Preliminary – comments most welcome

Redefining 'Foreign': an application to FDI spillovers

By

Lisbeth la Cour, Sara L. McGaughey and Pascalis Raimondos

May 16, 2017

Abstract: How should we define a firm as 'foreign' and/or as 'domestic' in our complex globalising world? Using a multi-country firm-level panel dataset we distinguish between foreign and domestic firms by using both the traditional IMF 10% *direct* foreign ownership definition and the control-based 50% *ultimate* foreign ownership definition. By comparing the two methods we find a large set of firms that are 'foreign' according to the ultimate ownership definition but 'domestic' according to the direct ownership definition. These firms turn out to be the most productive of all firms in our data. The implications of such a mis-classification is exemplified in the FDI spillover literature. When running FDI spillover regressions we find (consistent with the rest of the literature) no horizontal spillover effect when using the IMF definition of what is 'foreign', but positive and significant effects when using the ultimate ownership definition.

JEL Codes: F23, D24.
Keywords: Foreign direct investment, ultimate owner, spillovers, total factor productivity.

Authors Affiliations: LaCour: Copenhagen Business School. McGaughey: Griffith University. Raimondos: Queensland University of Technology, CEPR, and CESifo.

Corresponding Author: Pascalis Raimondos, School of Economics and Finance, QUT, 2 George st., Gardens Point, 4000 Brisbane, Australia (pascalis.raimondos@qut.edu.au)

Acknowledgments: We would like to acknowledge Federico Clementi for his comments in the initial phase of this paper. We also thank seminar participants at Chuo and Hitotsubashi universities. Natalie Kessler, Jacek Przybyszewski, Emilie Helberg Pedersen, and Astrid Vibe-Pedersen have helped us through the years to create the data used in this paper. The research was supported by a grant from the Danish Research Council (DFF 4003-00004).

1. Introduction

The recent United Nations World Investment Report (UNCTAD, 2016) makes a significant contribution in documenting the complexity of MNE ownership structures, including how popular indirect ownerships have become in our globalising world. Large MNEs increasingly utilize detailed and complicated ownership structures that, in part, seek to hide direct ownership patterns for tax and financial reasons. Complexity in MNE structures is further driven by the increasing growth and fragmentation of production that results in MNEs constantly reconfiguring their international supply chains, and by modalities of growth such as mergers and acquisitions, joint ventures and alliances between firms. The ownership structure of some MNEs is thus characterised by considerable vertical depth – that is, multiple steps from the ultimate owner to affiliate, often across multiple borders.

The current IMF (2009) definition of a foreign direct investment enterprise involves a single foreign investor *directly* owning at least ten per cent (10%) of shares in a company, with the purpose of gaining an effective voice in the management of the company. Whether the investor is 'foreign' is determined simply in terms of residency address. A seemingly domestic firm under this definition may conceivably be controlled by a foreign entity through series of ownership linkages, with no direct ownership of the local affiliate even at the 10% level (UNCTAD 2016). Who is the ultimate owner can, however, be non-obvious resulting in many empirical studies relying on direct ownership only. The growing complexity and depth in MNE ownership structures nonetheless raises important questions for research and policy alike. How prevalent really are 'foreign' firms in local economies? How relevant is the traditional 10% threshold single foreign owner IMF definition when studying the presence and consequences of foreign direct investment? Is there a better way to categorise foreign and domestic firms in empirical studies of FDI, and how does this affect existing understandings of MNE activity?

To investigate these questions, we devised a two-part study. In part A, we develop a categorisation of firms that distinguishes between foreign and domestic firms, and identifies their prevalence and characteristics. We then apply this in Part B to the literature of FDI-induced spillovers. Our categorisation aims to more fully recognise the ultimate owner and thus indirect ownership links, which the latest World Investment Report (UNCTAD 2016) emphasized as especially relevant in our globalizing world. Specifically, we define 'foreign firms' using both the IMF/OECD threshold of 10% direct ownership by a single foreign entity (which we call $FDI10$) and a 50% threshold (named $FDI50$) whereby 50% or more of the firm is owned directly or

indirectly by a foreign entity, i.e. an *ultimate* ownership definition.

Our use of 'ultimate ownership' is driven by an interest in who has the ultimate *control* over decision making. This clearly contrasts with the emphasis on *influence,* which is part of the $FDI10$ definition. We believe that technology and knowledge transfers from a parent to an affiliate are affected by whether the affiliate is controlled or not. With control comes a greater willingness by the MNE to transfer knowledge and technology to affiliates. If control matters in this way, then the possible spillover effect from foreign affiliates to domestic firms will be affected by what definition we use to categorize a firm as foreign.

Focusing on ultimate ownership also helps us to identify a 'domestic multinational' set of firms. Domestic MNEs (e.g. Phillips in Holland) operate as any other MNE in global markets and are thus able to secure productivity enhancements through internal mechanisms (e.g. within-the-firm labour and technology markets). Clearly, for many purposes, classifying these firms as 'domestic' is not appropriate. For example, in trying to measure the productivity spillover that foreign firms may exert to domestic firms, domestic MNEs should be excluded from the set of 'domestic' firms. Identifying these firms requires knowledge of whether these firms control firms in other countries and thus knowledge of ultimate ownership structures. Such knowledge can only be found by working with a multi-country firm-level data.

In the present paper we use the Amadeus dataset of all European firms and their time-variant ownership pattern for the $2001-2008$ period. An expectation could be that the IMF definition with a low 10% cut-off will pick up more foreign firms than the high 50% definition. However, we find the opposite (for a quick peek, see our "egg" figure 2 below): there are double as many firms that are ultimately controlled ($FDI50$) than what the IMF definition captures ($FDI10$). Moreover, these $FDI50$ firms turn out to be on average larger (employ more capital, labour, and materials) and more productive than $FDI10$ foreign firms. In particular, within this set of $FDI50$ firms there is substantial subset of firms that is not captured by the $FDI10$ set of firms because they are controlled by only indirect ownership links. These indirect $FDI50$ ($I-FDI50$) firms are found to be the most productive of all. Further, the domestic MNEs ($MNE50$) are almost just as many as other foreign firms, but far more productive than the $FDI10$ set of firms.

In Part B, we build on these insights and explore the implications of paying careful attention to what is 'foreign' and what is 'domestic', through an application to FDI-induced productivity spillovers. Productivity spillovers are defined as informal, involuntary, non-market transfers that occur when the activities of one firm affect the productivity of another firm in ways that are

not fully captured by the source firm (Eden, 2009). Worldwide, the trend in investment policy continues to be towards greater liberalization and promotion of FDI (UNCTAD 2016), in part to capture such spillovers. Measuring productivity spillover effects is, however, not an easy task. While the extensive literature on horizontal FDI spillovers is inconclusive, the general pattern of results shows that the presence of FDI seems more often than not to have no statistically significant productivity effects on domestic firms in the same (horizontal) industry – see, among others, Javorcik (2004). Further, empirical studies in this field have most commonly relied on an FDI10 definition, ignoring indirect and ultimate ownership. It is thus an area where our approach may make a significant contribution.

Common to virtually all papers in the FDI spillover literature, a firm is defined as 'foreign' using the IMF definition (i.e. the 10% direct ownership that the $FDI10$ set uses). This literature also defines a firm as 'domestic' if it is not 'foreign'.[1] Our conjecture, based on the findings of Part A, is that inclusion of the $I-FDI50$ firms in the domestic firm dataset upward biases the productivity of the domestic firms and downward bias the productivity of foreign firms. Similarly, including the $MNE50$ firms in the domestic firm dataset also upward biases the productivity of the domestic firms. Thus, previous studies have perhaps stacked the cards against finding positive spillover effects from the presence of foreign firms!

Our empirical strategy in running FDI productivity spillover regressions compares results from different firm sets. Appropriately categorising firms into the foreign and domestic groups will be our main contribution to the literature. In doing that, we also pay due attention on how to estimate total factor productivity. We adopt the ACF control function approach developed in Ackerberg et al. (2006, 2015) and applied in De Loecker and Warzynski (2012) and De Loecker et.al (2016). This approach is careful in dealing with the endogeneity of inputs problems that exist when calculating the residual of the output minus inputs component of productivity.

The results support our conjectures. Running FDI spillover regressions using the $FDI10$ definition we find positive effects that weaken and eventually disappear as more control variables are added. This is consistent with prior studies. In contrast, when we run regressions using the $FDI50$ definition of foreign firms, we find positive and robust spillover effects. These effects remain when we consider only the $I-FDI50$ firms. Overall it seems that considering a more modern approach to ownership in the definition of what is 'foreign' and what is 'domestic' has a

[1] Some papers have correctly removed the domestic MNEs from the set of domestic firms when estimating horizontal FDI spillovers; see Bekes et al. (2009).

significant impact in identifying positive spillover effects within the same (horizontal) industry. Our findings hold significant policy implications for the attractiveness of FDI by suggesting that more horizontal productivity spillovers from FDI may occur than previously thought.

The remainder of our paper is organised as follows. In section 2, we present Part A of our study. Specifically, we elaborate our concept of 'foreign', our data and provide summary statistics. In section 3, we present Part B. We review the existing literature and challenges related to the measurement of intra-industry FDI-induced productivity spillovers. We then present our empirical strategy and methods. We present our results in section 4. Section 5 deals with several robustness checks as (i) including the Herfindahl index of concentration, (ii) separating western european and eastern european countries, and (iii) alternative methods for estimating TFP. We conclude in section 6 with a discussion of the implications of our findings for policy and future directions in research on FDI-induced productivity spillovers, and speculate on other applications where our concept of 'foreign' can add value.

2. Part A: our concept of 'foreign' and the data

The historical development of what is 'foreign direct investment/enterprise' provides important context for our categorisation of firms (Part A of our study) and the evolution of the FDI-induced spillover literature (related to Part B). The IMF provided one of the earliest and most enduring attempts at proposing and refining such definitions in the context of the post war era in its Balance of Payments Manual. In particular, an emphasis on control was explicit in definitions provided in the early editions of the Manual (BPM1 1948, BPM2 1950). For example, the very first edition (IMF 1948, p. 47) defined foreign direct investment as comprising: (a) an enterprise in country Y which is a branch of an enterprise in country X; or (b) an enterprise in country Y that is a subsidiary of an enterprise in X – i.e. it is incorporated in Y but effectively controlled by residents in X – where control is inferred if 50% or more of voting stock is controlled by residents of X, or 25% or more of voting stock is concentrated in the hands of a single holder or organised group of holders in X, or a resident of X has a controlling voice in its policies; or (c) commercial real estate in Y owned by residents of X. The first edition even hinted at more complex ownership structures: "A direct investment may be owned by two or more countries jointly; similarly, a direct investment in Y may be owned by an enterprise in X which itself us a direct investment of an enterprise in Z (or even Y) (IMF 1948, p. 47). This definition remained in the second edition of 1950.

Elaborating on the notion of foreign direct investment, the third edition of the IMF Balance

of Payments Manual (BPM3) (1961) defines [foreign] direct investment as "investment made to create or expand some kind of permanent interest in an enterprise: it implies a degree of control [emphasis added] over its management. [...] It is characteristic of direct investment that the investor possesses managerial control over the enterprise in which the investment is made and he [sic] also makes available to it his technical knowledge (know-how)" (IMF 1961, p. 118). Direct investment continued to be distinguished from portfolio investment, where the investor "has no intention of playing a major role in the direction of policies of the enterprise." There emerged, however, considerable definitional ambiguity. The "exercise of an important voice" was used interchangeable with "direct control" (p. 120). Further, the third edition stated that it was not "desirable to give a rigid definition of the concept of the direct investment enterprise" and that "specific percentages suggested for determining whether a given enterprise is to be classified as a direct investment enterprise should be regarded as no more than rules of thumb" (p. 119). By the fourth edition, the foreign direct investor's purpose was to "have an effective *voice* [emphasis added] in the management of the enterprise" (IMF 1977, p. 128, 136).

The fourth edition included a survey of member country concepts and practices concerning direct investment flows, undertaken by IMF staff. Diverse practices among countries showed accepted evidence of FDI to range from 25 to 10 per cent foreign ownership, with a tendency to the low side (IMF 1977, p. 137). The survey also explicitly asked about indirect ownership whereby a foreign investor could exert an 'indirect voice' in the resident enterprise (p. 189). Indirect investment was not commonly considered by respondents at the time, with the direct ownership link typically being the only link registered in a country's national statistics. Nonetheless, the subsequent fifth edition (IMF 1993) for the first time defined a direct investment enterprise as one in which a direct investor, who is resident in another economy, owns 10% or more of the ordinary shares or voting power (or equivalent). It also made explicit that a direct investment enterprise is either directly or indirectly owned by the direct investor (IMF 1993, p. 86). This definition has been retained in the sixth and latest edition of the Manual (IMF 2009, p. 101), which was conducted in parallel with the OECD Benchmark Definition of Foreign Direct Investment and the System of National Accounts to maintain and enhance consistency between the three important standards.

Two aspects of the evolution in these definitions of FDI stand out. First, whereas the initial emphasis was on effective control with somewhat higher percentages of foreign ownership required to signify foreign direct investment, a shift towards influence or an important voice

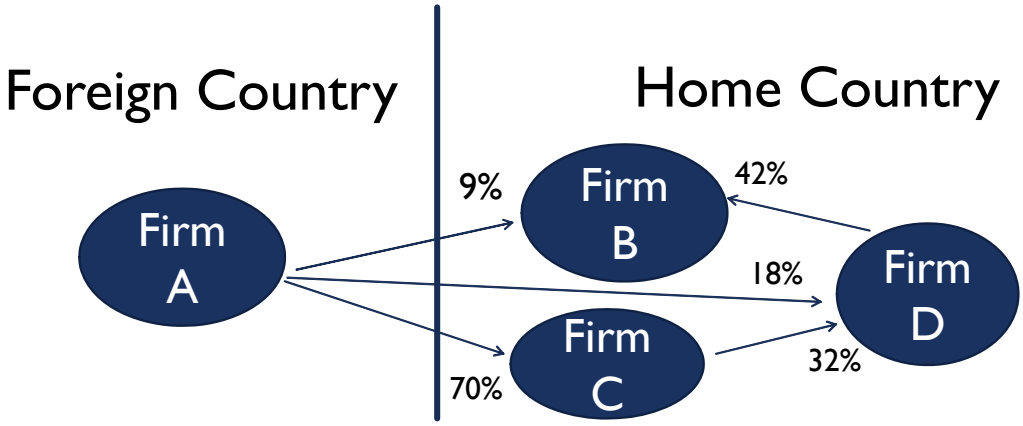
was evident from at least BPM3 in 1961. Related, a much lower threshold for ownership was reported in country practices in BPM4 (IMF 1977), with the minimum threshold of ownership being reduced to ten per cent (10%) in the BPM5 (IMF 1993) definitions. Second, in contrast to the early emphasis on direct ownership links, indirect ownership by a foreign direct investor was explicitly included in the definition of a direct investment enterprise as recently as BPM5.

On what aspects of the IMF definition empirical researchers will focus is, of course, dependent on the question at hand. For example, in the profit-shifting literature, a foreign affiliate is empirically identified by whether there exists an owner that controls 50% of the firm's shares; see among others Huizinga and Laeven (2006) and Dharmapala and Riedel (2013). Such a control may not only be exercised through direct ownership links but also through indirect ownership links. By combining the direct and the indirect ownership links the concept of ultimate ownership (UO) arises. Such a concept is directly linked to the independence of a firm, i.e. if the firm is independent it will have no ultimate owner, and vice versa. In contrast, virtually all published studies of FDI-induce spillovers consider only direct ownership links, and most commonly use 10% foreign ownership as the minimum threshold when identifying FDI.

Here is where a major contribution of the present paper lies. Instead of using only the minimum 10% influence-based definition of what is 'foreign', we use the minimum 50% control-based definition of what is 'foreign'. This empirical approach is made possible by virtue of improved data and its availability. However, we ground our approach in theory. We believe that technology and knowledge transfers from a parent to an affiliate are affected by whether the affiliate is controlled or not. With control comes a greater willingness by the MNE to transfer knowledge and technology to affiliates. Knowledge is widely recognised as a source of competitive advantage. Indeed, Teece (1998) argues that the essence of a firm is its ability to create, transfer, assemble, integrate and exploit knowledge. Knowledge is not, however, naturally scarce in the way that physical resources are. The multinational enterprise internalises markets in knowledge through foreign direct investment. Ultimate ownership provides a security that the residual income that the firm generates belongs to the ultimate owner and it can be transferred back to the ultimate owner. Such a security creates a level of trust that allows parent and affiliate firms to exchange knowledge and technology at a much higher level than any couple of firms with no controlled relationship. If control matters in this way, then the possible spillover effect from foreign affiliates to domestic firms will be affected by what definition we use to categorize a firm as foreign.

To visualize the difference between direct and ultimate ownership we draw the below example.

Figure 1: direct vs ultimate ownership



In the above figure we depict two countries, 'home' and 'foreign'. In the home country firms B, C, D are connected through ownership links. In the foreign country, Firm A has some direct ownership of all three firms in the home country. It is easy to see that Firm A controls directly Firm C by owning more than 50% of its shares. Moreover Firm A owns 18% of Firm D. Thus, firms C and D are categorised as foreign using the $FDI10$ definiton of what is foreign. Firm B will be categorised as domestic because the direct ownership links used in the $FDI10$ definition show a domestic owner (the legal address of Firm D is domestic). However, using the ultimate ownership definition of what is foreign gives a different picture. All three firms B, C, and D are controlled by Firm A by direct and indirect ownership links. Firm A is the ultimate owner of all domestically operating firms. Knowing the complete (direct and indirect) ownership tree of a firm will also help us identify whether a domestic firm is the ultimate owner of firms in other countries – that is, a domestic MNE (named here $MNE50$).

Our database is the Amadeus/ORBIS dataset owned by Bureau Van Dijk.[2] Amadeus is the European subset of the ORBIS database that the 2016 World Investment Report uses. We focus on the European dataset as it offers us the longest firm-level panel dataset within ORBIS. We use both the older Amadeus DVDs and the online ORBIS versions to supplement each other.

[2] We work with the large dataset where all firms with 5 or more employees are included. For a detail account of ORBIS see Kalemli-Ozcan et al. (2015).

Being careful of how we categorise a firm as domestic or foreign, we acquire DVDs with single releases of the data for the 2003 to 2010 period. We are thereby able to track the changes that have happen in firms' ownership structure. This allows us to create a consistent unbalanced firm-level panel dataset for approximately 2.5 million manufacturing firms between $2001 - 2008$ with full ownership and financial data. Appendix 1 describes the details of how we cleaned and prepared the dataset.

The invaluable advantage of the Amadeus dataset is that it provides the ultimate ownership (UO) variable that we need here. Bureau van Dijk has carefully collected this information and built it in their dataset. Ownership of an affiliate does not always reflect control. Shareholdings in affiliates provide the rights to not only dividends but also voting rights. Control requires the ability to affect strategic decisions through the exercise of voting rights (WIR 2016) and thus requires one to distinguish between voting and non-voting shares when considering ownership. The Amadeus database tracks control rather than merely ownership. Hence, when share categories are split into voting and non-voting, the ownership percentages recorded are those linked to the category of voting shares. From the 3 levels of ultimate ownership thresholds reported in Amadeus ($25\%, 50\%$, and 75%) we pick the 50% that secures control.

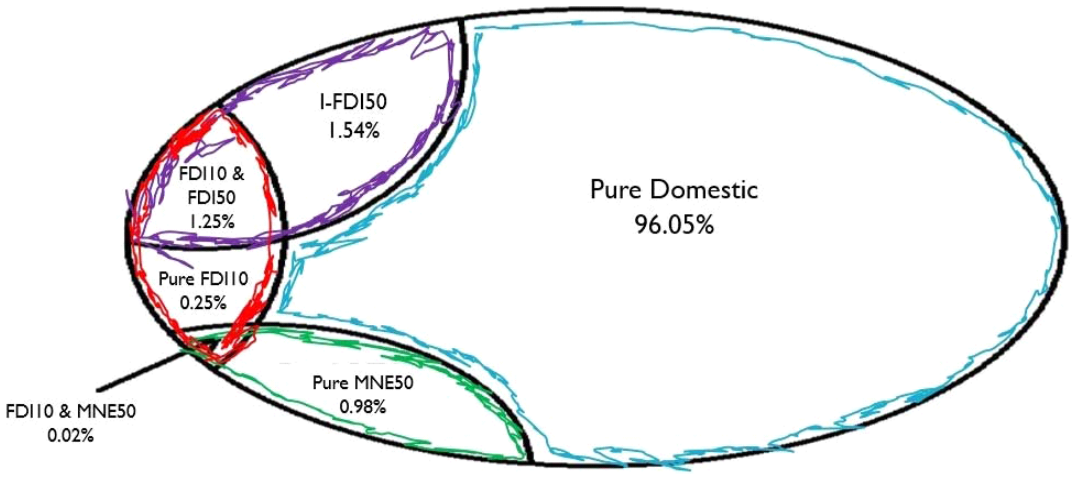
The exact definitions of the different firm sets that we use are as follows:

- $FDI10$: firms where a single foreign owner directly owns at least 10% of shares.
- $FDI50$: firms where a single foreign owner ultimately owns at least 50% of shares.
- $I - FDI50$: firms that are $FDI50$ but not $FDI10$.
- $MNE50$: firms that are not $FDI50$ and which ultimately own subsidiaries in another country.
- Pure domestic firms: firms that are neither $FDI10$ nor $FDI50$ nor $MNE50$.

Figure 2 below — the "egg" — illustrates the distribution of the different types of firms. This is based on a total of $2,344,488$ observations, which corresponds to roughly $600,000$ firms.[3]

[3]Note: the percentages in the figure are calculated based on data after cleaning and trimming but before TFP estimations have been performed. We have chosen to illustrate the spilt of the data according to observations and not firms as some firms change ownership status during the sample period. The focus on observations and not on firms avoids 'double-counting'.

Figure 2: Illustration of ownership data



Note: Own calculations using $2001 - 2008$ firm level data from AMADEUS.

As seen in Figure 2, the large majority of the observations are purely domestic firms (96.05%) – the set outlined in blue. While $FDI50$ observations (the purple set) make up about 3% of the data, $MNE50$ observations (the green set) are around 1.0% and the $FDI10$ observations (the red set) around 1.5%. An observation cannot be $FDI50$ and $MNE50$ at the same time. When focusing on the standard definition of 'foreign' ($FDI10$) we see a large overlap with our $FDI50$ definition. The overlap with $MNE50$ is negligible.

The activity data among the different sets of firms reveals an interesting pattern. The descriptive statistics are seen in Table 1 below.[4]

[4]For the categorization of the number of firms we have consistently classified a firm to a category based on the last year's information about ownership. This has been done to avoid double counting of firms that change ownership status during the sample period. Labour productivity is defined as sales over number of employees from the firm-level data and not as the ratio of columns 3 and 4.

Table 1: Activity data summary statistics

	Obs.	**Firms**	**Sales** (1000 USD)	**Labour**	**Capital** (1000 USD)	**Material** (1000 USD)	**Labour Productivity**
Total	2,343,495	575,844	9,303	49	1,653	5,183	140
FDI10	35,742 (1.5%)	13,007 (2.3%)	82,105	283	13,314	48,970	319
FDI50	65,475 (2.8%)	21,146 (3.7%)	103,350	340	16,757	61,785	366
I-FDI50	36,149 (1.54%)	6,014 (1.04%)	118,865	381	19,134	70,872	398
MNE50	20,787 (0.89%)	4,937 (0.86%)	208,961	564	30,813	118,219	342
Pure domestic	2,250,817 (96.05%)	555,033 (96.43%)	4,544	36	918	2,389	131

Note: own calculations using the 2003-2008 Amadeus database.

Table 1 shows that the purely domestic firms are on average considerably smaller and less productive than foreign firms, irrespective of how we define them. It also shows that the $FDI10$ firms seem to be smaller than other foreign firms. In particular, the $FDI50$ and the I-$FDI50$ firms are even larger and more productive (in terms of labour productivity; TFP will be derived below). The domestic multinationals (that is, the approx. 3,000 European MNEs' HQ) are by far the biggest firms in terms of activity data.[5]

3. PART B: AN APPLICATION TO THE FDI SPILLOVER LITERATURE

FDI-induced productivity spillovers take place when local firms learn about new technologies, marketing or management techniques by observing foreign affiliates (i.e. demonstration effects) or by hiring workers trained by foreign affiliates (i.e. labour market impacts), and in this way improve their performance. The 'fresh winds of competition' may also lead local firms to improve their efficiency and reduce their costs. However, competition can also reduce the scale of operation of the host firms and lead to negative productivity effects. With the overall spillover effect being theoretically ambiguous, numerous empirical studies have attempted to find and explain FDI-induced productivity effects.

However, measuring productivity spillover effects is not an easy task. The general pattern

[5] In contrast MNE50 seem to be less labour productive than FDI50 firms. The same result holds when calculating total factor productivity for the different firm categories. One should note, however, that for especially MNE HQs there will be an issue of profit shifting, i.e. not reporting the appropriate revenues to the HQ country where taxes are usually higher than taxes in small European countries. In a different but related project we focus on correcting productivity estimations by taking into account the extent of profit shifting a MNE will have.

of results shows that the presence of FDI seems more often than not to have no statistically significant productivity effects on domestic firms in the same (horizontal) industry. In some cases negative effects have been found in horizontal industries. For example, Aitken & Harrison (1999) and Javorcik (2004) showed that FDI has at worst a negative (and at best a zero) effect on the productivity of domestic firms within the same industry. Positive effects have been found in upstream industries and, as such, reflect supplier linkage effects rather than intra-industry technology transfer and learning effects. An extensive literature since then confirms the absence of positive effects within industries and the presence of positive effects between industries — see among others Görg and Strobl (2001), Görg and Strobl (2005), Görg and Greenaway (2004), Altomonte and Pennings (2009) and Javorcik & Spatareanu (2008).

The definitions used for foreign direct investment in the FDI-induced spillover literature have been very variable, with seemingly no common standard.[6] For example, using data drawn from Venezuela's National Statistical Bureau, Aitken and Harrison (1999) were able to distinguish between firms with less than 20% direct foreign ownership, with 20 to 49.9%, and 50% or more. In contrast, using Romanian data extracted from ORBIS, Altomonte and Pennings (2009, p. 1133) considered a firm foreign if more than 10% of its capital belongs to an MNE, and domestic otherwise. Temouri, Driffield and Higón (2008) uses ultimate owner variable from Amadeus to identify firm nationality in Germany, but then define a foreign firm using the IMF 10% minimum threshold. In contrast, Castellani and Zanfei's (2003) sample of 3932 firms across France, Italy and Spain distinguishes foreign-owned firms (using a foreign ultimate owner definition) and domestic firms, but does not distinguish between different sub-sets. In short, very few studies discuss or even hint at the existence or importance of indirect ultimate ownership, to which our study (Part A) points, or distinguish between various types of 'foreignness'.

As mentioned in the introduction, our contribution to this literature is in redefining 'foreign'. Following Javorcik (2004) we run FDI productivity spillover regressions using a domestic firm-level measure of TFP and a measure to indicate the degree of 'foreign' penetration in a market. The latter, which is defined below, is our explanatory variable in the FDI spillover regressions:

$$HP_{jct} = \frac{\sum_{i=1,\text{ in } jct}^{N} SALES_{it} * FDI_{it}}{\sum_{i=1,\text{ in } jct}^{N} SALES_{it}}$$

[6] Indeed, we found a number of FDI-induced spillover studies where what constitutes FDI is not even remarked upon.

Horizontal penetration/presence (HP) is defined as the share of sales of foreign firms in a given 3-digit industry j within a given country c and for a given year t. In this sense, a market is defined as an industry-country-year combination and for each of these combinations we derive an HP value.[7] The FDI_{it} indicator in the above formula is a binary variable that takes the value of 1 if the firm is foreign and 0 if the firm is domestic. Clearly, how we categorize firms will matter for the nominator of the above formula; the denominator will not be affected as this is the total sales in that particular industry-country-year combination. Our HP measure will be affected by whether we define firms to be foreign using the $FDI10$ definition, the $FDI50$ definition, or the $I - FDI50$ definition.

Table 2 below reports the distribution of the HP variable for each of the above definitions of what is a foreign firm.

Table 2: Distribution of HP for each definition of foreign

	Mean	**P10**	**P50**	**P90**	**SD**
HP_{FDI10}	0.0765	0	0.0338	0.1950	0.1078
HP_{FDI50}	0.1560	0.0061	0.0911	0.3924	0.1708
$HP_{I-FDI50}$	0.0924	0	0.0449	0.2480	0.1206

Note: own calculations using our 2003-2008 Amadeus database.

Both the mean and the median of the HP variable differs significantly depending upon the definition of 'foreign' used in the calculations. For example, while foreign presence is 7.6% under the $FDI10$ definition of 'foreign', it is 15.6% under the $FDI50$ definition. A focus on the median may be justified as the distributions of the HP measures are quite skewed.

Moving on to how we derive our TFP measure, we estimate a production function (a revenue-based Cobb Douglas production function in our case) using the procedure suggested by Ackerberg et al. (2006, 2015) and applied in De Loecker (2011), De Loecker and Warzynski (2012), and De Loecker et al. (2016). This procedure (henceforth referred to as the modified ACF procedure) is an extension of the procedures suggested by Olley and Pakes (1996) and Levinshon and Petrin (2003). While all three procedures are able to handle the potential endogeneity of the input variables, the ACF procedure is able to address collinearity problems present in OP

[7] The use of sales is sometimes in the literature substituted by employment levels. We have used both measures and found a correlation of 0.94 between an HP-sales and an HP-employment index. In what follows we use the HP-sales index.

and LP procedures. De Loecker (2011) augments the ACF procedure by allowing more variables (than just lagged productivity) appearing in the productivity law of motion equation.[8]

We estimate total factor productivity at the firm level for each of the years in our sample period (as the estimation method uses lagged values, we need the data for the first year of the sample to initialize the process). We perform TFP estimations for each NACE 2-digit manufacturing industry for each country. The main equation of the procedure is the production function equation, logarithmically transformed:

$$lnY_{it} = \beta_0 + \beta_1 lnL_{it} + \beta_2 lnK_{it} + \beta_3 lnM_{it} + \omega_{it} + \epsilon_{it} \quad (1)$$

where Y is sales or revenues, L is labour input (number of employees), K is capital input (in value terms), M is material input (in value terms),[9] ω is the unobserved (to the econometrician but not the firm manager) productivity and ϵ is the error term (unobserved to both the firm manager and the econometrician). The modified ACF procedure is stepwise and rests on a set of assumptions: (i) the production function has a scalar Hicks neutral productivity term, (ii) the coefficients of (1) are the same across all firms in each sub sample (country-industry combinations); (iii) input prices are autocorrelated. All input coefficients are estimated in the second step. The first step is needed to isolate productivity ω from the unobserved error term ϵ. To elaborate, the stepwise modified ACF procedure is as follows:

1. Assuming that labor input is decided before material input and that productivity evolves according to a Markov(1) process, it is possible by OLS to estimate a parametric approximation to the sum of the first five terms in (1) and therefore obtain estimates for the error terms, ϵ_{it}. From the estimates of the first five terms in (1) it is possible to obtain a first estimate of productivity by subtracting the estimated terms for L, M and K. To do that, the unobserved productivity term is replaced by an inverted function of the demand for materials (first used by Levinsohn and Petrin, 2003). This function will depend on labor and capital, their cross products and interactions, and also the contemporaneous value of HP.

[8] Admittedly, our data can only allow us to estimate revenue-based measures of productivity and not physical productivity. The fact that we are working with a multi-country firm-level panel dataset makes it hard in finding price data for each of the 20+ countries that we consider. De Loecker (2011) proposes an approach that can somewhat address this issue. However, the use of a multi-country dataset restricts the applicability of his approach. Gadhi et.al (2015) propose another remedy and we are in the process of implementing their code to our data.

[9] See Appendix 1 that describes how we prepare our data step by step.

2. We then specify the productivity law-of-motion function, i.e. a function that determines how productivity evolves as a function of lagged productivity and other lagged explanatory factors. We adopt a version of the law-of-motion that adds the lagged HP measure to the regressors, $\omega_{it} = g_t(\omega_{it-1}, HP_{jt-1})$. The intuition for doing that is that we believe managers know how much foreign horizontal penetration there is and thus they take this into account when employing labour and capital inputs. This law-of-motion allows us to derive changes in productivity that not even the management of the company can predict. These changes are by nature uncorrelated with input variables from the previous period and also with capital from the same period (due to the assumption about the order in which the decisions are made). Therefore the productivity innovations and the instruments (lagged input variables and contemporaneous capital) form the moment conditions on which the GMM estimation rests. The final coefficient estimates of lnL, lnK and lnM from (1) are then derived by GMM.

3. Retrieve the TFP measures (ω_{it}) as the 'known' part of the error term from (1) for industry-country combinations where the betas of lnL, lnK and lnM are non-zero.

All the above steps are written in the code that we use in our econometric analysis (see Appendix 2; only for referees).

Having explained how we derive our TFP measure for domestic firms and the horizontal penetration measure of foreign activity, we now present the preferred regression model:

$$TFP_{ijct} = \alpha_i + \beta_1 HP_{jct} + \beta_2 HP_{jct-1} + \beta_3 D_t + \beta_4 D_t x D_j + \beta_5 D_t x D_c + \varepsilon_{ijct} \qquad (2)$$

where i refers to a firm, j refers to a 2-digit industry, c refers to a country, and t refers to a year. In order to remove any influence from time invariant firm specific variables we estimate equation (2) using firm-fixed effects.[10]

We use both the contemporaneous and the lagged values of the HP variable as our explanatory factors of main interest. We do this recognizing that spillover effects may take time. The current specification of equation (1) that includes the lagged HP measure, allows for consistency with the Markov(1) assumption of the ACF-method for estimating TFP. Notice that by

[10] We have also experimented with including Herfindahl indices of competition as controls in our horizontal penetration regressions. As it turns out these indices never became significant, so we dropped them again (see the robustness section).

including both HP and lagged HP, the long-run effect of a change in HP will be the sum of the two beta coefficients $(\beta_1 + \beta_2)$. As HP and HP lagged are often highly correlated (often around 0.90), it may be difficult to obtain statistical significance for the individual coefficients while their combined significance can be tested by means of an $F-$test. In models with both HP and lagged HP we will report the result of such an $F-$test as well.

In the specification of (2) we allow for time fixed effects by using the D_t dummy. With 7 years of data it makes sense to allow for different means in TFP for each year in addition to the HP effects. We also include the interaction dummies for year and industry, and year and country to allow the effects of industry and country to vary over the years. Including the full set of fixed effects constitutes our most robust regressions.

In what follows, we present different estimations of (2) depeding of how we define 'foreign'. As we will see, the coefficients vary significantly. However, as pointed out earlier, the definition of 'foreign' affects both the sample of domestic and foreign firms and the actual estimation of TFP (as HP enters the productivity law-of-motion function). In separating the different effects, we start out by doing the obvious: adopt our preferred definition of 'foreign', viz. $FDI50$, and use it as a basis in all steps. We then do the same using the $FDI10$ definition of 'foreign' and we compare the estimates. Clearly, in doing that, we re-clasify some firms from domestic to foreign even if these firms are not 'foreign'. In a second set of regressions we perform the same comparison but on the set of domestic firms that always stay domestic, i.e. the *pure domestic* firms (see Figure 2) — these are the firms that one should be interested in to see whether there are any spillover effects. This set of pure domestic firms exclude the domestic multinationals (the $MNE50$ set).

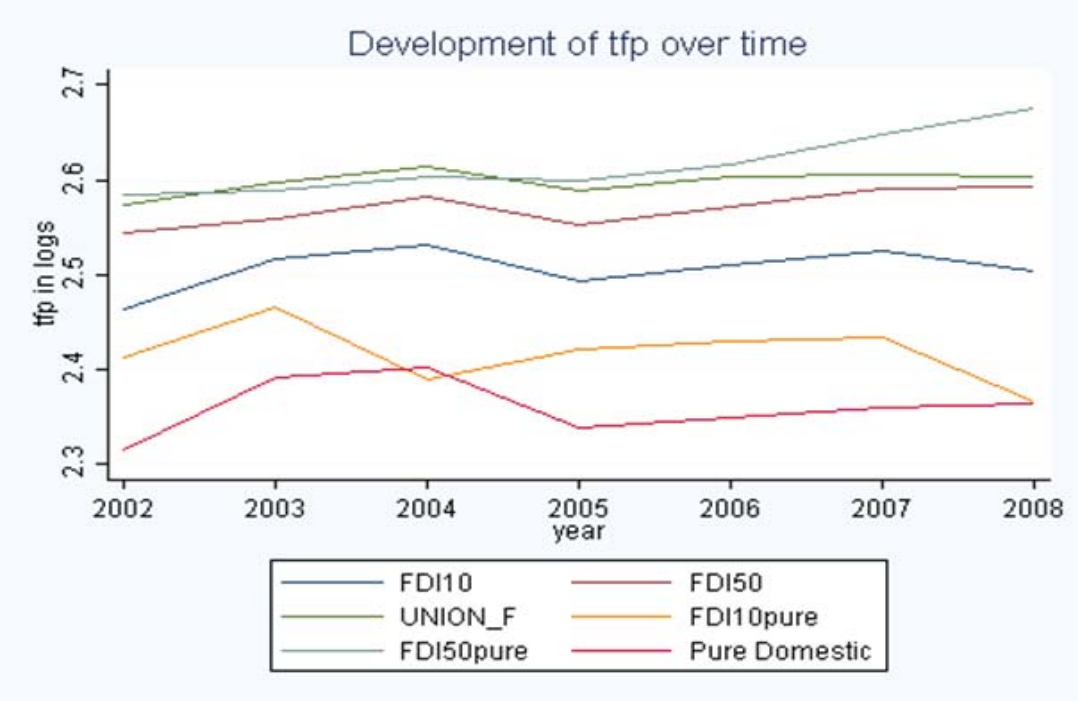
Finally, in a third set of regressions we keep the focus on pure domestic firms but now split the $FDI50$ set foreign firms to the two subsets depicted in Figure 1, viz. whether the firm was also included in the $FDI10$ set of firms definiton or not. The sum of these two new HP measures will add up to our original HP for the $FDI50$ firms. By including each of them in the spillover regression we can isolate the importance of identifying the $I - FDI50$ firms.

The above constititutes our 'benchmark' estimates; robustness checks that address other issues are presented in section XX.

4. RESULTS

Using the above method for estimating total factor productivity we report the evolution of the estimates in the different firm sets during the 2002-2008 period.

Figure 3



As seen, the most productive of all is the $I - FDI50$ set of firms (called $FDI50pure$ in the figure). The $FDI10$ set is clearly a less productive set of firms (the blue line). If from that set we remove the firms that also belong in the $FDI50$ set, then we get the $FDI10pure$ set (the yeloow line) that is just as productive as the pure domestic firms (the red line). Creating the union of all foreign firms ($FDI10$ and $FDI50$), called above $UNION\_F$ (dark green line), lifts the aggregate level. There is no clear trending behavior in any of the series.

Our first set of regressions are presented below in Table 3.[11] As mentioned above, we run spillover regressions with two different definitions of what is 'foreign'; the FDI10 and the FDI50. In doing that we start by defining 'domestic' what is not 'foreign'. Due to this, the number of observations changes between the FDI10 and the FDI50 regressions (it is higher in the FDI10 as

[11] We have also run the regressions in steps having only the contemporaneous and only the lagged HP variables without any major difference. For brevity, we report here only the regressions where both are included at the same time.

that definition categorizes too few firms as foreign). The first 5 columns use the FDI10 definition, while the last 5 columns use the FDI50 definition. Each colum is a different combination of the fixed effects included as controls. We focus in interpretating columns 5 and 10 that included all the fixed effects that are allowed.

Table 3: (around here)

As seen by the joint F-test values, the long run FDI10 coefficients of the spillover effect are not statistically different from zero. This is not the case for the long run FDI50 spillover coefficients that show a positive and statistical significant spillover effect.

Running the same regressions but now on the same set of domestic firms — the pure domestic firms — does not change significantly this result. Table 4 below reports the results.

Table 4: (around here)

Again the FDI10 regression, with all fixed effects included, shows no statistical significant long-run spillover effect to pure domestic firms. In contrast, the FDI50 regression does provide evidence that domestic firms are positively affected by the presence of foreign firms within the same industry.

In trying to highlight the importance of the $I-FDI50$ firms, i.e. the firms that were classified as 'foreign' due to indirect owenrship links and thus were missed by the FDI10 definition, we split the FDI50 set of firms in two subsets and run the same regressions. Table 5 below reports the results.

Table 5: (around here)

5. Robustness checks

(i) regressions including the Herfindhal index.

(ii) regressions splitting the country sample in West and East Europe.

(iii) regressions using a different TFP method (Gandhi et al., 2016).

(iv) .....

6. Conclusions

(to be written)

References

[1] Ackerberg, D.A., K. Caves, and G. Frazer (2006). Structural identiffication of production functions, mimeo.

[2] Ackerberg, D.A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators, *Econometrica* 83, 2411–245

[3] Aitken, B. J. and A.E. Harrison (1999). Do domestic firms benefit from direct foreign investment? Evidence from Venezuela. *The American Economic Review* 89, 605-618.

[4] Altomonte, C., and E. Pennings (2009). Domestic plant productivity and incremental spillovers from foreign direct investment. *Journal of International Business Studies* 40, 1131-1148.

[5] Blalock, G. and P.J. Gertler (2009). How firm capabilities affect who benefits from foreign technology. *Journal of Development Economics* 90, 192-199.

[6] Castellani, D. & Zanfei, A. (2003). Technology gaps, absorptive capacity and the impact of inward investments on productivity of European firms. *Economics of Innovation and New Technology* 12, 555-576.

[7] De Loecker, J. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity, *Econometrica* 79, 1407–1451.

[8] De Loecker, T. (2013). Detecting learning by exporting, *American Economic Journal: Microeconomics* 5, 1–21.

[9] De Loecker, J. (2011). Recovering markups from production data, *International Journal of Industrial Organization* 29, 350-355.

[10] De Loecker, J., P.K. Goldberg, A.K. Khandelwal, and N. Pavcnik (2016). Prices, markups, and trade reform, *Econometrica* 84, 445–510.

[11] De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status, *American Economic Review* 102, 2437-2471.

[12] Dharmapala, D. and N. Riedel (2013). Earnings shocks and tax-motivated income-shifting: Evidence from European multinationals, *Journal of Public Economics* 97, 95–107.

[13] Eden, L. (2009). Letter from the Editor-in-Chief: FDI spillovers and linkages. *Journal of International Business Studies* 40, 1065-1069.

[14] Gandhi, A., S. Navaro, and D. Rivers (2015). On the identification of production functions: how heterogeneous is productivity, mimeo.

[15] Görg, H. and D. Greenaway (2004). Much ado about nothing? Do domestic firms really benefit foreign direct investment? *The World Bank Research Observer* 19, 171-197.

[16] Görg, H. and E. Strobl (2001). Multinational companies and productvity spillovers: a meta-analysis. *The Economic Journal* 111, 723-739.

[17] Görg, H. and E. Strobl (2005). Spillovers from foreign firms through worker mobility: an empirical investigation. *Scandinavian Journal of Economics* 107, 693-709.

[18] Huizinga, H. and L. Laeven (2008). International profit shifting within multinationals: A multi-country perspective, *Journal of Public Economics* 92, 1164-1182.

[19] Javorcik, B.S. (2004). Does foreign direct investment increase the productivity of domestic firms? in search of spillovers through backward linkages, *American Economic Review* 94, 605–627.

[20] Javorcik, B.S. (2008). Can survey evidence shed light on spillovers from foreign direct investment?, *World Bank Research Observer* 23, 139–159.

[21] Javorcik, B.S. and M. Spatareanu (2011) Does it matter where you come from? vertical spillovers from foreign direct investment and the origin of investors, *Journal of Development Economics* 96, 126–138.

[22] International Monetary Fund (1948). Balance of Payments Manual, first edition (BPM1), Washington D.C., USA.

[23] International Monetary Fund (1950). Balance of Payments Manual, second edition (BPM2), Washington D.C., USA.

[24] International Monetary Fund (1961). Balance of Payments Manual, third edition (BPM3), Washington D.C., USA.

[25] International Monetary Fund (1977). Balance of Payments Manual, fourth edition (BPM4), Washington, D.C., USA.

[26] International Monetary Fund (1993). Balance of Payments Manual, fifth edition (BPM5), Washington, D.C., USA.

[27] International Monetary Fund (2009). Balance of Payments Manual, sixth edition (BPM6), Washington, D.C., USA.

[28] Kalemli-Ozcan, S., B. Sorensen, C. Villegas-Sanchez, V. Volosovych, and S. Yesiltas (2015). How to construct nationally representative firm level data from the ORBIS global database: NBER WP:21558.

[29] Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies* 70, 317–341.

[30] Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica* 64, 1263–1297.

[31] Temouri, Y., N. Driffield, and A. Higón (2008). Analysis of productivity differences among foreign and domestic firms: evidence from Germany. *Review of World Economics* 144, 32-54.

[32] United Nations Conference on Trade and Development (2012). *World Investment Report 2012, Towards a New Generation of Investment Policies,* Geneva.

[33] United Nations Conference on Trade and Development (2016). *World Investment Report 2016, Investor Nationality: Policy Challenges,* Geneva.

Table 3: Spillovers to different sets of domestic firms

	$FDI10$					$FDI50$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
HP_t	0.188^{***} (8.93)	0.147^{***} (7.21)	0.0381^{**} (3.15)	0.150^{***} (7.25)	0.0304^{*} (2.39)	0.0958^{***} (5.05)	0.139^{***} (6.92)	0.0314^{**} (3.06)	0.148^{***} (6.95)	0.0442^{***} (4.36)
HP_{t-1}	0.232^{***} (11.55)	0.112^{***} (5.89)	0.0226^{*} (1.99)	0.100^{***} (4.98)	-0.0048 (-0.42)	0.100^{***} (5.12)	0.105^{***} (5.85)	0.0007 (0.06)	0.113^{***} (5.85)	0.008 (0.79)
Year dummies	no	yes	yes	yes	yes	no	yes	yes	yes	yes
Year x industry dummies	no	no	no	yes	yes	no	no	no	yes	yes
Year x country dummies	no	no	yes	no	yes	no	no	yes	no.	yes
Obs.	1,565,835	1,565,835	1,565,835	1,565,835	1,565,835	1,559,197	1,559,197	1,559,197	1,559,197	1,559,197
R-squared	1.7%	4.7%	21.7%	6.5%	23.0%	0.3%	5.2%	24.9%	7.4%	26.4%
Joint F-Test p-value	184.3 0.000^{***}	81.74 0.000^{***}	14.08 0.000^{***}	72.38 0.000^{***}	2.301 0.129	46.12 0.000^{***}	46.12 0.000^{***}	3.975 0.0462^{*}	68.22 0.001^{***}	12.04 0.001^{***}

Note: t statistics in parentheses. The Joint F-test is a test for no long-run effect i.e. of the hypothesis that both coefficients of HP and HP_lagged are zero at the same time.

One star means significance at 5% level, two stars at 1% level and three stars at 0.1% level.

Table 4: Spillovers to *pure* domestic firms

	$FDI10$					$FDI50$				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
HP_t	0.195^{***} (8.93)	0.154^{***} (7.21)	0.0395^{**} (3.15)	0.151^{***} (7.25)	0.032^{*} (2.39)	0.100^{***} (5.16)	0.143^{***} (6.96)	0.0323^{**} (3.12)	0.150^{***} (6.88)	0.044^{***} (4.33)
HP_{t-1}	0.246^{***} (11.54)	0.121^{***} (6.01)	0.0236^{*} (1.97)	0.104^{***} (4.90)	-0.006 (-0.54)	0.101^{***} (5.05)	0.108^{***} (5.86)	0.001 (0.10)	0.114^{***} (5.80)	0.010 (0.98)
Year dummies	no	yes	yes	yes	yes	no	yes	yes	yes	yes
Year x industry dummies	no	no	no	yes	yes	no	no	no	yes	yes
Year x country dummies	no	no	yes	no	yes	no	no	yes	no.	yes
Obs.	1,511,473	1,511,473	1,511,473	1,511,473	1,511,473	1,511,473	1,511,473	1,511,473	1,511,473	1,511,473
R-squared	1.9%	5.0%	22.2%	6.7%	23.5%	0.3%	5.3%	25.5%	7.3%	27.0%
Joint F-Test p-value	184.7 0.00^{***}	84.27 0.00^{***}	13.69 0.00^{***}	67.90 0.00^{***}	2.058 0.15	46.56 0.00^{***}	73.36 0.00^{***}	4.12 0.04^{*}	66.13 0.00^{***}	12.39 0.00^{***}

Note: t statistics in parentheses. The Joint F-test is a test for no long-run effect i.e. of the hypothesis that both coefficients of HP and HP_lagged are zero at the same time.

One star means significance at 5% level, two stars at 1% level and three stars at 0.1% level.

.

Table 5: Focusing on $FDI50$ spillovers to *pure* domestic firms

	(1)	(2)	(3)	(4)	(5)
$HP_t\|_{I-FDI50}$	0.101^{***} (6.12)	0.135^{***} (7.72)	0.028^{**} (2.74)	0.131^{***} (7.49)	0.036^{*} (3.61)
$HP_{t-1}\|_{I-FDI50}$	0.038^{*} (2.28)	0.073^{***} (4.62)	0.011 (1.14)	0.073^{***} (4.63)	0.022^{**} (2.59)
$HP_t\|_{FDI10\cap50}$	0.241^{***} (10.35)	0.211^{***} (9.63)	0.0630^{***} (4.92)	0.204^{***} (9.11)	0.0551^{***} (4.47)
$HP_{t-1}\|_{FDI10\cap50}$	0.253^{***} (11.04)	0.143^{***} (6.86)	0.013 (1.11)	0.122^{***} (5.66)	-0.009 (-0.79)
Year dummies	no	yes	yes	yes	yes
Year x industry dummies	no	no	no	yes	yes
Year x country dummies	no	no	yes	no	yes
Obs.	1,511,473	1,511,473	1,511,473	1,511,473	1,511,473
R-squared	1.9%	5.1%	21.5%	6.9%	22.9%
Joint F-Test 1 p-value	37.40 0.00^{***}	76.73 0.00^{***}	8.152 0.01^{***}	71.87 0.00^{***}	20.50 0.00^{***}
Joint F-Test 2 p-value	200.90 0.00^{***}	127.4 0.00^{***}	19.29 0.00^{***}	103.7 0.00^{***}	7.421 0.00^{***}
Joint F-Test 3 p-value	45.03 0.00^{***}	13.08 0.00^{***}	6.170 0.00^{***}	11.21 0.00^{***}	2.147 0.14

Note: t statistics in parentheses. The Joint F-test is a test for no long-run effect i.e. of the hypothesis that bothcoefficients of HP and HP_lagged are zero at the same time. Here this is done for different combinations of the HP variables. Joint F-test 1 tests wehether $HP_t|_{I-FDI50} + HP_{t-1}|_{I-FDI50} = 0$. JointF-test 2 tests whether $HP_t|_{FDI10\cap50} + HP_{t-1}|_{FDI10\cap50} = 0$. Joint F-test 3 tests whether $HP_t|_{I-FDI50} + HP_t|_{FDI10\cap50} = 0$. One star means significance at 5% level, two stars at 1% level and three stars at 0.1% level.

Appendix 1: Data preparation

In this appendix we carefully report the different steps we went through to create the database that we use. We start with the variable list and the sample delimitations.

Table A.1. Variables

Variable	*Definition*
y (log of output)	Operating revenue deflated by the producer price index (PPI). We have used PPI at 2-digit NACE level. Sources: OPRE is from Amadeus, Orbis ; PPI from EUROSTAT. NACE revision 1 has been used for all countries but Romania. Coverage: 2001 - 2008
k (log of capital)	Tangible fixed assets deflated by a price index for capital. Sources: TFAS are from Amadeus, Orbis; price index for gross fixed capital formation is the average from five capital producing sectors from EUROSTAT. Coverage: 2001-2008
l (log of labour)	Number of employees. Sources: EMPL from Amadeus, Orbis Coverage: 2001-2008
m (log of materials)	Expenditures in intermediate inputs deflated by the producer price index (PPI). We have used PPI at 2 digit NACE level. Sources: MATE from Amadeus, Orbis; PPI from EUROSTAT. NACE revision 1 has been used for all countries but Romania. Coverage: 2001-2008
FDI10	A dummy being 1 if 10% direct single foreign ownership and 0 otherwise. Sources: Amadeus Coverage: 2001-2008
FDI50	A dummy being 1 if 50% ultimate foreign ownership or 50% direct single foreign ownership and 0 otherwise. Sources: Amadeus Coverage: 2001-2008
MNE50	A dummy being 1 if the company belongs to the country in question but ultimately owns with at least 50% affiliates in other countries and 0 otherwise. Sources: Amadeus. Coverage: 2001-2008
Herfindahl	Calculated as the sum of the squared market shares in a given 3 digit industry. Sources: based on OPRE from Amadeus, Orbis Coverage: 2001-2008
HP (Horizontal presence)	Calculated as the share of sales of foreign firms in a given 3-digit industry. Sources: Amadeus Coverage: 2001-2008

Some details from our data preparatory work follows. We start by describing our treatment of missing observations.

a. Economic activity variables

The AMADEUS DVDs are used in the following way: first, we collect as many accounting variables as possible for each for all years 2001 – 2008 from the most recent DVD in our possession (the 2010 DVD)[1]. We do this as we consider these data the most reliable as values for a given year may have been updated compared to earlier DVDs.

For our purpose we need data in their unconsolidated form. This is also basically what AMADEUS offers. However, in some cases – especially for large MNEs' headquarters – the data appear as consolidated. In those cases we first try to get hold of the true unconsolidated data by combining our AMADEUS data with data from ORBIS.

In case of missing values for our unconsolidated variables of interest we fill in from previous versions of the DVDs where such values are available (we have DVDs back to 2003). Our procedure runs as follows: in case of missing values for a certain year we first try to fill in by looking for that specific value on a DVD from a previous year. In case we are still unsuccessful, we rely on interpolated values provided certain conditions are fulfilled. For each of the variables we replace a missing (or negative or zero) value for a given year with the simple arithmetic average of the values from the years immediately preceding and following the year with the missing value. In cases where the jump to an existing value is more than one year no interpolation is performed. Table A.2 summarizes our retrieval of activity variables. From the table it is seen that we lose most observations due to 1) countries with missing observations for material costs or PPI (approx. 2.5 mill obs.) and 2) more randomly missing activity values (approx. 4.7 mill obs.).

Table A.2 Retrieving and interpolation of activity data in manufacturing sector.

	OPRE=Y	EMPL=L	TFAS=K	MATE=M	Total Obs.
Observations from AMADEUS DVD 2010	**4,525,518**	**3,993,344**	**5,328,883**	**3,285,210**	**9,742,272**
Observations filled in from previous versions of AMADEUS DVDs	757,418	532,718	770,989	508,272	939,713
Total after addition of observations from previous versions of AMADEUS DVDs	5,282,936	4,526,062	6,099,872	3,793,482	10,681,985
Obs. With missing ownership information deleted.	99,931	92,034	108,650	60,892	150,951
Total after missing ownership information deleted	5,183,005	4,434,028	5,991,222	3,732,590	10,531,034

[1] Due to the updating procedure of AMADEUS the most complete sample is often two years prior to the actual date of a DVD) hence we stop our sample in 2008.

Deleting inactive and non-manufacturing firms deleted	258,730	195,396	376,857	184,140	888,405
Total after inactive and non-manufacturing firms are deleted	4,924,275	4,238,632	5,614,365	3,548,450	9,642,629
Obs. set to missing due to data being consolidated	10,823	7,536	9,044	7,321	-
Obs. before filling with Orbis Data	4,913,452	4,231,096	5,605,321	3,541,129	9,642,629
Obs filled-in to substitute for consolidated data using Orbis	7,142	7,466	8,103	3,799	-
Total at this raw stage	**4,920,594**	**4,238,562**	**5,613,424**	**3,544,928**	**9,642,629**
Deleting sector 16					4,538
Deleting NACE Rev. 2 non-manufacturing firms from Romania					10,728
Total after deleting sector 16	4,909,017	4,230,206	5,600,015	3,534,187	9,627,363
Obs for countries without material costs or PPI	825,452	761,114	1,313,813	156,137	2,483,785
Total after dropping of countries without material costs or without PPI	4,083,565	3,469,092	4,286,202	3,378,050	7,143,578
Setting activity data equal to zero or negative to missing.	35,814	7,135	510,721	106,521	35,814
Total after setting data equal to zero or negative to missing.	4,047,751	3,461,957	3,775,481	3,271,529	7,143,578
Obs filled in when single years are missing	79,124	139,889	55,390	58,963	-
Total after fill-in when single years missing	4,126,875	3,601,846	3,830,871	3,330,492	7,143,578
Obs. Deleted if still missing activity data	1,664,093	1,139,064	1,368,089	867,710	4,680,796
Total after deleting all obs. With missing data	2,462,782	2,462,782	2,462,782	2,462,782	2,462,782
Obs deleted as outliers	119,287	119,287	119,287	119,287	119,287
Total obs before tfp estimations	**2,343,495**	**2,343,495**	**2,343,495**	**2,343,495**	**2,343,495**

Note: the light grey rows show total numbers of observations at a given stage in the process. The white rows show the changes to the number of observations for the given action.

b. Ownership variables

For the ownership variables we need the full set of DVDs to be able to allow ownership to vary over the years. For ownership variables we also face problems of missing values. To save

observations we fill in 'forward' based on the assumption that if a company has once been influenced by foreign ownership it will keep some knowledge for the years to come. Hence once a firm has had the value 1 for one of the 'foreign' dummies, the 1 is kept for future years as well.

Tables A.3 summarizes our retrieval of ownership observations. Fortunately it is only a relatively small share of the observations that are obtained by the fill-forward procedure. For the estimations we will have to use fewer observations due to the dynamic nature of our model equations.

Table A.3. Retrieving and interpolation of ownership variables

Stage	Tot. Obs	Total Filled Forward	Zeros Filled Forward	Ones Filled Forward
Raw Stage	9,642,629	124,790 (1.29%)	106,540 (1.10%)	18,250 (0.19%)
After Cleaning/ Before TFP estimation	2,343,495	38,073 (1.62%)	31,227 (1.33%)	6,846 (0.29%)

c. PPIs and deflation

As our sample period covers years where different versions of the EUROSTAT producer price indices (PPIs) exist, we make our deflation consistent by using the NACE revision 1 for most of the countries. For Romania where only the NACE revision 2 exists we use this deflator instead. For operating revenue (OPRE) and material costs (MATE) we use PPI with a base year in 2005. For capital costs (TFAS) we use the capital deflator for 2005. Following Javorcik (2004), the capital deflator is the simple average of the PPIs from the five capital equipment producing industries: machinery and equipment, office, accounting and computing machinery, electrical machinery and apparatus, motor vehicles, trailers and semi-trailers and other transport equipment. As we use number of employees as our measure of labor no deflation is needed for this variable.

d. Trimming of the data.

We trim our data to get rid of potential outliers by dropping the top and bottom 1% quantiles of the observations in each 2 digit NACE industry in each country in each year[2] based on a combination of growth rates and ratios considerations for the activity variables (we consider growth rates calculated as log changes of OPRE, EMPL, MATE and TFAS and ratios calculated as MATE/OPRE, TFAS/EMPL, OPRE/EMPL). Finally, we drop country-industry combinations with less than 100 observations available for TFP estimation; see Table A.4.

[2] We do the trimming at these levels as we estimate the total factor productivity for each NACE 2 in each country.

Table A.4: Loss of observations and industries-country combinations with less than 100 observations available to form a sample for the tfp estimation.

Definition of 'Foreign'	# of domestic obs. after cleaning	# Of industry-country combinations deleted (out of)	Obs. deleted	# of domestic obs. for tfp estimation
FDI10	2307753	52 (420)	2501	2305252
FDI50	2278020	54 (420)	2349	2275671
Union_F	2250817	56 (418)	2393	2248424
Pure FDI10	2337549	47 (421)	2122	2335427
Pure FDI50	2307346	51 (420)	2380	2304966

After the TFP estimations we also drop observations from country, industry combinations with negative coefficients for either labor, capital or material costs. As this procedure implies that we drop different numbers of observations depending on the choice of 'foreign' definition (because this choice affects the construction of the HP measures used in the law-of-motion), we summarize our loss of information due of negative coefficients in the production function in Table A.5.

Table A.5 Loss of observations and industries-country combinations due to negative coefficients in the production function estimations.

Definition of 'Foreign'	Neg. coeff. of labor		Neg. coeff. of capital		Neg. coeff. of material		Total Obs. Deleted
	# obs	#industry-country combinations	# obs	#industry-Country combinations	# obs	#industry-Country combinations	
FDI10	32127	15	96995	43	1081	3	128,871
FDI50	16835	17	83428	40	12204	3	99,525
Union_F	41068	12	57145	41	1414	4	98,314
Pure FDI10	35830	20	70762	43	896	3	106,781
Pure FDI50	18065	16	84796	42	1558	3	93,587