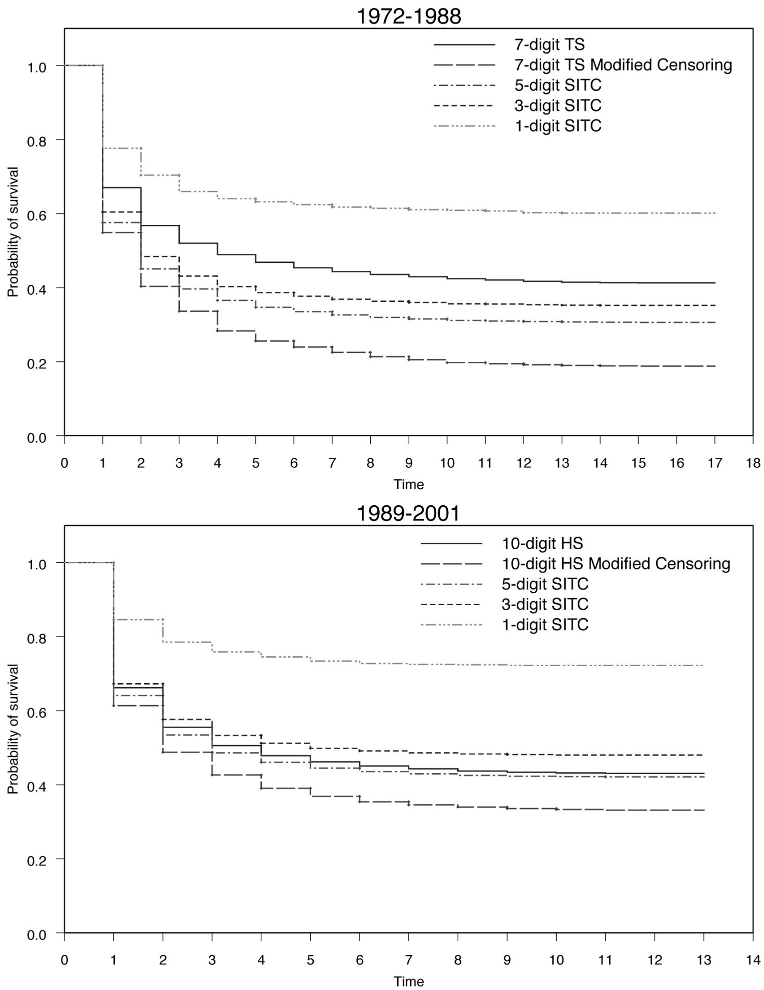

Figure 17.1: *Survival Functions for Product and Industry Level Data*[8]

[8] Based on Figure 3 in Besedeš and Prusa (2006a)

The brevity of duration times for SITC data is surprising. The other surprising result is that as we aggregate from the product level to the SITC industry level the estimated probability of survival decreases. This paradoxical result is related to the unique censoring problems (numerous product code changes) only present at the product level. If we compare the SITC industry results to the modified censoring results using the product level data we see that that aggregation works as expected (Figure 17.1).

Two important lessons emerge from the SITC analysis. First, our benchmark censoring approach is overly cautious; we classify too many relationships as censored when they actually are failures. Second, SITC data confirm that short duration is not a result of overly fine parsing of the trade data. The aggregation exercise confirms that the main finding is not an anomaly. Most trade relationships are short-lived.

We considered several alternative approaches toward the issue of multiple spells. First, we simply limit the analysis to relationships with a single spell only. We find very little difference between distributions for single spell and benchmark data, especially for TS data. The estimated survival function for single spell data has a similar pattern as benchmark data: high hazard in the first few years followed by a leveling off of the survival function. We do find that the single spell data have significantly higher survival than the benchmark results, but we find that most of the difference is explained by the greater fraction of relationships that are censored in the single-spell data. When we re-estimate single spell data using the modified censoring approach we find the median survival time is now three years as compared with two years in the benchmark data. We also explored limiting the analysis to first spells—relationships with just one spell and the first spell of relationships with multiple spells. The results are generally similar to the single spell results and are available upon request.

Second, we considered the possibility that some of the reported multiple spells are due to a measurement error. Specifically, if the time between spells is short, it may be that the gap is mis-measured and interpreting the initial spell as 'failing' is inappropriate. It may be more appropriate to interpret the two spells as one longer spell. To allow for such misreporting, we assume a one-year gap between spells is an error, merge individual spells, and adjust spell length accordingly. Gaps of two or more years are assumed to be accurate and no change is made. In comparison with benchmark data, the average spell length is less than a year longer. The 1-, 4-, and 12-year survival rates in gap-adjusted data are about 7 to 9 percentage points higher than in benchmark data.

4. DOES SHORT DURATION IMPLY POOR MATCHES?

While it might appear on average that bilateral trade patterns are stable, a closer look at individual product trade patterns reveals that trade is fraught with failure—about half of all relationships fail shortly after they get started. The results suggest that relationship-specific investments might be important. To explore this possibility, we apply results from the Rauch and Watson (2003) matching model to trade duration data.

Details of the model can be found in Besedeš (2008) and Besedeš and Prusa (2006b); we briefly sketch the idea here. The model begins with the realistic assumption that trade between parties does not just happen by chance but rather begins with a search—a domestic buyer searches for a foreign supplier. After paying a search cost and being matched with a foreign supplier the buyer immediately observes the supplier's efficiency. The buyer cannot immediately ascertain, however, whether the foreign supplier will be successful in fulfilling a large order. If the supplier turns out to be unreliable, relationship-specific fixed cost investments are lost and the buyer must search again. Because of the risk of losing the lump-sum investment, the buyer might forestall making the investment and instead just make several small-volume purchases in order to learn about the supplier's reliability. If the supplier proves to be reliable, the buyer makes the investment necessary for a large order.

The model implies there are three possible actions for the buyer who has just been matched with a foreign supplier: start big (which means the relationship-specific investment was made), start small (which means sampling in order to determine the quality of the match), or reject the supplier. Besedeš (2008) identifies five implications of the Rauch and Watson (date) model as applied to formation and duration of US import trade: (1) some relationships will start with small initial order while others will start with larger ones, with larger ones enjoying an advantage in the form of a longer duration; (2) higher supplier reliability will result in a larger initial order and longer lasting relationships; (3) lower search costs increase initial order and duration; (4) a relationship is most likely to fail in its early stage; and (5) a small fraction of relationships will end with a buyer switching to a new supplier. Besedeš (2008) studies these implications using data on US imports from developing as well as developed countries. Rauch and Watson (2003) developed their model with developed country buyers searching for developing country suppliers. Besedeš (2008) shows that many features of those relationships hold for those between developing countries as well.

Since the model is silent on what constitutes a small initial order, Besedeš (2008) divides relationships into five groups based on initial order: (1) under \$10,000; (2) between \$10,000 and \$50,000; (3) between \$50,000 and \$100,000; (4) between \$100,000 and \$1,000,000; and (5) those above \$1,000,000. More than a half of all US import relationships start under \$10,000, while only four per cent commence with more than \$1,000,000 indicating that many import relationships commence in a 'testing the water' phase. Estimated Kaplan–Meier survival functions and corresponding hazard functions support the model's implications as seen in Figure 17.2. Relationships starting with larger initial orders exhibit consistently higher survival probabilities. Regardless of initial size, hazard rates for all relationships are the highest in early years and continuously decline. However, while they approach zero as relationships mature they never fully decline to zero indicating that some mature and successful relationships end when buyers switch to new suppliers.

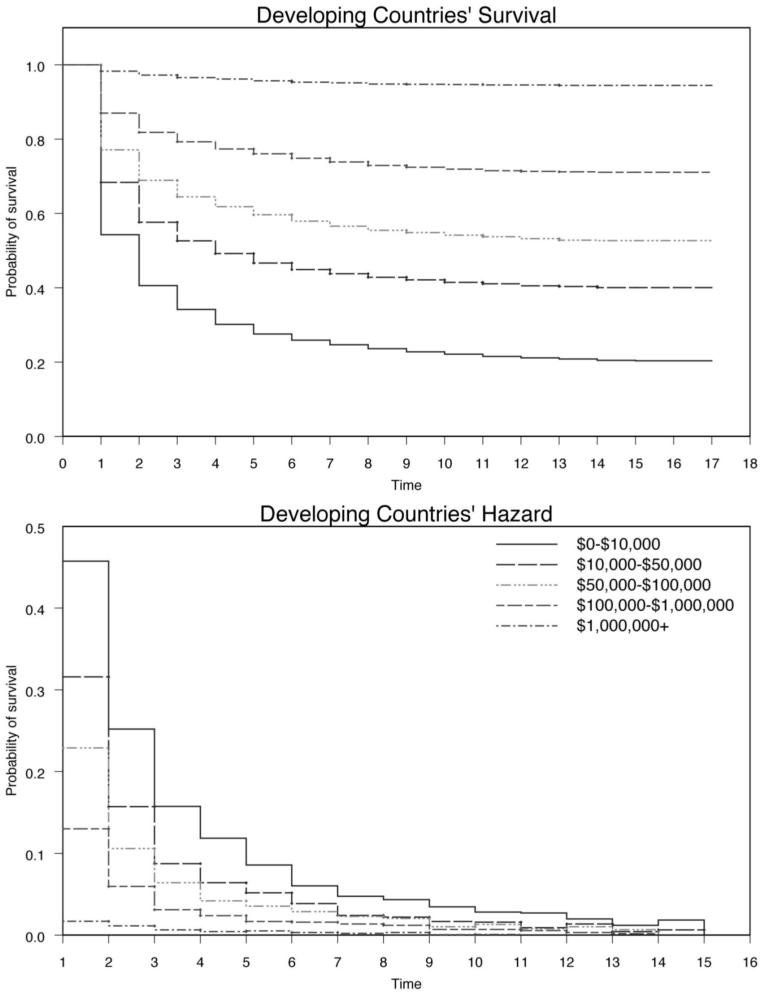

Figure 17.2: *Survival and Hazard Functions by Initial Size*[9]

Besedeš (2008) estimates the Cox proportional hazard model to examine the role played by supplier reliability and search costs as well as other factors. Supplier reliability is proxied by per capita GDP and a multiple-spell dummy. Search costs are proxied by distance, common language, contiguity, and the number of potential suppliers. Other explanatory variables include GDP, the percentage change in the real exchange rate, *ad valorem* transportation costs, an intermediate goods dummy, an agricultural goods dummy, the level of first year imports,

[9] Based on Figure 1 in Besedeš (2008)

and dummies for initial order size. Results are presented in Table 17.2. As is common in the survival literature, we present results in terms of hazard ratios. An estimated hazard ratio less (greater) than 1 implies the variable lowers (raises) the hazard rate. A ratio equal to 1 implies no impact on the hazard rate. Results indicate that higher supplier reliability and lower search costs decrease the hazard. Results also support the nonparametric estimates: as the size of the initial order increases, the hazard decreases. All else equal, relative to the smallest starting relationships, those starting with $10,000 to $50,000 have about a 30 per cent lower hazard, while the largest starting relationships have a roughly 92 per cent lower hazard.

Table 17.2: *Cox Proportional Hazard Estimates for 1972–1988, 7-digit TSUSA Data*[10]

		Developing Countries	Developed Countries
Distance	(unit = 1,000 kilometers)	1.015	0.964
Language dummy		0.970	0.811
Contiguous with USA		0.719	
Number of potential product suppliers		0.994	0.986
GDP per capita	(1995 US$, unit = $1000)	0.983	0.992
Multiple spell dummy		1.314	1.841
GDP	(1995 US$, unit = $100bil)	0.911	0.953
%Δ relative real exchange rate	(unit = 10%)	0.952	0.829
Ad-valorem transportation cost	(unit = 10%)	1.011	1.038
Intermediate goods		1.108	1.001*
Agricultural goods		0.895	1.030*
First year imports	(unit = millions $1987)	0.943*	0.969*
First year imports between $10,000 and $50,0000.692		0.636	
First year imports between $50,000 and $100,000		0.513	0.444
First year imports between $100,000 and $1,000,000		0.292	0.267
First year imports above $1,000,000		0.078	0.081
Observations		440,852	705,022
No. Subjects		193,855	230,382

Stratified by regions and 1-digit SITC industries
Note: * denotes estimates not significant at the 1% level

Besedeš and Prusa (2006b) examine the implications of the Rauch and Watson (2003) model for trade in homogeneous and differentiated products. They examine three implications: all else equal, (1) relationships starting with large orders will have longer duration; (2) a decrease in investment costs increases the probability that a relationship starts large; and (3) a decrease in search costs increases the likelihood that the buyer will opt to switch to a new supplier.

Homogenous goods (which are sold on organized markets) minimize the search cost the buyer is required to pay in order to find an appropriate supplier. Differentiated goods are not sold on organized markets and search costs will be considerably higher as the buyer has to go out and find an appropriate supplier. Likewise, it is reasonable to expect the relationship-specific investment will be

[10] Extract from Table 3 in Besedeš (2008).

smaller for homogeneous goods. These goods are standardized products which do not differ significantly across suppliers. Differentiated goods, with their multitudes of differences across many dimensions, will require the buyer to make larger relationship-specific investments. If one assumes that differentiated goods have higher search costs and require lower supplier-specific investments than homogeneous goods, then the model implies that holding initial purchase size constant, duration of relationships involving differentiated goods should be longer than those involving homogeneous goods. The model also implies that for each product type, duration of relationships starting with large orders should be longer than those starting with small orders.

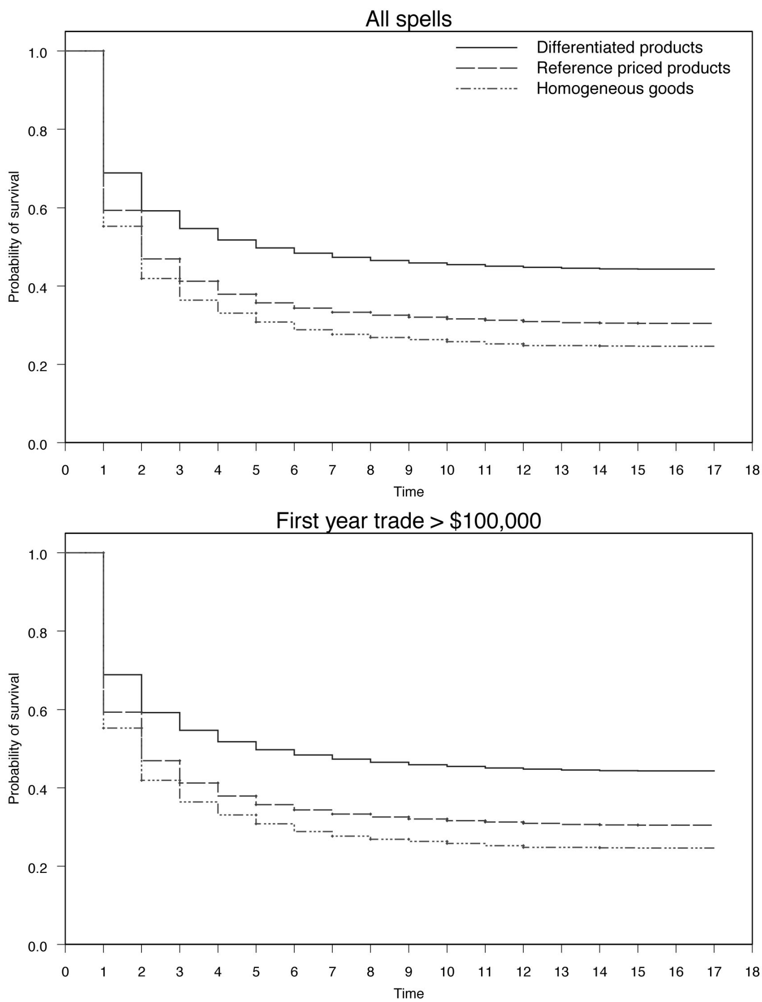

Figure 17.3: *Survival Functions by Type of Good*[11]

[11] Based on Figure 1 in Besedeš and Prusa (2006b)

We follow Rauch (1999) and classify commodities into three categories: homogeneous, reference priced, and differentiated. Rauch classified products traded on an organized exchange as homogeneous goods. Products not sold on exchanges but whose benchmark prices exist were classified as reference priced; all other products were deemed differentiated.

We begin by examining nonparametric Kaplan–Meier estimates of survival functions across product types. Estimates are graphed in Figure 17.3. As seen, median survival times are extraordinarily short: five years for differentiated products and two years for reference priced and homogeneous goods. Half of the trade relationships involving reference priced and homogeneous goods fail during the first two years. We report the nonparametric Kaplan–Meier estimates of survival functions across product types in Table 17.3. As predicted by the model, differentiated products dominate the other product types in their survival rates, at any stage of a relationship. In year one, 69 per cent of relationships involving differentiated goods survive to year two, while only 55 and 59 per cent of relationships involving homogeneous and reference priced goods do so. By year four, these rates decline to 52 per cent for differentiated and 33 per cent for homogeneous goods. Between years four and 12 survival rates are stable declining by just 7 percentage points for each product type. The differences in survival across product types are statistically significant. Similar results are found for the HS data (lower part of the table).

Table 17.3: *Kaplan-Meier Survival Rates*[12]

	Differentiated Products			Reference Priced Products			Homogeneous Goods		
Data	Year 1	Year 4	Year 12	Year 1	Year 4	Year 12	Year 1	Year 4	Year 12
1972-1988 (7-digit TSUSA)									
Benchmark	0.69	0.52	0.45	0.59	0.38	0.31	0.55	0.33	0.25
Obs>$100,000	0.92	0.86	0.83	0.80	0.66	0.60	0.69	0.49	0.41
1989-2001 (10-digit HS)									
Benchmark	0.66	0.48	0.44	0.65	0.46	0.40	0.62	0.40	0.35
Obs>$100,000	0.92	0.85	0.83	0.86	0.75	0.71	0.76	0.59	0.55

Note: The survival functions across the product types within each dataset are statistically significant at the 1% level using the logrank test

The model's predictions regarding starting size are also supported. In order to investigate whether small, valued spells are at greatest risk we filtered out small dollar-value observations; that is, we eliminated spells with trade in the first year below some minimum level. We then estimate survival functions for each product type after dropping the small-valued observations. In Table 17.3 we report survival rates based on dropping all observations where the value of trade in the first year of the spell was less than $100,000.

[12] Extract from Table 2 in Besedeš and Prusa (2006b).

First, as seen in the second chart in Figure 17.3, when we restrict ourselves to the relationships that start large (initial trade values greater than $100,000) we find better survival; this result is predicted by the model. Said differently, survival functions shift up as we drop small observations. Spells that begin with small trade value are at greatest risk. This is true for all product types. For example, for differentiated goods the one-year survival rate increases from 69 to 92 per cent. This pattern holds for each product type in both time periods studied. As implied by the model, the larger the initial purchase, the longer the duration for each product type.

Second, the estimates provide no evidence that differences among product types are driven by small observations. Differences among product types grow as we eliminate the smaller-trade observations. When we restrict the sample to only those spells with initial transactions exceeding $100,000 the one-year survival rate is 92 per cent for differentiated and 69 per cent for homogeneous goods which compare with 69 and 55 per cent for the benchmark.

A limitation of the Kaplan–Meier estimates is that we cannot control for a myriad of factors that might be influencing duration. To control for other possible explanations of survival, we estimate the Cox proportional hazard model. The basic estimation model includes regressors designed to control for country and product characteristics that might influence duration. Details of these exogenous variables are found in Besedeš and Prusa (2006b); here we note that we control for GDP, *ad valorem* measure of transportation costs, the tariff rate, the change in the relative real exchange rate, the coefficient of variation of unit values, multiple spells, agricultural products,[13] and country fixed effects.

The findings are reported in Table 17.4. In the first column we report the benchmark estimates based on product level data. We are primarily interested in the product type estimates. Letting differentiated products be the benchmark, reference priced products have a 17 per cent higher hazard and homogeneous goods a 23 per cent higher hazard. The estimates strongly support what Figure 17.3 suggested: namely, product type matters. When we filter out spells that start small we find that our results are not driven by small-value spells. Compared to differentiated products, homogeneous goods face a 71 per cent higher hazard at the $100,000 cutoff level; reference priced products face a 59–155 per cent higher hazard.

5. RELATED EMPIRICAL SUPPORT

Since the publication of Besedeš and Prusa (2006a) a growing literature has emerged analyzing the duration of export and import trade. Most of it follows the approach pioneered in Besedeš and Prusa (2006a) and summarized in this research note. The following is a brief synopsis of related work in this area.

[13] Agricultural products are generally classified as homogeneous products, and since agricultural products are more likely to be subject to weather or disease disruption we include an agricultural dummy.

Table 17.4: *Cox Proportional Hazard Estimates for 1972-1988 7-digit TSUSA Data*[14]

		Benchmark	Obs>$100,000
Ad-valorem transportation cost	(unit = 10%)	1.068	1.039
GDP	(unit = $100bil)	0.946	0.940
Tariff rate, 4-digit SITC	(unit = 1%)	0.979	0.945
%Δ relative real exchange rate	(unit = 10%)	0.906	0.897
Coefficient of variation of unit values		0.927	0.864
Multiple spell dummy		1.495	2.254
Agricultural goods		1.040	0.949*
Reference priced products		1.173	1.594
Homogeneous goods		1.226	1.712
Observations		1,140,896	356,141
No. Subjects		444,378	85,629

Country fixed effects inlcuded but not reported
Note: * denotes estimates not significant at the 1% level

Nitsch (2009) finds duration of German imports to be similarly short to duration of US imports. Using 8-digit product level data from he finds most trade relationships last only one to three years. Hess and Person (2009) find duration of imports of European Union members between 1962 and 2006 to be very short, with median duration of one year, and that imports from more diversified exporters, those exporting a greater number of products, exhibit a lower hazard. Hess and Person (2010) replicate the results of Besedeš and Prusa (2006b) using a discrete-time hazard model and make a methodological contribution showing that such models are better suited to trade duration data. This approach eliminates the reliance on the proportional hazard assumption of earlier papers (which is shown to be violated) while taking into account unobserved heterogeneity more robustly.

Besedeš and Prusa (2007) and Besedeš and Blyde (2010) show that duration of exports from a number of Central and South American—the Asian Dragons countries as well as the US and EU countries—is very short. Namely, many relationships fail in their first year resulting in most countries having median duration of an export relationship at only one or two years. Examining duration of exports for a large number of countries, Brenton, Saborowski, and von Uexkull (2009) find evidence that learning-by-doing decreases the hazard of exporting of developing countries, while Jaud, Kukenova, and Strieborny (2009) find that financial development improves export survival of developing countries by reducing the costs of external finance to firms. Fugazza and Molina (2009) examine duration of exports of almost one hundred countries between 1995 and 2006 finding that developed countries, differentiated products, export experience, and the volume of exports all decrease the hazard of exporting. Minondo

[14] Extract from Table 3 in Besedeš and Prusa (2006b).

and Requena (2008) examine duration of exports of regions of Spain finding the median duration for all regions to be just one year and probability of survival rapidly decreasing. Volpe and Carballo (2008) examine export survival of newly exporting Peruvian firms and find their median export duration to be just one year. They also examine the impact geographical and product diversification play for survival and conclude geographical diversification is more important. Görg, et al. (2008) use data on exports of Hungarian firms at the 6–digit HS product level and find that the median duration is between two and three years. In slight contrast to Besedeš and Prusa (2006a; 2006b), Nitsch (2009), and Volpe and Carballo (2008) who all find the hazard to be highest in the first year, Görg, et al. (2008)find the hazard of exporting initially increases and reaches its maximum between the third and fourth year, after which it decreases rapidly.

Using the same data as Görg, et al. (2008), Muraközy and Bekes (2008) examine differences between permanent and temporary trade, where temporary trade is defined as any trade relationship with duration under three years. They find that while the long-term survival rates for US and Hungarian trade relationships are similar, short-term survival rates are not. They offer a new explanation for temporary trade and short-lived relationships. Unlike Rauch and Watson (2003) and Besedeš (2008), where short duration is a consequence of uncertainty and 'testing the waters', Muraközy and Bekes find one-fifth of temporary trade to be the consequence of one-time asset and inventory sales. Caron and Anson (2008) examine duration of low-valued Brazilian exports. These are exports under $20,000 which can be exported using simplified export regulations available through many post offices in Brazil. Most of these exports are of short duration with the median of just one year. Fabling and Sanderson (2008) study duration of New Zealand's exports at the 5–digit SITC and 10–digit HS product level as well at the firm and firm-product level. They too find export duration to be short with median duration at one or two years across various levels of aggregation. They find duration at the firm level to be slightly longer than at the product level, which is likely due to firms changing the mix of products they export. Cadot, Iacovone, Rauch, and Pierola (2010) study the first year survival of exporters at the firm-product level from Malawi, Mali, Senegal, and Tanzania and find low survival rates. Survival improves as firms build experience both with the product they export as well as the destination where they export. In addition, agglomeration has a positive effect on survival – the more firms export the same product to the same destination, the higher the survival for every firm.

Eaton et al. (2008) use Columbian firm-level data to examine export dynamics. While they do not estimate a duration model, they find that about half of Columbian firms exporting in any year tend to be new exporters who export low volumes and most of whom do not survive the first year. Álvarez and Fuentes (2009) find qualitatively similar results for exports of Chilean firm recorded at the 8-digit HS level between 1991 and 2001, while Lederman, Clare, and Xu (2010) find that over 40 percent of new export activities on the part of Costa Rican firms between 1997 and 2007 do not survive the first year, and argue that low survival is one of the main impediments to higher export growth.

6. CONCLUDING COMMENTS

In a series of papers Besedeš and Prusa (2006a; 2006b) and Besedeš (2008) have provided a novel approach to examining international trade, which provided trade economists and policy makers with a set of interesting and surprising facts. We have discovered that countries tend to trade most products for short intervals of time. Reassuringly, as verified by us and a growing number of other authors, the findings appear to be quite robust. Trade relationships remain short when we change the way relationships are measured, and when we control for multiple spells and censoring in different ways. There are a large number of short trade relationships in every industry and for every country studied to date. Trade is of short duration, whether one looks at highly disaggregated product-level data or moderately aggregated industry-level data.

Studying the duration of trade has also provided additional empirical evidence that trade in differentiated and homogeneous products is different. We have shown that a search cost model of relationship formation does a good job of explaining the formation and duration of trade as well as observed differences in trade in differentiated and homogeneous products (for example, differences in initial purchase size, duration, and so on). Our analysis suggests survival in export markets will be longer if a country trades in differentiated good rather than homogeneous products. Of course, this does not imply that exporters should focus exclusively on differentiated products as the work to date does not offer a full theoretical model of trade duration or its welfare effects. Our work suggests economists should incorporate search cost–network approaches into the dominant models of trade.

Tibor Besedeš is an Assistant Professor of Economics at the School of Economics, Georgia Institute of Technology.

Thomas Prusa is a Professor at Rutgers – The State University of New Jersey.

BIBLIOGRAPHY

Álvarez, Roberto and J.Rodrigo Fuentes, (2009). Entry Into Export Markets and Product Quality Differences. Central Bank of Chile Working Papers, No. 536

Besedeš, Tibor, 2008. A Search Cost Perspective on Formation and Duration of Trade. *Review of International Economics*, 16(5):835–849

Besedeš, Tibor and Thomas J Prusa, 2006a. Ins, Outs, and the Duration of Trade. *Canadian Journal of Economics*, 39(1): 266–295

—2006b Product Differentiation and Duration of US Import Trade. *Journal of International Economics*, 70(2), 339–58

—2007 The Role of Extensive and Intensive Margins and Export Growth. NBER Working Paper No. 13628

Besedeš, Tibor and Juan Blyde, 2010. What Drives Export Survival? An Analysis of Export Duration in Latin America. Inter-American Development Bank, mimeo

Brenton, Paul, Christian Saborowksi, and Erik von Uexkull, 2009. What Explains the Low Survival Rate of Developing Country Export Flows? Policy Research Working Paper 4951, The World Bank

Cadot, Olvier, Leonardo Iacovone, Ferdinand Rauch, and Denisse Pierola, 2010. Success and Failure of African Exporters. World Bank working paper

Caron, Justin and José Anson, 2008. Trade facilitation for low-valued exports in Brazil: Lessons to be learned from simplified export declarations and the use of postal networks through 'Exporta Fácil.' International Postal Union, mimeo

Eaton, Jonathan, Marcela Eslava, Maurice Kugler, and James Tybout, 2008. The Margins of Entry into Export Markets: Evidence from Columbia. in Helpman, E, Marin, D, Verdier, T, Eds., *The Organization of Firms in a Global Economy*. Cambridge, MA: Harvard University Press

Fabling, Richard and Lynda Sanderson, 2008. Firm Level Patterns in Merchandise Trade. Ministry of Economic Development, Occasional Paper 08/03

Görg, Holger, Richard Kneller, and Balázs Muraközy, 2008. What Makes a Successful Exporter? CEPR Discussion Paper 6614

Fugazza, Marco and Ana Cristina Molina, 2009. On the Determinants of Exports Survival. HEI Working Papers 05–2009, Economics Section, The Graduate Institute of International Studies

Hess, Wolfgang and Maria Persson, 2009. Survival and Death in International Trade – Discrete Time Durations for EU Imports. Working Papers 2009:12, Lund University, Department of Economics

—2010. The Duration of Trade Revisited: Continuous-Time vs. Discrete-Time Hazards. Working Papers 2010:1, Lund University, Department of Economics

Jaud, Mélise, Madina Kukenova, and Martin Strieborny, 2009. Financial Dependence and Intensive Margin of Trade, Paris School of Economics Working Paper 2009–35

Lederman, Daniel, Andrés Rodríguez-Clare, and Daniel Yi Xu, (2010). Entrepreneurship and the Extensive Margin in Export Growth: A Microeconomic Accounting of Costa Rica's Export Growth During 1997—2007," World Bank working paper.

Melitz, Marc, 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71(6), 1695–1725

Minondo, Asier and Francisco Requena, 2008. The Intensive and Extensive Margins of Trade: Decomposing Exports Growth Differences across Spanish Regions. *FUNCAS Working Paper*Muraközy, Balázs and Gábor Békés, 2008. Temporary Trade. Institute of Economics-Hungarian Academy of Science, mimeo

Nitsch, Volker, 2009. Die Another Day: Duration in German Import Trade. *Review of World Economics*, 145(1), 133–154.

Rauch, James E, 1999 Networks versus markets in international trade. *Journal of International Economics* 48(1), 7–35

Rauch, James E and Joel Watson, 2003. Starting small in an unfamiliar environment. *International Journal of Industrial Organization* 21(7), 1021–1042

Volpe, Christian and Jerónimo Carballo, 2008. Survival of New Exporters in Developing Countries: Does It Matter How They Diversify? INT Working Paper 04, Inter-American Development Bank

18

Openness and Export Dynamics: New research directions

JAMES TYBOUT

INTRODUCTION

Most modern theories that link openness to patterns of trade flows and the gains from trade rely on the assumption of heterogeneous firms. The basic idea, elegantly laid out by Melitz (2003) and Bernard et al. (2003), is that reductions in foreign trade barriers create new opportunities for foreign sales. At the same time, reductions in domestic trade barriers create new competition from abroad and reduce the share of the domestic market captured by local firms. Efficient firms gain more from the former effect than they lose from the latter, so their size increases when all trade barriers come down. Inefficient firms, in contrast, don't find it profitable to export: the operating profits available to them in foreign markets are more than offset by the shipping and (or) foreign market entry costs that they would have to bear. Thus they are subjected to the latter effect without gaining from the former and they shrink or exit in consequence. With inefficient firms contracting and efficient firms expanding, economy-wide average productivity improves with openness.

This mechanism has been embedded in models too numerous to list, including variants that focus on multinational behavior (for example, Helpman, et al., 2004), multiproduct firms (for example, Bernard et al., 2010), transition dynamics (for example, Ghironi and Melitz, 2005; Constantini and Melitz, 2008), and endogenous innovation (for example, Atkeson and Burstein, forthcoming; Constantini and Melitz, 2008). But with few exceptions, this literature has treated trade costs in a very cursory way. Specifically, the costs of breaking into foreign markets and the fixed costs of staying in are assumed to be fixed parameters that are common across firms. Further, the variable costs of exporting are proportional to the physical volume of goods exported.

The treatment of trade costs is important because these costs govern the extent of the resource reallocation that accompanies changes in openness. Moreover, they determine the transition path between one trade regime and another and in doing so they determine the short-run current account deficit, which critically affects the political sustainability of reforms. In what follows I quickly review the

evidence from firm-level data on the nature these trade costs. Then I turn to new evidence and research directions based on shipment-level trade data collected from customs agencies and merged with information on the exporters and importers who are party to the shipments. I argue that the standard assumptions regarding trade costs are hard to reconcile with patterns found in the shipments data, and that an emerging literature focusing on the formation of buyer–seller relationships promises to yield richer models that do better.

1. EARLY EVIDENCE ON SUNK AND FIXED COSTS

Early contributions to the micro literature on export market participation were based on the notion that firms faced 'beachhead' (market entry) costs when they ventured into foreign markets for the first time (for example, Dixit, 1989; Baldwin and Krugman, 1989). These one-time costs were meant to capture the fact that, in order to begin exporting, firms must learn bureaucratic procedures, establish distribution channels, and repackage or even redesign their products for foreign consumers. Melitz (2003) incorporates such costs into his model, and many others have followed suit, either by including sunk entry costs or by assuming that per period fixed costs of exporting are significant.[1]

Many firm- and plant-level empirical studies claim to support the notion that fixed and (or) sunk costs matter. Most of these studies point to the theoretical models developed by Dixit (1989) and Baldwin and Krugman (1989) to motivate their tests. To summarize the empirical version of the Dixit–Baldwin–Krugman model, let the state of firm j in period t be given by its productivity, φ_{jt}, its time-invariant characteristics (like product appeal), z_j, the real effective exchange rate, e_t and a binary variable indicating its exporting status at the beginning of the period, y_{jt-1}. Then the exporting decisions of risk-neutral, profit maximizing firms can be characterized by the policy function $y_{jt} = y\left(\varphi_{jt}, z_j, e_t, \varepsilon_{jt}, y_{jt-1} \mid \theta\right)$ where θ is a vector of parameters and $\varepsilon_{jt} = (\varepsilon_{1jt}, \varepsilon_{2jt})$ captures transitory shocks to fixed or sunk costs. This function solves:

$$V\left(\varphi_{jt}, z_j, e_t, \varepsilon_{jt}, y_{jt-1} \mid \theta\right) = \max_{y(\cdot)} \quad E_t \sum_{\tau=t}^{t+H} \delta^\tau u\left(\varphi_{jt}, z_j, e_\tau, \varepsilon_{j\tau}, y_{j\tau-1}, y_{j\tau} \mid \theta\right).$$

where the net current-period payoff from exporting is:

$$u(\varphi_{jt}, z_j, e_t, \varepsilon_{jt}, y_{jt-1}, y_{jt} \mid \cdot) = \begin{cases} \pi(\varphi_{jt}, z_i, e_t \mid \cdot) - \gamma_F + \varepsilon_{1it} & \text{if} \quad y_{jt} = 1,\ y_{jt-1} = 1 \\ \pi(\varphi_{jt}, z_i, e_t \mid \cdot) - \gamma_F - \gamma_S z_i + \varepsilon_{2it} & \text{if} \quad y_{jt} = 1,\ y_{jt-1} = 0 \\ 0 & \text{if} \quad y_{jt} = 0 \end{cases}.$$

[1] In a steady state with time-invariant productivity shocks and stable demand, sunk entry costs and per period fixed costs play exactly the same role. The distinction between fixed costs and sunk costs becomes important when there is uncertainty about future market conditions. Then sunk costs make exporting a forward-looking decision because, once borne, they open the option to continue exporting in the next period without paying market entry costs.

Here $\pi(.)$ is gross current exporting profits, and γ_F and γ_S are average fixed and sunk exporting costs, respectively.

Sunk entry costs make firms want to avoid repeatedly starting and stopping foreign sales, so they induce firms to consider future market conditions when they decide whether to export today. In addition, since firms won't start exporting unless they expect to recoup their entry costs eventually, sunk costs place a lower bound on the expected gross export profit stream of a new market entrant. Fixed costs bound exporting profits too, but since they are borne each period that a firm exports, they do not in themselves create a role for expectations about future market conditions.

Early empirical studies tested for sunk costs by looking for evidence that current exporting status depended upon lagged exporting status, controlling for other sources of persistence in exporting behavior, both observed and unobserved. As Greenaway and Kneller (2007, F140) note in their survey, 'Exporting next period is strongly correlated with exporting this period, even when other determinants of persistence have been controlled for.' For many firms, having exported in the previous period increases the probability of exporting in the current period by more than 50 percentage points (Roberts and Tybout, 1997).

In a more recent study, Das et al. (2007) use the structure of the dynamic optimization problem sketched above to put dollar magnitudes on sunk entry costs and per period fixed costs. They find that on average, entry costs are large ($700,000) and fixed costs are modest ($15,000) for Colombian manufacturing firms.[2] They further show that, given their estimates, expectations about future market conditions can matter a lot for small-scale exporters. For example, a devaluation that is viewed as a transitory shock is predicted to induce about half as many firms to export as a devaluation that is perceived as a permanent change in the exchange rate process. However, most of the exports in a typical industry come from a handful of dominant firms, and for these exporters expectations about future market conditions are unimportant. Their current operating profits from exporting dwarf market entry costs, even during periods when markets are unfavorable to exporters. One implication is that the credibility of policy reforms has a much bigger impact on the number of exporters than on the value of total exports. Another implication is that export promotion schemes that subsidize market entry are much less efficient in terms of their impact on total exports than schemes that subsidize export sales.

2. NEW EVIDENCE ON EXPORTING COSTS: MARKETING, SEARCH AND LEARNING

The finding that sunk costs are large explains why only a minority of firms venture into foreign markets, and why the likelihood that a current exporter will

[2] 'Sunk entry costs are identified by differences in exporting frequencies across plants that have comparable expected profit streams, but differ in terms of whether they exported in the previous period.. 'Given profit streams and sunk costs, the frequency of exit among firms with positive gross profit streams identifies fixed costs.' (Das, et al. page 2007)

continue exporting is greater than the likelihood a non-exporter will enter foreign markets. But it is difficult to reconcile big sunk costs with a new set of stylized facts that are emerging from transactions-level data on international shipments. Specifically, in a typical year, one-third to one-half of all the commercial exporters observed in customs data did not export in the previous year. Most of these firms ship tiny amounts (worth several thousand dollars), and most will revert to exclusive reliance on domestic sales in the following year. [3] Figure 18.1 depicts these patterns over a 9-year period for the case of Colombian exporters.

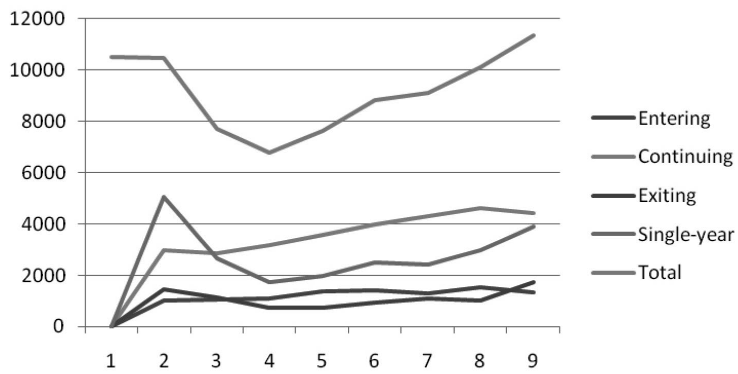

Figure 18.1: *Numer of Exporters by Type**

*Based on Eaton et al. (2008). For each year *t*, entering firms did not export *t*-1 but do export in t and *t*+1, continuing firms exported in *t*-1 and continue exporting in t and *t*+1, exiting firms exported in *t*-1 and continue to export in *t*, but will not export in *t*+1, and single-year exporters export in *t*, but not in *t*-1 or *t*+1.

Further clues about export dynamics emerge if one follows a cohort of new exporters through time as it matures. Defining the year *t* cohort to be the set of firms first observed to be exporting from Colombia to the United States in year *t*, Figure 18.2 shows the average number of cohort members, average total exports, and average exports per surviving cohort member as a function of cohort age. (Cohort *t* is one year old in year *t+1*, etc.) Notice that, although the number of cohort members drops dramatically after one year, the attrition rate is much slower thereafter, suggesting that new exporters go through a shakedown process. Also, among the cohort members who survive the early years, exports per firm grow very rapidly, so the total exports of a typical cohort expand despite rapid early attrition. Finally, successful members of each cohort typically have larger initial

[3] Das et al. (2007) model accommodates the frequent entry and exit of small-scale exporters by assigning large variances to the transitory shocks to sunk and fixed costs (that is, the 's). Thus many firms draw small sunk and fixed costs in some periods, and these firms quickly exit when their draws prove less favorable in the future. This explanation works only if one is willing to accept large year-to-year fluctuations in fixed and entry costs.

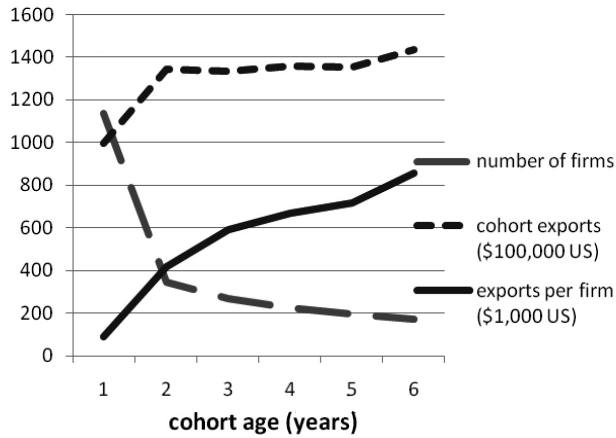

Figure 18.2: *New Cohort Maturation**

*Based on Eaton et al. (2009)

export sales than other cohort members, and as they expand, they do so both by increasing sales per buyer and by increasing the number of buyers they ship to in the US market (not pictured).[4]

Several theories are available for these findings, and for related earlier results based on 6-digit product-level trade flows (Besedes and Prusa, 2006). One was developed by Rauch and Watson (2003) before these patterns were documented. These authors model the behavior of developed-country buyers who engage in costly search for developing-country suppliers. Suppliers are heterogeneous in terms of their productivity and their capabilities to deliver specialized products, but only their productivity is costlessly observable. When a buyer meets a seller, they can either invest in training them to produce the desired product and place a major order immediately, or they can place a small trial order and thereby learn about the seller's capabilities, deciding afterward whether to reject them or invest in training them. The former strategy can lead to rapid order fulfillment, but it can also lead to investments in sellers who are incapable of delivering. The latter strategy reduces risk, but it is costly because it takes time to discover a seller's capabilities through trial orders. In equilibrium, buyers immediately reject sellers with low productivity; they place small trial orders with sellers who have moderate productivity (then either train or reject them when their capabilities are revealed), and they immediately invest in training sellers with high productivity.

The Rauch–Watson characterization of international buyer–seller matching accounts for several patterns in the shipments data that are inconsistent with the

[4] These statements are based on Eaton et al. (2009), who use US customs records to keep track of the identification of exporters and importers for shipments from Colombia to the US

simple sunk cost model. In particular it explains why (1) many exporters ship small amounts and then exit the destination market, (2) sellers who start with larger export volumes are more likely to remain in the export market, and (3) on average, the successful members of each cohort experience rapid export growth after their first year. However, the Rauch–Watson idea has some limitations. It treats suppliers as passive agents who do not search for buyers, and it does not account for the fact that buyers use multiple suppliers.

Arkolakis (2008a; 2008b) shifts the search process from buyers to sellers in his interpretation of the stylized facts mentioned above. More precisely, he argues that exporters can easily find a few customers in a destination market, but after the low-hanging fruit has been picked they must ferret out after increasingly hard-to-find buyers. Thus the costs of building a clientele rise more than proportionately with export sales. This characterization of marketing costs means that non-exporting firms who experience favorable productivity shocks should find it easy to establish a toehold in foreign markets, counter to the standard sunk cost specification. But the more market inroads these new exporters make, the tougher shedding becomes, and the more their growth rates slow down. By assuming that exporters' productivity shocks follow a Brownian motion, Arkolakis (date) is able to explain the large volume of short-lived, small-scale exporting episodes, the surge in shipments among a small set of successful new exporters, and the growth slowdown as firms' exporting relationships mature.

Eaton et al. (2009) also develop a model in which sellers search for buyers, but they add seller-side learning. As in Arkolakis's models, costs are convex in search intensity. However, each time an exporter meets a buyer, the seller receives a noisy signal about their product's appeal to consumers in the destination market. A large order from a new buyer signals that the product is likely to be popular with others, while a small order signals the opposite. Each time a match is made and a signal is conveyed, the exporter updates their priors concerning the product's appeal and adjusts their search intensity. Early signals are the most informative, so they result in the largest adjustments in search intensity. Matches between buyers and sellers endure from one period to the next, subject to an exogenous hazard of separation.

More formally, Eaton et al. (date) assume that firm j is able to discover new buyers at the rate λ_j when it spends $c(\lambda_j)$ on search activities, and that its existing matches break up at some exogenous rate. Further, they assume that j's expected profit stream from its next match can be written as $\tilde{\pi}(\bar{z}_j^n, \sigma_j^n, e_t, \varphi_{jt})$, where e_t is the real effective exchange rate, φ_{jt} is j's current productivity, and $\bar{z}_j^n$ and $\sigma_{z_j}^n$ are the mean and standard deviation of the posterior distribution that summarizes j's beliefs about its product's appeal after it has experienced n matches.[5] These beliefs are based on a standard Bayesian updating process which in turn reflects the size of the orders that have been placed by buyers whom seller j has previously

[5] The expected profit stream $\tilde{\pi}(\bar{z}_j^n, \sigma_j^n, e_t, \varphi_{jt})$ is related to single period export profits by the match separation hazard, the Markov processes for exogenous state variables, and the Bayesian updating rule.

met. Assuming that e_t and φ_{jt} follow first-order Markov processes, Eaton et al. (date) then find *j*'s optimal search intensity as the solution to:

$$V(\bar{z}_j^n, \sigma_j^n, \varphi_{jt}, e_t) = \max_\lambda \Big\{ -c(\lambda) + \lambda \cdot \tilde{\pi}(\bar{z}_j^n, \sigma_j^n, \varphi_{jt}, e_t) \\ + \frac{1-\lambda}{1+r} \int_{\varphi'}\int_{e'} V(\bar{z}_j^n, \sigma_j^n, \varphi_{jt}, e_t)\, dG\big(\varphi', e' \mid \varphi_{jt}, e_t\big) \\ + \frac{\lambda}{1+r} \int_{z^{n+1}} \int_{\varphi'}\int_{e'} V(\bar{z}_j^{n+1}, \sigma_j^{n+1}, \varphi_{jt}, e_t)\, dG\big(\varphi', e' \mid \varphi_{jt}, e_t\big) d\Phi\big(\bar{z}'^{n+1} \mid \bar{z}^n\big) \Big\}$$

Here G( ) and Φ() are transition densities, and the latter reflects Bayesian updating. Combined with realizations on the exchange rate, productivity levels, and buyer-specific demand shocks, the optimal search rule implies exporting patterns for each potential seller.

The Eaton et al. (2009) model explains the prevalence of tiny, short-term exporters as a consequence of low costs for low-level search. Such costs imply that lots of Colombian firms maintain a mild interest in the US market and experience occasional matches. These matches typically result in small orders; hence they typically discourage sellers from intensifying their search efforts. To the contrary, since they are early signals they receive heavy weight, and they often discourage further search altogether. Consonant with the stylized facts reviewed above, the Eaton et al. (2009) model also implies that a handful of the new matches will result in non-trivial orders, and that the sellers who are fortunate enough to encounter one of these relationships will tend to exhibit rapid subsequent export growth as they intensify their search for clientele. Eventually, however, their client base becomes so large that their search intensity is just sufficient to replace the clients who are separating.

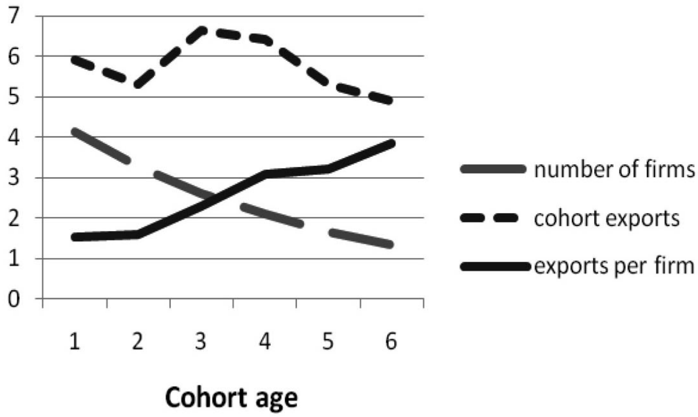

Figure 18.3: *Simulated Cohort Maturation**

*Based on Eaton et al. (2009)

Eaton et al. (2009) calibrate a prototype version of their model in order to examine these implications numerically. They find they can crudely replicate the patterns documented in Figure 18.2, except for the extraordinarily high exit rate among first-year exporters (Figure 18.3). (This limitation is easily rectified by allowing the match separation rate to depend upon the number of periods that the match has endured.) Their model also generates hysteresis effects and state dependence in trade flows, since exporters do not forget what they have learned about their product's appeal once they have generated a match history. For example, since search and learning should intensify during periods of favorable exchange rates, and since learning is irreversible, temporary devaluations can trigger permanent increases in export volumes.

3. RESEARCH DIRECTIONS

The recent shift of focus toward micro dynamics of business relationships opens up a number of new directions for research. In addition to econometrically estimating the model mentioned above, Jonathan Eaton, Marcela Eslava, Maurice Kugler, C J Krizan and I are planning to pursue several related exercises. First, we hope to generalize our model so that potential exporters learn about their product's appeal not just from their experiences in the destination market, but from their experiences in their home market and in other countries to which they have exported. Similarly, we hope to incorporate learning from the experiences of other exporters in similar industries. This will open the possibility that a few pioneer exporters might create demonstration effects, and thus induce industry-specific export surges similar to those described by Hausmann and Rodrik (2003) and Hausmann et al. (2007). It will also help us to understand contagion effects that induce exporters who experience success in some markets to begin exporting to other countries with similar demand features.[6]

A second issue we are pursuing is the question of which type of buyer tends to match with which type of seller, and once matched, which types of buyer-seller pairs tend to flourish. To this end we are merging plant-level information on Colombian exporters and on US importers from annual industrial surveys with the trade shipments data. Then we are borrowing techniques from the marriage literature to characterize the assortative sorting process. (One challenge here is that the marriage literature generally presumes monogamy, whereas buyers and sellers are often polygamous.) If successful, this exercise will shed light on the way business relationships are formed in general, and could conceivably identify promising exporting opportunities for particular types of firms that have not yet been exploited.

Third, to confirm the ideas behind the model, and to learn more about the way that managers think about building exporting relationships, we are planning interviews with potential exporters, exporters, and importers.

Others are also starting projects based on shipments data. Blum et al. (2009) have used shipments data between Chile and Colombia to determine which exporters use intermediaries, and which ship directly to their final buyers abroad.

They find, *inter alia*, that 'in virtually every Chilean exporter–Colombian importer pair, at least one of the parties is a large international trader (2009, abstract).' They go on to develop a theoretical model that explains firms' choices of intermediation technology, emphasizing the importance of the volume of the shipment as a determinant. Large sellers do not need to use intermediaries; rather, buyers find them. But small sellers do best to use trading firms to achieve visibility.

Finally, Drozd and Nosal (2009) develop a dynamic general equilibrium model in which exporters invest in building customer relationships abroad. Once a match is made, the terms of the sale are determined by a bargaining game, so different prices emerge for the same product in different countries or at different points in time. Among other things, the model provides a new interpretation for observed patterns of pricing to market.

James Tybout is a Professor in the Department of Economics at Pennsylvania State University.

REFERENCES

Arkolakis, Costas (2008a). Market Access Costs and the New Consumers Margin in International Trade, Working Paper, Department of Economics, Yale University

—(2008b). A Unified Theory of Firm Selection and Growth, Working Paper, Department of Economics, Yale University

Atkeson, Andrew and Ariel Burstein (forthcoming). Innovation, Firm Dynamics and International Trade, Journal of Political Economy.

Baldwin, Richard E and Paul R Krugman (1989). Persistent Trade Effects of Large Exchange Rate Changes. *Quarterly Journal of Economics*, 104: 635–54

Bernard, Andrew, Stephen Redding and Peter Schott (2010). Multi-product Firms and Product Switching, *American Economic Review* 100(1): 70-97.

Bernard, Andrew B, Eaton, Jonathan, Bradford Jensen, J and Samuel Kortum (2003). Plants and Productivity in International Trade. *American Economic Review*, 93: 1268–90

Besedes, Tibor and Thomas J. Prusa. (2006). Ins, Outs, and the Duration of Trade, *Canadian Journal of Economics* 39(1): 266–95

Blum, Bernardo, Sebastian Claro and Ignatious Horstmann (2009). Intermediation and the Nature of Trade: Theory and Evidence, Working paper, Rotman School of Management, University of Toronto.

Bustos, Paula (2007). Multilateral Trade Liberalization, Exports and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinean Firms, mimeo

Constantini, James A and Marc J Melitz (2008). The Dynamics of Firm-level Adjustment to Trade Liberalizations, in Elhanan Helpman, Dalia Marin and Thierry Verdier, (Eds.), *The Organization of Firms in a Global Economy*. Cambridge: Harvard University Press, 107–41

Das, Mita, Roberts, Mark and James Tybout (2007). Market Entry Costs, Producer Heterogeneity and Export Dynamics. *Econometrica*, 75(3): 837–74

Dixit, Avinish (1989). Hysteresis, Import Penetration, and Exchange Rate Pass-Through. *Quarterly Journal of Economics*, 104: 205–28

Drozd, Lucasz and Jaromir Nosal (2009) .Understanding International Prices: Customers as Capital. Working Paper, University of Wisconsin, Department of Economics

Eaton, Jonathan, Marcela Eslava, Maurice Kugler and James Tybout (2008). Export Dynamics in Colombia: Firm-level Evidence, in Elhanan Helpman, Dalia Marin and Thierry Vierdier, (Eds.). *The Organization of Firms in a Global Economy*. Cambridge: Harvard University Press

Eaton, Jonathan, Eslava, Marcela Kugler, Maurice, Krizan, Cornell J and James Tybout (2009). A Search and Learning Model of Export Dynamics. Working Paper, Department of Economics, Pennsylvania State University

Ghironi, Fabio and Marc J Melitz (2005). International Trade and Macroeconomic Dynamics with Heterogeneous Firms. *Quarterly Journal of Economics*, 120: 865–915

Greenaway, David and Richard Kneller (2007). Firm Heterogeneity, Exporting and Foreign Direct Investment. *Economic Journal*. 117 (517): F134–F161

Hausmann, Ricardo and Dani Rodrik (2003). Economic Development as Self-discovery. *Journal of Development Economics*, 72(2): 603-633

Hausmann, Ricardo, Dani Rodrik and J Hwang (2007). What You Export Matters, *Journal of Economic Growth*,. 12(1): 1-25

Helpman, Elhanan, Marc Melitz and Steven Yeaple (2004). Exports versus FDI with Heterogeneous Firms. *American Economic Review*, 94(1): 30–16

Melitz, Marc (2003). The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71: 1695–1725

Rauch, James and Joel Watson (2003). Starting Small in an Unfamiliar Environment. *International Journal of Industrial Organization* 21: 1021–42

Roberts, Mark and James Tybout (1997). The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs. *American Economic Review*, 87(4): 545–63

19

Market Penetration Costs and International Trade[1]

COSTAS ARKOLAKIS AND OLGA TIMOSHENKO

1. INTRODUCTION

For the past decade a large number of studies have documented important empirical regularities for the role that individual firms play in international trade. One predominant finding is that in any given year only a small fraction of firms export, typically less than 15 per cent.[2] In addition, the distribution of sales of exporters in a destination country is dominated by small exporters: for example the 25 per cent of French firms with the lowest sales in an export market sell less than $10,000 in that market, as pointed out by Eaton, et al. (2008).[3]

Related evidence has also been documented using data sets that report trade in goods at a very disaggregated level. An important finding relates to the growth of trade for individual goods during trade liberalization episodes. In particular, the percentage increase in the trade flows is higher for the goods that were traded in small but positive volumes prior to the liberalization, which implies a reallocation of market shares within traded goods. Arkolakis (2008a) studies the case of imported Mexican goods from the United States that were positively traded before the NAFTA liberalization episode. He finds that the least traded amongst these goods that accounted for the 15 per cent of total imports in 1990–92 increased their share to almost 25 per cent after the liberalization.[4]

Models of trade with heterogeneous productivity firms and a love for variety of preferences (constant elasticity of substitution (CES) Dixit–Stiglitz), such as as Melitz (2003) and Chaney (2008), have served as the benchmark specification in understanding the patterns of firm-level trade data. These models typically assume fixed per period costs of exporting, which imply that only the firms that are able to generate sufficiently high profit to overcome the fixed cost will become exporters. Thus, such models can explain the limited participation of

[1] We are grateful to Guido Porto for suggestions on this note. All remaining errors are ours.

[2] See Bernard, et al. (2007) for facts about the US manufacturing firms; Eaton, et al (2008) for Colombian firms; Eaton, et al. (2008) for French firms.

[3] This is true for each one of more than 100 exporting markets that Eaton, et al. (2008) study. The fact that sales of the most exporters are small compared to the total volume exported has also been confirmed by Arkolakis and Muendler (2007) among others.

[4] Similar findings are reported in Kehoe and Ruhl (2003) and Kehoe (2005).

firms in the export markets. Furthermore, since profitability of a firm is increasing in its productivity in these models, only the most productive firms export, thus explaining the productivity difference between exporters and non-exporters. Some important empirical regularities, however, remain puzzling for the fixed cost models: the dominance of many small exporters, and the growth in trade of individual goods in response to policy changes.

This note discusses the implications of a new formulation of market penetration costs developed by Arkolakis (2008a) and incorporated in a framework of firm productivity heterogeneity such as the ones of Melitz (2003) and Chaney (2008). In this formulation, the per period export costs depend on the number of consumers a firm chooses to reach in an export market and, therefore, are endogenous to the firm. A firm enters a market if it makes profits by reaching a single consumer there and pays an increasing marginal cost to access additional consumers. The formulation we discuss intends to capture broadly the marketing costs that the firm incurs in order to increase its sales in a given market.

The model improves significantly upon puzzling predictions of the fixed cost models and, therefore, has important implications for policy analysis. First, it reconciles the typically large estimates of the fixed cost with evidence on the existence of many firms exporting small amounts to particular markets through the extensive margin of consumers' mechanism.[5] It implies that even small market entry costs could exclude many firms from the market and imply low levels of market penetration of others. In the policy context, if we think of these market penetration costs as per consumer marketing costs, policies that could reduce this component of the costs (such as advertisement of foreign products, trade fairs, and so on) could be beneficial for entry of new exporters and for the growth of existing small ones.

Second, the endogenous formulation of market penetration costs and, as a result, the departure from the Dixit–Stiglitz demand structure, enables the model to account for the heterogeneous response of trade flows by goods to changes in variable costs of trade such as tariffs. The new framework implies a higher growth rate in trade for firms or goods with positive but little previous trade, a result which is largely consistent with various studies on trade liberalization such as Kehoe (2005), and Kehoe and Ruhl (2003). As a result, the model is well suited to analyze and predict the behavior of trade flows in response to policy changes.

The rest of the note is organized as follows. Section 2 discusses the limitations arising within models of trade with fixed entry costs. Section 3 presents the details of the theory of endogenous market penetration costs and Section 4 illustrates its predictions. Section 5 offers a discussion on how the theory of endogenous costs is useful in the context of industrial policy in developing countries. Section 6 summarizes current research on the sunk entry costs and Section 7 concludes.

[5] For example, Das, et al. (2005), examine a sample of Colombian exporters for the period of 1981 to 1991. Using a dynamic model, they estimate a (one time) fixed cost for new exporters ranging between $300,000 and $500,000 per firm. These estimates are rather large compared to the predominance of French exporters that sell less than $10,000 reported by Eaton, et al. (2008).

2. MODELS OF TRADE WITH FIXED COST AND TRADE FACTS

In this section we discuss the extent to which the fixed cost model can explain international trade facts. In this model the assumption of fixed per period costs allows the model to explain the limited participation of firms in the exporting markets, and the productivity difference between exporters and non-exporters. A number of empirical regularities, however, remain puzzling for the fixed cost model: the dominance of many small exporters, and the high growth in trade for goods with a low but positive volume of trade prior to trade policy changes. As shown in Arkolakis (2008a), a fixed cost model parameterized to match the fraction of exporters overpredicts the sales of the smallest percentiles of firms by a factor of 150. The problem partly arises from the indivisibility property of the fixed costs: large fixed costs are necessary to account for the small fraction of exporters, while small costs are consistent with the large number of exporters selling small amounts in the destination market. The mechanism of the model suggests that only the firms with exporting sales that are high enough to overcome the fixed costs will become exporters. On the one hand if the fixed costs are assumed to be high, very few firms will export with export sales being at least as high as the assumed value of the fixed costs. On the other hand if the fixed costs are assumed to be negligible, most of the firms will become exporters and many will export tiny amounts.

The extensions of the uniform fixed cost models that allow heterogeneous fixed costs across firms would alleviate the puzzling predictions regarding the predominance of small exporters. However, such an assumption by itself is not enough to resolve the predictions regarding the larger growth of goods with positive but little trade in foreign markets. This problem arises due to the assumption of the love for variety preferences (specifically CES Dixit–Stiglitz demand). Such a preference structure yields constant elasticity of trade flows of a good with respect to trade costs. Thus, a decrease in trade costs will cause the same percentage increase in trade volumes across the goods of different types.

Rather than arbitrarily specifying a fixed cost structure that would work, we will describe a procedure that starts from first principles and which models how firms reach foreign consumers. The result is a market penetration cost structure that not only alleviates the puzzling predictions of the fixed cost model, but also retains its desirable properties: the prediction that few firms export and the sales distribution of large exporters seem to be described very well by the fixed cost model.[6] We present such a structure in the next section.

3. A NEW THEORY OF MARKET PENETRATION COSTS: THE IDEA

The contribution of the recent work of Arkolakis (2008a) is to offer an alternative formulation of entry costs. This formulation postulates that market penetration costs are an endogenous choice of the firm. The more of a prespecified marketing

[6] See the discussion in Eaton, et al. (2008).

cost the firm pays the more it can sell to a given market. This abstract interpretation of market penetration costs can take more precise and deterministic representations which we describe below.

A first interpretation of these costs is that of marketing costs to reach foreign consumers. The main argument assumes that the marginal costs of reaching consumers are increasing in the fraction of consumers reached.[7] In this case a more productive firm, which generates more sales per consumer, will choose to reach more consumers and generate higher sales. A less productive firm will reach a small fraction of consumers and generate low sales. A firm with sales per consumer not high enough to cover the marketing costs of reaching the first consumer will choose not to participate in the export market. Thus, the model can explain both endogenous export participation (if firms optimally decide not to reach any consumer) and small sales (if firms optimally decide to enter a country but reach few consumers).

Increasing market penetration costs to reach consumers is not the only explanation however. The same logic applies if the model is reinterpreted as a model where marginal market penetration costs are constant but result in declining marginal revenues from market penetration. This interpretation can be precisely stated as declining marginal revenues from introducing new products as formalized by Arkolakis and Muendler (2007), and also from consumers with heterogeneous tastes. The common assumption of all these specifications is that there is another margin of firm sales that the firm can regulate, using payments to marketing costs versus simply reducing its price. Of course this assumption has a variety of relevant policy implications that are different from the assumption of fixed cost of market penetration.

4. IMPLICATIONS OF MODELING MARKET PENETRATION COSTS

The model of endogenous market penetration costs improves the predictions of the uniform fixed cost model in two important dimensions: size distribution of exporters, and growth in trade in response to policy changes. The assumption of endogenous increasing marginal cost of reaching consumers allows firms with potentially small sales to export into a destination market, thereby explaining the dominance of the distribution of sales by small exporters. The data on the sales of French firms in Portugal (similar results are true for the other markets where French firms sell) is plotted against the predictions of the endogenous cost model versus the fixed cost model in the top panel of Figure 19.1.[8] The data show

[7] Decreasing returns in marketing outlays may arise as (i) less responsive consumers are reached or the same consumers respond less to additional marketing efforts, or (ii) an increasing amount of effort has to be taken in order to reach a consumer that has not yet been reached as discussed in Bagwell (2007, 51). The analytical representation of these marketing costs builds on seminal contributions in the advertising literature such as those of Butters (1977) and Grossman and Shapiro (1984).

[8] The distribution of the sales of French firms in a destination market is robust across different the markets as established in Eaton, et al. (2008). Portugal is considered here as a representative example. The calibration procedure for the endogenous and fixed cost models is described in Arkolakis (2008a).

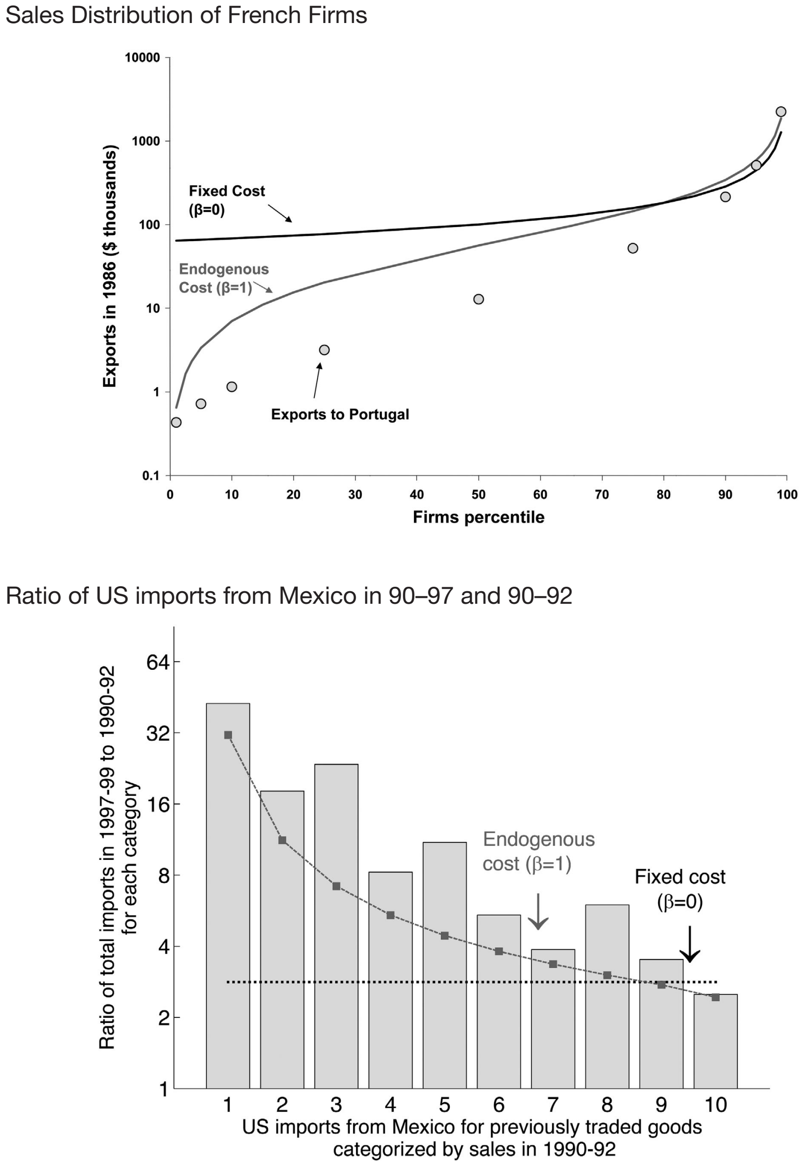

Figure 19.1: *Cross-sectional and Comparative Static Predictions of the Model*

Data Source: Eaton, Kortum and Kramarz (2008) and www.sourceoecd.org tabulated by Arkolakis (2008a).

that the sales distribution is dominated by small exporters: the individual sales of at least 25 per cent of French firms in Portugal are below $10,000. The fixed cost model overpredicts the sales of the smallest exporters, however it predicts well the sales of the largest exporters. The model with endogenous costs does

well where the fixed cost model does well, but, in addition, it also predicts the predominance of small exporting firms. The reason for the improved predictions is that firms with low productivity are able to penetrate the market by paying small endogenous market entry costs to reach only a few consumers and, as a result, attain low levels of export sales in the market.

The model of endogenous market penetration costs yields non-trivial predictions regarding the response of trade flows by goods to changes in trade policy. Since firms are allowed to make decisions regarding what fraction of consumers to reach, the demand function departs from the Dixit–Stiglitz demand used in the fixed cost models. As a result, the elasticity of demand with respect to trade costs is not constant across goods of different types. In particular, the elasticity of trade has two components. The first component (intensive margin) is the same as in the Dixit–Stiglitz context and originates from the per consumer sales elasticity. The second component (extensive margin) originated due to the fact that the number of consumers reached by the firm changes in response to a change in trade costs. Thus, as trade costs fall, the sales of a firm increase per consumer, as well as due to the increase in the number of consumers. The interplay of the intensive and extensive margins of trade is what allows the endogenous cost model to match the prediction of the trade flows more closely.

Arkolakis (2008a) shows that the implications of the endogenous cost model are consistent with the recent findings regarding trade liberalization episodes: goods with little, but positive, trade before a liberalization are the ones that experience the highest growth rates after the episode, as depicted in the bottom panel of Figure 19.1. The figure uses the data on US total imports from Mexico in 1997–98 and 1990–92. The previously traded goods are split into deciles depending on how much they were traded before the liberalization. The data reveal an interesting pattern: the growth rate of the goods is higher the less traded the goods were before the liberalization episode. This pattern is predicted by the calibrated model with endogenous market penetration costs. In addition, goods that were not traded before liberalization have a much smaller share in new trade, given that they are traded in very small amounts after the liberalization episode. In this case the calibrated model with fixed entry costs predicts identical changes in trade flows for each category of goods, while the endogenous cost model delivers a close match to the data. It captures very well the growth of the goods with least trade, since the model predicts higher trade elasticity with respect to trade costs for lower productivity firms.

In a follow up paper, Arkolakis (2008b) shows that adding market penetration costs is a fruitful assumption when considering individual firm growth. In particular, a model of productivity dynamics combined with endogenous market penetration costs would imply that new entrants are small and grow fast in accordance with the data. Both are predominantly empirical findings for individual exporters and firms overall as reported by Eaton, et al. (2008) and Dunne, et al. (1988). Instead, the fixed cost benchmark implies large entrants with growth rates that are smaller than the ones implied in the data. In the next section we discuss the implications of modeling endogenous market penetration costs for international development.

5. DEVELOPMENT APPLICATIONS

Direct evidence on marketing expenditures that exporters incur can give us an idea of the type of the entry costs that the formulation of endogenous market penetration costs could be capturing. These costs potentially include the costs incurred by a firm during the process of promoting its product and reaching consumers, as well as establishing the related distribution channels in order to sell its product. Evidence about the exact nature of these market penetration costs for the case of exporting is provided by Keesing (1983) and Roberts and Tybout (1997b). The authors discuss a number of costs reported by managers of exporting firms in a series of interviews. These data indicate that firms must research the foreign market by identifying and contacting the potential consumers of their good.

The additional insight that the modeling of market penetration costs offers is that these costs are not likely to be large for small exporters in a market. However, since penetration in foreign markets is increasingly difficult, firms can optimally choose not to enter a market or to enter and sell little. Moreover, the model implies that for small changes of these costs small firms can expand their market shares substantially. Therefore, policies that are targeted towards improving marketing technology of exporters (such as the advertisement of national products abroad, trade fairs, and so on) could be very beneficial for entry of new exporters and growth of existing small ones. In addition, to the extent that marketing technologies in developing countries are outdated this type of investment could enhance the ability of developing countries to attract foreign trade and investment. The theory also implies that these types of policies are likely to make a small difference for large exporters: to the extent that these exporters have already established large distribution channels, little change in their market shares is expected. Thus, a dollar spent in improving the marketing of small firms would be more beneficial for overall exports versus a dollar spent in improving the marketing of large firms.

In addition, the modeling of endogenous market penetration costs has implications for the growth of trade of firms or goods with respect to changes in the variable costs of trade, such as tariffs. Existing firms (or goods) with low levels of market penetration, and thus low sales, would increase their market share much faster in response to changes in these costs. This feature of the model can be used to explain the large growth of trade for goods with little trade before liberalization reported by Kehoe (2005), Kehoe and Ruhl (2003), and Arkolakis (2008a). Potentially, a mechanism like the one implied by the formulation of endogenous market penetration costs could be incorporated in an applied general equilibrium framework. This framework could then be used in predictions of the growth of trade during trade liberalization episodes.

6. CURRENT RESEARCH ON SUNK ENTRY COSTS

Arkolakis (2008b) has shown that a model with endogenous market penetration costs and firm productivity dynamics can go far in explaining dynamic facts on

exporter growth. In particular, such a model that is based on persistence in firm productivity shocks can generate persistence in exporting status: a firm that is currently exporting is more likely to export next year versus a firm not exporting. An outstanding question for the Arkolakis (2008b) formulation is whether the predictions of this model conflict with previous evidence on hysteresis behavior in export participation. Hysteresis behavior refers to the asymmetric decisions of firms regarding export participation in response to a positive or negative shock at the macro or at the firm level.[9]

To account for the hysteresis phenomenon, previous literature postulated the existence of sunk costs of exporting. Sunk costs of exporting are defined as one-time entry costs that a firm has to pay before it starts to export. These costs have to be paid every time the firm exits and reenters a market. A firm that is hit by a beneficial productivity shock pays the sunk cost and becomes an exporter. When the productivity reverts to its original level, a firm will continue to be an exporter in order to avoid paying the sunk costs in the future in response to another beneficial productivity shock. Thus, conditional on productivity, or in the empirical context – on sales, a firm that is an exporter will have a higher probability of continuing to be an exporter compared to a non-exporting firm. However, the discussion in previous sections regarding indivisibilities in exporting costs leads to the conclusion that sunk costs may not be a realistic representation of these exporting costs.[10]

Current research on entry costs is motivated by these findings and looks at sources of hysteresis in exporting behavior without assigning a particularly large role to indivisibilities in entry costs. One mechanism that can create hysteresis in exporting behavior is learning, or experience accumulation. As a firm enters a destination market, learning occurs: the longer a firm exports the more information it gains about the appeal of its product.[11] In a reduced form evaluation of the learning mechanism in the form of age-dependent sales, Timoshenko (2009) finds that age in the destination market is an important predictor of the future probability of exporting, and the effect of sunk costs, typically captured by the one period lagged exporting status, declines by almost one half when controlled properly for age. Controlling for the depreciation of the sunk costs, Timoshenko (2009) finds that those costs do not depreciate as quickly when the effect of age in the destination market is taken into account. These findings substantially weaken the support for the sunk entry cost and, rather, provide supporting evidence that the effect of the accumulated knowledge on sales makes firms continue exporting their products.

[9] For a greater discussion on hysteresis phenomenon see Dixit (1989), Baldwin and Krugman (1989),

Roberts and Tybout (1997a).

[10] The assumption of sunk costs of entry is also inconsistent with the observation that the average sales size of the exits and the new entrants is approximately the same Eaton, et al. (2008). As suggested by the analysis above, in a model of sunk costs exits are typically smaller than entrants. In addition, as shown by Ruhl and Willis (2008), the sunk cost model performs poorly in predicting the survival rate of exporters.

[11] See for example Eaton, et al. (2008) and also the recent work of Ruhl and Willis (2008).

The above implies that the learning mechanism is a promising avenue for future research related to entry costs. The avenue of modeling learning in a model of international trade has been recently adapted by Eaton, et al. (2008), and Ruhl and Willis (2008).[12] Of course an outstanding challenge is to relate and connect this research to the findings of the theory of endogenous market penetration costs as is done for example by Eaton, et al. (2008).

7. CONCLUSION

In this note we have briefly discussed the implications of a new theory of entry costs to foreign markets. This theory postulates that entry costs are endogenous rather than fixed, in the sense that paying higher costs allows firms to increase their market share in a country. We have shown that such a formulation of entry costs is consistent with a number of important regularities on international trade. In addition, we argue that this new modeling of entry costs can have valuable practical implications related to the way we think about adjustments to trade opportunities.

Costas Arkolakis is an Assistant Professor in the Department of Economics of Yale University.

Olga Timoshenko is a Ph.D. candidate in the Department of Economics of Yale University.

[12] Arkolakis and Papageorgiou (2009) also model learning in a model of firm heterogeneity.

BIBLIOGRAPHY

Arkolakis, C (2008a). *Market Penetration Costs and the New Consumers Margin in International Trade*, NBER Working Paper 14214

—(2008b). A Unified Theory of Firm Selection and Growth, mimeo, Yale University

Arkolakis, C, and M-A Muendler (2007). The Extensive Margin of Exporting Goods: A firm level analysis, mimeo, University of California San Diego and Yale University

Arkolakis, C, and T Papageorgiou (2009). Selection, Growth, and Learning, mimeo Penn State University and Yale University

Bagwell, K (2007). The Economic Analysis of Advertising, in *Handbook of Industrial Organization*, M Armstrong, and R Porter, (Eds) vol. 3, 1701-1844. North-Holland, Amsterdam

Baldwin, R, and P Krugman (1989). Persistent Trade Effects of Large Exchange Rate Shocks, *Quarterly Journal of Economics*, 104(4): 635–54

Bernard, A, B. Jensen, Redding, S and P Schott (2007). Firms in International Trade, *Journal of Economic Perspectives*, 21(3): 105–30

Butters, G (1977). Equilibrium Distributions of Sales and Advertising Prices, *Review of Economic Studies*, 44(3): 465–91

Chaney, T (2008). Distorted Gravity: The Intensive and the Extensive Margins of International Trade, *American Economic Review*, 98(4), 1707–21

Dixit, A. (1989). Entry and Exit Decision Under Uncertainty, *Journal of Political Economy*, 97(3): 620–38

Dunne, T, Roberts, M J and L Samuelson (1988). Patterns of Firm Entry and Exit in US Manufacturing Industries, *RAND Journal of Economics*, 19(4): 495–515

Eaton, J, M. Eslava, M, Krizan, C J M Kugler M and J Tybout (2008). A Search and Learning Model of Export Dynamics, mimeo, New York University and Pennsylvania State University

Eaton, J, Eslava, M, Kugler, M and J Tybout (2008). The Margins of Entry Into Export Markets: Evidence from Colombia, in *Globalization and the Organization of Firms and Markets*, E Helpman, D Marina, and T Verdier. (Eds) Harvard University Press, Massachusetts

Eaton, J, Kortum, S and F Kramarz (2008). *An Anatomy of International Trade: Evidence from French Firms*, NBER Working Paper 14610

Grossman, G M, and C Shapiro (1984). Informative Advertising with Differentiated Products, *Review of Economic Studies*, 51(1): 63–81

Keesing, D B (1983). Linking Up to Distant Markets: South to North Exports of Manufactured Consumer Goods, *American Economic Review*, 73(2): 338–42

Kehoe, T J (2005). An Evaluation of the Performance of Applied General Equilibrium Models of the Impact of NAFTA, in *Frontiers in Applied General Equilibrium Modeling*, T.J Kehoe, Srinivasan, T and J Whalley, (Eds) 341-377. Cambridge University Press, New York

Kehoe, T J and K J Ruhl (2003). How Important is the New Goods Margin in International Trade, Federal Reserve Bank of Minneapolis, *Staff Report*, 324

Melitz, M J (2003). The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6): 1695–1725

Roberts, M J, and J R Tybout (1997a). The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs, *American Economic Review*, 87(4): 545–64

—(1997b). 'What Makes Exports Boom? The World Bank, Washington, DC Ruhl, K J, and J. Willis 2008: New Exporter Dynamics, mimeo New York University

Timoshenko, O (2009). *State Dependence in Export Market Participation: Does Exporting Age Matter?*, Working Paper, Yale University

20

Taking Advantage of Trade: The role of distortions

KALA KRISHNA

1. INTRODUCTION

One of the central themes of trade theory is that, overall, trade is a force for good. Trade people will readily agree that while there may not be Pareto improvements from trade, potential Pareto improvements are possible. This mantra, that the benefits of integration into the world economy outweigh the costs, is often invoked and is quite widely accepted. Yet, not all countries seem to have gained equally from their own liberalization or from that of their trading partners. It might even be that some have lost. Even when developed countries have offered preferential access to the poorest of the developing countries, there has been a differential ability on the part of these poorest countries to take advantage of the preferences. In fact, the very poorest seem to be the least able to take advantage of such preferences with the result that, among the less well off, gains tend to accrue to those who need them the least, rather than the most.

Here I argue that the existence of distortions broadly speaking, and their interactions with trade liberalization might well be one reason for the above to occur.[1] I should make it clear that the term distortions will be used very broadly: maybe too broadly. It will include product market distortions related to the exercise of monopoly power, as well as factor market distortions that involve paying workers other than their marginal product (whatever be the reason). It will also cover situations that might not be thought of as distortions like the existence of holdup problems (both due to corruption and technology), to legal constraints that prevent incentives from operating, and (or) prevent efficient organizational forms from being used, to the lack of infrastructure that causes power cuts and long delays, and high costs in transportation. If existing distortions broadly speaking are (at least part of) the reason why some countries may have failed to gain from greater integration with the world market, then it is important to identify these distortions and alleviate them before, or at least along with, urging such countries to liberalize. Research directed towards identifying and empirically documenting the existence of such distortions is then a precursor to good policy advice.

[1] This is a standard result in the theory of the second best.

The idea I want to push is simple and is perhaps best made through an analogy. Think of an island where all the natives are effectively one legged because they are required by law to strap up one leg. Despite this handicap relative to the two legged, they manage to subsist on the flora and fauna of the island. Although they have some trouble climbing trees to pick fruit, and maybe getting around while hunting is slow, they manage. Now think of what happens when other humans not subject to this stricture arrive. Being faster, they get to the fruit and the prey before the one legged do so that there is nothing left for the one legged who then starve.

Is there some wage at which the one legged can be employed? Maybe not. If the two legged can always beat the one legged to the spoils, the one legged will be quite useless (unless there is an excess supply of fruit trees or prey, relative to labor, so that this is not the case). What does this have to do with trade? Well, think of the one legged as the individuals (or firms) operating in the domestic distorted economy and the two legged as foreign individuals (or firms). Think of trade as opening up your economy to the two legged. Before trade, even if your productivity was low, your people survived. After trade, they are helpless against the foreigners. This does not mean that trade should be abjured. Rather, it should be taken as a call to get the law that required one leg to be strapped up revoked. In other words, domestic distortions should be fixed before opening up to trade, as trade can often make them much more pernicious.

I will proceed as follows. First, I will lay out the received wisdom on trade effects in second best situations. I will begin with some simple ideas from the theory of the second best. Then I will outline some insights that are obtained from work in two directions, namely the presence of labor market distortions with indivisibilities in consumption, and the presence of credit constraints. Finally, I will speculate on some as of yet under-researched ideas about why countries may be able to exploit access to world markets differentially. These include ideas on the role of infrastructure and the importance of sunk costs in product choice, which suggest some promising directions for future research.

2. LESSONS FROM THE THEORY OF THE SECOND BEST

The main result from the theory of the second best is that in the presence of existing distortions, relaxing a single distortion, or a group of distortions need not raise welfare. What does this result have to do with trade? The inability to trade can be thought of as a distortion resulting in domestic prices not being aligned with world prices. In the presence of other market distortions, removing or reducing this distortion (that is, opening up to trade or liberalizing trade) could easily reduce welfare. This insight, in and of itself, is not particularly useful. What one wants to know for this insight to have any remote policy relevance is which distortions are likely to be aggravated by trade. In general, the following presumption is not too far from being true: product market distortions are less likely to be made worse by trade than are factor market ones.

2.1 Product market distortions

Since trade integrates product markets, distortions due to market power in the product market are likely to be diminished by trade. This is the thrust of much of the (now) older work in trade with imperfect competition and increasing returns with free entry. Opening up a country to trade raises the total number of (domestic and foreign) firms in the market thereby reducing their market power as reflected in price cost margins, as well as raising the output of each firm, thereby allowing them to exploit economies of scale better and reduce average costs.[2]

However, even in this case there is an important caveat: welfare is the sum of producer surplus, consumer surplus, and net government revenue; when profits are present trade could well have adverse welfare effects due to profit shifting to foreign firms. Although there might be gains from lower prices for consumers, the losses due to profit shifting of trade liberalization could outweigh these gains. This was the thrust of the strategic trade policy literature of the 1980s. However, the empirical relevance of this literature was soon seen as limited. Two exceptions to this latter statement are particularly worth noting. First, the presence of other distortions can amplify the extent of the welfare effects of strategic trade policy. For example, in his work on automobiles Dixit (1988) shows that in a simple model calibrated to the US automobile industry in the 1980s, the potential gains from strategic trade policy would be at most in the millions of dollars. However, once labor rents accruing to workers (due to the presence of trade unions who jack up wages) are accounted for, this number can morph into the realm of billions of dollars. Second, in situations with a small number of players, policies that seemed to be marginal, like setting non-tariff barriers (NTB) at supposedly non-binding levels, could end up being far from such. Thus, even quotas set at the free trade level of imports could end up restricting trade. The reason is that in strategic environments it could be worthwhile for one party to make the NTB bind on the other, and this could result in seemingly innocuous policies affecting trade flows. This is the thrust of the work on NTBs in strategic environments.[3]

2.2 Factor market distortions

In trade, work on factor market distortions (FMD) has been targeted for the most part to the effects of minimum wages. See for example, the classic work of Brecher (1974a; b) which looks at the effect of a minimum wage distortion on an open economy[4]. The more recent work of Davis (1998) builds on this work and looks at the effects of trade between an economy with a minimum wage distortion (Europe) and one without it (the United States). Davis argues that trade may simultaneously prop up US wages and cause greater unemployment in Europe). Thus, opening up to trade could well make Europe much worse off. It is an

[2] This is very well laid out in Dixit and Norman (1980).

[3] See Krishna (1990) for an overview of this area.

[4] The minimum wage distortion in this paper is exogenously specified. Brecher (1992) develops an efficiency wage model with an endogenous factor market distortion which results in unemployment.

excellent example of the second best principle at work. In contrast to Davis's work, the endogenous distortion in the work reported below results in resource misallocations instead of unemployment.[5]

2.2.1 A Simple Ricardian setting

Rather than use a minimum wage distortion, think of the FMD as an inability or unwillingness of firms, in at least part of the economy, to distinguish between the ability of heterogeneous workers, even though all markets are otherwise perfectly competitive. In the former socialist economies, the state-owned sectors (the distorted sector) usually paid a flat wage per worker which was only loosely related to ability.[6] If other sectors are un-distorted and pay a productivity based wage, the best workers will be attracted to the un-distorted sector while the lower ability ones flock to the distorted sector.

Similarly, in developing economies, agriculture is run along family farm lines so that workers in agriculture (the distorted sector) can be thought of as obtaining a fixed wage rather than the value of their marginal product. When workers differ in their abilities, this leads to higher ability workers leaving agriculture. Of course, the higher the wage, the higher the average quality of worker attracted to the sector offering a uniform wage per worker.

Suppose that there is this distortion in one sector, while in other sectors of the economy this distortion does not operate and workers are paid according the their marginal product. This is depicted in Figure 20.1. Think of γ as the effective units of labor in a worker of type γ. In the un-distorted sector, the unit requirement of effective labor is unity and this sector's output (Y) is taken as the numeraire. A worker can therefore always make γ by working in the un-distorted sector so that if the wage per worker is w, all workers of type less than w will work in the distorted sector making good X, while the remaining will choose to make Y.

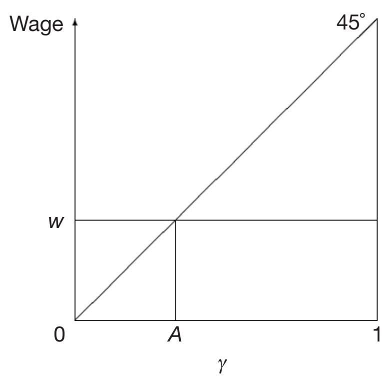

Figure 20.1: *The Allocation of Labor Between Sectors in the Distorted Economy*

[5] The discussion below is based on Krishna and Yavas (2005) and Krishna et al. (2005).

[6] Jefferson (1999) argues that 'the inability of state enterprises to monitor and reward high quality labor is likely to create an adverse selection problem in which the most skilled and motivated workers exit from the state sector...'.

It is easy to see from Figure 20.1 that the distortion results in a firm paying all but the marginal worker willing to work for the firm's offered wage more than their opportunity cost. This overpayment, in turn, raises the firm's costs above the level that would prevail if the firm could distinguish between differentially able workers.

The effect of the distortion on production (and hence consumption) in autarky is straightforward: as the cost is high, so is the price (as this equals cost) and consequently, demand is low. Thus, relative to the first best, too little of the distorted good is made and consumed. In our Ricardian setting, since there is no unemployment and a single factor, effective labor, the economy remains on the production possibility frontier (PPF), but at the wrong point on it: at *B* in Figure 2, not *A*, which is the first best.[7]

In autarky, the effect of the distortion on welfare depends on the extent of substitutability in consumption. The greater the substitutability, the greater the deleterious effects of the distortion in autarky; since the price of the distorted good is higher than in an un-distorted economy, consumers subtitute away from it a lot when substitutability is high, causing far too little of the distorted good to be produced (as compared to the efficient level). In the extreme, when the goods are perfect complements, the consumption levels are the same as in an un-distorted economy.

Under trade, consumption and production are de-linked. The distorted economy has a comparative disadvantage in the distorted sector: thus, it is imported. Hence, opening up to trade involves making even less of the distorted good: in Figure 2, the economy specializes in *Y* at *C*. This effect reduces welfare at the production point as moving from *B* to *C* at given prices can only reduce welfare.

However, having access to cheaper *X* makes the price line steeper and this effect raises welfare, as the economy is specialized in *Y*. The more substitutable the goods are in consumption, the greater the welfare gain via this price effect.

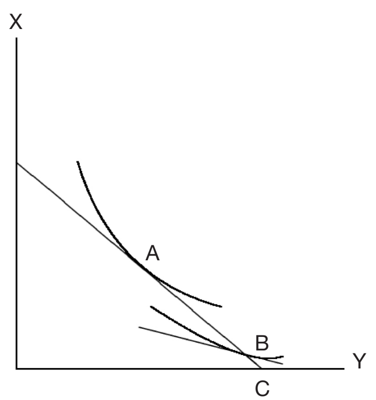

Figure 20.2: *Autarky and Trade*

[7] Factor market distortions (FMD) could place the economy on the wrong point on the production possibility frontier, or, even with full employment, put the economy inside the frontier when there are many factors.

The important thing to note is that FMDs tend to be made worse by opening up to trade. The FMD raises the cost, and hence the price of the good made by the distorted sector. As a result, in the absence of trade, too little of this good is produced and consumed relative to the first best. But when this country opens up to trade, it will have a comparative disadvantage in the good made by the sector with the FMD. Thus, it will import it and reduce its own production of the good. But since too little of the good was being produced to begin with, this will make the production point deviate even more from the first best resulting in possible losses from trade.

Note that if all sectors are distorted, then, even with heterogeneous workers, there is no place for high ability workers to be paid according to their ability. Thus, workers will not sort across sectors but instead each sector will get workers of average ability, just as it would if it paid workers according to their ability. Thus, in general equilibrium, there is no misallocation of workers across sectors. However, if some sectors become market driven, then the distortion hurts. This suggests that it is the economies in transition with both state run and free enterprise sectors that would be most likely to suffer losses from trade. This may help explain why transition countries are slow in reaping the full extent of gains from trade. They may even be hurt by trade.

The kinds of issues that arise with such FMDs are amplified by the existence of indivisibilities. Think again of a two-good world. One of the goods, which can be thought of as a lumpy consumer good like a refrigerator or a car, is indivisible in consumption: either zero or one unit of it can be consumed.[8] Moreover, this indivisible good is highly valued, meaning that consumers are much better off if they can afford the good, than if they cannot. As the good is indivisible, only consumers with incomes high enough (above the price) can afford the good.[9]

Think of this indivisible good (*X*) as being made in the state run sector where all workers are paid the same, independent of their productivity. Of course, the higher the wage offered in this sector, the more workers choose to work there. As explained above, paying workers a wage independent of their ability raises production costs above what they would be otherwise.[10] But now, this is not the end of the story. If workers are well paid, they can afford the indivisible good, and because of this, the demand for the indivisible good is high, (in fact, everyone buys the indivisible in this good equilibrium), which makes the demand for the workers in that sector high and keeps their wage high. On the other hand, if workers are poorly paid, this chain of causation works in reverse. If wages are low, demand for indivisibles is low (in fact only the ablest working in the un-distorted

[8] The good itself can differ according to the level of development and particular needs. In poorer settings, this might be an item as small as a water purifier, or a wood or gas cooking stove, or a radio or TV.

[9] Although goods can be made divisible by renting or sharing, an essential indivisibility remains since it is usually much more costly to rent than buy. Why not share the good then? This seems hard as there is moral hazard problems involved in sharing.

[10] To fix ideas, think of a mixed economy like India before reforms or the former Soviet Union: much of the economy is state run, though there is a private sector.

sector can afford the indivisible), which keeps demand for workers and wages in the state run sector too low for workers there to afford the indivisible. In this way, an FMD together with indivisibilities can give rise to two equilibria: a good one where all agents can afford the indivisible and a bad one, where only a select few can do so.

Where are such multiple equilibria likely to occur and when is there a unique equilibrium? Multiple equilibria tend to occur when the economy is productive, but not too productive. If the economy is too unproductive, then it just has not got the resources to produce enough to meet high demand levels so that the good equilibrium cannot exist. If the economy is too productive, then costs are so low that everyone can always afford the indivisible[11] so that only the 'good' equilibrium exists in autarky. In between lies the realm of multiple equilibria. Thus, it may be that in developing countries, like China in the past, the good equilibrium might not have been viable at all. However, the good equilibrium might have been a possibility in the former Soviet Union, where despite all the inefficiencies of the system, the standard of living was reasonably high before its collapse.

In such a setting, there is reason to expect trade to have large adverse welfare effects for the distorted economy. Why? Well, think of this distorted closed economy in the good equilibrium. Even though costs are high in the distorted sector, the high wages in the state run sector (where wages per worker have to (where wages per worker have to be lower than those in the non-distorted sector as all workers could choose to work in the latter) ensure that everyone has the ability to buy a stove, or refrigerator or scooter,....[12] Now think of opening the economy up to trade. As the distorted economy has higher costs, and hence a comparative disadvantage in indivisibles, this good equilibrium cannot survive after opening up to trade. Indivisibles will be imported, at lower prices, but this will not help the less able who have lost their well paying jobs. Of course, the abler will be better off but society as a whole will be worse off when the loss of welfare of the less able is taken into account. If the world price under trade is close to the autarky cost of indivisibles in the distorted economy (as it would be if the distorted economy was 'large' so that its autarky prices prevailed) then there could even be a weak Pareto loss in welfare from opening up to trade.

While these models are special, the ideas that emerge from them make sense at a basic level. Maybe the citizens of the former Soviet Union did not have access to the luxuries and the quality of products in the West, but before its fall they had a pretty decent standard of living. Once the economy opened up, no one wanted

[11] If the economy is too unproductive, then this good equilibrium where everyone can afford the indivisible cannot be supported: it takes more workers than are available in the economy to make enough of the indivisible for everyone.

[12] Although the economy is distorted, the less able gain a lot from this distortion, though the abler lose as they pay higher prices for the indivisible.

Soviet style goods, so these sectors closed down, with consequent devastating effects on incomes and welfare of those working there. Of course, those who succeeded in the new Russia were much better off, but it is hard to argue that on average this was the case, especially in the early years of the transition.

These ideas also have some relevance for development economics. Family farming results in workers earning the average rather than the marginal product in agriculture. When workers are identical in ability and marginal product is diminishing, as has been assumed in this literature, average product exceeds marginal product so that too many workers remain in agriculture. In the development literature this distortion has been linked with the concept of disguised unemployment (see Sen 1960). However, when labor varies in ability, as in the above model, only lower-ability labor remains in agriculture. The marginal worker in agriculture in effect subsidizes all other workers in the sector, as he obtains a wage below his marginal value product in the sector. As a result too little effective labor remains in agriculture rather than too much, which is the opposite of what is predicted in the classic work on disguised unemployment.[13] The effect of the distortion on output is the same. In autarky, too little of the distorted good is made and its price is too high. As a result, the distorted economy has a comparative disadvantage in the distorted good which is imported when the economy is opened up. This reduces the output of the distorted good and worsens the distortion. On the other hand, trade results in the usual price effects that raise welfare. Thus, welfare may rise or fall as a result of trade liberalization. However, a large distorted economy always loses from trade as, by definition, it does not reap any beneficial price effects.[14]

2.3 Credit constraints

Credit constraints operate in both static and dynamic contexts. In the static context, credit constraints will tend to limit the size of the sector that is dependent on credit. In a now familiar manner, one could argue that to the extent that credit constraints raise the costs of the credit-intensive sector relative to the rest of the economy, the economy will make too little of the credit-intensive sector's output. Moreover, trade will likely make this output distortion worse as the economy has a comparative disadvantage in the credit intensive sector.

More interesting and unique are the dynamic effects. Here I will draw upon Krishna and Chesnokova (2009) and Chesnokova (2007). The former paper looks at steady state equilibria only. It shows that in a general equilibrium setting, where the acquisition of skills by heterogeneous workers is explicitly modeled, one can think of steady state supply much as we do in a static model. In a steady

[13] This is in line with the observation that the young and able are disproportionately represented in those migrating from rural areas, with children, women, and the elderly staying behind.

[14] This is perfectly in line with the literature on the theory of the second best (see Lipsey and Lancaster,1956) where a recurring theme is that in the presence of existing distortions, a reduction or removal of a distortion can lower welfare. See, for example (Ethier 1982).

state, if credit constraints are not binding, steady state relative supply is shown to be increasing in price as in static settings. If credit constraints are binding however, steady state relative supply need not be increasing in price, and consequently, multiple steady state equilibria may exist. Non-monotonicity of relative supply may even result in trade equilibria where the country ends up importing the industrial good (which is intensive in its use of credit) at prices higher than its autarky price. As a result, there are potential losses from opening up to trade.

The latter paper is particularly fascinating. At the modeling level, it takes the classic paper of Banerjee and Newman (1993) (which deals with occupational choice in an overlapping-generations type model, in the presence of credit constraints, in a closed economy setting) and extends it to an open economy. It is the first paper that provides a clean explanation of how 'immizerizing deindustrialization' can occur. It has been argued (see, for example the deindustrialization debate in India under British rule) that particular trade and (or) domestic polices in India under the British resulted in the irreversible loss of certain industries, and that this was bad for welfare.

In her work, Chesnokova (2007) shows how trade can not only cause deindustrialization, but how this deindustrialization can reduce welfare. Think of an economy making two goods: the industrial good which needs up-front investment, and the agricultural good which does not. Trade can cause deindustrialization by bringing the price of the industrial good down below the level at which bequests are large enough to permit the next generation to invest in the sector where large investments need to be made up front. In the presence of credit constraints, the inability to make such bequests causes the industrial sector to shrink over time, which explains how deindustrialization can occur. But why should this deindustrialization be immizerizing? Deindustrialization cannot be welfare reducing if the price of the industrial good remains low. However, as the industrial sector shrinks, the price of the output of this sector will rise. Yet, the sector will be unable to expand the way it shrank, as the absence of credit will prevent agents from moving into the sector freely. In this way, she argues, not only can trade destroy the industrial sector, but it can also raise its price above that prevalent under autarky. This is what makes its demise immizerizing. She shows that this immizerizing deindistrialization cannot happen if the agricultural sector is productive enough, but it can occur if it is not very productive because this is when agricultural agents cannot leave bequests to their offspring that are large enough for them to become industrialists.

While this work is quite stylized, there is enough truth in it for this to be a cautionary tale for developing countries where the operation of credit markets tends to be relatively poor. Interest rates in such countries can often be well over 100 per cent per annum and this could well make the loss of sectors, especially the ones where up-front investments are needed, very hard to reverse.

3. THE SECOND BEST IN DISGUISE

A concern that is often expressed these days is that the countries that seem to have gained the most from trade are the ones that have needed it the least. The poorest and most mismanaged ones have been the slowest to reap the benefits of globalization. An interesting take on this can be found in the work of Demidova (2008). Her paper looks at two features of globalization, productivity improvements, and falling trade costs and considers their effect on welfare in a monopolistic competition model with heterogenous firms (a la Melitz (2003)) and technological asymmetries. Productivity improvements are interpreted as having a better productivity distribution from which to draw.[15] She shows that improvements in a partner's productivity always hurt a country. Her reason is not the usual one that relies on adverse terms of trade effects resulting from an improvement in productivity.[16] Rather, it is really the second best theorem in disguise. In her work there are two sectors: the first which makes a homogeneous good and is competitive which we can call agriculture. The other sector (call it manufacturing) makes differentiated goods and is monopolistically competitive. In this sector firms are heterogeneous *ex post* but homogeneous *ex ante* in their productivities as they only discover their productivity after they make a fixed and sunk investment. Firms choose where to set up shop and there is free entry.

The logic of her result is that as there is market power in manufacturing, firms hold back supply to raise price and hence, too little is produced relative to the first best. Trade further reduces output in the country with the worse productivity distribution as firms are attracted to locations where they can get a better productivity draw. Thus, trade reduces the entry and hence the output of industry in the technologically backward country, which results in losses from trade. She also shows that falling trade costs result in disproportionate gains to more technologically advanced countries, as a fall in trade costs puts the country with lower costs in a better position to exploit its advantage.

Her results seen in this light are no mystery, yet the implications for policy are quite profound once one asks from where these differences in productivity distributions might be coming. It is likely that they arise from differences in infrastructure and institutions broadly speaking.[17] Onerous labor regulations that prevent the firing of workers, or poor roads and ports will in effect reduce the productivity of a firm in the poorly managed country making it a bad place to locate. This could explain why many Indian multinationals choose to headquarter, not in Mumbai or Delhi, but in Singapore. Similarly, the vast difference in infrastructure investment in India and China could be a large part of the reason for their differences in their current per capita incomes.

At the micro micro level, these differences in infrastructure could be the reason for the difficulty poor countries seem to have in accessing world markets. This

[15] Better means hazard rate stochastic dominance.

[16] Productivity improvements, in essence, shift the relative supply of the exportable outwards, thereby reducing its price.

[17] See Krishna (2007) for more on this and the constraints on Indian development prospects.

difficulty is both via the obvious direct effect of limited port facilities which creates long delays in ships berthing as well as corruption or inefficient labor practices, or more subtly via holdup problems. It is clear that farmers in developing countries are unlikely to get rich growing staples like wheat or rice. Rather, the money seems to be in making high value specialty products like fresh flowers or exotic fruit for which the developed world is willing to pay what seems like a fortune to the poor.

So why does this not happen? One might argue that the reason is that the farmers have no idea that this demand is out there and need to engage in discovery. But this can be, at best, the smaller part of the story. Even if the farmers knew what to make, they would probably not make it. Why? All the products mentioned above are perishable. This makes sellers open to the classic holdup problem. Once the flowers are ready, or the fruit is ripe, the farmer cannot consume it himself. Nor can he credibly hold on to it if the buyer tries to take advantage of its perishability to offer a low price *ex post*. This makes the farmer wary of getting into this business *ex ante*. After all, if the price of wheat is low, he can store it and eat it. At least he will not starve. But this is not the case if he is growing flowers. The situation would would be made worse by bad roads that limited the number of buyers coming to the village leading to markets that are very thin. This will clearly exacerbate the holdup problem as it will be harder for a farmer to find an alternative buyer if he is held up by his original one.

Why not contract on the price of the product *ex ante*? This will do little to alleviate the holdup problem if the courts are overloaded or corrupt so that it is hard to sue for breach of contract. In this manner, poor judicial systems and corruption also make this holdup problem worse. As a way around such problems, innovative contracts arise. For example, in India, it used to be the norm that mangoes were sold to the buyer after they were ripe. This left farmers open to holdup as discussed above. In recent times, a new contractual form has sprung up where farmers rent out the mango trees to the buyer once the flowers have formed and a reliable estimate of yield is possible. The renter then takes care of the tree and all else for the season with full rights to the fruit. As the farmer gets the payment up front, there is no holdup problem on his side which makes him willing to make the long-term investments in trees needed to increase production. There may be inefficiencies here as the farmer may be better placed to monitor and protect the trees, but at least it helps solve the holdup problem.

All of the above suggests that a broad interpretation of the theory of the second best calls for government to look actively for ways to deal with such issues. Whether these ways involve creating cooperatives that offer fair prices and distribution networks for perishable products, or building roads to distant places, or investing in port facilities and removing onerous regulations, removing these impediments is vital to accessing world markets and reaping the gains from globalization.

Kala Krishna is Liberal Arts Research Professor in the Department of Economics at The Pennsylvania State University.

BIBLIOGRAPHY

Banerjee, Abhijit and Andrew Newman (1993). Occupational Choice and the Process of Development. *Journal of Political Economy*, 101: 274–97

Brecher, Richard (1974a). Optimal Commercial Policy for a Minimum Wage Economy. *Journal of International Economics*, 42: 139–49

Brecher, Richard (1974b). Minimum Wage Rates and the Pure Theory of International Trade. *Quarterly Journal of Economics*, 88(1): 98–116

Brecher, Richard (1992). An Efficiency-Wage Model with Explicit Monitoring: Unemployment and Welfare in an Open Economy. *Journal of International Economics*, 32(1): 179–91

Chesnokova, Tatyana and Kala Krishna (2009). Skill Acquisition, Credit Constraints, and Trade, *International Review of Economics & Finance* special issue on Research in trade: Where do we go from here? 18(2): 2009

Chesnokova, Tatyana (2007). Immiserizing Deindustrialization: A Dynamic Trade Model with Credit Constraint. *Journal of International Economics*, 73: 407-420

Demidova, Svetlana (2008). Productivity Improvements and Falling Trade Costs: Boon or Bane?, *International Economic Review*. November 2008, 49(4): 1437–1462

Davis, Donald (1998). Does European Unemployment Prop Up American Wages?: National Labor Markets and Global Trade. *American Economic Review*, 88(3): 478–94

Dixit, Avinash and Victor Norman (1980). *Theory of International Trade: A Dual, General Equilibrium Approach*. Cambridge University Press

Dixit, Avinash (1988). Optimal trade and industrial policies for the US automobile industry, in *Empirical Methods in International Trade* (Ed.) Robert Feenstra, Cambridge, MA: MIT Press, 141–65

Ethier, Wilfred (1982). Decreasing Costs in International Trade and Frank Graham's Argument for Protection. *Econometrica*, 50(5): 1243–68

Jefferson, G, (1999). Missing market in labor quality: The role of quality markets in transition,.William Davidson Institute Working Paper No. 260, University of Michigan

Krishna, Kala (1990). Trade Policy with Imperfect Competition: A Selective Survey, with M. Thursby in *New Developments in Trade Theory: Implications for Agricultural Research*, Colin Carter et al. (eds.), Oxford: Westview Press, 1990: 9–35

Krishna, Kala, Abhiroop Mukhopadhyay and Cemile Yavas 2005.Trade with Labor Market Distortions and Heterogeneous Labor: Why Trade Can Hurt? in G S Heiduk and K-Y Wong (Eds.), *WTO and World Trade: Challenges in a New Era*, Heidelberg, Germany: Physica-Verlag, 2005

Krishna, Kala (2007). Review of Dilip Mookherjee, The Crisis in Government Accountability: Essays on Governance Reform and India's Economic Performance. *Journal of Economic Literature*, 45: 211-214

Krishna, Kala and Cemile Yavas (2005). When Trade Hurts: Consumption, Indivisibilities and Labor Market Distortions, with C Yavas. *Journal of International Economics*, 67(2): 413–27

Lipsey, R G and K Lancaster (1956). The General Theory of Second Best. *Review of Economic Studies*, 24: 11–32

Melitz, Marc J (2003). The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity,. *Econometrica* 71(6): 1695-1725

Sen, A, (1960). *Choice of Technique*. Oxford: Blackwell

21

Credit Constraints and the Adjustment to Trade Reform

KALINA MANOVA

1. INTRODUCTION

In standard trade theory, gains from international trade result from countries exploring their comparative advantage as shaped by consumer preferences, factor endowments, and production technologies. The more recent trade literature emphasizes the role of firm heterogeneity, and suggests that reallocations across sectors and across firms within a sector are equally important for the aggregate gains from trade. As powerful as these frameworks are, they abstract from market frictions that may arise from agency problems, and presume that entrepreneurs can freely enter any industry or expand production. In practice, however, firms require external financing for their operations, and frequently face borrowing constraints.

A growing literature on trade and finance has established that credit constraints are an important determinant of global trade patterns. This literature has largely examined the effects of financial frictions on countries' trade flows in steady state, without studying the response to changes in trade policy. This article surveys these recent theoretical and empirical developments, and discusses their implications for the role of credit constraints in the adjustment to trade reform. It also identifies promising directions for future research in this area.

Why should credit constraints matter for international trade flows? Firms often incur substantial up-front costs which they can profitably recover only after realizing sales revenues. These costs may be either sunk, in the sense that they need to be paid only once upon entry into an industry, market, or product line, or recurrent fixed per-period costs. For example, firms may have to engage in research and development (R&D) and product development, marketing research, advertising, or investment in fixed capital equipment. Although some variable costs such as employment compensation and equipment repairs and (or) depreciation are typically paid at the end of a production cycle, other variable expenses may also be up-front, such as intermediate input purchases, advance payments to salaried workers, and land or equipment rental fees.

Exporting is associated with additional sunk, fixed, and variable costs that make production for foreign markets more costly than manufacturing for the

home country. Sunk and fixed trade costs include learning about the profitability of potential export markets; making market-specific investments in capacity, product customization, and regulatory compliance; and setting up and maintaining foreign distribution networks. The additional variable costs of trade comprise shipping, freight insurance, and applicable trade duties. As with production costs, most of these trade expenses may have to be incurred before export revenues are realized.

Firms are not always able to meet their liquidity needs with retained earnings or cash flows from operations, and routinely rely on external financing for their production and export expenditures. This financing often comes in the form of bank loans or bank-provided trade credit. In addition, private parties on two sides of a transaction may also offer trade credit to each other. For instance, final-good purchasers (importers) may secure trade financing for their suppliers (exporters), and final-good producers (exporters) may extend trade credit to their intermediate input suppliers. In practice, these types of buyer or supplier trade credit require bank guarantees and ultimately also depend on firms' access to bank financing.

In the presence of financial frictions—because of imperfect contractibility or a limited pool of available financial capital in an economy—credit-constrained firms may not be able to become exporters or to export to their full potential. Evidence suggests that sectors differ greatly in their requirement for external finance and their availability of assets that can be collateralized—assets that can be used to secure credit. For these reasons, borrowing constraints can reduce a country's aggregate exports and affect their sectoral composition by limiting the investment opportunities open to producers with insufficient private capital.

The trade and finance literature has indeed documented large economic effects of credit constraints on trade. The main finding in this literature is that financially developed countries export greater volumes and a wider range of products to more destinations. Moreover, these patterns are especially pronounced in financially vulnerable sectors. Section 2 below reviews this country-level evidence and some recent firm-level studies. It also presents a useful theoretical framework for thinking about the role of credit constraints in the context of international trade.

Most papers on the link between trade and finance have focused on the exporting country's domestic financial development. New evidence suggests that foreign capital flows can compensate for an underdeveloped domestic financial system, though this topic remains underexplored. Section 3 discusses recent work on the role of equity market liberalization and foreign direct investment in alleviating credit constraints and stimulating trade flows.

Building on the intuition and results developed for the effects of credit constraints on trade in steady state, Section 4 explores the role of financial frictions in the adjustment to trade reforms. The goal of this section is to provide informed priors and outline an agenda for future research. Since trade policy changes are most relevant for developing and emerging economies where credit constraints are most acute, the discussion focuses on the response of financially underdeveloped countries to trade liberalization. The last section concludes and considers the merits of concurrent and sequential trade and financial sector reforms.

2. THE EFFECTS OF CREDIT CONSTRAINTS ON TRADE

2.1 Theoretical framework

The literature on trade and finance has offered a number of theoretical frameworks to rationalize the effects of credit constraints on trade. A common prediction of these models is that financially developed countries have a comparative advantage in financially vulnerable sectors.[1] This subsection closely follows Manova (2007), who incorporates credit constraints in a heterogeneous-firm model of trade à la Melitz (2003).

In the model, credit constraints affect firms in different countries and sectors differentially. For technological reasons, firms in some sectors have greater liquidity needs and must finance a bigger share of their export costs externally. Industries also differ in their endowment of tangible assets that can serve as collateral. Thus, entrepreneurs find it more difficult to start exporting in financially vulnerable sectors since they need to raise more outside finance, or because potential investors expect a lower return in case of default. In addition, credit constraints vary across countries because contracts between firms and investors are more likely to be enforced at higher levels of financial development. When a financial contract is enforced, the borrowing firm makes a payment to the investor; otherwise, the firm defaults and the creditor claims the collateral. Firms thus enjoy easier access to external finance in countries with stronger financial contractibility.

In the absence of credit constraints, all firms with productivity above a certain cut-off level would become exporters, as in Melitz (2003). Credit constraints, however, interact with firm heterogeneity and reinforce the selection of only the most productive firms into exporting: Because more productive firms earn bigger revenues, they can offer creditors a higher return in case of repayment, and are thus more likely to secure the outside capital necessary for exporting. This pattern is consistent with evidence in the corporate finance literature that smaller firms tend to be more credit constrained.[2]

The model implies that the productivity cut-off for exporting will vary systematically across countries and sectors. It will be higher in financially vulnerable industries which require a lot of external finance or have few assets that can be collateralized and lower in countries with high levels of financial contractibility. Importantly, the effect of financial development will be more pronounced in financially vulnerable sectors. Credit constraints thus preclude potentially profitable firms from exporting and result in inefficiently low levels of trade participation.

Note that, for a given distribution of productivity across firms, the lower the productivity cut-off for exporting, the more firms can sell in foreign markets. If

[1] See Kletzer and Bardhan (1987); Beck (2002); Matsuyama (2004); Ju and Wei (2005); and Becker and Greenberg (2007) among others. The Ricardian, representative-firm nature of these models delivers the counterfactual prediction that either all or no producers in a given sector will become exporters.

[2] See, for example, Beck et al. (2008); Beck et al. (2005); and Forbes (2007).

the productivity cut-off for exporting to a particular destination is too high, no firms will be able to enter that market and there will be zero exports at the country level. Therefore, countries will be more likely to export to any given destination in a financially vulnerable sector if they are more financially developed. Given positive exports, in financially vulnerable sectors more firms will participate in trade when located in financially advanced economies. If firms produce differentiated goods, this will also be reflected in greater product variety in country exports.

When firms need to raise outside funds to finance their variable production and trade costs, credit constraints will affect not only firms' decision to export, but also their scale of operations. While the most productive (and least constrained) exporters will be able to trade at first-best levels, less productive firms will only be able to export if they ship lower volumes than would be optimal in the absence of financial frictions. Such firms are able to secure less outside credit than would be necessary to trade at first-best levels, and use it to support lower export quantities which entail lower variable costs. The extent of this distortion will vary systematically across countries and sectors. In particular, firms located in financially developed countries will be able to export greater volumes, especially if they are active in a financially vulnerable sector.

To summarize, credit constraints affect both the extensive (number of firms exporting; number of export destinations) and the intensive (firm-level exports) margin of trade. In the aggregate data, this will manifest itself in financially developed countries having a comparative advantage in financially vulnerable industries. Countries with strong financial contractibility will ship greater quantities of exports to more destinations, especially in sectors with high external finance dependence and sectors with few tangible assets.

This theoretical framework abstracts from sunk costs as well as cost or demand shocks associated with exporting. There is, however, evidence of hysteresis in countries' and firms' participation in international trade, which has been ascribed to substantial sunk costs of entry into foreign markets. At the same time, recent product- and firm-level studies find frequent exit and re-entry into exporting. This churning suggests that either sunk costs are not as large as previously believed, or shocks to profitability are very volatile.

A dynamic model with sunk costs, idiosyncratic shocks, and credit constraints remains a fruitful area for future research. *A priori*, easier access to external financing should help firms to cover their sunk costs and enter into exporting. The effects of credit constraints on firm survival and continued exporting in a given market, however, would likely be ambiguous. On the one hand, as Manova (2007) shows, firms in financially developed countries would be able to overcome cost shocks more easily and continue selling in a foreign market. On the other hand, greater availability of credit would reduce the option value of continuation since it would be easier to finance the sunk cost of entry, should the firm exit and decide to re-enter in the future. The net effect of financial frictions on firm dynamics could thus be theoretically ambiguous. Finally, a richer dynamic model of exporting would also allow credit-constrained firms to retain earnings and accumulate resources until they enter into exporting at the optimum time.

2.2 Country-level evidence

There is by now strong and robust empirical evidence that credit constraints are an important determinant of global trade patterns. Most of this evidence comes from the analysis of country-level trade flows that exploits the variation in financial development across countries and the variation in financial vulnerability across sectors.[3] In particular, a number of studies have found that financially advanced economies export relatively more, especially in sectors with greater requirements for external capital, and sectors with few assets that can be collateralized (Beck, 2002; 2003; Svaleryd and Vlachos, 2005; Hur et al., 2006; Becker and Greenberg, 2007; Manova, 2007). Moreover, Manova (2007) shows that financial frictions have a sizeable effect on trade above and beyond their direct impact on domestic production. This is important given the results in the finance and growth literature that financially vulnerable sectors are larger and grow faster in financially developed countries. In the data, the distortion to domestic production is responsible for only 20 to 25 per cent of the total effect of credit constraints on trade.

A number of different indicators of financial development have been used in the literature. Two common proxies are stock market capitalization and the amount of credit extended by banks and other financial institutions to the private sector, both as a share of GDP. These outcome-based measures reflect the actual availability of external financing in a country, but they may be subject to certain endogeneity concerns. Reassuringly, authors frequently present robust results using institutional indices for accounting standards, creditor rights protection, and contract enforcement. These indices reflect the potential of an economy to maintain financial contracts, and capture the notion of financial contractibility in the theoretical framework above.

The empirical proxies for sectors' financial vulnerability also stay close to the model. External finance dependence is typically measured by the fraction of capital expenditures not financed with internal cash flows from operations. Asset tangibility is similarly defined as the share of net plant, property, and equipment in book value assets. Both variables are usually constructed from US firm-level data. Researchers point to the much larger variation in these measures across sectors than among firms in an industry as an argument that the indices capture technologically determined sector characteristics exogenous to individual firms. Empirically, all that is required for identification is that the relative rank ordering of industries be preserved across countries, even if the exact measures deviate from those for the United States.

Using these standard country and industry indicators, Manova (2007) has confirmed that credit constraints reduce countries' total exports by affecting all margins of trade. Financially developed countries are more likely to export to any given destination, and thus transact with more trade partners. Conditional on

[3] This approach was introduced in the finance and growth literature (see Rajan and Zingales, 1998 among others).

entry into a given export market, financially advanced economies also export more to that country. All of these effects are more pronounced in financially vulnerable sectors. These results are consistent with the idea that firms incur trade costs in each market they enter and face credit constraints in their financing. Analyzing bilateral trade flows also provides more convincing evidence for the importance of external credit because it permits the inclusion of destination or destination sector-specific fixed effects which control for differences in demand and trade costs.

The effect of financial frictions on bilateral trade flows can be further decomposed into an extensive margin (number of exporting firms) and an intensive margin (firm-level exports). Ultimately, analyzing firms' participation in international trade will reveal the exact mechanism through which credit constraints affect exporting and allow more specific policy recommendations. In the absence of systematic firm-level data across countries and sectors, Manova (2007) examines the number of products that countries export as an imperfect proxy for the extensive margin of trade. She finds that financially advanced countries ship a broader range of goods to any given market, and this pattern is stronger in financially vulnerable sectors.

Manova (2007) also adopts the two-stage estimation procedure developed by Helpman et al. (2008). This approach exploits the information contained in the data on both zero and positive bilateral exports to infer what fraction of domestically active firms export to each destination. The results suggest that a third of the effect of financial development on trade values is attributable to firm selection into exporting, while two-thirds is due to reductions in firm-level exports. This indicates that firms face binding credit constraints in the financing of both fixed and variable export costs. By contrast, if firms required outside capital to cover only the fixed costs of trade, financial market imperfections would have distorted only the extensive margin of exports.

What is the economic magnitude of these effects? Recall that financial frictions both reduce countries' aggregate exports and alter their sectoral composition. Evaluating the former effect has been difficult because it requires the estimation of cross-country regressions of trade on a country-level measure of financial development. The high correlation between financial development and other country characteristics, however, presents a challenge for estimating the causal impact of credit constraints. On the other hand, using interaction terms makes it possible to establish causality by showing that financial development has a differential effect on trade flows across sectors.

Comparative statics based on this difference-in-difference approach indicate that the effect of credit constraints is substantial and comparable to that of traditional sources of comparative advantage, such as cross-country differences in factor endowments. For example, if the Philippines, a country at the first quartile in the distribution of financial development, were to improve to the level of the third quartile (Italy), the Philippines could increase its textile exports (highly dependent on external finance, third quartile) by 19 percentage points more than its mineral products exports (intensive in internal funding, first quartile). Similarly, exports of low tangibility sectors (other chemicals, first quartile) would

grow 17 percentage points faster relative to sectors with harder assets (wood products, third quartile). These effects are equally important for the extensive and intensive margins of trade. A one standard-deviation increase in financial development, for example, would raise export product variety by 10 percentage points more in a financially vulnerable sector (third quartile) relative to a less dependent industry (first quartile).[4]

In this context, Chor (2009) develops a methodology for quantifying the importance of different sources of comparative advantage for country welfare. He estimates mean welfare losses of −2.1 per cent if cross-country differences in financial development were not allowed to generate comparative advantage and gains from trade. By comparison, the corresponding numbers for physical and human capital are −2.8 per cent and −3.1 per cent, respectively. While these calculations do not reflect the welfare loss from the overall reduction in trade due to credit constraints, but capture only the resulting distortions to the sectoral composition of countries' exports, they do suggest far from negligible effects.

In addition, Manova (2007) shows that credit constraints affect not only the number of a country's trade partners, but also their characteristics. An implication of the theoretical model in Section 2.1 is that export markets can be ranked by their profitability, which increases with market size and falls with transportation costs (distance). To maximize export profits, firms will therefore first export to the most profitable destination and enter additional markets in decreasing order of profitability, until they hit their credit constraint. This pattern is also predicted to hold in the aggregate data. Indeed, results suggest that the size of the largest export destination does not vary systematically across exporting countries and sectors. In other words, all countries export to the biggest markets in the world, such as the United States, Germany, and Japan. On the other hand, the smallest destination that financially advanced economies trade with has lower GDP, and this pattern is particularly pronounced for financially vulnerable sectors.

The welfare implications of this pecking order of trade have yet to be examined. If economies experience unsynchronized business cycles, exporting to more markets may provide hedging opportunities and smooth firms' export profits over time. Understanding the importance of credit constraints in this context is a topic for future research.

Finally, Manova (2007) examines the effects of credit constraints on countries' export dynamics. Over time, countries are more likely to continue exporting a given product to a specific trade partner if they are more financially developed. This effect is more pronounced for goods in sectors with greater reliance on external finance or fewer tangible assets. These results indicate that credit constraints matter in the presence of stochastic costs or other disturbances to export profitability, and are an important determinant of export dynamics. Further research could explore whether, in addition to stimulating overall trade flows, financial development improves aggregate welfare by reducing product churning and the volatility of exports.

[4] Comparative statics from Manova (2007).

2.3 Firm-level evidence

A few recent studies have used firm data to provide micro-level evidence for the importance of credit constraints in trade. These studies shed more light on the ways in which financial frictions restrict firms' export participation, and indicate that firm-level analysis is a prominent avenue for future research.

Muûls (2008) exploits firm data for Belgium, which provides information on both bilateral exports and firms' access to external financing. In particular, she uses an annual firm indicator of credit worthiness developed by a large French credit insurance company, Coface International. The Coface score is based on firm size, profitability, leverage, and liquidity, as well as industry and macroeconomic characteristics. While this score is not directly affected by firm export behavior, it is clearly endogenous to a firm's overall performance. Thus, the results capture the correlation between firm credit constraints and exporting instead of a causal effect, although the regression analysis conditions on firm size and productivity.

Muûls (2008) finds that liquidity-constrained firms are less likely to become exporters and export to fewer destinations. Conditional on participating in international trade, firms also earn greater export revenues and export more products when they have easier access to financing. Finally, less constrained firms go further down the pecking order of destinations, and the size of the smallest market they enter is systematically lower. These results are consistent with the theoretical framework above and with the idea that firms require external funds to overcome both fixed and variable costs of exporting. They also speak to the growing literature on multi-product firms. In fact, the model in Section 2.1 could be extended to firms that manage multiple product lines with varying profitability. Firms would then also follow a pecking order of products and add new goods to their export mix until they exhaust their available credit.

Berman and Héricourt (2010) find similar results using data for 5,000 firms in nine developing and emerging economies. They examine three proxies for the extent to which firms face financial constraints: the ratio of total assets to total debt, the ratio of cash flows to total assets, and the share of tangible assets in total assets. The most robust result is that credit-constrained firms are less likely to become exporters, even controlling for firm size and productivity. While less consistent across specifications, there is also some evidence that, conditional on exporting, liquidity-strapped firms ship lower values and are more likely to stop exporting from one year to the next. These findings suggest that overcoming the sunk costs of trade may be the greatest challenge credit-constrained firms face. By contrast, the recurrent fixed and variable costs of trade appear to cause smaller distortions.

A challenge for these firm-level studies is obtaining firm measures of financial constraints that are exogenous to the exporting decision. One concern is that banks may be more willing to provide loans and trade credit to firms which export, because exporting is associated with a higher revenue stream and signals high levels of unobserved firm productivity. Indeed, Greenaway et al. (2007) find

that the financial health of UK firms improves after they start exporting. At the time of entry into exporting, however, future exporters do not appear to be financially healthier than firms serving only the domestic market. By contrast, the two studies described above show that lagged credit constraints affect current firm export participation.

Further work is needed to understand the interdependencies between financial frictions and trade at the firm level better, and will likely require analyzing firms' export dynamics with panel data. Alternatively, one could examine the causal effect of credit constraints on firm export performance by exploring the variation in financial vulnerability across sectors, or by exploiting shocks to the availability of external finance that have a differential impact across firms.[5]

Berman and Héricourt (2008) go in this direction and show that country financial development may have a differential effect across firms. In particular, it allows the most productive of the financially constrained firms to become exporters. This speaks to the allocative efficiency of well-functioning financial markets and the distributional effects of policies aimed at improving the availability of external finance. While this topic warrants further research, addressing it will require firm measures of credit constraints that are exogenous to the export decision.

3. THE ROLE OF FOREIGN FINANCIAL FLOWS

Most of the literature on trade and finance has focused on the effects of countries' domestic financial development on their export performance. This interest is motivated by the importance of local bank financing for most firms. Companies, however, may also use alternative sources of funding to meet their liquidity needs. Indeed, recent evidence suggests that foreign portfolio and direct investments alleviate firms' credit constraints and stimulate trade flows. These findings not only provide further confirmation for the effects of credit constraints on trade, but also indicate the potential for financially underdeveloped economies to improve their export performance by liberalizing their capital account.

Manova (2008) explores the effects of equity market liberalizations, and finds that they increase countries' exports disproportionately more in sectors intensive in external finance and intangible assets. These results are not driven by cross-country differences in factor endowments, and are independent of simultaneous trade policy reforms. They suggest that pre-liberalization, trade was restricted by financial constraints, which foreign portfolio investments relaxed to a certain degree. Conceptually, equity market reform should result in resources flowing from capital-abundant developed countries, where expected returns are low, to capital-scarce emerging countries, where expected returns are high. Opening stock markets has indeed been shown to reduce the cost of capital in liberalizing economies, increase investment and output, and promote an efficient resource allocation. The new evidence indicates that it also boosts trade flows.

[5] See Amiti and Weinstein (2009), Chor and Manova (2009) and Iacovone and Zavacka (2009) for evidence on the effects of financial crises on firms' and countries' export performance.

The effects of equity market reform on trade appear highly economically significant. Within three years, a liberalizing country's exports of financially vulnerable sectors (75th percentile) increase 13 to 17 percentage points faster than exports of less financially dependent sectors (25th percentile). These results are comparable to a 20 to 40 per cent improvement in domestic financial development, as measured by private credit or equity market capitalization.

Manova (2008) also explores how the effects of financial liberalization vary with the size and activity of the domestic stock market. Conceptually, countries with a well functioning capital market may benefit more from allowing foreign flows, since they already have the financial infrastructure in place to allocate new resources. At the same time, financially underdeveloped countries stand to gain the most at the margin. In the data, equity liberalizations boost exports more in economies with less active stock markets prior to reform, as measured by initial market turnover or value traded as a share of GDP. This suggests that foreign portfolio flows may compensate for a weak domestic financial system.[6]

In addition to international equity flows, foreign direct investment (FDI) may also channel resources into countries with underdeveloped capital markets and help alleviate firms' credit constraints. The literature on the operations of multinational companies (MNCs) has found that foreign affiliates raise finance in the host country when possible. When that financing is not sufficient, however, subsidiaries receive additional funds from their parent company. For this reason, MNC affiliates enjoy easier access to external capital than domestic firms in the same host country.

Indeed, Manova, et al. (2009) show that foreign-owned firms and joint ventures in China have superior export performance relative to private domestic companies. Moreover, this advantage is especially pronounced in financially vulnerable sectors which require more external finance, have few assets that can be collateralized, or rely more on trade credit. This holds for all export margins at the firm level: total exports, number of export destinations, bilateral exports, number of exported products, and number of products exported to a specific market. These results suggest that firms face credit constraints in the financing of destination-product-specific fixed and variable costs. They also provide micro-level evidence that foreign ownership affects firm export performance by relaxing credit constraints. Finally, they suggest that financial considerations shape the spatial and sectoral composition of MNC activity.[7]

The policy implications of these results are clear: Financially underdeveloped countries may be able to improve firms' access to external credit by relaxing restrictions on foreign portfolio and direct investment. The fact that domestic fi-

[6] The effects of equity market reform do not appear to vary across countries with different stock market capitalization or private credit as a share of GDP. This is consistent with the idea that active stock markets redistribute resources, and suggests that market activity may be a better indicator of an economy's potential to provide external financing than market size.

[7] See Antras et al. (2009) for a model of multinational activity with relationship-specific investments by local affiliates who face credit constraints. MNCs emerge in equilibrium to monitor the local affiliates and incentivize local investors to finance their investment. The parent company may also partly fund its affiliates.

nancial development remains an empirically important determinant of global trade patterns suggests that international financial flows do not (yet) fully compensate for it. Whether they may do so in the absence of any restrictions on foreign investment remains a topic for future research. In particular, it would be important to establish if the same firms that have easier access to domestic financing are also favored by foreign investors.

Note that the distributional consequences of FDI policy depend on the relative prevalence of greenfield investment and foreign mergers and acquisitions. While both modes may bring new capital into a financially underdeveloped economy, mergers and acquisitions would benefit some existing host firms, possibly at the expense of others. By contrast, greenfield FDI would have no direct effect on domestic enterprises but may worsen their credit constraints by increasing competition in the local credit or final goods market. Further research is needed to evaluate the aggregate and distributional implications of capital account reforms for domestic firms' export performance.

4. CREDIT CONSTRAINTS AND THE ADJUSTMENT TO TRADE REFORM

4.1 Towards informed priors

The existing literature offers no direct evidence on the role of credit constraints in the adjustment process to trade reform. Ideally, this question would be explored with panel data on firms' export performance for a country that underwent a trade liberalization episode. This approach would be valid even if the liberalization were anticipated, although in that case the effects of financial frictions may be underestimated if firms responded in advance of the reform date. In the absence of direct evidence from trade policy changes, this section provides informed priors based on the results and intuition developed in the trade and finance literature.

When credit constraints are immaterial, a country would increase its aggregate exports if its trade partners removed or relaxed their import restrictions. This may result from more firms being able to export as well as existing exporters expanding their foreign sales. The latter effect can be further decomposed into two components: the number of products firms export and the value of exports per product. Finally, exports may increase faster in some sectors than others, and may even decrease in the country's comparative disadvantage industries.

When firms require external finance to fund their export activities, credit constraints would likely affect all of these margins of adjustment to trade reform. Continuing exporters would not be able to expand exports as quickly or introduce as many new product lines because of the associated fixed and variable costs of production and trade. It would also be more difficult for new firms to begin exporting or for surviving exporters to enter new markets because such expansion would require substantial sunk costs. All of these distortions would be larger at lower levels of financial development and in financially vulnerable sectors that need more outside capital or have few assets that may be collateralized.

One goal for future work should be to establish the magnitude and welfare implications of these distortions.

It is likely that any short-run response to trade liberalization would result from adjustments on the intensive margin, while the extensive margin would react with a lag. Experienced exporters should find it relatively easier to finance the variable costs of expanding exports of already traded products to tested export destinations. By contrast, introducing new products and entering new markets would entail additional fixed and sunk costs, which should be more difficult to fund. This matters because a delayed extensive-margin response to trade reform may be more costly than slower adjustment on the intensive margin, even holding the level of aggregate exports fixed. There is in fact evidence that, in response to trade liberalization, the reallocations across firms within an industry and across products within a firm are as important for aggregate productivity and welfare gains as reallocations across sectors.

A related concern is that financially underdeveloped countries not only have less external credit available, but also tend to allocate these funds less efficiently. In the framework of Section 2.1, more productive firms are less credit-constrained because they are better positioned to incentivize investors and raise outside capital. In practice, firm productivity is imperfectly observed and poor financial contractibility is often associated with corruption and nepotism. These two forces may magnify the distortions caused by financial frictions in the adjustment to trade reform. If resources are sub-optimally directed towards firms with lower export potential, this will further reduce both the extensive and the intensive margins of trade.

The discussion so far ignores the fact that firms may invest in better technology to improve their productivity and export performance. Similarly, there may be market-specific quality standards which firms need to meet, or firms may choose to increase product quality to enhance their competitiveness. Such technological and quality upgrading is associated with sizeable sunk costs, which credit-constrained firms may not be able to incur. These aspects of the export decision and their welfare implications have yet to be explored in the trade and finance literature.

Another avenue for future research concerns the role of retailers and wholesalers. When liquidity-constrained firms are unable to export directly to a foreign market, they may use the services of an intermediary who specializes in international trade, without engaging in manufacturing. These intermediaries likely have lower export cost because they can rely on established distribution networks. Moreover, trading companies may have easier access to external financing from domestic and foreign banks, as well as to trade credit from similar wholesalers in other countries with whom they have a standing trading relationship. This suggests that the (endogenous) presence of wholesalers could partially alleviate the distortions caused by credit constraints.

What about the effects of trade reform on a liberalizing country with underdeveloped financial markets? Heterogeneous-firm models predict that the least efficient domestic producers would exit because of increased foreign competition in the local market. While more productive firms would survive, they would face

lower demand and reduce output. Credit frictions should be irrelevant for firm exit since they constrain only expansion, but not downsizing. If the adjustment on the extensive margin increases the availability of external financing for surviving companies, they may in fact be able to expand production. In addition, if the liberalizing economy has strong financial contractibility, surviving firms may be able to invest in better or higher quality technology that makes them more competitive. Lastly, the removal of import restrictions may give final-good producers access to cheaper imported intermediate inputs. This could lower their production costs, and stimulate their domestic sales and foreign exports.

4.2 Relevant empirical evidence

Although no study has specifically examined the role of credit constraints in the adjustment to trade reform, a few papers offer some indirect evidence.

While Manova (2008) focuses on the effects of equity market liberalization on exports, she also confirms their robustness to controlling for trade reforms using the Wacziarg and Welch (2003) binary indicator.[8] These sensitivity checks indicate that when countries reduce trade barriers, their exports rise relatively more in financially vulnerable sectors. This highlights a parallel between trade and financial liberalization: While trade reforms reduce trade costs for a given level of financial frictions, equity market reforms relax credit constraints for a given level of trade costs. In either case, the sectors that benefit most are those intensive in external finance or intangible assets, where credit constraints are more acute. Moreover, it appears that opening stock markets has a greater impact when trade policy is more restrictive. Since the proxy for trade openness is extremely rough, however, these results are only suggestive.

A few studies have also argued that credit constraints restrict firms' and countries' ability to respond to export opportunities arising from exchange rate depreciations. Becker and Greenberg (2007), for example, find that in financially advanced countries total exports are more sensitive to exchange rate movements than in countries at lower levels of financial development. Berman and Berthou (2009) provide similar evidence, and show that these effects are more pronounced in sectors which require more external finance. Desai et al. (2008) instead explore the advantage that foreign affiliates have over domestic firms because the former can access internal financing from the parent company. They confirm that the affiliates of US multinationals abroad increase sales, assets, and investment significantly more than local firms during and immediately after sharp depreciations. Unfortunately, data limitations do not allow the authors to directly examine how firm exports respond to currency crises. Further work on the adjustment to exchange rate fluctuations could shed more light on the importance of financial frictions for the response to trade reforms.

[8] A country is labeled effectively closed to trade if average tariff rates are at least 40 per cent; non-tariff barriers cover at least 40 per cent of trade; a black market exchange rate exists and is on average depreciated at least 20 per cent relative to the official exchange rate; the state holds a monopoly on major exports; or there is a socialist economic system.

5. CONCLUSION

Credit constraints and underdeveloped financial institutions have been shown to affect aggregate trade flows and to restrict firms' participation in international trade. This suggests that financial frictions would likely also play an important role in firms', sectors' and countries' adjustment to trade reform. Future research is needed to understand these effects and to establish their consequences for aggregate welfare.[9]

The results and intuition presented in this paper indicate that the gains from trade liberalization may be larger at higher levels of financial development. This implies that countries would respond more to export opportunities resulting from falling trade barriers if they strengthened their domestic financial institutions and liberalized their equity markets and FDI policies. Trade flows may also be stimulated by government-provided subsidized loans and trade credit to exporting firms. Similarly, countries relaxing their own import restrictions might alleviate the effects of increased competition on local producers by pursuing financial sector reforms and facilitating credit access. In either case, the sooner the availability of external financing improves, the faster the response of aggregate trade and output would likely be, and the smoother the transition. While intuitive and appealing, these policy prescriptions remain tentative and their confirmation requires further research.

Kalina Manova is an Assistant Professor in the Stanford Department of Economics.

BIBLIOGRAPHY

Amiti, M and D Weinstein (2009). Exports and Financial Shocks. Columbia University mimeo

Antràs, P, Desai, M and F Foley (2009). Multinational Firms, FDI Flows and Imperfect Capital Markets. Quarterly Journal of Economics 124(3):.1171–1219.

Beck, T 2003. Financial Dependence and International Trade. *Review of International Economics* 11, 296–316

Beck, T 2002. Financial Development and International Trade. Is There a Link? *Journal of International Economics* 57, 107–31.

Beck, T, Demirgüç-Kunt, A, Laeven, L and R Levine 2008. Finance, Firm Size, and Growth. *Journal of Money, Banking, and Finance* 40(7), 1371–405

Beck, T, Demirgüç-Kunt, A and V Maksimovic 2005. Financial and Legal Constraints to Firm Growth: Does Size Matter? *Journal of Finance* 60 (1), 137–77

Becker, B and D Greenberg 2007. Financial Development, Fixed Costs and International Trade. *Harvard Business School mimeo*

Berman, N and A Berthou 2009. Financial Market Imperfections and the Impact of Exchange Rate Movements on Exports. *Review of International Economics* 17(1), 103–20

[9] Do and Levchenko (2007) have suggested that financial markets may in fact develop endogenously in response to trade openness.

Berman, N and J Héricourt (2010). Financial Factors and the Margins of Trade: Evidence from Cross-Country Firm-Level Data. *Journal of Development Economics* (forthcoming).

Chor, D 2009. Unpacking Sources of Comparative Advantage: A Quantitative Approach. *Singapore Management University mimeo*

Chor, D and K Manova (2009). Off the Cliff and Back? Credit Conditions and International Trade during the Global Financial Crisis. Stanford University mimeo

Desai, M, Foley, F and K Forbes 2008. Financial Constraints and Growth: Multinational and Local Firm Responses to Currency Depreciations. *Review of Financial Studies* 21(6), 2857–88

Do, Q-T and A Levchenko 2007. Comparative Advantage, Demand for External Finance, and Financial Development. *Journal of Financial Economics* 86(3), 796–834

Forbes, K 2007. One Cost of the Chilean Capital Controls: Increased Financial Constraints for Smaller Traded Firms. *Journal of International Economics* 71(2), 294–323

Greenaway, D, Guariglia, A and R Kneller 2007. Financial Factors and Exporting Decisions. *Journal of International Economics* 73(2), 377–95

Helpman, E, Melitz, M and Y Rubinstein 2008. Estimating Trade Flows: Trading Partners and Trading Volumes. *Quarterly Journal of Economics* 123, 441–87

Hur, J, Raj, M and Y Riyanto 2006. Finance and Trade: A Cross-country Empirical Analysis on the Impact of Financial Development and Asset Tangibility on International Trade. *World Development* 34(10), 1728–41

Iacovone, L and V Zavacka (2009). Banking Crises and Exports: Lessons from the Past. World Bank Policy Research Working Paper 13984

Ju, J and S-J Wei 2005. Endowment vs. Finance: A Wooden Barrel Theory of International Trade. CEPR Discussion Paper 5109

Kletzer, K and P Bardhan 1987. Credit Markets and Patterns of International Trade. *Journal of Development Economics* 27, 57–70

Manova, K, Wei, S-J and Z Zhang (2009). Firm Exports and Multinational Activity under Credit Constraints. Stanford University mimeo.

Manova, K 2008. Credit Constraints, Equity Market Liberalizations and International Trade. *Journal of International Economics* 76, 33–47

—2007. Credit Constraints, Heterogeneous Firms and International Trade. *Stanford University mimeo*

Matsuyama, K 2005. Credit Market Imperfections and Patterns of International Trade and Capital Flows. *Journal of the European Economic Association* 3, 714–23

Melitz, M 2003. The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica* 71(6), 1695–725

Muûls, M 2008. Exporters and Credit Constraints. A Firm Level Approach. *London School of Economics mimeo*

Rajan, R and L Zingales 1998. Financial Dependence and Growth. *American Economic Review* 88, 559–86

Svaleryd, H and J Vlachos 2005. Financial Markets, the Pattern of Industrial Specialization and Comparative Advantage: Evidence from OECD Countries. *European Economic Review* 49, 113–44.

22

Standards, Trade and Developing Countries

MIET MAERTENS AND JO SWINNEN

1. Increasing and Tightening Standards in Trade

Standards are increasingly dominating world trade and production. This is particularly important in sectors such as food and agricultural exports (Jaffee and Henson, 2005). Over the past decades food standards have increased with new regulations and requirements from national and international governments as well as from private actors and with standards focusing on different issues such as product quality, food safety, and increasingly also ethical and environmental concerns. At the international level, food standards are set by the Codex Alimentarius, the International Plant Protection Convention (IPPC) and the World Organization for Animal Health (OIE); and regulated by the WTO Sanitary and Phytosanitary (SPS) agreement and the Technical Barriers to Trade (TBT) agreement. Under these agreements WTO member states still have the right to adapt and deviate from international standards, as long as it is in the interest of human, plant, and animal health and based on scientific principles. Most national and regional governments have their own food laws and regulations and apply their own food standards that are often stricter than international requirements.

In addition to international and national public regulations, many large food companies, supermarket chains, and NGOs have engaged in establishing private food standards—that are often stricter than public requirements—and have adapted food quality and safety standards in certification protocols. Examples include GlobalGAP (formerly EurepGAP), the British Retail Consortium (BRC) Global Standards, Ethical Trading Initiative (ETI), Tesco Nature's Choice, Save Quality Food (SQV) Program and so on. Although private standards are legally not mandatory they have become de facto mandatory because of commercial pressure as a large share of buyers in international agri-food markets require compliance with such private standards (Henson and Humphrey, 2008). Private standards often go beyond food quality and safety specifications and include ethical and environmental considerations as well.

Food standards are also tightening with more stringent and stricter requirements, especially for phytosanitary and hygiene requirements such as maximum residue levels and levels of contamination. For example, notifications of new SPS

measures to the WTO have increased exponentially over the past 10 years (Henson, 2006). The EU General Food Law of 2002 introduced new food safety requirements and health issues such as the precautionary principle—that measures to protect human health are permissible on the ground of reasonable food safety concerns, even if scientific support is lacking—and traceability, implying the identification of the origin of feed and food in order to facilitate the withdrawal of produce in case of food safety hazards. Also EurepGAP requirements, which rapidly expanded internationally over the past couple of years as GlobalGAP standards, became more stringent with more and stricter compliance criteria. For example, the number of EurepGAP compliance criteria increased from 145 to 199 in just three years (EurepGAP, 2009).

Food standards have mainly emerged from high-income countries and regions, such as the EU and the United States. A number of factors contribute to explaining the increase in food standards in global agri-food trade. A series of major food safety hazards in high-income countries has increased consumer and public concern on food-borne health risks and created an increased demand for food safety. In addition, rising income levels and changing dietary habits have increased the demand for high-quality food. Consumers are also increasingly (made) aware of ethical and environmental aspects related to food production and trade, which has increased the need for specific standards related to these aspects. But also the increased trade in fresh food products such as fruits, vegetables, fish, and meat, which are prone to food safety risks and subject to specific quality demands by consumers, have increased the need to regulate trade through standards. In addition, the increased dominance of supermarkets in food chains contributes to explaining the increased importance of food standards. Large retail chains put much emphasis on freshness, product quality, and food safety as a product-differentiation strategy or so as to reduce food safety risks and the costs related to the risk of selling unsafe food.

2. CONCERNS FOR DEVELOPING COUNTRIES

The increased proliferation and tightening of food standards has cast doubt on the beneficial effect of trade liberalization for poor countries. Major concerns are that:

1. standards act as new non-tariff barriers diminishing the export opportunities of the poorest countries who face multiple constraints in complying with stringent standards and upgrading their supply chains;
2. poor farmers and smallholder suppliers are excluded from high-standards food supply chains because of their inability to comply with high standards; and
3. these farmers are exploited in the chains because stringent standards decrease the bargaining power of small farmers vis-à-vis large food exporters and multinational food companies, and increase the possibilities for rent extracting in the chain.

Standards are therefore often seen as barriers to trade and barriers to development for poor countries. However, the arguments are subject to debate and also empirical studies have come to diverse conclusions about the effects of increasing and tightening food standards on trade and development. In the next two sections we further discuss these arguments and present some empirical evidence on the implications of increasing food standards for developing country food exports and for economic growth, rural development, and poverty reduction in these countries. In the last section we specifically present case-study evidence on the implications of standards in horticultural export sectors in poor Sub-Saharan African countries.

3. STANDARDS AS BARRIERS OR CATALYSTS TO TRADE?

Standards have most often been discussed to act as new non-tariff barriers to trade, diminishing especially the export opportunities of developing countries (Augier et al., 2005; Brenton and Manchin, 2002; Ferrantino, 2006). First, public regulations and standards can potentially be used as protectionist tools to bar imports and protect domestic farmers and companies (Maertens and Swinnen, 2007). Increased trade liberalization itself might create incentives for countries that see quotas removed and tariffs reduced to (ab)use standards to bar imports (Neff and Malanoski, 1996). There is indeed evidence of effective use of parallel standards. For example, Mathews et al. (2003) find that several countries effectively discriminate by having zero-tolerance for salmonella on imports of poultry products from developing countries, while not attaining or monitoring this standard for domestic supplies—which has contributed to a number of disputes raised at the WTO. Also Jaffee and Henson (2005) note an example from Australia prohibiting imports of sauces from the Philippines on the basis of containing benzoic acid, while permitting imports from New Zealand of similar products containing that additive. Moreover, Desta (2008) argues that the EU Food Safety Law with its precautionary principle results in effective discrimination against imports of livestock products from East Africa. Despite such anecdotal examples of effective discrimination, Jaffee and Henson (2005) argue that there is no systematic evidence of the discriminatory use of standards as protectionist tools by industrial countries to bar developing country imports, and that many of these anecdotal cases involve at least partially legitimate food safety and agricultural health issues.

Developing countries confronted with supposed discrimination often lack the scientific and institutional capacity for WTO dispute settlement. In recent years, however, the participation of developing countries in the WTO institutional processes has improved and the number of SPS related notifications by developing countries has increased (Roberts, 2004).

Second, standards can act as barriers to trade because of the high costs of compliance and certification. Such costs might be high specifically for developing countries that generally lack the infrastructure, institutional, technical, and scientific capacity for food quality and safety management and who face a wide di-

vergence between national food quality and safety norms and international standards. The empirical evidence on this issue is limited and mixed. Some authors find evidence of high compliance costs and point to the fact that certification for developing country producers can only be maintained through massive donor support. However, some studies have estimated that the costs of compliance with standards are only a small fraction of total production costs and conclude that compliance cost is much lower than generally assumed. For example Aloui and Kenny (2005) estimate the cost of compliance with SPS measures to be 3 per cent of the total cost of export tomato production in Morocco. Cato et al. (2005) have estimated the cost to implement compliance to quality and safety standards to be less than 3 per cent and the cost to maintain this compliance less than 1 per cent of the total value of shrimp exports from Nicaragua.

Third, the inability of developing countries to comply with stringent standards can be very costly and trade-distorting. The inability to comply with standards can at first lead to border detentions and ultimately result in trade restrictions such as import bans for specific products. For example, in the period January–May 1999, the US Food and Drug Administration reported almost 3,000 border detentions of imported fruits and vegetables and more than 1,500 detentions of fishery products, mostly from developing countries, on the basis of contamination, pesticide residue violation, and failure to meet labeling requirements (Henson et al., 2000; Unnevehr, 2000). In addition, in 1997 the EU banned fish exports from Kenya on grounds of food safety risks and from Bangladesh on the basis of noncompliance with hygiene norms in processing plants.

Trade restrictions and import bans are extremely costly; in the short run in terms of immediate forgone export earnings and in the long run in terms of damaging a country's reputation and eroding its export competitiveness. For example, the EU ban on fish exports from Kenya decreased export earnings by 37 per cent (Henson et al., 2000), and US border detentions of vegetable shipments from Guatemala made this country lose $35 million annually in the period 1995–97 (Julian et al., 2000).

This shows that the empirical evidence of standards acting as barriers to trade because of the discriminatory use of standards against developing countries, and because of the high costs of compliance specifically for developing countries, is rather limited. In addition, standards can act as catalysts to trade. Standards are most often in the interest of public health and can facilitate trade between countries with diverging norms. As such, standards and certification schemes can help to reduce transaction costs, promote consumer confidence in food product safety, and increase developing countries' access to international markets (Henson and Jaffee, 2008). In fact, standards provide a bridge between producers in developing countries and consumer preferences in high-income markets and could be used as catalysts for upgrading and modernization of developing countries' food supply systems and improving their competitive capacity.

Some developing countries have indeed been successful in complying with increasing food standards and upgrading their export sectors as a basis for long term export growth. Jaffee and Henson (2005) note that the most successful coun-

tries and/or sectors have used high quality and safety standards to (re)-position themselves in competitive global markets. A key element in attaining these benefits is to be proactive in food quality and safety and facilitate business strategic responses (Jaffee and Henson, 2005; Henson and Humphrey, 2008).

4. STANDARDS AS BARRIERS OR CATALYSTS TO DEVELOPMENT?

Understanding the link between standards on the one hand, and export competitiveness and performance of developing countries on the other hand is crucial in the design of a broader development agenda, as integration in global markets is generally believed to benefit economic growth (Bhagwati and Srinivasan, 2002; Dollar and Kraay, 2002; 2004). Yet, there is a concern that the poor may not benefit proportionately from high-standards international trade. Hence, another critical policy issue is to understand the link between standards, export chains, and rural incomes in developing countries. The proliferation of stringent private and public standards has caused dramatic changes in the way food production and trade are organized and governed with important implications for producers and rural households (McCullough et al., 2008; Swinnen, 2007; Wilson and Abiola, 2003). Hence, the costs and structural changes associated with standards compliance can cause significant redistribution of welfare—not only across countries but also along supply chains and in rural societies (World Bank, 2005). A key issue in understanding the local welfare implications of increasing high-standards food trade is the way in which supply chain structures and governance systems respond to increasing standards. In this section we first discuss issues of supply chain organization and governance, and then turn to the development implications of increasing standards.

4.1 Supply chain organisation and governance

The main structural changes in food supply chains that are induced by increasing standards include:

1. the increasing levels of vertical coordination (VC) in global supply chains, and
2. the ongoing consolidation of the supply base with large food companies—often multinationals—dominating the chains.

First, compliance with increasingly complex and stringent food standards and monitoring of this compliance throughout the supply chains requires tighter VC in the chains. In order to ensure large and consistent volumes of high-quality and safe produce, food traders and processors increasingly procure from preferred suppliers or specialized wholesale markets, often on a contract basis, and thereby push the food distribution system towards more VC. Also, upstream the supply chain VC is increasing. Traditional spot-market trading systems with middlemen are generally not effective in high-standards trade because of high trans-

action costs related to monitoring compliance with standards. Faced with increased standards, agro-industrial food companies, and exporters are increasingly changing their procurement system towards more VC (Swinnen, 2005). This can occur through different forms of contract-farming or in the most extreme case through complete ownership vertical integration. The latter implies a shift from smallholder contract-based production towards large-scale vertically integrated estate production by agro-processing and food-trading companies. This large-scale vertically integrated way of production increases the scope for standardized production and for meeting high standards at low transaction costs. However, this also entails additional risks and costs for the agro-industry—for example, labour supervision costs.

There are a substantial number of empirical studies demonstrating these supply chain developments. Some studies have pointed to the development of comprehensive VC schemes with extensive monitoring and complex contracting between large food companies and developing country producers as a result of increasing food standards (for example, Gulati et al., 2007; Jaffee, 2003; Minten et al., 2006; Swinnen, 2005). Other authors have presented case-study evidence of a shift towards vertical integration in high-standards export production (Dolan and Humphrey, 2000; Maertens and Swinnen, 2009; Minot and Ngigi, 2004; Gibbon, 2003). How far-reaching the shift from small-scale contract-based production to large-scale vertically integrated industrial production varies considerably across sectors and countries and has major implications for the way in which producers and local households benefit (see further).

Second, food standards pose specific challenges—arising from financial, technical, and institutional constraints—for small agro-food businesses and exporters in developing countries to stay in business in export markets. Although in general the cost of compliance with standards might be low, relative to the total export value, this cost might be very high relative to the means of small firms, leading to the market exit of small and less capitalized firms (Reardon et al., 1999). In addition, smaller businesses might be disadvantaged in contracting with overseas importers and large retail chains because they cannot guarantee the volumes required by these large buyers. Moreover, multinational holdings increasingly seek vertical integration through establishing subsidiaries in developing countries. Foreign direct investment (FDI) in food export, processing, and retailing sectors in developing countries have increased tremendously in the past decade (Colen et al., 2009), which might push smaller firms with less access to capital, poorer technologies, and less access to market information out of high-standards export sectors. Increasing food standards and increased FDI in the food sectors of developing countries could lead to weaker players exiting profitable export markets, and hence to consolidation at the export node of the supply chains.

The empirical literature has presented evidence of ongoing consolidation in agricultural export production in low-income countries (Dolan and Humphrey, 2000; Jaffee, 2003). For example, Maertens, Dries, Dedehouanou and Swinnen (2007) report that in the bean export sector in Senegal the number of exporting companies has dropped from 27 in 2002 to 20 in 2005 with mainly smaller ex-

porters leaving the market and resulting in an increased market share for the three largest companies from slightly less than half to two thirds over the same time period.

4.2 Local welfare implications

As mentioned before, among the main concerns for developing countries are the issues of smallholder suppliers – and especially the poorest ones—being either excluded from or exploited in high-standards supply chains.

Concerning the issue of exclusion of smallholder producers, there is mixed evidence in the literature. Some studies argue that small farmers, and especially the poorest ones, are being squeezed out from high-standards export production because of high compliance costs and increasing levels of vertical coordination (Gibbon, 2003; Reardon and Barrett, 2000; Reardon et al., 1999). First, also at the farm level the individual cost-of-compliance to strict public and private standards might be prohibitively high for smallholder producers to (continue to) engage in high-standards production, especially when credit markets are imperfect and credit is rationed.

Empirical studies that have actually calculated the cost-of-compliances and certification for individual smallholder producers in high-standards supply chains mostly come to different conclusions. For example, Henson (2009) estimates that the initial non-recurrent investment costs of EurepGAP certification (type II, group certification) for smallholder contract-farmers in the Ghana pineapple sector represent less than 2 per cent of sales value and the recurrent cost of maintaining certification less than 1 per cent of sales value. In addition to these relatively small compliance costs the variable production costs were found to be lower due to EurepGAP certification, mainly because of better use of pesticides and other chemicals, and general improvements in agronomic practices.

Second, increasing levels of vertical coordination in food supply chains may result in a bias against the smallest and poorest farmers, either because they are excluded from contract-farming schemes with agro-processors and traders or because smallholder production is replaced by estate production in vertically integrated agro-industries. Contract-farming schemes may be biased towards relatively larger and better endowed farms because of smaller transaction costs—especially for monitoring conformity with standards, and smaller investment costs in terms of farm extension and financial assistance by the contractor firm (Key and Runsten, 1999).

However, standards are themselves instruments for harmonizing product and process attributes over suppliers, and can as such also reduce transaction costs in dealing with a large number of small suppliers. Moreover, well-specified contracts include farm extension and assistance programs that can alleviate the financial and technical constraints small farmers face in meeting stringent standards. In fact, high-standards contract-farming with tight contract-coordination and intensified farm assistance programs could provide a basis for constrained small farmers to participate in high-value export production. In addition,

firms might prefer to contract with smaller farms because they might have a cost advantage—especially if it concerns labour intensive production—or because contract enforcement might be less costly with small suppliers.

The actual evidence on smallholder involvement in high-standards supply chains and the changes in this induced by standards is very mixed. There are cases of complete vertical integration with hardly any smallholder involvement; for example in the tomato export sector in Senegal (Maertens et al., 2008) and the fruit and vegetable export sectors in Zambia (Legge et al., 2006). On the other hand, there are also many examples of cases where export production—destined for markets where standards are high and increasing—remains dominated by smallholders; for example the vegetable export sector in Madagascar (Minten et al., 2006) and Ghana (Legge et al., 2006). In most cases of high-standards export production there is a mix between smallholder contract production and large-scale agro-industrial production (Maertens et al., 2009). Some studies have pointed to sharp reductions in the share of smallholder production as standards increase; for example Jaffee (2003)—recognizing the limitations of the available data—estimates that for export vegetables in Kenya the share of smallholder production decreased from 45 per cent in the mid 1980s to 27 per cent in 2001–02. More recent studies, however, state that these early estimates exaggerate the problem of increased smallholder exclusion and that in fact more smallholders are involved in Kenyan high-standards horticulture exports than previously estimated (Asfaw et al., 2007; Mithoefer et al., 2006). In general the figures on horticulture sectors in Africa seem to point out that there is actually much more smallholder involvement in high-standards production than would be assumed based on the above arguments. Similar observations were described for several agricultural sectors in transition countries in Eastern Europe and Central Asia (Swinnen, 2005).

Concerning the issue of exploitation of smallholder producers, it has repeatedly been argued that the gains from high-standards agricultural trade are captured by foreign investors, large food companies, and developing country elites and that standards lead to a more unequal distribution of the gains from trade (for example, Dolan and Humphrey, 2000; Reardon et al., 1999). On the one hand, consolidation of the export supply base and vertical coordination in the supply chains are said to amplify the bargaining power of large agro-industrial firms and food multinationals, displace decision-making authority from the farmers to these downstream companies, and strengthen the capacity of these companies to extract rents from the chain to the disadvantage of poor farmers and local households (Warning and Key, 2002). On the other hand, vertical coordination schemes provide a basis for farmers to access the credit, inputs, and technology they need for upgrading their production in terms of productivity and quality and to increase their incomes.

Recent empirical studies have demonstrated a beneficial effect for smallholders participating in high-standards contract production. Demonstrated benefits include productivity gains, increased household income, reduced volatility, and more stable incomes, technology spillovers and so on. For example, Dries and Swinnen

(2004) show that small dairy farmers gain in terms of productivity from contract production with large foreign milk processors. Gulati et al. (2005) provide similar evidence for smallholder animal production in Southeast Asia. Maertens and Swinnen (2009) show that contract-farming in the Senegalese horticulture export sector leads to very high and significant increases in household income. Minten et al. (2006) show that the vegetable exports from Madagascar to the EU are completely based on smallholder contract production, leading to more income stability for local households and technology spillovers on rice production.

Moreover, an important—and much overlooked—argument in the welfare analyses of high-standards trade is that poor households may benefit through employment effects. High-standards trade creates new employment opportunities in processing and handling of produce, and on vertically integrated estate farms and large contracted farms. Some recent empirical studies show that high-standards trade creates substantial employment that is well-accessible for the poor, leading to increased rural incomes and reduced poverty rates (Maertens et al., 2008). For examples, Maertens and Swinnen (2009) estimate that local poverty reduced by 12 percentage points as a result of high-standards bean export in Senegal, mainly through employment effects.

5. CONCLUSION

The main conclusion is that increasing and tightening food standards may be both barriers and catalysts for the participation of poor countries in international agricultural trade and for development in these countries. In this paper we have summarized a series of studies and have documented arguments about why standards are not necessarily non-tariff barriers to trade. We also argue that the widely held belief that high-standards trade is non-inclusive and inequitable may need to be revised. There is substantial evidence that high-standards trade can benefit poor countries and smallholder farmers and rural households in those countries.

Pro-poor effects may arise because high-standards trade is also typically high-value trade and thus allows better returns for those who can participate. This, in turn, provides incentives for exporting companies to develop extensive (vertically coordinated) contracting schemes with developing country producers, which includes technology transfer and input provisions. In addition, contracting problems for exporters typically lead to premia to local producers to ensure their supplies and sufficient quality of production. Rural households in developing countries may benefit either as smallholder contract farmers or through the labor market because of enhanced wages and employment opportunities in rural areas.

Johan Swinnen is Professor in the Department of Economics and Director of LICOS – Centre for Institutions and Economic Performance, K.U.Leuven.

Miet Maertens is Assistant Professor in the Department of Earth and Environmental Sciences, K.U.Leuven.

BIBLIOGRAPHY

Aloui, O. and L. Kenny (2005). *The Cost of Compliance with SPS Standards for Moroccan Exports: A case study*. Agricultural and Rural Development Discussion Paper, Washington DC: The World Bank

Asfaw, S., D. Mithoefer and H. Waibel (2007). *What Impact are EU Supermarket Standards Having on Developing Countries' Export of High-value Horticulture Products?* Evidence from Kenya. 105th Seminar of the European Association of Agricultural Economists, Bologna, March 8–10, 2007

Augier, P, M. Gasiorek and C. Lai Tong (2005). "The Impact of Rules of Origin on Trade Flows", *Economic Policy* 20, no. 43: 567–623

Bhagwati, J. and T.N. Srinivasan (2002). "Trade and Poverty in the Poor Countries", *American Economic Review*, 92(2), 180–83

Brenton, P. and M. Manchin (2002). *Making the EU Trade Agreements Work: The role of rules of origin*. CEPS working document no. 183, Brussels: Centre for European Policy Studies

Cato, C.J., S.W. Otwell and A. S. Coze (2005). *Nicaragua's Shrimp Subsector: Developing a production capacity and export market during rapidly changing worldwide safety and quality regulations*. Agricultural and Rural Development Discussion Paper, Washington DC: The World Bank

Colen, L, M. Maertens and J.F.M. Swinnen (2009). "Foreign Direct Investment as an Engine for Economic Growth and Human Development: A review of the arguments and empirical evidence", *Human Rights and International Legal Discourse*, 3(2)

Desta, M. G. (2008). "EU Sanitary Standards and Sub-Saharan African Agricultural Exports: A case study of the livestock sector in East Asia", *The Law and Development Review*, vol. 1, issue 1

Dolan, C. and J. Humphrey (2000). "Governance and Trade in Fresh Vegetables: The impact of UK supermarkets on the African horticulture industry", *Journal of Development Studies* 37, no. 2: 147–76

Dollar, D. and A. Kraay (2002). "Growth is Good for the Poor", *Journal of Economic Growth*, 7(3), 195–225

—(2004). Trade, Growth and Poverty. *Economic Journal*, 114 (493), 22–49

Dries, L, and J.F.M Swinnen (2004). "Foreign Direct Investment, Vertical Integration and Local Suppliers: Evidence from the Polish dairy sector", *World Development* 32, no. 9: 1525–44

EurepGAP (2009). http://www.eurepgap.org/

Ferrantino, M. (2006). "Quantifying the Trade and Economic Effects of Non-tariff Measures". *OECD Trade Policy Working Papers*, no. 28, OECD Publishing. doi:10.1787/837654407568

Gibbon, P. (2003). "Value-chain Governance, Public Regulation and Entry Barriers in the Global Fresh Fruit and Vegetable Chain into the EU", *Development Policy Review* 21, no. 5: 615–25

Gulati, A, N. Minot, C. Delgado and S. Bora (2007). "Growth in High-value Agriculture in Asia and the Emergence of Vertical Links with Farmers", In J F M Swinnen (Ed.), Global *Supply Chains, Standards and the Poor*. Oxford, CABI publishing

Henson, J.S. (2006). *The Role of Public and Private Standards in Regulating International* Food Markets. Paper presented at the IATRC Symposium. Bonn

Henson, S. (2009). *Standards in Agricultural Trade: Some strategic considerations for Africa*. E-learning course on Food and Agricultural Trade in Africa, World Bank Institute, April 2009Henson, S. and S. Jaffee (2008). "Understanding Developing Country Strategic Responses to the Enhancement of Food Safety Standards", *World Economy* 31(4): 548–68

Henson, J.S., A.-M. Brouder and W. Mitullah (2000). "Food Safety Requirements and Food Exports from Developing Countries: The case of fish exports from Kenya to the European Union", *American Journal of Agricultural Economics* 82, no. 5: 1159–69

Henson, S. and J. Humphrey (2008). *Understanding the Complexities of Private Standards in Global Agri-food Chains*. Paper presented at the international workshop: Globalization, Global Governance, and Private Standards, Leuven Centre for Global Governance Studies, 4–5 November, 2008

Jaffee, S. (2003). *From Challenge to Opportunity: Transforming Kenya's fresh vegetable trade in the context of emerging food safety and other standards in Europe*. Agricultural and Rural Development Discussion Paper, Washington DC: The World Bank

Jaffee, S. and S. Henson (2005). "Agro-food Exports from Developing Countries: The challenges posed by standards", in *Global Agricultural Trade and Developing Countries*, edited by M A Aksoy and J C Beghin, Washington DC: The World Bank

Julian, W.J., G.H. Sullivan and G.E. Sanchez (2000). "Future Market Development Issues Impacting Central America's Nontraditional Agricultural Export Sector: Guatemala case study", *American Journal of Agricultural Economics* 82: 1177–83

Key, N. and D. Runsten (1999). "Contract Farming, Smallholders, and Rural Development in Latin America: The organization of agroprocessing firms and the scale of outgrower production", *World Development* 27 no. 2: 381–401

Legge A, J. Orchard, A. Graffham, P. Greenhalg and U. Kleih (2006). *The Production of Fresh Produce in Africa for Export to the United Kingdom: Mapping different value chains*. Natural Resource Institute, United Kingdom

Maertens, M. and J.F.M Swinnen (2007). "Standards as Barriers and Catalysts for Trade and Poverty Reduction", *Journal of International Agricultural Trade and Development* vol. 4 (1)

—(2009). "Trade, Standards and Poverty: The case of vegetable exports from Senegal", *World Development* vol. 37(1), 161–78

Maertens, M., L. Dries, F.A. Dedehouanou and J.F.M Swinnen (2007). "High-value Supply Chains, Food Standards and Rural Households in Developing Countries", In: Swinnen, J F M (Ed) *Global Supply Chains, Standards and the Poor*, CABI publishing

Maertens, M., B. Minten and J.F.M. Swinnen (2009). *Growth in High-value Export Markets in Sub-Saharan Africa and its Development Implications*. LICOS Discussion Papers no. 244, Leuven, LICOS

Maertens, M., L. Colen and J.F.M Swinnen (2008). *Globalization and Poverty in Senegal: A worst case scenario?* LICOS Discussion Papers no. 217, Leuven, LICOS

Mathews, H. K., J. Bernstein and J.C. Buzby (2003). "International Trade of Meat/Poultry Products and Food Safety Issues". *International Trade and Food Safety: Economic Theory and Case Studies*, in J C Buzby (Ed.), Agricultural Economic Report no. AER 828, Economic Research Service, USDA

McCullough, E., P. Pingali and K. Stamoulis (2008). *The Transformation of Agri-food Systems: Globalization, supply chains and smallholder farmers*. The Food and Agricultural Organization of the United Nations and Earthscan, London

Minten, B., L. Randrianarison and J.F.M. Swinnen (2006). *Global Retail Chains and Poor Farmers: Evidence from Madagascar*. LICOS Discussion Papers no. 164, Leuven: LICOS

Minot, N. and M. Ngigi (2004). *Are Horticultural Exports a Replicable Success Story? Evidence from Kenya and Côte d'Ivoire*. EPTD/MTID discussion paper, IFPRI, Washington, DC: IFPRI

Mithöfer, D., E. Nang'ole and S. Asfaw (2006). Smallholder Access to the Export Market: The case of vegetables in Kenya. *Outlook on Agriculture*

Neff, A.S. and M.A. Malanoski (1996). "Trade Reform Increase Likelihood of Challenges to Food Standards", *Food Review*

Reardon, T. and C. B. Barrett (2000). “Agroindustrialisation, Globalization and International Development: An overview of issues, patterns, and determinants”, *Agricultural Economics* 23: 195–205

Reardon, T. et al. (1999). “Global Change in Agrifood Grades and Standards: Agribusiness strategic responses in developing countries”, *International Food and Agribusiness Management Review* 2, no. 3: 421–35

Roberts, D. (2004). *The Multilateral Governance Framework for Sanitary and Phytosanitary Regulations: Challenges and Prospects*. Washington DC: The World Bank

Swinnen, J.F.M. (2005). *The Dynamics of Vertical Coordination in Agrifood Chains in Eastern Europe and Central Asia*. Washington DC: The World Bank

Swinnen, J.F.M. (2007). *Global Supply Chains, Standards and the Poor*, CABI publishing

Unnevehr, L. (2000). “Food Safety Issues and Fresh Food Product Exports from LDCs”, *Agricultural Economics* 23, no. 3: 231–40

Warning, M. and N. Key (2002). “The Social Performance and Distributional Impact of Contract Farming: An equilibrium analysis of the Arachide de Bouche Program in Senega”l. *World Development* 30, no. 2: 255–63

Wilson, S. J. and V. Abiola (2003). *Standards and Global Trade: A voice for Africa*. Washington DC: The World Bank

World Bank (2005). *Food Safety and Agricultural Health Standards: Challenges and opportunities for developing country exports*. Poverty Reduction and Economic Management Trade Unit and Agricultural and Rural Development Department, report no. 31207, Washington DC: The World Bank

PART D

ADJUSTMENT PROGRAMS

23

Notes on American Adjustment Policies for Global-integration Pressures[1]

J DAVID RICHARDSON

1. OVERVIEW

The Trade Adjustment Assistance (TAA) program of the United States is nearing its fiftieth year. Like many 50-somethings it has expanded and contracted, sometimes in regretful ways. But it has also gained in experience and seasoning; and, we will argue, current relevance. Much has also changed over 50 years in global goods markets, including their volatility and its consequences. Goods-market volatility is sure to continue to confront firms, workers, and communities even after macroeconomic normalcy is restored.

Features of TAA's somewhat unshapely structure may in fact have promise for today's stressful labor markets, such as its conditional link of extended income support to a worker's retraining initiative. Looking further ahead, growing instabilities in worker earnings, as well as growing reasons for extended bouts of structural unemployment, suggest the need for a reshaped program. A reshaped program would recognize the many ways that globally enabled dynamism exposes workers to the same instability and displacement as does trade—even when such dynamism seems more narrowly conceived as technological and organizational innovation.

Both the regrets and the seasoning of historic TAA can provide a foundation for radical reform—a widening of its remit to structural adjustment assistance (SAA), a reemphasis on its reemployment goals for workers and training goals for firms, refinement of some of its features, in particular its insurance options (going well beyond its current wage insurance), and adoption of innovative new firm-worker-civic stakeholder partnerships, especially for training.

Such reshaped and regrounded structural assistance initiatives will require creative reform and innovation in domestic policies designed to underpin the ini-

[1] These notes draw on well over a decade of research by my colleagues and me at the Peterson Institute for International Economics; in particular Howard F Rosen, and also C Fred Bergsten, Kimberly Ann Elliott, J Bradford Jensen, Lori G Kletzer, Howard Lewis III, Catherine L Mann, and Matthew J. Slaughter. Because my understanding of my mandate for this World Bank project was American policy, I have consciously disregarded the extensive European and broader OECD literatures on adjustment policies, including the recently revised and expanded European Globalization Adjustment Fund (EGF).

tiatives. Their ultimate objective is to empower large numbers of Americans to do two things better simultaneously:

- to prosper from inter-linked global opportunity and technological dynamism, and
- to manage their risks and challenges more effectively.

2. NOTES ON HISTORICAL PATCHWORK TAA CONTEXT

American Trade Adjustment Assistance was designed originally in the 1960s to meet three implicit and often-conflicting objectives: efficiency, equity, and (political) compensation. These words described deliberate and fair relocation coupled with income support for those bearing excessive burdens on behalf of broader public policy—cross-border trade liberalization that enabled more imports. Adjustment assistance was originally linked only to explicit decisions to lower policy barriers at the border, and to their incremental injurious effects.

These strictures and very tight insistence that imports be the major cause of dislocation led to no awards of TAA (out of 25 petitions) in the 1960s. There was widespread dissatisfaction that the program was no more than symbolic tokenism. In preparation for the WTO's Tokyo Round negotiations, and in the throes of the global economic 'reordering' of the 1970s, Bergsten and others within the US Executive Branch pressed for a more genuine program, linked importantly to import growth itself, not just to incremental import growth from trade agreements, and extended to firms and community economic development.[2]

Though these changes breathed life into the TAA program, the worker petitions that were subsequently granted were largely focused only on income support, much like unemployment insurance (UI), and were rarely linked to firm or community assistance. The 1970s program subsequently lost much of its popular support from the confluence of three politically unsupportable trends: TAA recipients turned out to be increasingly recalled to former employers, and were disproportionately in high-paid unionized jobs.[3] TAA's budget costs soared by a factor of five when President Jimmy Carter ordered autoworker petitions expedited during his reelection campaign. A knowing public and their Congressional representatives saw that TAA was providing no 'adjustment', only 'assistance' to those who had weak warrants for it compared to more marginal workers.

TAA during most of the 1980s was starved and haphazard...unsurprisingly, given the lull between the WTO's Tokyo Round and the Uruguay Round, and given the standing of labor interests under the Reagan Administration. TAA was actually high on the list of programs for elimination in the early Reagan years.

[2] See Rosen (2006) for the most comprehensive retrospective. Among similar retrospectives, US GAO (2000), Baicker and Rehavi (2004), and Bown and McCulloch (2005; 2007) provide more recent but more limited historical reviews.

[3] See Richardson (1982), reporting on a program evaluation conducted by Mathematica Policy Research, and based on a survey of TAA recipients in the late 1970s. See Rosen (2006, 91–3) for the astounding surge in coverage and budget outlay.

Though the program returned to late-1970s usage in the late 1980s, changes were quite minor (for example, explicit authorization of energy-exploration workers). A NAFTA-TAA clone was introduced in 1993, and lasted until 2002. It anticipated—in the NAFTA context only—the expanded eligibility criteria to come in 2002 to the general TAA program, into which it was then folded.

3. The More Expansive 2002 and 2009 Evolutions

Major TAA reform took place around the turn of the millennium as both the late Clinton and early Bush administrations struggled to get Congress to re-launch authority for global and regional trade agreements. The reforms focused especially on workers, making their assistance more reemployment-oriented and training-contingent.

Under the Trade Act of 2002,[4]

- eligibility was broadened to include secondary workers displaced upstream or downstream from an importantly impacted group; impact was broadened to include not only imports but shifts in production to any countries with which the United States had a preferential trade agreement, and a small TAA program for farmers and fisherman was introduced with different criteria (it has since accounted for roughly two per cent of TAA spending).

Under the 2002 Act, TAA recipients could receive:

- up to 130 weeks of training, which needed to be pursued full-time, including 104 weeks of vocational training and 26 weeks of remedial training (such as for English as a Second Language or for language literacy);
- up to 78 weeks of extended income support, after the 26 weeks of standard UI was exhausted, if enrolled in training;
- job search and relocation assistance;
- a 65 per cent, payable in advance, refundable Health Coverage Tax Credit (HCTC) to help offset the cost of maintaining health insurance during the period of unemployment; and
- a targeted program of wage insurance called Alternative Trade Adjustment Assistance (ATAA). ATAA provided workers aged 50 and older who became reemployed within 26 weeks and earned less than $50,000 half of the difference between their new and old wages, for up to two years subject to a $10,000 maximum.

Wage insurance was probably the most innovative, market-based labor market adjustment program to be introduced in the United States over the last several decades. Although the take-up rate was low for the early years, initial anecdotal reports suggested that many workers benefited from the program.

Wage insurance and the HCTC are two illustrations of a commendable shift in worker assistance—from traditional income transfers to support that is arguably

[4] The account in the following paragraphs comes from Rosen (2008, 3) and from Kletzer et al. (2007, 12–13).

more targeted and cost effective. Both, for example, serve as an implicit subsidy for workers to take a job with a new employer, whose costs of on-the-job training (OJT) are implicitly subsidized to the degree that workers are attracted to jobs they might have turned down in the absence of wage insurance (jobs with low wages or long vesting periods for benefits).[5] Yet the degree of success and its exact cost-effectiveness are still matters of controversy. There is yet to be any systematic evaluation of either program, even though the Trade Act of 2002 called for it.

Although the annual number of petitions fell from roughly 3600 to 2200 between 2003 and 2007, program take-up rates among eligible workers increased, and the proportion of petitions accepted rose slightly from mid-50s per cent to 65 per cent Import-related displacement accounted for roughly half of the accepted petitions, shifts in production abroad for 40 per cent, and spillover from upstream and downstream supply chain effects for 10 per cent.[6]

The Trade and Globalization Adjustment Assistance Act of 2009 strongly scaled up and accelerated the momentum of expansiveness initiated by the Trade Act of 2002. Eligibility was expanded, almost every benefit was made more generous, and many contingencies were removed. Specifically,

- eligibility was extended for the first time explicitly to service-sector and public-agency workers;
- dislocation from shifts in production to any country now warranted consideration for support, not just to preferential-trade-agreement partners;
- training support was increased uniformly by 26 weeks;
- workers no longer needed to contribute 10 per cent 'co-pays' to job-search and relocation allowances;
- workers could receive both wage insurance, re-christened Reemployment Trade Adjustment Assistance (RTAA presumably, instead of ATAA) and training support, removing the 2002 Act's one or the other contingency.

Yet an even-more expansive TAA may be promising for today's economic challenges, as discussed in the remainder of these notes.

4. THE CURRENT SLUMP: REALITY EVOLVES TOWARD EVOLVING PATCHWORK

It is a truism that life imitates art. Something similar is happening today regarding Trade Adjustment Assistance. The American version of the global downturn, triggered and fueled by the global financial crisis, has generated an environment that is, ironically, ripe in principle for an expansion of vocational and remedial training, supported by more generous income replacement allowances and by wage insurance, somewhat in the spirit of the Post-World War II GI bill, only aimed at the soldiers in a war against depression.

[5] See Brainard et al. (2005).
[6] Rosen (2008, 2–3); US Department of Labor (2009)

But none of that characterizes traditional American unemployment insurance (UI), which remains very similar in structure, finance, and administration to its founding mid-twentieth century self. Modest income support facilitates job search, but not much else. There are no training or retraining mandates, and no wage insurance. And state and federal outlays for traditional UI have always been a huge double-digit multiple of the fairly stable annual $1 billion or so spent on TAA,[7] and an especially large multiple during deep recessions.

So is a TAA or a traditional UI system better for today? Kletzer and Rosen (2005; 2006) vote strongly for the former. They argue that the whole American workforce should become eligible for TAA-style adjustment assistance as a better, more relevant and effective program for today than UI. But many observers go even further than this. They believe that today's globally integrated environment is even riper for something even better than expansive up scaling of TAA, something more radically addressed to the way that traditional trade pressures have been:

- amplified by the radical fusion of traditional trade and investment-based production shifts with changes in technology and in business organization and supply chains;
- borne increasingly:
  - by firms rather than by industries;
  - by occupations rather than by broad worker skill-groups; and
  - by those firms and workers who are somehow less advantaged compared to their peers.

We can think of the first of these features as the 'integration of (many types of) integration' and the second as the downward devolution of adjustment burdens to precise micro agents, rather than groups of agents in so-called industries and skill groups. We turn in the remainder of these notes to what these two features might imply for future adjustment policies, keeping an American focus.

5. NOTES ON A TWENTY-FIRST CENTURY ADJUSTMENT POLICY—POLICY DESIGN

5.1 The twenty-first century context: 'Integrated integration' for 'micro-units'[8]

Two new trends in global integration and its understanding shape future adjustment policies: the interwoven character of many types of integration and the central importance of micro-level agents in accounting for losses and gains. Traditional accounts of globalization explain how trade, investment, and migration are prompted by a country's resource endowments and comparative advantage interacting with their global counterparts. Trade, investment, and migration in

[7] Kletzer et al. (2007, 13)

[8] This section is an updating and rearrangement of Richardson (2005b, 12–14). Documentation and references to the literature have been severely abridged for the purposes of these notes.

turn change domestic rewards to broad groups of resource-owners, such as skilled and less-skilled workers, and owners of productive physical and intangible capital. These accounts remain valid today, though their empirical implementation has always revealed only modest impacts on measures of dislocation and adjustment. Instead, today's empirical action turns out to be increasingly at the level of micro units and to be hard to differentiate from globally enabled technological and organizational innovation, as described below.

The past three decades of American and global economic integration have increasingly featured interwoven drivers of change, with variegated adaptation by heterogeneous micro units within traditional groups. The modern period has been punctuated with:

- revolutionary change in information and communications technology, and with associated job shifts toward knowledge workers adept in forensics, diagnostics, problem-solving and complex communication;
- rapid product and process innovation, including creative standardization (for example, electronic components), differentiation, and customization, as well as radical change in intellectual property law and administration to protect such design innovation;
- aggressive deregulation, downsizing, and fragmentation of conglomerates and vertically integrated production relationships;
- the advent of what some call the global business model—lines of business dedicated to a global market for their core competencies (equivalent to corporate comparative advantage), reliant on other businesses for key inputs and services that can be as finely defined as tasks (for example, payroll management), integrated both globally and with upstream and downstream suppliers and distributors, including suppliers of innovation to them and users of their own innovation.

These trends are interwoven. They might be called 'integrated integration'—the integration of many different types of integration:

- integration across national borders, corporate borders, marketing borders (for example, finely differentiated products), and temporal borders (for example, successive upgrades of a product);
- integration across precisely defined tasks in the production process, or across finely differentiated input types (for example, standardized and sophisticated reading of X-rays).

Only one of these many facets of integrated integration concerns traditional international trade, "...and it is impossible to isolate it even conceptually from all the other types of integrated integration."[9] We will argue that twenty-first cen-

[9] For example, to be viable, much technological and structural change needs global access to ideas and markets. Traditional global integration thus facilitates productive innovation in products, production, organization, and the management of vertical supply chains and even labor relations. Many of the drivers of change over the past few decades are interdependent, and cannot be distilled into a pure essence of globalization, to be distinguished antiseptically from technology and structural evolution.

tury adjustment policies should not try to isolate it either, in the spirit of several recent treatments of these integrated trends.[10]

Integrated integration has generated enormous material benefits. Productivity growth and growth in American and emerging-economy standards of living surged in the middle 1990s and more or less persisted through severe regional crises and slowdowns, before becoming erratic in 2008–9. There are only a few reasons to think that productivity growth will not return to something near its handsome recent rate as economies work through their current slump. Many sectors have shared in this sustained productivity surge, not merely manufacturing.

Yet until recently many of the traditional measures of national inequality were trending up over these same three decades, mildly in the late 1990s and more strongly otherwise. With a few exceptions, within-economy inequality has trended up both across and within traditional categories.[11] In the United States inequality trended up within and across educational groups, within and across regions, and for women as well as for men. The growth in inequality within well-defined categories poses a special challenge for analysis and policy design, since the usual explanations of trends in inequality focus on between-category determinants. And a concomitant of the growth of within-category inequality is the growth of individual income volatility.

Researchers studying integrated integration have thus had to expand traditional perspectives synthetically[12] to feature diversity—heterogeneity within groups—in many dimensions:

- heterogeneity across firms in productivity, product differentiation, job attributes, and innovation's costs and rewards;
- heterogeneity across workers in ambition, adaptability, creativity, collegiality, and other hard-to-measure personal workplace traits;
- heterogeneity across regions and communities in infrastructure, business climate and culture, and in openness to other cultures and communities.

The important implication of the blended synthesis is that global integration can have both traditional impacts and effects on the distributional shape of outcomes across heterogeneous firms, workers, and communities. The growing dispersion of those outcomes is one of the most important aspects of that distribution's shape.[13]

[10] For example, Mann (2006); Aldonis et al. (2007).

[11] On the growth of so-called residual inequality, inequality within categories, inequality that is not easily correlated with (explained by) any observable fundamentals, and on the resulting growth in expected income volatility, see pioneering research by Gottschalk and Moffitt (1994), authoritative recent research by Violante (2002) and a huge ongoing and supportive literature on the same themes.

[12] The synthetic blend of traditional and newer perspectives is discussed in more detail in Bernard et al. (2007a), in Bernard et al. (2007b), and in the Appendix to Part 2 of Richardson (2005b). Helpman et al. (2009) is a path-breaking generalization of these syntheses and of the pioneering work of Davidson and Matusz (2004). The Helpman et al. (date) paper brings together theoretically heterogeneous firms, heterogeneous workers, and equilibrium structural unemployment for an open, globally engaged economy.

[13] So also are the skewness and kurtosis that describe whether the upper, lower, or middle ranges of that distribution are unusually densely concentrated.

The growing body of empirical micro-data research for the United States suggests that globally integrated integration widens the dispersion of outcomes among American workers, firms, and communities, sifting and sorting among the advantaged who gain more, the less-advantaged who gain less (or lose), and those in the middle who are often propelled toward either extreme.[14] If this remains an accurate summary of the micro-trends both during and after the current macro troubles, then American adjustment assistance policies need radical reshaping, not mere rescaling.[15] Yet, ironically, these same micro trends can provide the resources and innovation to fund the radical reshaping.

6. NEW RESEARCH ON THE GAINS FROM GLOBAL INTEGRATION AND WHO GETS THEM

Recent research provides a consensus on both the reasons and the resources for reshaping adjustment policies. In fact, the reasons and the resources are opposing faces of integrated integration, and underlie the need to pair innovation in integration always with innovation in adjustment policies.

6.1 Research consensus 1: Globally integrated integration generates large gains

'Twins' research on American micro units shows that, compared to measurably matched peers, globally integrated firms enjoy higher growth and lower failure rates. Their workers enjoy faster employment growth in more stable jobs paying higher rewards. The communities that host them enjoy tax bases that grow faster and more stably.

This has salutary results for industries and overall economies. Globally integrated integration facilitates sifting and sorting among heterogeneous firms. Firms with higher productivity and other advantages find themselves able to select into integrated integration of all types. Then as they grow faster and fail less often than their less-advantaged and lower-productivity peers, they represent larger and larger shares of any industry. Their advantaged workers likewise represent growing shares of worker-group employment, and their host communities account for growing shares of regional and national output. Overall populations are increasingly represented by their fittest members.

The large gains from this process are not limited in sectoral scope: these same patterns apply to service firms, and to service occupations, as well as to manufacturing. 'Tradable occupations' reward their workers with better wages (full-

[14] The last finding pertains to workers in particular, and seems to suggest fatter distributional tails. The more accurate summary of this trend is a fatter tail at the very top of the distribution, and a less dense concentration of gains from integrated integration among the working poor. But the non-working poor seem to have gained proportionately from integrated integration due to cheaper goods and services (see, for discussions of these productivity-induced price effects, (Broda et al. (2009)).

[15] As does adjustment assistance in countries with similar trends.

time; frequency).[16] And the gains seem to accumulate. The most integrated of the globally integrated firms, those workers in tradable occupations and industries, and the most restructuring-minded firms that are trade and investment and technology engaged all seem to enjoy multiple, possibly multiplicative, performance premiums.

6.2 Research consensus 2: There is an unbalanced distribution of those gains[17]

But the opposite face to the large gains described above is the tenuous survival of the less-fit. Firms with lower productivity and other disadvantages grow slowly, shrink, and die. Their workers face grimmer workplaces and workplace opportunities (for example, for on-the-job training and for promotion), and their host communities lose tax base to others. These heightened adjustment pressures on the less-productive and less-advantaged are an inescapable downside of the sifting and sorting gains generated by deeper integrated integration.

Economic mobility can in principle ease the heightened adjustment pressures. Lower-productivity firms and their workers can be absorbed by high-performing, globally integrated firms. Workers themselves can seek to move between employers of varying fitness, seeking to make the best possible match.[18] But overall trends in American economic mobility are negative,[19] and structural impediments to worker mobility remain prominent in the American economy. New policies are needed to help.

6.3 Underlying policy mindset for the twenty-first century context

A new policy mindset is also needed. Further deepening of America's strongly beneficial engagement of globally integrated integration needs newly creative, newly effective domestic policy. The recipe for American success involves pairs of ingredients always, complementary cognates, and dynamic global-integration initiatives paired with creative domestic adjustment initiatives. With domestic policy reform and innovation to diffuse the benefits and to increase the typical American worker's capability to engage global dynamism, deeper future global integration will be more sustainable, even perhaps widely welcome.

[16] See, for example, Jensen and Kletzer (2005; 2008) and Mann (2006), as well as the forthcoming Jensen, Kletzer, and Mann book.

[17] Richardson (2004; 2005a) contains additional detail and documentation.

[18] See Andersson et al. (2005) for impressive evidence that among comparably disadvantaged, low-wage American workers, there are enormous gains to landing a job with a high-productivity firm (compared to an average firm). Micro-agent matches matter a great deal.

[19] This is the briefest thumbnail summary of results from the ongoing Economic Mobility Project, supported by the Pew Charitable Trusts (www.economicmobility.org).

7. CONCRETE POLICY IMPLICATIONS: WIDER MANDATE, NEW ACTORS

American Trade Adjustment Assistance has already expanded its scale, as described above. There is an urgent need now to reshape it, and to expand its scope and its constituency. In scope, traditional adjustment assistance needs to expand to cover 'structural' dislocation in addition to traditionally narrow trade-related versions of structural dislocation.[20] The distinction between structural and cyclical dislocation is well-established in macro and labor economics, and could be codified into eligibility criteria that are at least as persuasive as in current TAA decision-making. One nuance that might help to bind criteria for awarding workers such adjustment assistance is that their structural dislocation should be linked to global integration (that is, integrated integration, in the parlance of this paper), thereby maintaining continuity with the historic TAA program (for example, outsourcing, even domestically, would be covered; natural catastrophes like floods and fires would not). A structural expansion of scope would match the reality of the twenty-first century's multiple integrated forms of integration. It would also shift structural adjustment assistance in a healthy no-fault direction from an ineffective and politically volatile blame-trade mindset.

The detail of structural adjustment assistance—SAA, say—could build on TAA with minor modifications. Rules for petitioning and training and reemployment-oriented income support (including wage and benefits insurance) could be very similar to those currently used. Effective design would require serious—mandated and funded—monitoring and evaluation of results and cost-effectiveness, including longitudinal surveys of program participants. Effective design would also abjure some suggested TAA reforms that have never yet formally been implemented, and industry certifications for example, in which SAA petitions from an entire sector or occupational group were accepted and processed. In the twenty-first century world this would be wastefully ineffective. Within every sector and occupation are high-performing workers and firms who need and deserve the chance to move up and expand at the expense of others; smart adjustment assistance targets these others, not the universe, and helps them move toward the high-performance skills and practices of the successful micro-units.

In constituency, traditional adjustment assistance needs to expand its constituency—stakeholders—to include natural and new American institutions. Among the natural constituents are labor unions and community colleges, both of which are beginning to turn from their normal audiences toward benefits management, skills-upgrading, and job-search training. Among new constituents should be not-for-profit social service firms[21] and even for-profit businesses themselves, because of the empirically proven value of on-the-job training com-

[20] This expansion of scope reflects a growing consensus in the research community. See Brainard et al. (2005); Mann (2006); Aldonis et al. (2007); Lawrence (2008); and Kletzer et al. (2007).

[21] In the late 1990s, Australia contracted out its job-matching and other services to its long-term unemployed. The so-called Job Network still exists and has been modeled by Oslington (2005), but project evaluations and empirical assessments remain to be accomplished.

pared to any other variety, as long as free riding can be disciplined by making OJT training incentive-compatible for both employers and employees (see below). Insurance companies in particular should be interested in new and incentive-compatible forms of 'worker-asset' insurance, backstopped perhaps by government as reinsurer.[22]

The detail of constituency-expanding reform might include:

- mandatory but refundable human-investment payroll taxes[23] for on-the-job training. Both workers and firms would contribute. Workers could or would receive their tax back once (and only if) they reach a given tenure threshold in a new job. Employers in that case (but not otherwise) would be allowed to deduct their share of the worker's cumulated training taxes from corporate income. Firms would have much stronger incentives to become training mediators. Unions and community colleges would have incentives to become firms' training sub-contractors (in addition to the training role they serve naturally). Arrangements like these are attractive for the prominent centrality of OJT that works and for their pay-or-stay incentive compatibility. The administrative costs are low in charging the payroll tax system to be overseer.
- insurance refinements. Current adjustment insurance, including standard UI, TAA or SAA, and wage and benefit insurance is too narrowly construed as a worker entitlement and an employer tax burden. The true stakes and stakeholders are much broader, and could be made concretely visible by refinements such as:
  - giving workers the opportunity to finance individual insurance accounts or voluntary supplements from privately provided add-ons to existing programs;[24]
  - rebalancing employer and taxpayer premium contributions, and adding worker contributions, all to enhance incentive compatibility;
  - adding (and sometimes adding back) principles of sound insurance management: deductibles, co-pays, and caps, all of which would be burdensome and unpopular by themselves, but which would be counter-balanced in principle by more generous training opportunities and dislocation or wage insurance.
  - participation mandates with narrowly construed default-option provisions for both workers and their employers.
- insurance innovation. Current adjustment insurance is focused almost exclusively on income flows; worker assets (skills, mobility, lifetime health, and per-

[22] Insurance companies are in fact already interested, as documented in Kletzer et al. (2007).

[23] Catherine L Mann has recommended these in a number of places. Human investment tax credits would supplement the way that wage insurance provides an implicit subsidy for training a firm's newest employees, as described above.

[24] Chile has been experimenting with a mixed public-individual UI system for several years. See Acevedo et al. (2006), summarized briefly in World Bank (2005, 154), and Senbruch (2004) for a more skeptical evaluation. See also a large number of assessments of the Danish 'flexicurity' system, publicly administered but privately chosen—or not—by Danish workers.

sonal assets) are inappropriately neglected. Innovation in workers' asset-value insurance would make adjustment to dislocation significantly less burdensome. For example:

- ○ workers' housing equity could be insured against specified types of structural-dislocation catastrophes;[25]
- ○ workers' educational and training investments might be similarly insured (and, perhaps, financed as well as insured, as suggested by the concept of training mortgages);[26]
- ○ even the community's tax base could be insured against structural trends beyond the control of its resident employers.[27]

Expanding the number of stakeholders in structural adjustment programs and aligning their incentives appropriately to minimize chronic contention, forms of cheating, and free riding is the key to twenty-first century SAA reform. Compared to the three motives for historic TAA, efficient adjustment becomes primary, and distributional equality and political compensation recede as motives.

Finally, the chances of success for the structural-adjustment initiatives described above are much enhanced when underpinned by two types of foundational civic infrastructure:

- best-practice public education[28] involving measurable upgrading in the American context, and
- best-practice active and passive labor-market policies (for example, raising national thresholds for core labor rights in the direction of international best practice).[29]

[25] Scheve and Slaughter (2001) show how popular support for border openness is lower among homeowner voters in communities with high import penetration, other voter characteristics being held equal.

[26] Insurance companies are always a blend of a mutual firm, pooling risk among their members, and a financial firm, taking in cumulated past premiums and paying out current and future claims. The recent global financial crisis has given a bad name to unsupervised financial innovation, but not to insurance innovation of the sort illustrated by weather insurance for crops and outdoor entertainment events, catastrophe bonds, and other types of creative, customized insurance products.

[27] Lawrence and Litan (1986, 119–122) made precisely this recommendation for trade-impacted communities in the context of historic TAA, and Lawrence has repeated it recently for a broader set of structural risks in Aldonis et al (2007, 48–9).

[28] See Richardson (2005b) for a description of the under-appreciated labor-market coping and adjustment benefits of improved primary, secondary, and higher education, including community colleges and training institutes. Worker-oriented innovation and reform in basic and higher education is a 'virtuous staircase', creating an ascent toward higher productivity and wages and toward broader footings from which to recover one's balance when confronted with change. In individual micro data, educational attainment is clearly correlated with upward quintile-to-quintile economic mobility, as documented in the Economic Mobility Project (www.economicmobility.org). And there is an intriguing negative micro-data correlation reported by Abowd et al. (2009) between the frequency of mass layoffs and (or) firm failure and the level of education (and skills and experience) of the firm's workforce—education may inhibit unanticipated change!

[29] See Richardson (2000) for ideas on a beginning agenda.

8. RECAP: NEW MOTIVES, NEW CONCEPTIONS FOR TWENTY-FIRST CENTURY WORKER ADJUSTMENT

Widely shared, globally engaged, efficient dynamism is the new motive for American adjustment initiatives. The policy refinements and innovations sketched here are not merely redistributive, not merely compensatory. They actually improve an economy's overall performance and welfare. They enhance its capacity to adapt to structural change and to negotiate deeper global integration by facilitating both new opportunity and risk management. A successful domestic policy infrastructure of the type discussed here is at least partially self-financing, with fewer distorting and unpopular burdens on taxpayers than one might naively expect.

'Adjustment services' are the new concept, rather than 'adjustment policies'. Americans are the most ingenious service providers in the world. The ideas described here are best conceived as services aimed at sharpening and broadening Americans' capabilities to engage in global dynamism. To re-conceive policies as services is not an academic exercise. In reality, firms and markets provide services, though often facilitated by policy. Firms include unions, schools, cooperatives, and not-for-profit organizations. If as services these ideas can be refined to work effectively, then they become one more American service sector with profitable comparative advantage, promising jobs, and global growth potential.

Inclusion is the deeper underlying motive for refining American approaches to adjustment assistance—inclusion of middle voters. Without their support, the nation will sacrifice vital future momentum in its standard of living because current internal, domestic policies are too weak to diffuse even large gains from global integration and structural dynamism widely across American society.

J. David Richardson is Professor of Economics and International Relations at the Maxwell School of Citizenship and Public Affairs, Syracuse University.

BIBLIOGRAPHY

Abowd, J.M, K. L. McKinney and L. Vilhuber (2009). The Link between Human Capital, Mass Layoffs, and Firm Deaths, in Dunne, initial et al., (eds.) 2009. Producer Dynamics: *New Evidence from Micro Data*. Chicago: University of Chicago Press. Conference on Research in Income and Wealth Studies in Income and Wealth No. 68

Acemoglu, D. (2009). *The Crisis of 2008: Structural Lessons for and from Economics*, CEPR Polii Insight No. 28, London: Centre for Economic Policy Research, January

Acevedo, G., P. Eskenazi and C. Pages (2006). *The Chilean Unemployment Insurance: A New Model of Income Support Available for Unemployed Workers?* Washington DC: World Bank

Aldonas, G.D. R.Z. Lawrence and M.J. Slaughter (2007). Succeeding in the Global Economy: A New Policy Agenda for the American Worker, *The Financial Services Forum*, 26 June *(http://www.hks.harvard.edu/fs/rlawrence/publicatgions.html)*

Andersson, F., H.J. Holzer and J. I. Lane (2005). *Moving Up or Moving On: Who Advances in the Low-wage Labor Market?* New York: Russell Sage Foundation

Baicker, K. and M. Rehavi (2004). Policy Watch: Trade Adjustment Assistance, *Journal of Economic Perspectives*, 18 (2, Spring), 239–55

Bergsten, C. F. (2005). *The United States and the World Economy: Foreign Economic Policy for the Next Decade*. Washington, D.C.: Peterson Institute for International Economics, January.

Bernard, A., S.J. Redding and P.K. Schott (2007a). Comparative Advantage and Heterogeneous Firms, *Review of Economic Studies*, 74 (1), 31–66.

Bernard, A., J.B. Jensen, S.J. Redding; and P.K. Schott (2007b). "Firms in International Trade," *Journal of Economic Perspectives*, 21 (3, Summer), 105–130.

Bown, C.P. and R. McCulloch (2005). "U.S. Trade Policy and the Adjustment Process," *IMF Staff Papers*, 52 (Special Issue), 107–28.

—(2007). "Trade Adjustment in the WTO System: Are More Safeguards the Answer?," *Oxford Review of Economic Policy*, 23 (3), 415–39.

Brainard, L., R.E. Litan and N. Warren (2005). "A Fairer Deal for America's Workers in a New Era of Offshoring," *Brookings Trade Forum 2005*, Washington, D.C.: The Brookings Institution.

Broda, C., E. Leibtag and D.E. Weinstein (2009). "The Role of Prices in Measuring the Poor's Living Standards," *Journal of Economic Perspectives*, 23 (2, Spring), 77–97.

Davidson, C. and S.J. Matusz (2004). *International Trade and Labor Markets: Theory, Evidence, and Policy Implications*. Kalamazoo, Michigan: W. E. Upjohn Institute for Employment Research.

Davis, S.J. and J. A. Kahn. (2008). "*Interpreting the Great Moderation: Changes in the Volatility of Economic Activity at the Macro and Micro Levels," Journal of Economic Perspectives*, 22 (4), Fall,. 155–180.

Decker, P. and W. Corson (1995). "International Trade and Worker Displacement: Evaluation of the Trade Adjustment Assistance Program," *Industrial and Labor Relations Review*, 48 (4, July), 758–774.

Gottschalk, P. and R. Moffitt (1994). "*The Growth of Earnings Instability in the U.S. Labor Market," Brookings Papers on Economic Activity*, 2, pp. 217–272.

Helpman, E., O. Itskhoki and S. Redding (2009). "*Inequality and Unemployment in a Global Economy*." Manuscript, May 19.

Jensen, J.B. and L.G. Kletzer (2005). "Tradable Services: Understanding the Scope and Impact of Services Offshoring," *Brookings Trade Forum 2005*, Washington, D.C.: The Brookings Institution.

—(2008). "*'Fear' and Offshoring: The Scope and Potential Impact of Imports and Exports of Services,"* Washington, D.C.: Peterson Institute for International Economics, Policy Brief No. PB08–1, January.

Jensen, J.B., L.G. Kletzer and C.L. Mann (forthcoming). *Trade in Services*. Washington, D.C.: Peterson Institute for International Economics.

Kletzer, L.G. and H. F. Rosen (2005). "Easing the Adjustment Burden on U.S. Workers," Ch. 10 in Bergsten et al. (2005).

—. (2006). "*Reforming Unemployment Insurance for the Twenty-First Century Workforce*," Hamilton Project Discussion Paper No. 2006–06. Washington, D.C.: The Brookings Institution, September.

Kletzer, L.G., J.D. Richardson and H. Rosen (2007). "*Preparing Workers and Communities for the 21st Century Economy: Designing Public and Private Adjustment Programs,*" Washington, D.C.: Peterson Institute for International Economics.

Lawrence, R.Z (2008). *Blue-Collar Blues: Is Trade to Blame for Rising U.S. Income Inequality?*, Washington, D.C.: Peterson Institute for International Economics, Policy Analysis No. 85, January.

Mann, C.L. with J.F. Kierkegaard (2006). *Accelerating the Globalization of America: The Role for Information Technology*. Washington, D.C.: Peterson Institute for International Economics.

Oslington, P. (2005). "Contracting-Out of Assistance to the Unemployed: Implications of the Australian Experiment," *The Economic Record*, 81 (March, No. 252), pp. 30–37.

Richardson, J. D. (1982). "Trade Adjustment Assistance Under the United States Trade Act of 1974: An Analytical Examination and Worker Survey," in Jagdish N. Bhagwati, ed., *Import Competition and Response*, Chicago: University of Chicago Press.

—(2000). "The WTO and Market-Supportive Regulation: A Way Forward on New Competition, Technological, and Labor Issues," *Federal Reserve Bank of St. Louis Review*, 82 (July/August), pp. 115–126.

—(2004). "Some Measurable Costs and Benefits of Economic Globalization for Americans," in John O'Loughlin, Lynn Staehli, and Edward Greenberg, eds., *Globalization and Its Outcomes*, New York: Guilford Press, Ch. 9, pp. 182–92.

—(2005a). "*Uneven Gains and Unbalanced Burdens? Three Decades of American Globalization,*" Ch. 3 in Bergsten et al. (2005).

—(2005b). *Global Forces, American Faces: U.S. Economic Globalization at the Grass Roots*. Unpublished Manuscript. Washington, D.C.: Peterson Institute for International Economics.

Rosen, H. (2006). "Trade Adjustment Assistance: The More we Change, the More It Stays the Same," Ch. 5 inMichael Mussa, ed., *C. Fred Bergsten and the World Economy*. Washington, D.C.: Peterson Institute for International Economics, December.

—(2008). "*Strengthening Trade Adjustment Assistance*". Washington, D.C.: Peterson Institute for International Economics, Policy Brief No. PB08–2, January.

—(2008). *Designing a National Strategy for Responding to Economic Dislocation*." Washington, D.C.: Peterson Institute for International Economics. June 24. Available at http://www.petersoninstitute.org/staff/author_bio.cfm?author_id=101. Accessed 5/27/09.

Scheve, K.F. and M.J. Slaughter (2001). *Globalization and the Perceptions of American Workers*. Washington, D.C.: Institute for International Economics, March.

Senbruch, K. (2004). "Privatized Unemployment Insurance: Can Chile's New Unemployment Insurance Scheme Serve as a Model for Other Developing Countries?", University of California, Berkeley, Center for Latin American Studies, Working Paper No. 12, December

US Department of Labor, Employment and Training Administration (2009). "*TAA Statistics*," http://www.doleta.gov/TradeAct/taa_stats.cfm. Accessed 5/18/09

US General Accounting Office (2000). *Trade Adjustment Assistance: Trends, Outcomes, and Management Issues in Dislocated Worker Programs*. Washington, D.C.: GAO-01–59, October

Violante, G. (2002). "Technological Acceleration, Skill Tranferability, and the Rise in Residual Inequality," *Quarterly Journal of Economics*, 117 (), pp. 297–338

World Bank. 2005. *World Development Report*. Washington, D.C.

24

Compensation Payments in EU Agriculture

JOHAN F M SWINNEN AND KRISTINE VAN HERCK

1. INTRODUCTION

Total spending in the European Union (EU) on the Common Agricultural Policy (CAP) in 2008 was in excess of €52 billion and spending on direct aids alone was almost €37 billion, a large share of the total EU budget (Table 24.1). These payments were initially introduced as compensation payments to farms, when the EU lowered import tariffs and price support. The payments have been reformed since—in the process of which the word 'compensation' was dropped. To understand this we need to take a brief historical tour on agricultural policies in the EU.

The CAP was designed in the late 1950s and introduced in the late 1960s. The official objectives as stated in Article 33 (39) of the EC Rome Treaty (1958) are to:

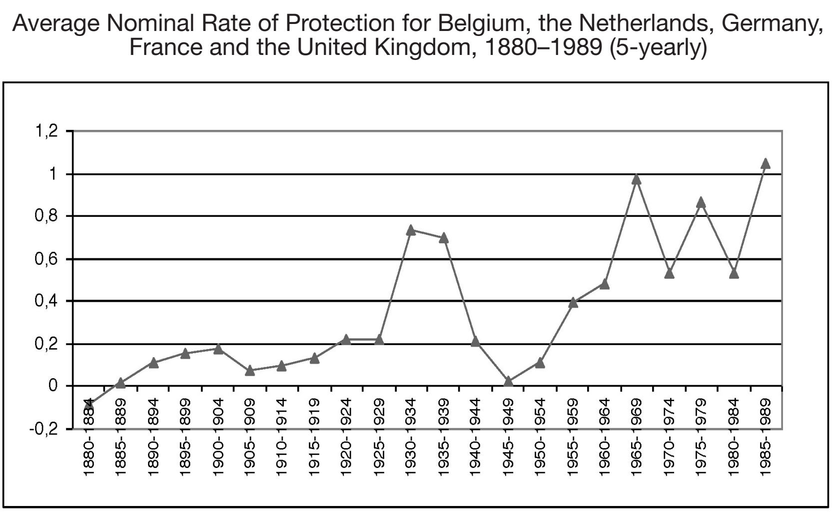

Figure 24.1: *The Growth of Agricultural Protection in Europe*

Source: Swinnen (2009b)

Table 24.1: *CAP Budget*

Heading	Appropriations 2008		Appropriations 2007		Outturn 2006	
	Commitments	Payments	Commitments	Payments	Commitments	Payments
Administrative expenditure of agriculture and rural development policy area	130 325 016	130 325 016	125 674 851	125 674 851	109 489 381,55	109 489 381,55
Interventions in agricultural markets	4 032 371 000	4 033 571 000	4 941 694 000	4 938 759 000	8 066 747 919,52	8 066 747 919,52
Direct aid	36 832 000 000	36 832 000 000	37 066 533 000	37 066 533 000	34 051 330 746,02	34 051 330 746,02
Rural development	12 926 551 889	11 379 281 817	9 897 556 092	9 657 686 782	11 931 312 505,15	11 328 848 347,59
Pre-accession measures in the field of agriculture and rural development	85 300 000	385 000 000	48 300 000	265 900 000	299 820 000,–	213 755 071,87
International aspects of agriculture and rural development policy area	6 230 000	6 230 000	6 161 000	6 161 000	5 817 680,62	6 185 630,64
Audit of agricultural expenditure	–342 500 000	–342 500 000	–86 500 000	–86 500 000	–275 097 022,97	–275 108 092,85
Policy strategy and coordination of agriculture and rural development policy area	31 450 000	34 060 500	41 174 000	41 149 756	36 557 969,78	37 214 746,82
Administrative support for Agriculture Directorate-General						
Total	53 701 727 905	52 457 968 333	52 040 592 943	52 015 364 389	54 225 979 179,67	53 538 463 751,16

Source: European Commission

1. increase agricultural productivity by promoting technical progress and ensuring the optimum use of the factors of production, in particular labour;
2. ensure a fair standard of living for farmers;
3. stabilize markets;
4. assure the availability of supplies; and
5. ensure reasonable prices for consumers.

The CAP resulted from the integration of various pre-EU member state policies which were introduced to protect EU farmers' income and employment from foreign competition and market forces. Figure 24.1 shows the long-term evolution of agricultural protection in Europe and clearly illustrates how protection increased very rapidly in the post-World War II decades. Political economists have explained this growth in protection by the decline in farm incomes compared to rapidly growing incomes in the rest of society, as well as the declining opposition of consumers and industry to tariff protection for agricultural commodities (Swinnen 2009a). Hence, the main objective of agricultural policies in the EU, and the main determinant of the level of agricultural protection was provision of compensation and support to a sector in (relative) economic decline in order to protect incomes and employment from market forces.

The mechanism of support was through high income tariffs, export subsidies, and fixing prices. While this created much stability on the EU market (directly related to Objective 3 of the CAP objectives) it created much instability on world markets, and considerable distortions throughout the economy.

Since the integration of agriculture in the GATT (WTO) the CAP instruments have undergone major reforms, including the introduction of compensation payments in the 1990s and the move to decoupled payments with the 2003 and 2008 reforms. The reforms in the 1990s and 2000s have substantially reduced the trade distortions of the CAP, in particular through the decoupling of the single farm payments (SFP) which are currently applied in the EU–15, and which are to be implemented by the New Member States (NMS) in the coming years.

However, what is important is that the level of these payments is still very much influenced by the initial objective of supporting incomes and employment in agriculture. To understand this, we briefly review the initial policies and the reforms since the start of the CAP.

2. THE HISTORY OF CAP POLICY INSTRUMENTS AND REFORMS

When the CAP was designed at the end of the 1950s and initially implemented in the 1960s the essence was a system of government interventions in the market to support a minimum price for farmers. This domestic intervention system was accompanied by trade measures to make it work: variable import tariffs (levies) and export subsidies (refunds) were set to isolate this system from international markets. The system (and the names given to the various instruments) differed between commodities. The most profound interventions occurred in mar-

Table 24.2: *The Main Instruments Used for the Implementation of the CAP—Selected products*

	Cereals[1]	Sugar	Dairy[2]	Beef/Veal	Sheep meat	Fresh fruit and Vegetables[2]	Processed fruit	Wine[3]
Intervention	X	X	X	X	X	X[4]	X[5]	X
Storage aid			X	X	X			X
Direct aid	X[6]		X	X	X	X[7]	X[8]	
Import levies and export refunds	X	X	X	X	X[9]	X	X[10]	X[11]
Co-responsibility levies	X	X	X					
Guarantee threshold	X						X[12]	
Production quota		X	X					

1 Except rice
2 Arrangements generally applicable only in periods of large-scale marketing
3 Only table wines are subject to the prices and intervention systems
4 Intervention only in a ‘crisis situation’; otherwise, ‘withdrawal’ of surpluses at a low price
5 No levies on imports
6 For durum wheat produced in certain regions of Italy, Greece, and France
7 For citrus fruit
8 Aid for processing selected products, in some cases with a quantitative ceiling. The products concerned are various tomato derivates, dried figs, raisins, a particular type of prune, and preserves in syrup (cherries, peaches, and William pears)
9 In the case of voluntary export restraints, levies may not exceed amounts laid down in the agreements
10 For a limited number of products
11 Provided that the import price is not lower than the relevant reference price, there are no levies on imports
12 For aid for processing tomatoes
Source: Rosenblatt 1988

kets of sugar, beef, dairy, wine, cereals, and oilseeds. Table 24.2 gives an overview of the main instruments used for the implementation of the CAP in different agricultural productions.

Intervention prices were set considerably above market prices. For some commodities, such as butter and white sugar, EU prices were four times the price on the world market, but also for other commodities EU prices largely exceeded world market prices. Table 24.3 gives the differences between the EU price and the world market price for selected commodities in 1967–68. This price structure resulted in a large increase in agricultural production. Between 1973 and 1988, the volume of agricultural production increased by 2 per cent per annum whereas internal consumption increased only by 5 per cent. This resulted already at the end of the 1970s in a high degree of self sufficiency (Figure 24.2) and the EU shifted from a net import to a net export position in agricultural and food products. In combination, these contributed to rapidly growing budgetary expenditures (for market intervention, storage, export subsidies, and so on) and distortions of international markets. Both resulted in pressures for reforms. These reforms and also the future of the CAP are still dominated by the early outline of the CAP, not only in terms of dealing with the surplus production and the environmental problems caused by intensive farming practices, but also regarding farmers' attitudes towards price policy (Fennell 1997).

Reforms of the CAP were proposed soon after its introduction. As early as 1968, Commissioner Mansholt proposed a plan to accelerate structural change in the agricultural sector. The main proposals in the plan were to implement monetary incentives to encourage about half of the farming population to leave the sector

Table 24.3: *Prices for Certain Agricultural Products in the EU Compared to the World Market Price Level in 1967–1968*[a]

	EU Common price ECU/ 100kg (1)	World market price ECU/ 100kg (2)[c]	(1) as a percentage of (2)
Soft wheat	10.7	5.8	185
Hard wheat[b]	16.1	8.1	200
Husked rice	18.0	15.3	117
Barley	9.1	5.7	160
Maize	9.0	5.6	160
White sugar	22.3	5.1	438
Beef	68.0	38.8	175
Pig meat	56.7	38.6	147
Poultry meat	72.3	55.0	131
Eggs	51.1	38.7	132
Butter	187.4	47.2	397
Olive oil	115.6	69.8	166
Oilseeds	20.2	10.1	200

[a] reference price differs for various products
[b] including direct production aids
[c] wholesale entry price
Source: Fennell 1997

Figure 24.2: *Self Sufficiency in the EU—Selected products*

Source: European Commission (2007)

by taking early retirement or engaging in alternative employment. In addition, he proposed that aid only should be provided to farmers who had a sufficient scale or farmers who engaged in a large jointly managed holding. However, strong opposition from farmers caused governments and the European Council to reject the proposal (Stead 2007).

To deal with the growing surpluses, in the mid 1980s production quotas were imposed in the sugar and dairy markets to control supply (and thus the budgetary effects and market distortions), while maintaining high support prices. Extension of this system to the cereals market was considered but the transaction costs (for monitoring, administration, and enforcement) of the system were deemed too high to be practical in the cereals market.

With the integration of agriculture in the GATT, pressure from trading partners also grew. Ultimately, a new approach was decided by lowering support prices to reduce market distortions and compensating farmers through compensation payments—later referred to as direct payments—linked to the area used (for example, cereals and oilseeds) or to animals (for beef). This was the most important part of the so-called MacSharry reforms in 1992.

The Agenda 2000 reforms basically represented a deepening and extension of the 1992 reforms (Ahner and Scheele 2000). Price support was reduced further for cereals and beef and the direct payments in these sectors were increased to compensate farmers at least partially for the price cuts. A similar reform with price cuts and the introduction of direct payments was initiated in the milk sector, but only from 2005 onwards. The reform was necessary for several reasons (Swinnen 2002; Van Meijl and Van Tongeren 2002). First, the enlargement of the EU with

10 Eastern European countries, which still had a relatively high share of agricultural produce in total production and employment, would have unsustainable budget implications if the CAP was not reformed. Second, without additional reforms the EU would not fulfill the commitments made under the GATT Uruguay Round Agreement on Agriculture (URAA). The combined result of the MacSharry and Agenda 2000 reforms implied, at least for the sectors concerned, a major shift from support through price and market interventions to farm support through direct payments.

The relative share of the EU agricultural budget in the total EU budget has declined somewhat (Figure 24.3). However, Table 24.4 shows that the total amount of support to agriculture has not declined. Moreover, some argue that support to agriculture has increased more than indicated by the numbers in Table 24.4, because compensation through direct payments was based on gross revenue declines, while net incomes have declined much less.[1]

Figure 24.4 also indicates the growth in expenditures on rural development. The Agenda 2000 decisions imply the consolidation of the EU rural development policy under the CAP (Ahner and Scheele 2000). While the budgetary allocations remain moderate compared to the other expenditures; the growing importance of rural development follows from the official reference to it as the 'second pillar of the CAP'.

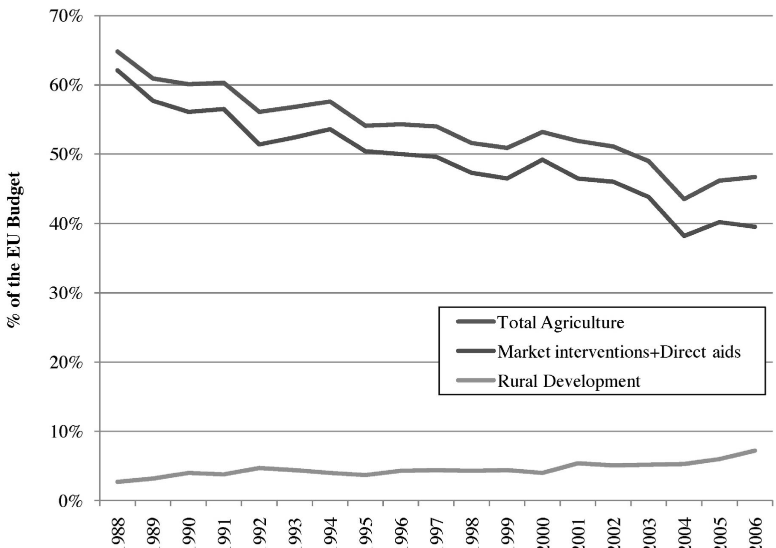

Figure 24.3: *EU Agricultural Budget as a Percentage of the Total EU Budget*

Source: European Commission

[1] For example, OECD calculations on transfer efficiencies of OECD agricultural policies suggest that the average net income gains from market and price support in OECD countries was only 20 per cent (OECD 1997). This means that, after factor markets and so on have adjusted to the new situation, a gross income decline, of say, € 100 is causing a smaller net income decline. Hence compensation based on gross income decline is overcompensating, the extent of which depends on the transfer efficiency of direct payments, which are also less than 100 per cent.

Table 24.4: *Support to EU Agriculture (Total and distribution)*

	1986–88	1989–91	1992–94	1995–97	1998–00	2001–03	2004–06
TOTAL SUPPORT (PSE%)*	41	34	37	35	37	35	34
Of which (in per cent)							
Market price support and payments based on output	91	85	72	61	64	56	52
Payments based on area planted/ animal numbers	3	7	19	31	29	34	14
Payments based on input use	5	7	7	7	7	9	10
Payments based on historical entitlements, input constraints and farm income**	1	1	2	1	0	1	14

* PSE%: Producer Subsidy Equivalent; measures total support to agriculture as a percentage of the production value

** For the 1991–1993 average, this category also includes miscellaneous payments (approximately 2 per cent of total)

Source: OECD

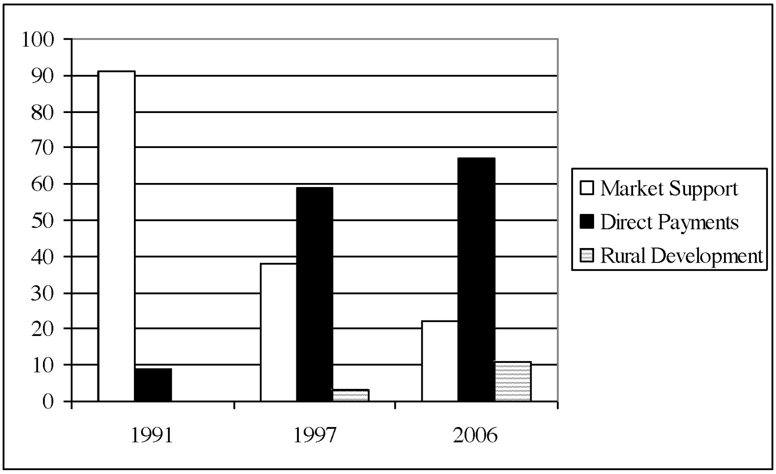

Figure 24.4: *Distribution of the EU Agricultural Budget (1991–2006)*

There were several reforms prepared and implemented over the two terms when Franz Fischler was Commissioner for Agriculture and Rural Development of the EU, which spanned almost a decade (1996–2004). Some of these reforms, such as the Agenda 2000 package, were important. However, his name is most associated with the reform of 2003, which, at the time, was generally referred to as the 'Mid-term Review', a term which, in hindsight, does not do justice to the extent and substance of the reform package that was decided in 2003. Those reforms were assessed as the most radical reform of the CAP since its creation (Olper 2008; Swinnen 2008).

The 2003 Fischler reforms contain the following key elements:

1. The key innovation was the introduction of the Single Farm Payment (SFP) on the basis of historical entitlements (although with some flexibility of application), decoupling a large share of CAP support from production. This reform essentially ensured that farms would continue to receive the amount of payments they received in the past, but no longer linked to their production activities, but as a single 'decoupled' payment.
2. New instruments called 'cross-compliance' and 'modulation' were introduced. Cross-compliance requirements are to ensure that SFP is only paid to farmers who abide by a series of regulations relating to environment, animal welfare, plant protection, and food safety. Modulation refers to the shift of funds to rural development policies (that is, from Pillar I to Pillar II) by limiting payments to the largest farms.
3. The reforms introduced changes in several market organizations, in particular in the dairy and rice sectors, by increasing dairy quotas and reducing price-support policies, replacing them by direct support to be integrated in the SFP.

The 2008 'Health Check' reform introduced relatively minor changes except for a substantial reform in the dairy sector.

3. THE WTO AND THE CAP

Since the conclusion of the URAA in 1992, EU subsidies to agricultural production and exports are constrained by World Trade Organization (WTO) rules. Among others, there are restrictions on the total support to agriculture and on both the amount of export subsidies and the volume of exports that can be subsidized.[2]

Several observers argue that the implementation of the WTO did not directly cause major trade and policy liberalization in the EU (Josling and Tangermann, 1999; Swinbank, 1999). Yet the URAA is an important factor for the CAP for several reasons. First, the URAA brought the link between the domestic policy aspects of the CAP and its international trade implications to the top of the policy agenda, something which was new in the EU at the time but which has since fundamentally changed CAP decision-making. Second, the URAA provided the key initiatives for the 1992 MacSharry reforms and, given the Eastern enlargement interactions with the WTO, for Agenda 2000 reform.[3] Third, the URAA provided a framework for future negotiations. A continuation of reductions in the

[2] Under the URAA a considerable amount of support in both the United States and the EU was classified as 'blue box' or 'green box' support. The green box is a category of so-called 'non- or minimally trade distorting' support policies. These green box support policies are not restricted under WTO rules. The blue box includes the EU direct payments which were introduced under the MacSharry and Agenda 2000 reforms. See for example, Burrell (2000); Josling and Tangermann (1999); and Swinbank (1999) for more extensive discussions and analyses.

[3] See Swinnen (2002) for more details.

subsidies under the next negotiation round could cause much more serious implications for the CAP. Many claim that the anticipation of this outcome was a crucial element in the 2003 Fischler reforms (Swinnen, 2008).

4. CAP AND ADJUSTMENT

A key question is whether the CAP payments in the past have been effective in achieving their objectives of ensuring a 'fair standard of living' and 'stabilizing markets'. When one looks at the short run, the answer on the income question is obviously: yes. Annual payments do increase farms' incomes—how can they not? And farm accounts and statistics will show that they can amount to a substantial share of farm net incomes for a given year, depending of course on the location and the specialization of the farms and the market situation of the particular year. However, this is a very unsatisfactory way of answering this question. One should look at how the CAP payments affect (relative) farm household incomes in the long run. And then the answer is much less obvious.

First, studies generally show that farm incomes (narrowly defined) are still behind average incomes but that farm household incomes are roughly the same (and sometimes higher) than average household incomes in the EU. The reason is that non-farm incomes make up an increasingly larger share of farm household incomes with the improved integration of rural areas in the rest of the economy.

Second, the reason why farm household incomes have grown is mostly due to the integration of rural areas and rural (output and factor) markets in the general economy over the past decades. Integration of rural capital markets has reduced the cost of capital; integration of rural labor markets has improved access to non-farm employment opportunities for farm households; and integration of services has improved both incomes and the quality of living in rural areas. Note that none of these factors has much to do with CAP payments.[4]

Another indicator of the effectiveness of CAP payments in terms of supporting agricultural incomes and employment is to look at the evolution of agricultural employment. The employment effects can also be interpreted as a rough indicator of relative incomes (through revealed preferences: if people had a good income from farming, they would stay in agriculture).

Figures 24.5 and 24.6 illustrate the decline in agricultural employment in the EU (we used data from some of the member states because EU total averages are strongly affected and (or) distorted by enlargements). Over the past two decades, despite the CAP, employment in agriculture fell by 35 per cent to 50 per cent. Although one cannot draw definite conclusions from such visual analysis without looking at the counterfactual, the data do confirm that agricultural employment in the EU has declined very strongly over the past decades, despite the large CAP support.

[4] The same conclusions and mechanisms apply to other countries including the United States (see various papers by Bruce Gardner).

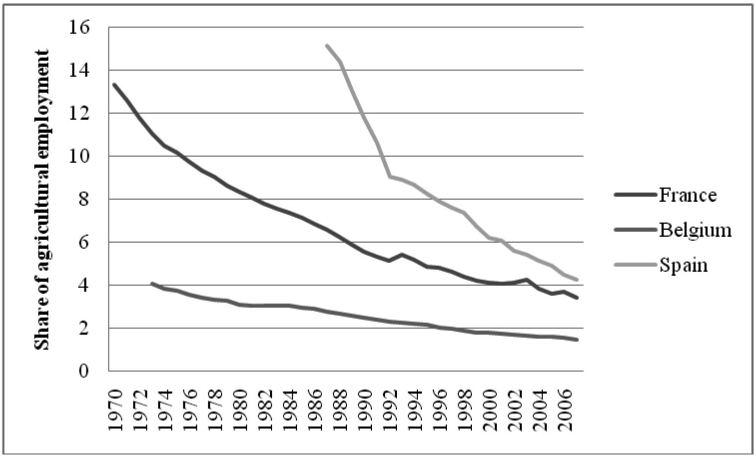

Figure 24.5: *Evolution of the Share of Agricultural Employment*

Source: ILO; Eurostat

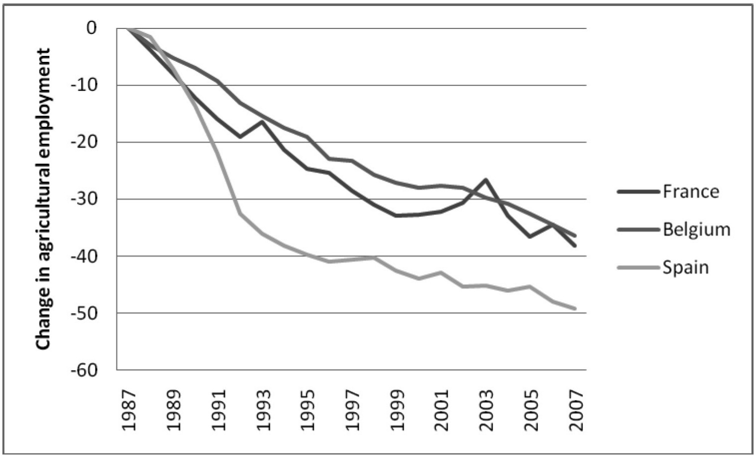

Figure 24.6: *Change in Agricultural Employment (per cent)*

Source: ILO; Eurostat

Long-run studies using much more detailed data and sophisticated statistical techniques largely confirm this conclusion: that CAP payments either had no effect or only a minor effect on employment.[5] In fact, what is interesting is that OECD data show, first, that over the past two decades (the 1987–2007 period) there is no positive relationship between (changes in) agricultural employment and (changes in) agricultural support (captured by the PSE indicator[6]) across the OECD countries (see Figures 24.7a and 24.7b). Moreover, over this period, there is actually a negative correlation between the change in agricultural support and the change in agricultural employment (see Figure 24.7c)—which is inconsistent with the notion that agricultural support has a significant impact on agricultural employment in the long run.

The reason why CAP payments have limited impact on relative farm incomes and employment is because of a combination of policy-rent dissipation and poor targeting. OECD studies showed that the net income effects for farmers of commodity price supports (the old CAP) were around 20 per cent, meaning that 80 per cent of the payments ended up with non-farm groups, including input-supplying companies and landowners (and reduced prices to non-EU consumers and producers). This rent dissipation has improved (that is, has been reduced) with the shift to area/animal payments and to single farm payments, but only so far, and not as much as the improvement in terms of output-market distortions. The main reason is that these payments are still linked to land use and are driving up land prices. For example, with the accession to the EU, land market prices and rents have increased very strongly in the NMS (between 100 per cent and 300 per cent—see Figure 24.8[7]). While the current payments in the EU-15 are decoupled from production, they are not decoupled from land use and, thus, continued dissipation of policy rents from farms to landowners should be expected[8].

Another factor is that much of the support goes to larger and typically better managed and more dynamic farms, often located in the richer areas of the EU. Notice that the shift from price support to direct payments (either area or SFP) has not changed this outcome, because the payments are based on historical CAP benefits.[9] Hence, the farms that have the lowest incomes in the EU typically receive least of the CAP payments.

Some conclusions from this analysis are as follows. Farm household incomes have caught up with those in the rest of society, but mostly because of other factors than CAP payments, that is, the integration of rural areas in factor markets and the rest of the economy. Agricultural protection under the CAP (and the di-

[5] See Tweeten (1979) and Barkley (1990) on the large employment reductions in the 1950s through the 1980s in the United States despite massive government subsidies to agriculture; Glauben et al. (2006) for Germany; other studies focusing on the nature of the farm subsidies find mixed, but generally small, effects (see for example, Dewbre and Mishra (2002); Serra et al. (2005).

[6] The PSE measures the share (in per cent) of the gross value of agricultural output which is due to government support.

[7] See Swinnen and Vranken (2009) on the impact of accession on NMS land markets.

[8] See Salhofer and Schmid (2004) for a formal analysis.

[9] This argument depends on the implementation of the SFP (regional versus historical model) and modulation mitigates this effect somewhat.

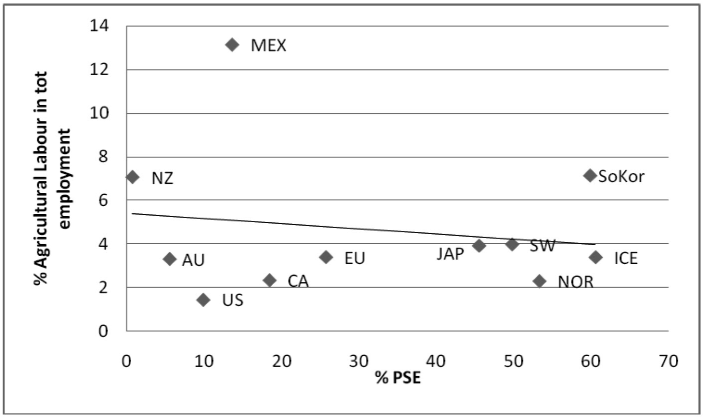

Figure 24.7a: *Share of Agricultural Labour in Total Employment and Percentage PSE in 2007*

Source: ILO, national statistics

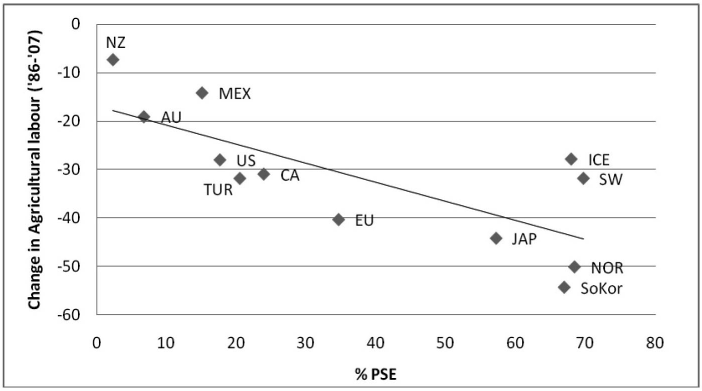

Figure 24.7b: *Change in Agricultural Labour and Percentage PSE (1987–2007)*

Source: ILO, national statistics

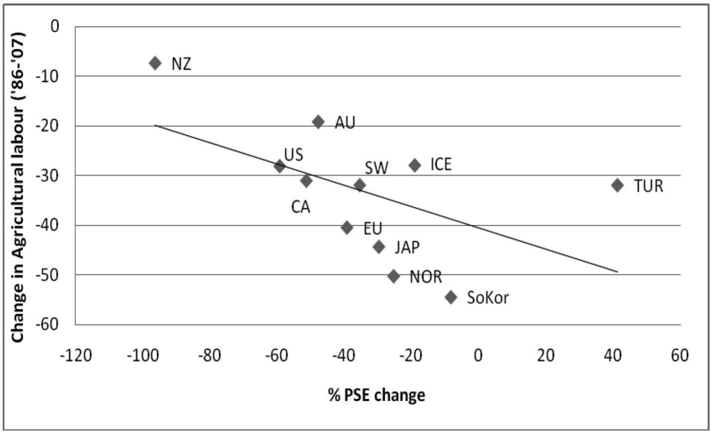

Figure 24.7c: *Change in Agricultural Labour and Change in Percentage PSE (1987–2007)*

Source: ILO, national statistics

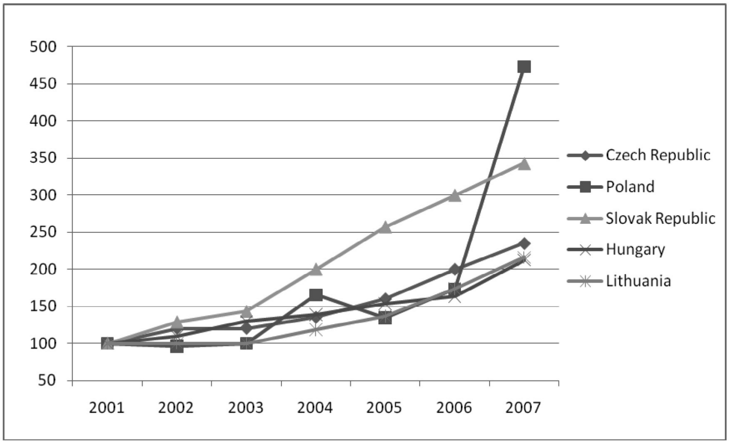

Figure 24.8: *Change in Land Rental Prices in the NMS*

Source: Swinnen and Vranken (2008)

rect payments) has not been effective at protecting EU agricultural employment in the long run. However, another interpretation of the same observation is that in a long-run perspective CAP payments have not created major distortions in the economy in terms of keeping labor in agriculture that otherwise would have been employed more productively in the rest of the economy.

These observations can be reconciled with each other in a political economy framework forwarded by the Berkeley–Cornell school (in particular by Gordon Rausser and Harry de Gorter and their collaborators). They interpret the joint determination of agricultural support and investments in productivity-increasing investments and activities as a mutually reinforcing decision. As people active in agriculture are hurt from productivity growth in agriculture (with inelastic demand) and in the rest of the economy, continued support for productivity growth (which is efficiency-enhancing) needs to be complemented with support for sectors in relative decline (such as agriculture) in order to be politically sustainable.[10]

Finally, this brief historical review and the (political) economic analysis points at some crucial elements and fundamental arguments in the discussion on the future of the CAP payments. Many of the reports and studies which focus on the so-called 'new objectives' of the CAP seem to ignore (accidentally or deliberately) the fundamental fact that the amount of CAP payments that are currently spent are a direct consequence of the history of the CAP and its reforms. The introduction and size of compensation payments was to compensate farmers for income losses due to the removal of price distortions that existed under the 1970s and 1980s CAP. Since these instruments and the derived payments were introduced with a main objective to support farm incomes, employment, and protect EU farmers against foreign competition, one should first address whether this is no longer an objective—and if not, ask why we need to continue the level of payments which has mostly been determined by these objectives.

4.1 Stabilizing markets and incomes?

Another important issue is the role that CAP subsidies play in stabilizing markets and incomes. As explained above stabilizing markets was one of the initial formal objectives of the CAP. The dramatic changes (both increases and decreases) in commodity and food markets over the past two years has raised concerns regarding the importance of addressing risk and uncertainty for farmers and other agents active in agricultural and food markets. Many of the reports on the future of the CAP also mention the importance for intervention to provide stability to markets, farm incomes, and to provide a (social) safety net.

[10] See for example Rausser (1992); de Gorter et al. (1992); and Swinnen and de Gorter (2002). In addition, Foster and Rausser (1993) argue that support instruments such as price supports that benefit the most efficient farms can be an efficient instrument from the perspective of reducing political opposition to growth-enhancing investments, while inducing the least efficient producers to leave the sector and the most efficient producers to continue. The CAP instruments, including the direct payment and historical SFP system—which are based on historical, that is price-support-determined levels of support are consistent with these arguments (see also Harvey 2004).

The impact on stabilization is more nuanced. It is important first to point out that reducing variability of prices, of incomes, and providing a safety net are not the same objective (they may even be conflicting). The old CAP system of government price interventions reduced price variability on the internal EU market, but at a huge cost in terms of inducing market distortions (both internally and on the world market), and it did not provide a good safety net as most of the benefits went to larger farms and much less support went to farms with low incomes.

The current direct payment system has less or no impact on price variability, but does reduce income variability and reduces risk in farming households by providing a guaranteed source of income.[11] In terms of risk reduction and insurance provision, there are a variety of private sector instruments available, and the question is (a) whether direct payments do a better job at providing insurance than market-provided instruments, and (b) why such instruments should be focused on agriculture and not on other sectors of the economy which are also facing problems of variability in markets—for example from energy prices.

In addition, the fact that direct payments provide an income guarantee does not imply that direct payments are an effective instrument to provide a social safety net—at least not under the current implementation. In order to provide a safety net at the EU level, the level of income support should increase when farm incomes fall below a certain threshold level. However, the direct payments are historically determined, based on the previous level of support which, at the farm-level, has little correlation with the likelihood of the farm household's income falling below a certain income level. In fact, given the historical distribution of farm support among regions and farms, the opposite is more likely to be the case: the most productive farms in regions where the most subsidized commodities were produced are most likely to have the highest level of payments. If direct payments were to serve as a safety net, they would have to be linked to the level of income.

5. THE FUTURE

We are at an historic moment in time, both in terms of policy timing and in terms of the challenges that face us. This forces us to raise more fundamental questions regarding the CAP.

Successive reforms of the CAP have been successful in reducing the market distortions caused by the CAP, from the price and market intervention system to the decoupled single farm payments. The question that we are facing now is whether the SFP system, either in its current form or in a modified form is likely to address the key challenges in the future. The most daunting challenges appear to be reducing and (or) mitigating climate change and producing sufficient, safe, and high-quality food.

[11] And as such, they may have an impact on production as they affect farm decisions in uncertain environments although the size of the effect is likely to be relatively small (see for example Hennessy 1998; Goodwin and Mishra 2006; Sckokai and Moro 2006).

Table 24.5: *List of CAP Objectives Proposed by Bureau and Mahé (2008)*

1.	To foster the economic performance and the competiveness of the farm and food policy chain
2.	To provide a buffer against extreme market or natural conditions and exceptional price falls; and to assist in the development of self-sustained schemes to reduce income volatility
3.	To ensure the availability of food supplies and to contribute to food security
4.	To ensure that food products reach consumers at competitive prices
5.	To meet consumer demand for safety and high quality food
6.	To preserve the natural resources of rural areas and to control pollution, with specific attention to environmentally sensitive and high-value portions of rural territories, to biodiversity and to ecosystems (note that the idea of considering organic farming according to its social benefits should be more explicitly mentioned)
7.	To encourage a degree of farming activity in areas with natural handicaps
8.	To ensure that fiscal resources devoted to agriculture and rural programs are effective and that the CAP is consistent with EU priorities and with other EU policies
9.	To harmonize the effectiveness of support with equity among individuals and with cohesion across regions and member states
10.	To require methods and processes of food production to be consistent with European values and ethics
11.	To ensure a fair standard of living and to expand earning opportunities for rural populations
12.	To ensure that the poorest and most deprived sections of the population have guaranteed access to food
13.	To preserve the European heritage of food variety
14.	To preserve the rural heritage of EU member states

Source: Bureau and Mahé (2008)

Past reforms have introduced some new official objectives in the CAP. In line with the requirements of EU citizens, the following factors have taken on greater importance, according to the European Commission (2007): improving the quality of Europe's food and guaranteeing food safety (standards); looking after the well-being of rural society; supporting the multifunctional role of farmers as suppliers of public goods to society, and ensuring that the environment is protected; providing better animal health and welfare conditions; doing all this at minimal cost to the EU budget.[12] This additional list of new factors or objectives is reflected in pillar II priorities and the so-called cross-compliance regulations, that is the conditions farms have to satisfy in order to receive the payments.

Regarding the future CAP, several task forces and reports have developed an even larger set of adjusted objectives for the CAP. For example, Bureau and Mahé (date) present a list of 13 policy objectives for their future CAP model (see Table 24.5). In contrast, the IEEP report (Baldock et al. 2008) presents two main new objectives: (1) to maintain the EU's capacity to produce food and maintain a renewable resource base in the longer term, and (2) to provide environmental benefits (including biodiversity, valued landscapes, and so on).

Needless to say, the extension of the list of objectives makes the entire exercise of identifying precise objectives and developing targeted instruments not easier—which is recognized by some of the authors of the reports—who then also list the need for simplicity and low transaction costs as additional factors to take into consideration.

[12] As listed in European Commission (2007), *The Common Agricultural Policy Explained*, Directorate for Agriculture, European Communities, Brussels. Available online: http://ec.europa.eu/agriculture/publi/capexplained/cap_en.pdf

In Swinnen (2009a) I review the objectives which are most often presented and which seem to be the ones with the most important budgetary and policy implications: food security and environmental benefits. I conclude that EU direct payments generally are not an effective way of dealing with these challenges. Food safety and quality objectives are addressed by other policies and direct payments have a very limited role to play in this.

In terms of providing a sufficient quantity of agricultural output, major challenges appear on the horizon. Even without government support for biofuels, demand for agricultural commodities for bio-energy purposes is likely to increase strongly in the long run—as we should expect oil prices to recover in the coming years. Similarly, the growth in food and feed demand from emerging countries, such as India and China, is likely to continue. Both fundamental developments are affected by the current financial and economic crises in the world economy, but in the longer term one should expect them to resume their critical importance. On the production side, productivity trends in the EU and other developed countries face declining growth rates. These fundamental trends will cause an upward pressure on agricultural and food prices.

Furthermore, climate change is likely to have a significant impact on EU agriculture. Although it may actually have a positive effect on aggregate EU output in the medium term, it is likely to imply major relocations and the need to adjust production systems. Vice versa, EU agriculture continues to contribute importantly to GHG emissions.

From a policy perspective all this has important implications.[13] One implication is that real agricultural market prices are likely to increase in the future. As a result, there are fewer arguments for governments to support farm incomes. This in itself has major implications for the use of direct payments, since their history and level have been determined by the perceived need and political demand for farm income support.

Direct payments can play some role in reducing income variation and household risk in the future, but they would have to be reformed fundamentally in order to become a real safety net. Moreover, their effectiveness in terms of risk reduction and providing insurance have to be compared with private sector instruments; and their effectiveness in terms of a social safety net has to be compared with that of an economy-wide social policy system, which provides a safety net across sectors. In both cases, policy and private sector instruments focused not on agriculture but on the entire economy are likely to be more efficient.

Given the daunting challenges to produce more agricultural commodities for food and non-food purposes, in combination with the challenges imposed by climate change, and the lagging productivity growth rates in the EU, there is a strong case for support and investments in R&D and technology development and diffusion: (a) to improve the lagging productivity of agricultural production, (b) to reduce the pressure of bio-energy on food prices, (c) to reduce the negative

[13] These are in addition to potential consumer policies, such as advising a less meat-intensive diet.

aspects of the relationship between agriculture and climate change, (d) to reduce energy-dependency in agricultural production, and (e) to pursue these efficiency objectives while taking into account important (additional) environmental constraints and objectives.[14]

In this perspective, the EU should consider instead of spending the budget on direct payments, to reallocate a substantial part of the CAP budget to stimulate the development and implementation of a series of new and improved (green) technologies to stimulate the EU rural–food–bio-economy.[15] It appears that such strategy could have major spillover effects on the rest of the economy in potentially leading to overall productivity gains and improved environmental conditions.

Johan Swinnen is Professor at and Director of LICOS – Centre for Institutions and Economic Performance, K.U.Leuven.

Kristine Van Herck is a researcher at the LICOS Centre for Institutions and Economic Performance, K.U.Leuven.

BIBLIOGRAPHY

Ahner, D and M Scheele (2000). *The Common Agricultural Policy: An overview from where we come, where we are, tasks ahead*, a paper presented at the International Postgraduate Course on European Agricultural Policy in Transformation, Wageningen University, 15 September 2000

Baldock, D, Cooper, T, Hart, K and M Farmer (2008). *Preparing for a New Era in EU Agricultural Policy*, Institute for European Environmental Policy, London

Barkley, A P (1990). The Determinants of the Migration of Labor out of Agriculture in the United States, 1940–85, *American Journal of Agricultural Economics*, 72(3): 567–73

Bureau, J C and L Mahé (2008). *CAP Reform Beyond 2013: An idea for a longer view*, Notre Europe, Paris

Burrel, A (2000). The World Trade Organization and EU agricultural policy, in Burrell, A and A Oskam (eds), *Agricultural Policy and Enlargement of the European Union*, Wageningen University Press, 91–110

de Gorter, H, Nielson, D J and G C Rausser (1992). Productive and Predatory Public Policies: Research Expenditures and Producer Subsidies in Agriculture, *American Journal of Agricultural Economics*, 74: 27–37

Dewbre, J and A Mishra (2002). Farm Household Incomes and US Government Program Payments, paper selected for presentation at the 2002 AAEA meetings, Long Beach, California, July 28–31

[14] This issue is becoming more important as agricultural commodity prices are linked more strongly to energy prices then they were in the past—because there is now both a supply (cost) and a demand (bio-energy) link between energy and agricultural commodity prices. This increases the demand for farming technologies which are less energy-intensive or energy-related.

[15] Another important policy issue in this framework is whether biotechnology should be part of such EU policy for the future. If the political objectives of biotechnology use in EU agriculture and the bio-economy remain too strong, the need for the search for and investment in alternative technologies is even stronger.

European Commission (2007). *The Common Agricultural Policy Explained*, European Commission, Directorate–General for Agriculture and Rural Development

Fennell, R (1997). *The Common Agricultural Policy: Continuity and change*, Clarendon Press Oxford

Foster W E and G C Rausser (1993). Price-distorting Compensation Serving the Consumer and Taxpayer Interest, *Public Choice*, 80: 173–89

Glauben, T, Tietje, H and C Weiss (2006). Agriculture on the Move: Exploring regional differences in farm exit rates in Western Germany, *Review of Regional Research*, 26: 103–18

Goodwin,B and A Mishra (2006). Are Decoupled Farm Program Payments Really Decoupled? An Empirical Evaluation, *American Journal of Agricultural Economics*, 88(1): 73–89

Harvey, D R (2009). Agriculture and Multifunctionality, paper to GEM, Science Po conference: 2020 European Agriculture: challenges and policies, Paris, January

Hennessey initial (1998). The Production Effects of Agricultural Income Support Policies under Uncertainty, *American Journal of Agricultural Economics*, 80: 46–57

Josling, T and S Tangermann (1999). Implementation of the WTO Agreement on Agriculture and developments for the next round of negotiations, *European Review of Agricultural Economics*, Vol. 26(3): 371–88

Olper, A (2008).Constraints and Causes of the 2003 EU Agricultural Policy Reforms, in Swinnen, J F M (ed.) *The Perfect storm: The political economy of the Fischler reforms of the common agricultural policy*, CEPS publications, Brussels

Rausser, G C (1992). Predatory versus Productive Government: The Case of US Agricultural Policies, *Journal of Economic Perspectives*, 6:133–58

Rosenblatt, J (1988). *The Common Agricultural Policy of the European Community*, International Monetary Fund, Washington DC

Salhofer, K and E Schmid (2004). Distributive Leakages of Agricultural Support: Some empirical evidence, *Agricultural Economics*, 30 (1): 51–62

Sckokai, P and D Moro, Modeling the Reforms of the Common Agricultural Policy for Arable Crops under Uncertainty, *American Journal of Agricultural Economics*, 88(1): 43–56

Serra, T, Goodwin, B K and A M Featherstone (2005). Agricultural Policy Reform and Off-farm Labour Decisions, *Journal of Agricultural Economics*, 56(2): 271–85

Stead, D R (2007). The Mansholt Plan Forty Years On, *EuroChoices*, Vol. 6(3): 40–45

Swinbank, A (1999). CAP reform and the WTO: compatibility and developments, *European Review of Agricultural Economics*, Vol. 26(3): 389–407

Swinnen, J F M (2002). Transition and Integration in Europe: Implications for agricultural and food markets, policy and trade agreements, *World Economy*, 25(4): 481–501

Swinnen, J F M and H de Gorter (2002). On Government Credibility, Compensation, and Under-investment in Public Research, *European Review of Agricultural Economics*, 29(4): 501–22

Swinnen, J F M (2008) (ed.). *The Perfect Storm: The political economy of the Fischler reforms of the common agricultural policy*, CEPS publications, Brussels

Swinnen, J F M and L Vranken (eds.) (2008), *Land & EU Accession: Review of the transitional restrictions by new member states on the acquisition of agricultural real estate,* CEPS Publications, Brussels

Swinnen, J F M (2009a). The Growth of Agricultural Protection in Europe, In Kym Anderson (ed.), *The Political Economy of Global Agricultural Distortions*, World Bank Publications, Washington D C

Swinnen, J F M (2009b). The Growth of Agricultural Protection in Europe in the 19th and 20th Centuries, *The World Economy*, 32(11):

Tweeten, L. (1979). *Foundations of Farm Policy*, 2nd ed. Lincoln NE, University of Nebraska Press

Van Meijl, H and F Van Tongeren (2002). The Agenda 2000 CAP Reform, World Prices and GATT–WTO Export Constraints, *European Review of Agricultural Economics*, 29(4): 445–70