

IO FOR EXPORT(S)*

Monika Mrázová[†]

University of Geneva,
and CEPR

J. Peter Neary[‡]

University of Oxford,
CEPR and CESifo

June 3, 2019

Abstract

We provide an overview and synthesis of recent work on models of monopolistic competition with heterogeneous firms in international trade, paying particular attention to pass-through, selection effects, competition effects, and matching endogenous with exogenous distributions. A recurring theme is that CES preferences are extremely convenient for deriving analytic results, but also extremely restrictive in their theoretical and empirical implications. We introduce the class of “constant-response demand functions” to describe some related families of demand functions that provide a unifying principle for much recent work that explores alternatives to CES demands.

Keywords: Heterogeneous Firms; Pass-Through; Quantifying Effects of Globalization; Super- and Sub-Convexity; Supermodularity.

JEL Classification: F12, L11, F23

*This is a revised version of a plenary lecture presented at EARIE 2018, the 45th Annual Conference of the European Association for Research in Industrial Economics, Athens, September 2018. We are grateful to participants on that occasion, at the 11th Research Workshop on International Trade in Villars, Switzerland, and at a seminar at University College Dublin, as well as Paolo Bertolotti, Luis Cabral, Toshi Okubo and John Vickers, for helpful comments and discussions. Monika Mrázová thanks the Fondation de Famille Sandoz for funding under the “Sandoz Family Foundation - Monique de Meuron” Programme for Academic Promotion. Peter Neary thanks the European Research Council for funding under the European Union’s Seventh Framework Programme (FP7/2007-2013), ERC grant agreement no. 295669.

[†]Geneva School of Economics and Management (GSEM), University of Geneva, Bd. du Pont d’Arve 40, 1211 Geneva 4, Switzerland; e-mail: monika.mrazova@unige.ch.

[‡]Department of Economics, University of Oxford, Manor Road, Oxford OX1 3UQ, UK; e-mail: peter.neary@economics.ox.ac.uk.

1 Introduction

In this paper we aim to provide an overview of recent research in international trade, from the perspective of the theory of industrial organization (henceforth IO). There are two aspects to this perspective, mirroring the two different interpretations of our title. On the one hand, we are interested in “IO for *Exports*”: what is the structure of export markets?; and how does it affect which firms export, whether or not trade fosters competition, and a host of other similar questions? On the other hand, we want to think about “IO for *Export*”: which IO models are used in trade? As is well-known, the answer to the latter question is not any of the many varieties of oligopoly used in IO itself, but rather the monopolistically competitive model of Chamberlin (1933), as formalized by Dixit and Stiglitz (1977); especially with CES preferences, introduced as a special case by Dixit and Stiglitz (1977), brought to center stage by Dixit and Norman (1980) and Krugman (1980) among many others, and extended to firm heterogeneity by Melitz (2003). Ironically, though monopolistic competition originated in IO, it is used relatively rarely there, so it could be described as “IO for export *only*.”¹

While monopolistic competition is by far the dominant paradigm in international trade, it is by no means the only approach. First there is a large literature on “strategic” trade policy, dating from the 1980s, that uses off-the-shelf IO models of oligopoly to address standard issues of trade policy, mostly at the level of a single industry.² Second, there have been some attempts to explore the implications of oligopoly in general equilibrium for trade questions, both of the traditional kind such as trade patterns, gains from trade and pricing-to-market,³ and also with applications to topics such as cross-border mergers and multi-product exporters.⁴ Finally, there is a small literature on the theory of superstar firms that compete oligopolistically, while coexisting with a monopolistically competitive fringe.⁵

¹See Neary (2003b) for further discussion.

²See Brander (1995) for an overview, and Mrázová (2011), Mrázová et al. (2013), and Mrázová (2015) for applications to trade agreements.

³See, for example, Neary (2003a), Atkeson and Burstein (2008), Neary (2016), Gaubert and Itskhoki (2016), and Nocke and Schutz (2018).

⁴See, for example, Neary (2007), Eckel and Neary (2010) and Eckel et al. (2015).

⁵See, for example, Neary (2010), Shimomura and Thisse (2012), Parenti (2018), and Cabral (2018).

However, there are good reasons why the monopolistically competitive paradigm has remained dominant in trade. The availability of large data sets on exporting firms, and the desire to allow for entry and exit and to take account of general-equilibrium feedbacks to and from factor markets, make bespoke modelling of individual sectors infeasible, and a “one-size-fits-all” approach to the choice of market structure very convenient.

Two general themes emerge from our analysis, one old, one new. First, the widespread assumption of CES preferences, implying demand functions that are log-linear in price, is extremely convenient for deriving analytic results, but also extremely restrictive in its theoretical and empirical implications.⁶ Second, we introduce the class of “constant-response demand functions” to describe a group of related families of demand functions, all of which nest CES demands, that imply an equal absolute or relative response of some firm-level variable to an exogenous change in marginal cost. As we shall see, this class of demands provides a unifying principle for much recent work, including our own, that explores alternatives to CES demands.

The plan of the paper is as follows. Section 2 introduces the notation and also discusses recent evidence that questions a key implication of CES preferences. Section 3 considers the issue of how cost changes are passed through to goods prices, and introduces an approach to comparing different demand functions that we use throughout. Sections 4 and 5 turn to consider the central topics of selection and competition effects respectively. Section 6 explores the relationship between the distribution of firm productivity, and the distributions of the level and growth rate of sales. Finally, Section 7 provides an overview of the constant-response demand functions introduced in earlier sections, while Section 8 concludes. Proofs and further details can be found in our other papers: Mrázová and Neary (2017) for Sections 3 and 5, Mrázová and Neary (2019) for Section 4, and Mrázová et al. (2015) for Section 6.

⁶For convenience, with a slight abuse of terminology, we refer to these demands as “CES demands”.

2 Preliminaries

2.1 A Core Model

We begin by presenting a simple core model of a monopoly firm. This serves to introduce the notation and setting that we will use throughout the remainder of the paper.

Consider a monopoly firm facing an inverse demand function $p(x)$, giving price as a strictly decreasing function of firm output x . This is consistent with monopolistic competition where the firm takes the “perceived” demand function as given. In general equilibrium, the demand function has extra arguments: we will see examples of that later. As for technology, we adopt the simplest specification popularized by Dixit and Stiglitz (1977): all firms incur a fixed cost f , often assumed to be constant across firms, and a constant marginal cost of production c .⁷ Profit maximization therefore implies simple first- and second-order conditions:

$$p + xp' = c \quad \text{and} \quad 2p' + xp'' < 0 \tag{1}$$

In words, marginal revenue must equal marginal cost, and marginal revenue must be declining in output at the optimum. It will prove convenient to write these conditions in terms of two key demand parameters:

$$\varepsilon(x) \equiv -\frac{p(x)}{xp'(x)} > 0 \quad \text{and} \quad \rho(x) \equiv -\frac{xp''(x)}{p'(x)} \tag{2}$$

Here $\varepsilon(x)$ is the demand elasticity and $\rho(x)$ the convexity of demand; these are unit-free measures of the slope and curvature of the demand function, respectively. Reexpressing the first- and second-order conditions in (1) in terms of these parameters yields, first, a simple expression linking the price-cost margin or markup, p/c , to the elasticity of demand, and,

⁷Dixit and Stiglitz should not be held responsible for the widespread use of CES preferences: they presented them only as an example, albeit with the unique implication of constrained efficiency (see their Section I), and each of them is on record as expressing reservations about them: see Dixit (2004) and Stiglitz (2004). However, they deserve credit for popularizing, and perhaps making respectable, the very simple $\{f, c\}$ technology. Ironically, this is another example of IO for export only: it is now ubiquitous in trade, whereas more complicated cost functions are often used in the IO literature.

second, a restriction that convexity must be less than two:

$$\frac{p}{c} = \frac{\varepsilon}{\varepsilon - 1} \quad \text{and} \quad \rho < 2 \quad (3)$$

A key benchmark is the CES special case, where the inverse demand function and the markup take the special forms:

$$p(x) = \beta x^{-1/\sigma} \quad \Rightarrow \quad \frac{p}{c} = \frac{\sigma}{\sigma - 1} \quad (4)$$

This illustrates two key implications of CES preferences: the elasticity of demand is constant, equal to σ , and the markup is the same for all firms, equal to $\sigma/(\sigma - 1)$.

2.2 Evidence on Variable Markups

It has been recognised for some time that the prediction of CES demands that markups are always the same, and depend only on the preference parameter σ , is unsatisfactory. However, it is only recently that persuasive evidence against this prediction has become available. A key reference is De Loecker et al. (2016), who calculate markups for a sample of Indian firms without making prior assumptions about market structure or the form of demand. Their results show that the distribution of markups from is very far from being concentrated at a single value. A possible explanation is that such markup heterogeneity arises from aggregation across sectors with different elasticities of substitution. However, Lamorgese et al. (2014), who use data on Chilean firms, show that even greater heterogeneity is observed when the data are disaggregated by sector. Taken together, this evidence suggests that markup distributions are far from the Dirac form implied by CES preferences, but to date little attention has been paid to seeking explanations for the pattern of markups revealed in these new studies.

3 Pass-Through

The first substantive question we address is that of pass-through from costs to prices. The topic is of interest in itself but also as an illustration of the differences in how essentially the same problem is addressed in IO and in international economics.

3.1 Pass-Through in Industrial Organization

The archetypical pass-through question posed in industrial organization is both simple and important: if marginal cost rises by a euro, by how much will a profit-maximizing firm raise its price; i.e., how large is dp/dc ? The answer is easy to calculate from the first-order condition in (1) above:

$$p + xp' = c \quad \Rightarrow \quad \frac{dp}{dc} = \frac{1}{2 - \rho} \quad (5)$$

Thus a shock to costs is passed through to prices by more the more convex is the demand function. Moreover, the threshold for full, or “euro-for-euro”, pass-through is that the degree of convexity is exactly one:

$$\frac{dp}{dc} - 1 = \frac{\rho - 1}{2 - \rho} = 0 \quad (6)$$

Demand functions implying constant pass-through are members of the Bulow-Pfleiderer family, due to Bulow and Pfleiderer (1983):⁸

$$p(x) = \beta \left(x^{\frac{1-\kappa}{\kappa}} + \gamma \right), \quad \beta \frac{1-\kappa}{\kappa} < 0 \quad (7)$$

This implies a constant pass-through equal to κ : $dp/dc = \kappa$. Special cases of this include linear demand, with convexity ρ equal to zero and a constant pass-through of $\kappa = 1/2$; and log-linear direct demand, $\log x(p) = \gamma + \delta p$, corresponding to the limiting case of (7) as κ

⁸This is more familiar when written, following Bulow and Pfleiderer, as $p = \alpha - \beta x^\delta$. We write it differently to facilitate comparison with other demand functions to be introduced later. Bulow and Pfleiderer considered only the cases where δ is either strictly positive or asymptotically zero, when the function becomes $p = \alpha - \beta \log x$.

approaches one. The latter demand function has convexity $\rho = 1$, and so is the demand function implied by the condition in (6) holding globally. In general, values of $\kappa > 1$ and $\rho > 1$ correspond to log-convex direct demand, and values of $\kappa < 1$ and $\rho < 1$ to log-concave demand.

3.2 Pass-Through in International Economics

Pass-through is also widely studied in international economics, both pass-through from tariffs to domestic prices in international trade, and pass-through from exchange rates to domestic prices in international monetary economics. However, the approach is totally different from that in industrial organization: compare for example Gopinath and Itskhoki (2010) with Weyl and Fabinger (2013), both important papers that refer to large literatures but have only a single reference in common. It is not merely the answers that differ between the two fields, but the questions too. In IO the focus is on absolute or euro-for-euro pass-through as we have seen, whereas in international economics the question posed relates to proportional pass-through: how large is $d \log p / d \log c$? Both approaches make perfect sense in their respective contexts: IO scholars are typically analysing the effects of a cost or tax change on a single industry, whereas in international economics the focus is on a change in policy or exchange rates at the economy-wide level. The first question is a partial-equilibrium one, so it is natural to work with absolute changes, whereas the second is a general-equilibrium one, mandating a focus on relative prices and proportional changes. However, it is important to understand the differences between the two.

Once again, the proportional pass-through coefficient can be derived from the first-order condition:

$$\frac{d \log p}{d \log c} = \frac{\varepsilon - 1}{\varepsilon} \frac{1}{2 - \rho} \quad (8)$$

Unlike the corresponding term for absolute pass-through, (5), this depends on the elasticity of demand as well as its convexity. So, the threshold for full (100%) proportional pass-through

is:

$$\frac{d \log p}{d \log c} - 1 = \frac{\varepsilon \rho - \varepsilon - 1}{2 - \rho} = 0 \quad (9)$$

This must be equivalent to the demand function taking the CES form, as we have already seen that only in that case is the ratio p/c constant. This is confirmed by calculating explicitly the convexity of CES demand and substituting into (9):

$$p(x) = \beta x^{-1/\sigma} \quad \Rightarrow \quad \varepsilon = \sigma, \quad \rho = \frac{\sigma + 1}{\sigma} \quad \Rightarrow \quad \frac{d \log p}{d \log c} = 1 \quad (10)$$

This suggests that pass-through will be more than 100% if and only if the demand function at the initial equilibrium is more convex than a CES demand function. We will return to this below.

We saw in the previous sub-section that an important benchmark is the Bulow-Pfleiderer demand family, which implies constant absolute pass-through. So it is natural to ask if there are demand functions that imply constant proportional pass-through (CPPT): $d \log p / d \log c = \kappa$. The answer to this question is the CPPT family, introduced in Mrázová and Neary (2017):⁹

$$p(x) = \frac{\beta}{x} \left(x^{\frac{\kappa-1}{\kappa}} + \gamma \right)^{\frac{\kappa}{\kappa-1}}, \quad \beta > 0 \quad (11)$$

This family also has the property that it implies a constant proportional response of markups to marginal costs: $d \log(p/c) / d \log c = \kappa - 1$. Casual inspection suggests that these demand functions are very different from the Bulow-Pfleiderer demands in equation (7). However, it is difficult to compare the two directly: each has three parameters and they have very different functional forms. Indeed, comparing any two demand functions is difficult in general. To address this problem, we digress in the next sub-section to review the approach introduced in Mrázová and Neary (2017) that provides a way of doing this.

⁹The parameter γ must be positive to ensure that the elasticity of demand exceeds one.

3.3 Digression: How to Compare Demand Functions?

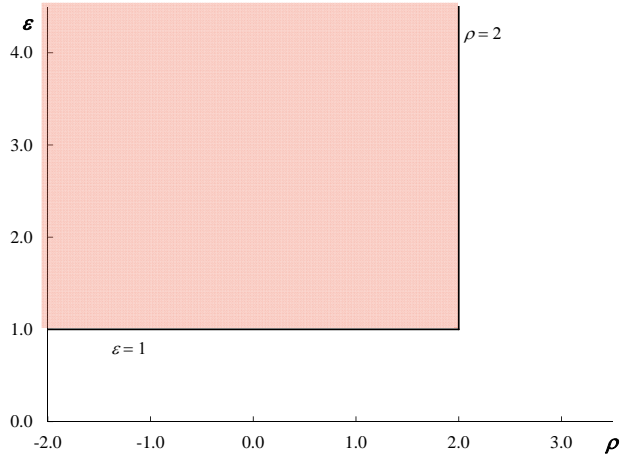


Figure 1: The Admissible Region in Convexity-Elasticity Space

With three parameters each, it is not clear how we can compare the two demand functions in (7) and (11). A clue is that for many questions of interest – both absolute and proportional pass-through as here, as well as other questions later in the paper – only two parameters implied by each demand function matter: the elasticity and convexity of demand already introduced above. Hence it makes sense to illustrate different demand functions in the space of these two parameters, as in Figure 1. The shaded area represents the admissible region implied by the restrictions in equation (3): at a profit-maximising equilibrium, a monopoly firm must be at a point on its demand function where the elasticity is greater than one, and the convexity is less than two.¹⁰

The space illustrated in Figure 1 may seem unfamiliar at first, so it is desirable to have some landmarks. One such is the linear demand curve: it implies that convexity is zero, and so it is represented by the vertical dashed line at $\rho = 0$ in Figure 2. It turns out that a more important benchmark is that of CES demands. Note first that this is the only exception

¹⁰Of course, the admissible values of ε are unbounded above and those of ρ are unbounded below. The range of values illustrated in the figures is broadly in line with the available empirical evidence: see for example Mrázová and Neary (2017), footnote 8 and Section IV.A.

to the general rule that both ε and ρ vary with sales: as we saw in equation (10), their values are uniquely determined by the elasticity of substitution σ . Eliminating σ from those expressions gives the relationship between elasticity and convexity implied by CES demands:

$$\varepsilon = \frac{1}{\rho - 1} \quad (12)$$

This relationship is illustrated by the curve labelled “SC” in Figure 2; it ranges from an asymptotically infinite value of ε as ρ approaches one to a value of ε equal to one when ρ equals two. Each point on this curve represents a particular CES demand function corresponding to a particular value of σ ; for example, the point $\{\rho = 2, \varepsilon = 1\}$, just on the boundary of the admissible region, corresponds to the Cobb-Douglas special case.

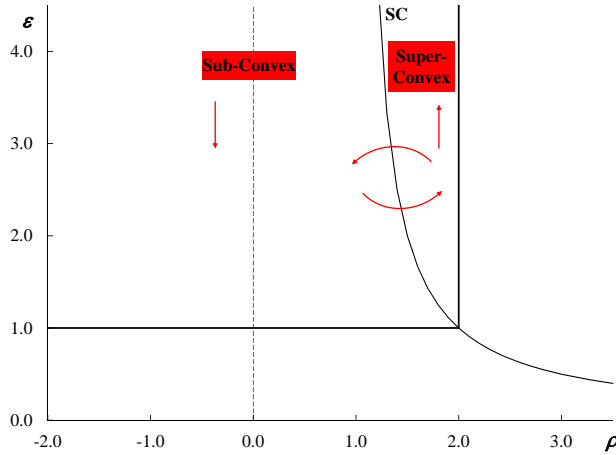


Figure 2: The Sub- and Super-Convex Regions

This locus is not just of interest in itself, it also relates to the argument already made that pass-through is less than 100% when demand is less convex than a CES demand function with the same elasticity, i.e., when $\rho < (\varepsilon + 1)/\varepsilon$.¹¹ Mrázová and Neary (2019) call demand functions with this property “subconvex”, and in Figure 2 they are represented by points

¹¹Formally, subconvexity is a local property, and is defined as $\log p(x)$ being concave in $\log x$ at a point. Mrázová and Neary (2019) show that this property is equivalent to those stated in the text.

in the region to the left of the SC locus.¹² A key property they exhibit is an elasticity of demand that is decreasing in output:

$$\varepsilon_x = \frac{\varepsilon}{x} \left[\rho - \frac{\varepsilon + 1}{\varepsilon} \right] \quad (13)$$

which is negative when ρ is less than its CES value. When the demand curve exhibits this property, larger firms face a lower elasticity of demand and hence have higher markups in the cross section, and all firms enjoy higher profit margins as they move down their demand curves over time. (The arrows in the figure indicate how elasticity changes as firm output rises.) In the converse case, of superconvexity, all these properties are reversed. However, subconvexity is theoretically more plausible, both on introspective grounds (see Marshall (1920));¹³ and also in that it leads to more plausible predictions (see Dixit and Stiglitz (1977) and Krugman (1979)). It is also consistent with a substantial body of empirical evidence that finds pass-through to be less than 100%.

The linear and CES demand functions are not the only ones that can be illustrated in $\{\rho, \varepsilon\}$ space. As shown in Mrázová and Neary (2017), any demand function can be illustrated in this space by a smooth curve, which they call the “demand manifold” corresponding to the demand function. (The only exception is CES demands, where as we have seen the manifold is a single point.) Figure 3 illustrates the manifolds for three other widely-used demand functions: the translog (which from a firm’s perspective is observationally equivalent to the “Almost Ideal” system of Deaton and Muellbauer (1980)), the Stone-Geary LES

¹²It is possible for the same demand function, with given parameters, to be subconvex for some levels of output and superconvex for others. (For an example, see Mrázová and Neary (2017), Online Appendix B10.) However, this property is rare, at least as far as most widely-used demand functions is concerned. Moreover, as the argument in the text shows, the manifold implied by such a demand function must be horizontal at the point where it crosses the SC locus, since it is falling in output to the left and rising to the right: see the horizontal arrows in Figure 2.

¹³In Chapter 3 of Book 4 of his *Principles*, Marshall introduced his “law of demand” – in modern parlance, demand curves slope downwards – while in Chapter 4 he introduced what he called “the general law of variation of the elasticity of demand”: “The elasticity of demand is great for high prices, and great, or at least considerable, for medium prices; but it declines as the price falls; and gradually fades away if the fall goes so far that satiety level is reached.” Hence the property is sometimes called “Marshall’s Second Law of Demand.”

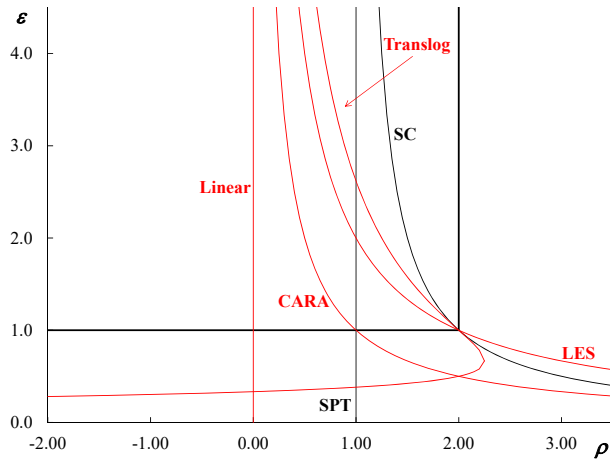


Figure 3: Demand Manifolds for Some Common Demand Functions

(“linear expenditure system”), and the negative exponential or CARA (“constant absolute risk-aversion”) system.¹⁴ (Explicit expressions are in Appendix B.)

There are a number of advantages to considering demand functions through the lens of their manifolds. First, different demand functions can be directly compared with each other. For example, Figure 3 shows that, despite their very different origins and system-wide properties, all three of the CARA, Stone-Geary and translog have properties at the firm level that are midway between those of linear and CES demands. In particular, they are all subconvex, though less so than linear demands. Second, demand manifolds make it possible to relate assumptions about demand directly to comparative statics properties: we will see examples of this later. Finally, though the location of a demand function in $\{x, p\}$ space is affected by the values of all its parameters, the location of its manifold in $\{\rho, \varepsilon\}$ space is often unaffected by, or invariant to, the values of these parameters. Mrázová and Neary (2017) call this property “manifold invariance”, and provide necessary and sufficient conditions for it to hold. Not all demand functions exhibit this property with respect to

¹⁴The non-CES demand functions whose manifolds are shown in Figure 3 have been explored in monopolistic competition with heterogeneous firms by a number of authors. Melitz and Ottaviano (2008) consider linear demands; Simonovska (2015) the Stone-Geary LES; Behrens and Murata (2007) the negative exponential; and Feenstra and Weinstein (2017) the translog.

some of their parameters, far less to all of them. This is true of the CES for example: the CES point-manifold is invariant with respect to changes in the level parameter β but not to changes in the elasticity of substitution σ . However, many demand functions have manifolds that are invariant to some of their parameters, and some (including all those other than the CES in Figure 3) have manifolds that are invariant to all of them. This makes it much easier to think about the implications of different demand assumptions using the demand manifold than the demand function itself, since the latter is never invariant to any parameter changes.

3.4 Back to Pass-Through

Armed with this way of comparing demand functions in general, we can now return to the question of how to compare the demand functions from Sections 3.1 and 3.2 in particular. This also allows us to compare visually the conditions that imply constant absolute and proportional pass-through. Direct calculations show that the demand manifold implied by the Bulow-Pfleiderer demand function in equation (7) is given by:

$$\rho = 2 - \frac{1}{\kappa} \quad (14)$$

while that implied by the CPPT demand function in equation (11) is given by:

$$\bar{\rho}(\varepsilon) = 2 - \frac{1}{\kappa} \frac{\varepsilon - 1}{\varepsilon} \quad (15)$$

These manifolds are illustrated for different values of κ in the two panels of Figure 4.

It is clear that the manifolds are much easier to compare than the corresponding demand functions. Whereas each demand function has three distinct parameters, each manifold is invariant with respect to two of these parameters, and depends only on the pass-through coefficient κ . The manifolds show clearly how demand adjusts as output rises (which in the subconvex region is equivalent to the elasticity falling). Bulow-Pfleiderer demands have constant convexity (given the one-to-one relationship between demand convexity and absolute

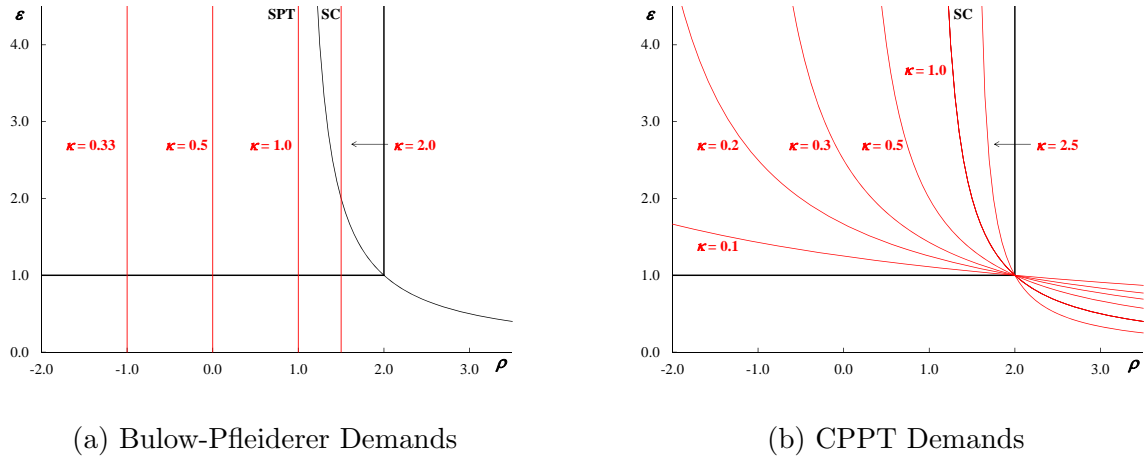


Figure 4: Constant Pass-Through Demand Manifolds

pass-through from (5) and (14)), while CPPT demands become more convex as elasticity falls.

The demand manifolds in Figure 4 are not only the manifolds implied by the Bulow-Pfleiderer and CPPT demand functions, they are also contour lines corresponding to different values of κ , the degree of either absolute or proportional pass-through. As a result, they can be combined with the demand manifolds for any other demand functions to read off the rates of pass-through implied by those demands. For example, recalling the three new demand manifolds in Figure 3, we can see that all three imply a higher level of absolute pass-through for large firms; Stone-Geary demands imply a constant rate of proportional pass-through for all firms, equal to 0.5, whereas CARA demands imply that the rate of proportional pass-through decreases with firm size, and translog demands imply that it increases.

4 Selection Effects

We turn next to one of the most distinctive features of models with heterogeneous firms: their ability to predict which firms will select into which activities. Consider first the question of which firms will choose to export. Here the key prediction of the Melitz (2003) model is that

more productive firms select into exporting. It turns out that this is an extremely robust result. It requires only that ex post profits π are decreasing in c , which must be true in all models of this kind. To see this, write the maximized value of operating profits as follows:

$$\pi(\tau, c) \equiv \max_x \tilde{\pi}(x, \tau, c), \quad \tilde{\pi}(x, \tau, c) = (p(x) - \tau c)x \quad (16)$$

where τ is an iceberg trade cost: τ units must be produced and shipped to the export market in order for one unit to arrive. For simplicity, assume that $\pi(\tau, c)$ is continuous in $\{\tau, c\}$. Differentiating (16) and invoking the envelope theorem, we can see that operating profits are unambiguously decreasing in c :

$$\pi_c = \tilde{\pi}_c = -\tau x < 0 \quad (17)$$

This result is independent of the form of the demand function; in particular, it is not sensitive to whether demands are CES or not. Hence ranking firms by their productivity, or inverse marginal cost, unambiguously predicts their ranking by total profits: $\Pi = \pi - f$. So there must be a unique threshold level of productivity, with all firms more productive than the threshold choosing to export and all those less productive choosing not to.

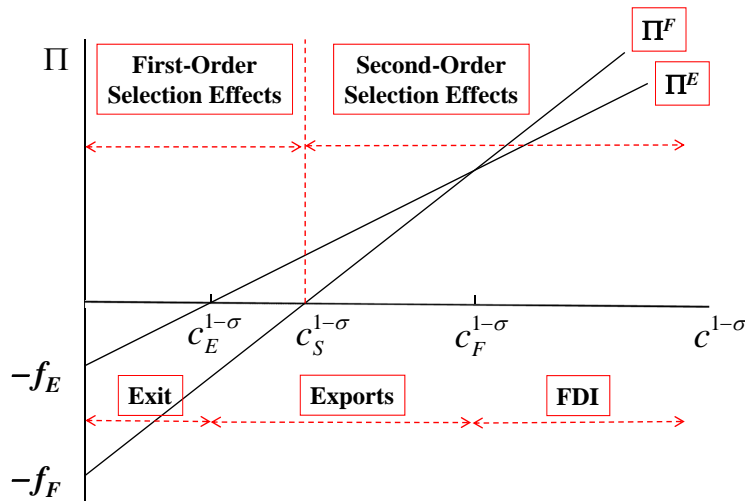


Figure 5: Selection into Exports versus FDI

Consider next the case where each firm has the option of choosing to serve a market in two alternative ways, either by exporting (E) or by engaging in foreign direct investment (FDI). Firms therefore face a “proximity-concentration trade-off”: FDI entails operating a foreign plant and selling to foreign consumers without incurring the trade cost, so $\tau = 1$; however, it foregoes the benefits of concentrating production at home, since it incurs a fixed cost f_F that is higher than the fixed cost of exporting f_E :

$$\Pi(\tau, c) = \max \begin{cases} \Pi^F(c) = \pi(1, c) - f_F \\ \Pi^E(\tau, c) = \pi(\tau, c) - f_E \end{cases} \quad (18)$$

The outcome under CES demand has been shown by Helpman et al. (2004) and is illustrated in Figure 5. Profits under both regimes are linear in an increasing power function of productivity; profits under FDI are lower than under exporting for the lowest-productivity firms, since FDI incurs a higher fixed costs; and the two profit schedules intersect once as shown.¹⁵ As the lower set of labels indicates, there is now three-way selection, with the least efficient firms not serving the market as before, firms of intermediate efficiency exporting, and the most efficient firms engaging in FDI.

However, Mrázová and Neary (2019) show that this predicted selection pattern is less robust than that for the choice between exporting or not. To see why, consider the upper set of labels in Figure 5. These categorise productivity levels not in terms of their predicted outcomes but in terms of the choices they offer to firms. Low-productivity firms, with costs higher than c_S (represented by points to the left of $c_S^{1-\sigma}$ in the figure), face the same choice as before: FDI is not profitable, so the only decision they face is whether to export or not. Mrázová and Neary (2019) call the outcomes of such a decision a “first-order selection effect”: firms face a choice between one activity, exporting, whose profits vary with productivity, and another, not exporting, whose profits are the same for all firms. This is in contrast to the

¹⁵To ensure that both exports and FDI are observed we require that the two curves intersect in the positive quadrant, which is equivalent to the “freeness” of trade (the variable trade cost adjusted by the degree of substitutability) being greater than the ratio of fixed costs: $\tau^{1-\sigma} > f_E/f_F$.

“second-order selection effect” that arise from the choices of high-productivity firms with costs lower than c_S . For such firms both exports and FDI are profitable, and their profitability in both modes varies with productivity. Hence, the conventional sorting, where more efficient firms select into FDI, is only guaranteed if the two curves have a single intersection as in Figure 5; in the language of monotone comparative statics, if they exhibit the single-crossing property.¹⁶

When might we expect the single-crossing property to hold? One consideration which suggests that second-order selection effects may be less robust than first-order ones comes from Appendix A. We show there that CES demands are the only case in which variable profits are a power function of marginal cost, and so total profits are affine in some simple transformation of marginal cost as in Figure 5. Thus, with any demands other than CES, we cannot be sure without further investigation whether the single-crossing property holds and so whether the second-order selection effects will be as illustrated in Figure 5.

Checking directly whether the single-crossing property holds in any particular application is not easy. However, it is easy to check for an important sufficient condition. Recall that π is continuous in c , and assume also that the two profit functions Π^F and Π^E have different slopes at any point where they intersect. Given this, the single crossing property is equivalent to Π^F being steeper than Π^E at any productivity level where they intersect, the configuration shown in Figure 5. This suggests a sufficient condition for second-order selection effects which is much easier to check: if Π^F is steeper than Π^E at *any* level of productivity, then single crossing must hold. This condition is equivalent to $\pi_c(1, c) < \pi_c(\tau, c)$ at every value of c , which, provided π is twice differentiable in $\{\tau, c\}$, is equivalent to a restriction on its second derivative: $\pi_{c\tau}$ is positive, or $\pi(\tau, c)$ is *supermodular* in $\{\tau, c\}$. A necessary and sufficient condition for $\pi_{c\tau} > 0$ can be found in turn by differentiating (17) with respect to τ :

$$\pi_{c\tau} = -x - \tau \frac{\partial x}{\partial \tau} = -x \left(1 + \frac{d \log x}{d \log c} \right) \quad (19)$$

¹⁶See for example Milgrom and Roberts (1990) and Vives (1990).

Hence $\pi(\tau, c)$ is supermodular in $\{\tau, c\}$ if and only if $-d \log x / d \log c > 1$; i.e., if and only if the elasticity of output with respect to marginal cost is greater than one in absolute value. (In what follows we use the more convenient mnemonic “EMR” for the elasticity (of output) with respect to marginal revenue.) To understand this condition, note that supermodularity is a restriction on a second derivative, that is, on a difference in differences: not whether higher trade costs reduce profits, but whether they do so by less for lower-productivity firms. The final term on the right-hand side of (19) shows that they are more likely to do so the more elastic is the volume sold to an increase in costs. Offsetting this is the first term, equal to one: this reflects the fact that higher trade costs reduce profits by less for a firm that is selling less to begin with. Whether or not supermodularity obtains depends on the balance between these two forces.

An alternative way of writing the condition for supermodularity in (19) allows us to illustrate it in demand-manifold space. It can be expressed in terms of demand parameters by noting that it equals the elasticity of demand times the elasticity of price with respect to marginal cost, already given in (8) above:

$$-\frac{d \log x}{d \log c} = \varepsilon \frac{d \log p}{d \log c} = \frac{\varepsilon - 1}{2 - \rho} \quad (20)$$

Supermodularity therefore requires that the following expression be positive:

$$-\left(1 + \frac{d \log x}{d \log c}\right) = \frac{\varepsilon + \rho - 3}{2 - \rho} \quad (21)$$

Thus, from (19), selection into FDI by large firms is only guaranteed if the EMR is greater than one in absolute value, which from (21) is equivalent to the demand function being sufficiently elastic and/or sufficiently convex that ε and ρ sum to more than three.

To understand these results, it is helpful first to consider Figure 6, which shows the demand and marginal revenue functions for three canonical special cases. Panel (a) illustrates the CES case. It is easy to check from (20), using the CES values for ε and ρ from (10), that

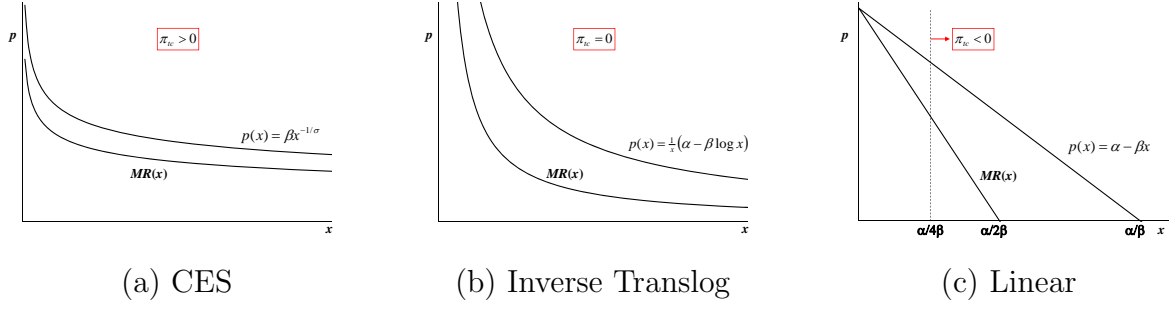


Figure 6: Second-Order Selection Effects and Demand

the EMR in this case equals σ , which is greater than one. Hence, if we compare two firms facing the same CES demand function, one with marginal costs 10% lower than the other, the more productive firm will have an output more than 10% higher than the less productive one. This helps explain the result of Helpman et al. (2004), which exhibits what is often called the “Matthew Effect” (“to those who have, more shall be given”): more efficient firms enjoy higher profits when they engage in FDI.

Panel (c), by contrast, illustrates the case of linear demands. Now the EMR is less than one for larger firms. Intuitively, for such large firms the additional trade cost they must pay if they export is small, since their production cost is low to begin with, and trade costs are proportional to production costs; moreover it reduces their sales very little, since the marginal revenue curve is not steep. Hence, supermodularity fails for larger firms under linear demands, and it can be shown that the largest ones will indeed exhibit reverse selection effects, choosing to export rather than to engage in FDI.

Finally, Panel (b) illustrates a third special case with the property that the EMR equals one. As a result, the proximity-concentration trade-off yields the same outcome for all firms irrespective of their productivity: the Π^X and Π^F curves in Figure 5 are parallel. Hence, no firms will find it profitable to engage in FDI since it incurs a higher fixed cost, and so no selection will be observed. Mrázová and Neary (2019) show that this demand function is the inverse translog, $p(x) = \frac{\beta}{x}(\log x + \gamma)$. It is a special case of a family of demand functions that exhibit a constant EMR: $d \log x / d \log c = -\kappa$. Mrázová et al. (2015) call this the “CEMR”

family:¹⁷

$$p(x) = \frac{\beta}{x} \left(x^{\frac{\kappa-1}{\kappa}} + \gamma \right) \quad \beta(\kappa - 1) > 0 \quad (22)$$

The corresponding CEMR manifold is $\bar{\rho}(\varepsilon) = 2 - \frac{1}{\kappa}(\varepsilon - 1)$, which implies a family of affine curves in $\{\rho, \varepsilon\}$ space, some of which are illustrated for different values of κ in Figure 7. Of these, the inverse translog, where the EMR is unity, corresponds to the boundary between the sub- and supermodular regions in $\{\rho, \varepsilon\}$ space, the curve labelled “SM” in Figure 7.

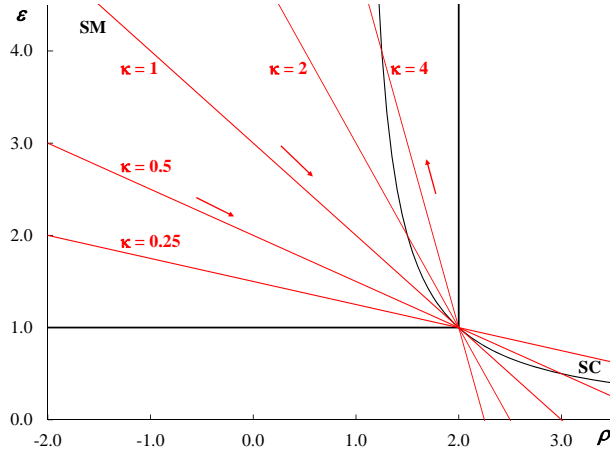


Figure 7: CEMR Demand Manifolds

5 Competition Effects

While selection effects are the most distinctive and novel feature of heterogeneous firms models, competition effects, and especially how they differ across firms, are also a key feature. The term “competition effect” can be used in two different senses in this context. On the one hand, competition effects often refer to the fact that markups vary across firms in the cross-section. In this sense, as we have already seen in Section 3, CES demands rule out competition effects, and subconvex demands are necessary and sufficient for monopolistic

¹⁷ Mrázová and Neary (2019), who first considered these demands in this context, called them the Inverse PIGL (“price-independent generalized linear”) demand family, drawing on Muellbauer (1975). The inverse translog is the limiting case of the CEMR family as $\kappa \rightarrow 1$, as can be shown using l’Hôpital’s Rule.

competition to exhibit the empirically plausible case where larger firms have higher markups. On the other hand, there is a different sense in which competition effects arise following any exogenous shock to industry equilibrium: such shocks typically affect firms both directly, and also indirectly by changing the competitive environment in which they operate. An important and interesting example of this is the effects of globalization, in the sense of an increase in the size of the global economy. Here too subconvexity is crucial: it is necessary and sufficient for a “Matthew Effect” on the profile of profits across firms, whereby larger firms expand but smaller ones are squeezed.

To see how competition effects arise following a globalization shock, we need to go beyond the “firm’s eye view” we have adopted so far, and consider the determination of industry equilibrium.¹⁸ Also, since we want to allow for firms to sell in many markets, we distinguish for the first time between their total output y and the amount they sell to an individual consumer x . Following Krugman (1979), we take the simplest setting, where all consumers are identical, there are L of them in each country, and the world consists of k identical countries, so $y = kLx$. In addition, though this is not needed for all our results, it is convenient to assume that preferences are additively separable, so the demand function for a typical good is given implicitly by the consumer’s first-order condition $u'(x) = \lambda p$, where λ is the marginal utility of income. With these assumptions, we can write the ex post profits of a typical firm as follows:

$$\pi(\underline{c}, \underline{\lambda}, \underline{k}) \equiv \max_y [p(y, \lambda, k) - c]y \quad \text{where} \quad p(y, \lambda, k) = \lambda^{-1} u'(y/kL) \quad (23)$$

In previous sections we considered the demand function from a firm’s perspective only. Now we want to solve explicitly for industry equilibrium, so we need to consider demand determinants additional to their sales. In (23) these are k , the exogenous globalization shock, equal to the number of identical countries in the world, and λ , the marginal utility of income

¹⁸This section is based on Mrázová and Neary (2017), Section III.B. See also Zhelobodko et al. (2012), Bertolotti and Epifani (2014), Mrázová and Neary (2014), and Bache and Laugesen (2015).

under additive separability, which plays the role of the level of competition in the industry.¹⁹ This is exogenous to each firm, but is determined endogenously in industry equilibrium by the free-entry condition, which implies that an entrant has zero expected value:

$$\bar{v}(\lambda, k) \equiv \int_{\underline{c}}^{\bar{c}} v(c, \lambda, k) g(c) dc = f_e, \quad v(c, \lambda, k) \equiv \max[0, \pi(c, \lambda, k) - f] \quad (24)$$

Here $g(c)$ is the distribution of firm marginal costs. Prospective firms know this but do not know their own marginal cost; to find that out they have to pay a sunk entry cost f_e in order to make a draw from the distribution. This sunk cost must therefore equal the expected return to sampling from $g(c)$, which with rational expectations and risk neutrality equals $\bar{v}(\lambda, k)$ as given in (24).

We can solve for the effects of globalization on every firm's profits by totally differentiating (23):

$$\frac{d \log \pi(c)}{d \log k} = \underbrace{\frac{\partial \log \pi(c)}{\partial \log k}}_{(M)} + \underbrace{\frac{\partial \log \pi(c)}{\partial \log \lambda} \frac{d \log \lambda}{d \log k}}_{(C)} \quad (25)$$

where the change in the level of competition is determined in turn by (24):

$$\frac{d \log \lambda}{d \log k} = - \frac{\partial \log \bar{v}}{\partial \log k} \bigg/ \frac{\partial \log \bar{v}}{\partial \log \lambda} \quad (26)$$

(25) shows that globalization is a two-edged sword from the perspective of firms. On the one hand it has a market-size effect, given by (M), which tends to *raise* each firm's profits. on the other hand it has a competition effect, given by (C): because *all* firms' profits rise at the initial level of competition, the latter must increase to ensure that the expected value of a firm remains equal to the fixed cost of entry; this in turn tends to *reduce* each firm's profits.

The net outcome is indeterminate in general. However, with additive separability, equation

¹⁹Pollak (1972) introduced the case of generalized additive separability, where inverse demand for each good depends on its own quantity consumed and on a single aggregate λ : $p(y, \lambda, k)$. Many of our results hold in this case, though their interpretation is simpler when we specialize to additive separability, which implies that λ equals the marginal utility of income.

(25) takes a particularly simple form:

$$\frac{d \log \pi(c)}{d \log k} = \underbrace{1}_{(M)} - \underbrace{\frac{\varepsilon(c)}{\bar{\varepsilon}}}_{(C)} \quad (27)$$

Now the market expansion effect is just one: other things equal, a 1% increase in market size leads to a 1% increase in the profits of each firm. As for the competition effect, it depends on the ratio of the demand elasticity the firm faces to $\bar{\varepsilon}$:

$$\bar{\varepsilon} \equiv \int_{\underline{c}}^{\bar{c}} \frac{v(c, \lambda, k)}{\bar{v}(\lambda, k)} \varepsilon(c) g(c) dc \quad (28)$$

Here $\bar{\varepsilon}$ is the firm-value-weighted average elasticity of demand across all firms, which we can interpret as the elasticity faced by the average firm. Higher competition reduces an individual firm's profits by $\varepsilon(c)$, while it reduces the average value of a firm by $\bar{\varepsilon}$. Recalling from (13) that elasticity falls with output if and only if demand is subconvex, we see from (27) that the ambiguity in the general case is resolved in a very simple way when preferences are additively separable: globalization raises profits of larger firms and reduces those of smaller firms if and only if demands are subconvex.

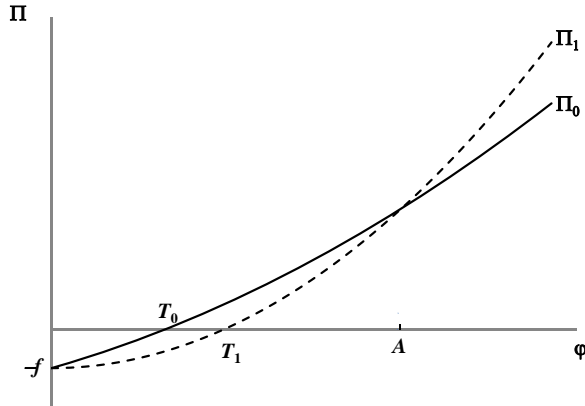


Figure 8: The Matthew Effect of Globalization

Figure 8 illustrates these effects in the case of subconvexity. The curve labelled Π_0 shows the initial profit profile across firms of different productivity, while that labelled Π_1 shows the profit profile after globalization, assuming that demand is strictly subconvex. The productivity level denoted T_0 is that corresponding to threshold firms in the initial equilibrium: they make zero profits and so are indifferent between producing or not; while that denoted A corresponds to the average firm that faces the demand elasticity $\bar{\varepsilon}$. As can be seen, in line with equation (28), the direct, market-expansion, effect dominates for firms of above-average size, $\varepsilon(c) < \bar{\varepsilon}$, so they expand; the indirect, competition, effect dominates for firms below average size, $\varepsilon(c) > \bar{\varepsilon}$, so they contract; and threshold firms cease to be profitable and exit. Thus with subconvex demand, globalization has a clear Matthew Effect. An important consequence of these responses is that the average productivity of active exporters rises. Finally, recall that all these effects are reversed with superconvex demands. Moreover, the CES case is very special: the direct and indirect effects of globalization exactly offset each other, so the profit profile in Figure 8 is unaffected.

6 Matching Productivity and Sales Distributions

The final substantive topic we want to discuss goes to the heart of firm heterogeneity in models of monopolistic competition. It concerns the relationship between the shape of the underlying distribution assumed for firm productivities, and the shape of the predicted distribution of firm outcomes, whether the level or the rate of growth of output, sales or employment.²⁰

There is a relatively small number of results of this kind. First is the canonical result in the field, due to Helpman et al. (2004) and Chaney (2008): in the Melitz (2003) model of heterogeneous firms, if firm productivities follow a Pareto distribution and demands are CES, then firm sales must also follow a Pareto distribution. Second, going beyond the canonical result, Head et al. (2014) retain the assumption that demands are CES, and show that if

²⁰This section draws on Mrázová et al. (2015).

firm productivities follow a Lognormal distribution instead of a Pareto then firm sales must also follow a Lognormal distribution. This result is important as there is abundant evidence that the distribution of firm sales is approximately Pareto for the “long tail” of large firms, but is more closely approximated by a Lognormal distribution for the whole population of firms, especially smaller ones: see, for example, Bee and Schiavo (2018).²¹ A third result concerns the distribution of the rate of growth of sales rather than their level. The classic result of Gibrat (1931), source of the eponymous Law, predicts that a firm’s rate of growth should be independent of its size. Empirical evidence is broadly in favour of this, at least for larger and older firms: see, for example, Haltiwanger et al. (2013).²² While most theoretical studies of Gibrat’s Law have explored it under perfectly competitive assumptions, recent work by Luttmer (2007) and Arkolakis (2010, 2016) shows that it also holds in models of monopolistic competition with heterogeneous firms and CES demands. There are in fact two distinct results of this kind: first, a uniform improvement in all firms’ productivity will raise their growth rate by an equal proportion; and second, more in line with Gibrat’s original result, if firm productivities evolve as a result of idiosyncratic shocks that lead to an asymptotically Lognormal distribution of productivity growth, then the growth rate of sales will also follow a Lognormal distribution.

All these results share two features. First, they assume CES demands, and so imply that all firms have the same markups which do not vary over time. Second, they exhibit a property that Mrázová et al. (2015) call “self-reflection”: the distribution predicted for the endogenous random variable takes the same form as, or “reflects,” the distribution assumed for the exogenous random variable. Self-reflection is an attractive property even in the cross-section: for example, it is desirable to understand in what circumstances the analytically tractable Pareto distribution can be used. It is particularly attractive for the evolution of

²¹Some authors have explored models of heterogeneous firms with other productivity distributions, including generalizations or combinations of the Lognormal and Pareto such as mixtures of thin- and fat-tailed Pareto or piecewise Lognormal-Pareto distributions, although these are typically not analytically tractable. See, for example, Luttmer (2007), Eaton et al. (2011), Edmonds et al. (2012), and Nigai (2017).

²²Sutton (1997) points out that different authors have interpreted Gibrat’s Law as applying to different firm aggregates, whether sales, employment or assets. Here we focus on sales.

firm size over time: there is no theory that is consistent with Gibrat’s Law in monopolistic competition if self-reflection from the distribution of productivity shocks to the distribution of firm growth rates does not hold. Hence, it is very desirable to know whether there are demand functions other than the CES that are consistent with this property. Putting this differently, is Gibrat’s Law consistent with demand functions that allow for variable markups within and between firms?

Mrázová et al. (2015) show that the answer to this question is “yes” by deriving conditions for self-reflection of two distributions, where the firm-level variables are linked by a functional relationship. Assume that the distributions of both variables are members of the same family, drawn from a wide class of distributions that includes Pareto, Lognormal and Fréchet. It then follows that the relationship between the two variables must take a simple power-law form. Moreover, these three conditions are necessary and sufficient for each other in the sense that any two imply the third. In the present context, the variables of interest are productivity and sales revenue, so the requirement for self-reflection is that sales revenue is a power-law function of productivity. Productivity is the inverse of marginal cost, so this is equivalent to a constant elasticity of sales revenue with respect to marginal cost: $d \log r / d \log c = -\kappa$. Since marginal cost equals marginal revenue in equilibrium, this condition can also be expressed as a “Constant Revenue Elasticity of Marginal Revenue,” whence the acronym “CREMR”.

The final step is to derive the demand functions that exhibit the CREMR property. It is straightforward to express the elasticity in terms of ε and ρ , using the fact that $r(x) = xp(x)$ and the expression for the elasticity of output with respect to marginal cost from (20):

$$\frac{d \log r}{d \log c} = \frac{\varepsilon - 1}{\varepsilon} \frac{d \log x}{d \log c} = -\frac{(\varepsilon - 1)^2}{\varepsilon(2 - \rho)} \quad (29)$$

Mrázová et al. (2015) show that the demand functions which ensure that this is constant are given by the following:

$$p(x) = \frac{\beta}{x} (x + \gamma)^{\frac{\kappa}{\kappa+1}}, \quad \kappa > 0, \quad x + \gamma(\kappa + 1) > 0, \quad \beta > 0 \quad (30)$$

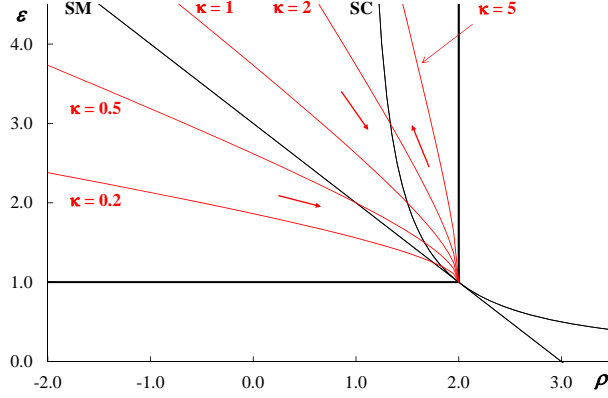


Figure 9: CREMR Demand Manifolds

Once again, this is a generalization of CES demands, and reduces to them when γ is zero. The results of Mrázová et al. (2015) imply that it is the only demand function that is consistent with self-reflection of the productivity and sales distributions in monopolistic competition. Hence it is also the only demand function that is consistent with Gibrat's Law, either in the sense that a uniform shock to all firms' productivities generates a uniform change in their sales; or in the sense that idiosyncratic productivity shocks cumulate to give a Lognormal distribution of sales, with firm growth independent of firm size in the long run. Mrázová et al. (2015) show that it also has other desirable properties: it implies distributions for both sales and markup that are theoretically plausible and empirically realistic, in that they provide a better fit for these distributions than other widely-used demand functions. The CREMR demand manifold is given by the following:

$$\bar{\rho}(\varepsilon) = 2 - \frac{1}{\kappa} \frac{(\varepsilon - 1)^2}{\varepsilon} \quad (31)$$

It is illustrated in Figure 9. Once again, these are very different from the demand manifolds for standard demand functions, as shown in Figure 3.

	Response	Demand Function $p(x; \beta, \gamma, \kappa)$	Demand Manifold $\bar{\rho}(\varepsilon; \kappa)$
Bulow-Pfleiderer	$\frac{dp}{dc} = \kappa$	$\beta \left(x^{\frac{1-\kappa}{\kappa}} + \gamma \right)$	$2 - \frac{1}{\kappa}$
CPPT	$\frac{d \log p}{d \log c} = \kappa$	$\frac{\beta}{x} \left(x^{\frac{\kappa-1}{\kappa}} + \gamma \right)^{\frac{\kappa}{\kappa-1}}$	$2 - \frac{1}{\kappa} \frac{\varepsilon-1}{\varepsilon}$
CEMR/Inverse PIGL	$\frac{d \log x}{d \log c} = -\kappa$	$\frac{\beta}{x} \left(x^{\frac{\kappa-1}{\kappa}} + \gamma \right)$	$2 - \frac{1}{\kappa}(\varepsilon - 1)$
CREMR	$\frac{d \log r}{d \log c} = -\kappa$	$\frac{\beta}{x} (x + \gamma)^{\frac{\kappa}{\kappa+1}}$	$2 - \frac{1}{\kappa} \frac{(\varepsilon-1)^2}{\varepsilon}$

Table 1: Constant-Response Demand Functions

7 Constant Response Demand Functions

Table 1 summarizes the demand functions that have been discussed in the paper. Each of them is a “constant-response demand function,” which is necessary and sufficient for an exogenous change in marginal cost to induce an equal absolute or relative response in some firm-level variable, whether price, output or sales. Each of them nests CES demands, and reduces to the CES case when the parameter γ is zero.²³ Putting this differently, only CES demands imply constant responses to costs of all these variables, and each of the four demand functions in Table 1 allows for variable markups if and only if γ is non-zero.

The explicit expressions for the different demand functions given in the third column of the table are suggestive. Yet even though they are written in ways that makes them as comparable as possible, their behavior relative to each other is not immediately obvious. It is easier to compare their properties using their manifolds, as shown in the fourth column, and as illustrated in Figures 4, 7, and 9. While each demand function depends on three parameters, each of the manifolds depends only on κ , the parameter that measures the

²³The CES special case also imposes restrictions on κ . Setting ε and ρ in the final column of Table 1 equal to their CES values (given in (10)), the implied values of κ with CES demands are: $\sigma/(\sigma - 1)$ for Bulow-Pfleiderer, one for CPPT, σ for CEMR, and $\sigma - 1$ for CREMR, where σ is the CES elasticity of substitution and demand.

response of the variable in question to the change in marginal cost. Thus, as we saw in Section 3.4, each manifold is a locus along which the relevant response to an increase in marginal cost is constant.

Except for the Bulow-Pfleiderer case, the other three demand functions in Table 1 have manifolds that are mostly quite different from those of the better-known and more widely-used demand functions shown in Figure 3. If we concentrate on the subconvex case, we can see that these demand functions allow for much more concavity than most widely-used demands. This is especially true for firms that are small and relatively unresponsive to shocks (i.e., for high values of ε and low values of κ). Mrázová et al. (2015) point out that this ability to allow small firms to face very concave demand may be why CREMR demands give a better fit to both sales and markups than other demand systems. By contrast, for larger firms (i.e., for lower values of ε), Bulow-Pfleiderer demands have constant convexity by construction, whereas the others become steadily more convex: CPPT and CEMR asymptote to Cobb-Douglas, and CREMR to a general CES with elasticity of demand σ equal to $\kappa + 1$.

While the manifolds in Table 1 are invariant with respect to the other two demand parameters β and γ , this does not imply that these have no role. Consider first the non-CES parameter γ : its sign is a key determinant of whether a given demand function is sub- or superconvex. More specifically, if, for a given value of κ , a demand function can be either sub- or superconvex, then it is superconvex if and only if γ is positive. Table 3 in Appendix C illustrates the possible configurations in detail. As for the parameter β , its role in these demand functions is similar to the one it plays in the CES itself: it affects only the level of demand, and is taken as given by firms, but is determined endogenously in general equilibrium. We saw in Section 5 how this gives rise to competition effects. In general, any exogenous shock to an industry equilibrium will have both a direct effect on individual firms and an indirect competition effect working through the free entry condition and the demand function. As we saw in Section 5, the competition effect operates in a particularly simple way when demand is generated by additively separable preferences, since it affects only β ,

which is inversely proportional to the marginal utility of income. Finally, a further role played by the β parameter arises if we want to move from demands to preferences. All of the demand functions in Table 1 can be integrated to get the corresponding preference system, which allows us to solve for the endogenous value of β . This is relatively straightforward when cross-price effects take a simple form, as in the case of additive separability or Kimball preferences. (Mrázová et al. (2015) discuss this further for CREMR demands.) Hence, these demand functions can be used in explicit calculations of the general-equilibrium welfare effects of policy and other shocks.

8 Conclusion

In this paper we have reviewed and extended recent work by ourselves and others on monopolistic competition in benchmark models of international trade. We have not discussed trade under oligopoly, which provides an alternative explanation for the evidence of variable markups cited in Section 2.2. (See, for example, Neary (2003, 2016) and Atkeson and Burstein (2008).) Nor have we discussed the extensive recent literature in trade that uses calibration methods to quantify the effects of exogenous shocks, often in CES-based frameworks. (See Costinot and Rodríguez-Clare (2014) for an overview.) Instead, we have concentrated on the implications of monopolistically competitive models for a number of important results, including pass-through, selection effects, competition effects, and the size distribution of firms.

The fact that the monopolistically competitive paradigm has become the dominant one in international trade may seem surprising from an IO perspective. However, there are good reasons for this. The availability of large data sets on exporting firms, and the desire to take account of general-equilibrium feedbacks between goods and factor markets, make a “one-size-fits-all” approach to the choice of market structure very convenient. Moreover, as we have shown, relaxing the assumption of CES preferences in monopolistic competition makes

its view of firms more realistic and its predictions more flexible. In this context, a unifying principle that we have highlighted is the class of constant-response demand functions that provide a parsimonious suite of alternatives to the CES case. Each of these functions adds a single parameter to the CES, so avoiding the counter-factual prediction of constant markups, while preserving one or other of its key properties.

Of course, while leaving the comfort zone of CES goes part of the way to a more realistic view of imperfectly competitive markets, it raises new problems. To paraphrase Tolstoy (1878), CES demand functions are all alike, every non-CES demand function is non-CES in its own way.²⁴ The CES benchmark is sure to retain a central place in this field, both because of its tractability and also because it alone exhibits all the constant-response properties that we have discussed.

In conclusion, we have already noted some ironies in the selective way in which international trade has borrowed from industrial organization in how it models firms in open economies. In particular, Dixit and Stiglitz proposed using additively separable preferences as a way of modelling product differentiation, and viewed CES preferences as just a special illustrative case; whereas it is the latter that has become the workhorse model for exploring a huge range of issues in trade. There is a further irony. The substantive issue they addressed was the efficiency of the market equilibrium, and in particular whether an unregulated monopolistically competitive market would lead to too many or too few varieties. This issue has not been addressed much in trade, but in recent years there has been a revival of interest in it: see for example Feenstra and Kee (2008) and Dhingra and Morrow (2019). In ongoing work with Mathieu Parenti (see, for example, Section 5 of Mrázová et al. (2015)) we show how the tools from Section 6 above can be used to throw light on this issue.

²⁴From the opening sentence of *Anna Karenina*: “Happy families are all alike; every unhappy family is unhappy in its own way.”

Appendices

A Constant Response of Profits Demand Functions

We would like to characterize the demand functions with the property that operating profits have a constant proportional response to marginal cost: $\frac{d \log \pi}{d \log c} = \kappa$. We know that this holds in the CES case, where $\pi = Bc^{1-\sigma}$.²⁵ To find the necessary condition for this property, note that, since $\pi = (p - c)x$, we have in general:

$$d \log \pi = d \log (p - c) + d \log x \quad (32)$$

Equation (20) gives the proportional response of output to changes in marginal cost, $d \log x / d \log c$. The next step is to consider $d \log (p - c)$. Consider first $d(p - c)$. From the firm's first-order condition this equals $d(-xp') = -p'dx - xp''dx = -(p' + xp'')dx$. Hence:

$$d \log (p - c) = \frac{d(-xp')}{-xp'} = \frac{(p' + xp'')dx}{xp'} = \frac{p' + xp''}{p'} d \log x = (1 - \rho) d \log x \quad (33)$$

Collecting terms gives, from (20) and (32):

$$\frac{d \log \pi}{d \log c} = \frac{d \log (p - c)}{d \log c} + \frac{d \log x}{d \log c} = (2 - \rho) \frac{d \log x}{d \log c} = -(\varepsilon - 1) \quad (34)$$

Clearly the CES case is the only one that yields a constant value of this. Hence CES demands are necessary and sufficient for a constant proportional response of operating profits to marginal cost.

	Demand Function	Demand Manifold
CARA	$x(p) = \gamma + \delta \log p$	$\bar{\rho}(\varepsilon) = \frac{1}{\varepsilon}$
Stone-Geary/LES	$p(x) = \frac{\beta}{\gamma x + 1}$	$\bar{\rho}(\varepsilon) = \frac{2}{\varepsilon}$
Translog/AI	$x(p) = \frac{1}{p} (\gamma + \delta \log p)$	$\bar{\rho}(\varepsilon) = \frac{3\varepsilon - 1}{\varepsilon^2}$

Table 2: Some Common Demand Functions and Their Manifolds

B Manifolds for Some Common Demand Functions

Table 2 gives some common demand functions and their manifolds, as illustrated in Figure 3. Further details are given in Mrázová and Neary (2017). As elsewhere in the paper, the demand functions are written from the perspective of an individual firm in monopolistic competition. They also depend on aggregate expenditure and on indices of the prices of all goods, including the good in question; these are taken as given by the firm, since it is of measure zero, and are subsumed into the demand parameters. (See Dixit and Stiglitz (1993) for further discussion.) All three manifolds are invariant with respect to these other variables. To take a specific example, Stone-Geary demands are more familiar when written as: $x(i) = \alpha(i) + (p(i))^{-1} (I - \int_{i' \in \Omega} p(i') \alpha(i') di')$, which, when $\gamma = (-\alpha(i))^{-1}$ and $\beta = \gamma (I - \int p(i') \alpha(i') di')$, is equivalent to the expression in Table 2. The latter expression is the special case of the CPPT demands from (11) when $\kappa = 1/2$.

C The Role of γ in Determining Superconvexity

Table 3 illustrates schematically the conditions under which each of the non-CES constant-response demand functions is either sub- or superconvex, depending on the values of the γ and κ parameters. In four of the eight cases, the demand functions can only be either sub-

²⁵Here B is exogenous from the firm's perspective, but it depends on the price index and on aggregate income, and so is endogenous in general equilibrium. See, for example, Melitz (2003), equation (4).

$\gamma < 0$		$\gamma > 0$	
$\kappa < 1$	BP:	n/a	BP: Subconvex
	CPPT:	n/a	CPPT: Subconvex
	CEMR:	n/a	CEMR: Subconvex
	CREMR:	Subconvex	CREMR: Superconvex
$\kappa > 1$	BP:	Subconvex	BP: Superconvex
	CPPT:	n/a	CPPT: Superconvex
	CEMR:	Subconvex	CEMR: Superconvex
	CREMR:	Subconvex	CREMR: Superconvex

Table 3: Possible Configurations of Sub- and Superconvexity

or superconvex, irrespective of the sign of γ . The remaining cases are represented in Figures 4, 7, and 9 by manifolds with two branches, one in the subconvex region and the other in the superconvex one. It is in these cases that the sign of γ is crucial, with a positive value corresponding to the superconvex branch.

References

- ARKOLAKIS, C. (2010): “Market Penetration Costs and the New Consumers Margin in International Trade: Appendix,” web appendix, Yale University.
- (2016): “A Unified Theory of Firm Selection and Growth,” *Quarterly Journal of Economics*, 131, 89–155.
- ATKESON, A. AND A. BURSTEIN (2008): “Pricing-to-Market, Trade Costs, and International Relative Prices,” *American Economic Review*, 98, 1998–2031.
- BACHE, P. A. AND A. LAUGESSEN (2015): “Monotone Comparative Statics for the Industry Composition,” Working Paper, Aarhus University.
- BEE, M. AND S. SCHIAVO (2018): “Powerless: Gains from Trade when Firm Productivity is not Pareto Distributed,” *Review of World Economics*, 154, 15–45.
- BEHRENS, K. AND Y. MURATA (2007): “General Equilibrium Models of Monopolistic Competition: A New Approach,” *Journal of Economic Theory*, 136, 776–787.
- BERTOLETTI, P. AND P. EPIFANI (2014): “Monopolistic Competition: CES Redux?” *Journal of International Economics*, 93, 227–238.
- BRANDER, J. A. (1995): “Strategic Trade Policy,” in G.M. Grossman and K. Rogoff (eds.): *Handbook of International Economics, Volume 3*, Amsterdam: Elsevier.
- BULOW, J. I. AND P. PFLEIDERER (1983): “A Note on the Effect of Cost Changes on Prices,” *Journal of Political Economy*, 91, 182–185.
- CABRAL, L. M. (2018): “Standing on the Shoulders of Dwarfs: Dominant Firms and Innovation Incentives,” CEPR Discussion Paper No. 13115.
- CHAMBERLIN, E. H. (1933): *The Theory of Monopolistic Competition: A Re-Orientation of the Theory of Value*, Cambridge, Mass: Harvard University Press.

- CHANEY, T. (2008): “Distorted Gravity: The Intensive and Extensive Margins of International Trade,” *American Economic Review*, 98, 1707–1721.
- COSTINOT, A. AND A. RODRÍGUEZ-CLARE (2014): “Trade Theory with Numbers: Quantifying the Consequences of Globalization,” in G. Gopinath, E. Helpman and K. Rogoff (eds.): *Handbook of International Economics*, Elsevier, 197–261.
- DE LOECKER, J., P. K. GOLDBERG, A. K. KHANDELWAL, AND N. PAVCNİK (2016): “Prices, Markups and Trade Reform,” *Econometrica*, 84, 445–510.
- DEATON, A. AND J. MUELLBAUER (1980): “An Almost Ideal Demand System,” *American Economic Review*, 70, 312–326.
- DHINGRA, S. AND J. MORROW (2019): “Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity,” *Journal of Political Economy*, 127, 196–232.
- DIXIT, A. AND V. NORMAN (1980): *Theory of International Trade*, Cambridge University Press.
- DIXIT, A. K. (2004): “Some Reflections on Theories and Applications of Monopolistic Competition,” in S. Brakman and B.J. Heijdra (eds): *The Monopolistic Competition Revolution in Retrospect*, Cambridge: Cambridge University Press, 123–133.
- DIXIT, A. K. AND J. E. STIGLITZ (1977): “Monopolistic Competition and Optimum Product Diversity,” *American Economic Review*, 67, 297–308.
- (1993): “Monopolistic Competition and Optimum Product Diversity: Reply [to Yang and Heijdra],” *American Economic Review*, 83, 302–304.
- EATON, J., S. KORTUM, AND F. KRAMARZ (2011): “An Anatomy of International Trade: Evidence from French Firms,” *Econometrica*, 79, 1453–1498.

- ECKEL, C., L. IACOVONE, B. JAVORCIK, AND J. P. NEARY (2015): “Multi-Product Firms at Home and Away: Cost- versus Quality-Based Competence,” *Journal of International Economics*, 95, 216–232.
- ECKEL, C. AND J. P. NEARY (2010): “Multi-Product Firms and Flexible Manufacturing in the Global Economy,” *Review of Economic Studies*, 77, 188–217.
- FEENSTRA, R. AND H. L. KEE (2008): “Export Variety and Country Productivity: Estimating the Monopolistic Competition Model with Endogenous Productivity,” *Journal of International Economics*, 74, 500–518.
- FEENSTRA, R. C. AND D. E. WEINSTEIN (2017): “Globalization, Markups, and U.S. Welfare,” *Journal of Political Economy*, 125, 1040–1074.
- GIBRAT, R. (1931): *Les inégalités économiques: applications d’une loi nouvelle, la loi de l’effet proportionnel [Economic Inequalities: Applications of a New Law, the Law of Proportionate Effect]*, Librairie du Recueil Sirey.
- GOPINATH, G. AND O. ITSKHOKI (2010): “Frequency of price adjustment and pass-through,” *Quarterly Journal of Economics*, 125, 675–727.
- HALTIWANGER, J., R. S. JARMIN, AND J. MIRANDA (2013): “Who Creates Jobs? Small versus Large versus Young,” *Review of Economics and Statistics*, 95, 347–361.
- HEAD, K., T. MAYER, AND M. THOENIG (2014): “Welfare and Trade Without Pareto,” *American Economic Review, Papers and Proceedings*, 104, 310–316.
- HELPMAN, E., M. J. MELITZ, AND S. R. YEAPLE (2004): “Export Versus FDI with Heterogeneous Firms,” *American Economic Review*, 94, 300–316.
- KRUGMAN, P. (1980): “Scale Economies, Product Differentiation, and the Pattern of Trade,” *American Economic Review*, 70, 950–959.

- KRUGMAN, P. R. (1979): “Increasing Returns, Monopolistic Competition, and International Trade,” *Journal of International Economics*, 9, 469–479.
- LAMORGESE, A., A. LINARELLO, AND F. WARZYNSKI (2014): “Free Trade Agreements and Firm-Product Markups in Chilean Manufacturing,” Tech. rep., Aarhus University.
- LUTTMER, E. G. (2007): “Selection, Growth, and the Size Distribution of Firms,” *Quarterly Journal of Economics*, 122, 1103–1144.
- MARSHALL, A. (1920): *Principles of Economics, An Introductory Volume*, Eighth Edition, London: Macmillan.
- MELITZ, M. J. (2003): “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, 71, 1695–1725.
- MELITZ, M. J. AND G. I. OTTAVIANO (2008): “Market Size, Trade, and Productivity,” *Review of Economic Studies*, 75, 295–316.
- MILGROM, P. AND J. ROBERTS (1990): “The Economics of Modern Manufacturing: Technology, Strategy, and Organization,” *American Economic Review*, 80, 511–528.
- MRÁZOVÁ, M. (2011): “Trade Agreements when Profits Matter,” Discussion Paper, Department of Economics, University of Surrey.
- MRÁZOVÁ, M. (2015): “Regionalism as a Path to Free Trade,” mimeo., University of Geneva.
- MRÁZOVÁ, M. AND J. P. NEARY (2014): “Together at Last: Trade Costs, Demand Structure, and Welfare,” *American Economic Review, Papers and Proceedings*, 104, 298–303.
- MRÁZOVÁ, M. AND J. P. NEARY (2017): “Not So Demanding: Demand Structure and Firm Behavior,” *American Economic Review*, 107, 3835–3874.

- (2019): “Selection Effects with Heterogeneous Firms,” *Journal of the European Economic Association*, forthcoming.
- MRÁZOVÁ, M., J. P. NEARY, AND M. PARENTI (2015): “Sales and Markup Dispersion: Theory and Empirics,” Discussion Paper No. 774, Department of Economics, University of Oxford; revised May 2019.
- MRÁZOVÁ, M., D. VINES, AND B. ZISSIMOS (2013): “Is the GATT/WTO’s Article XXIV bad?” *Journal of International Economics*, 89, 216–232.
- MUELLBAUER, J. (1975): “Aggregation, Income Distribution and Consumer Demand,” *Review of Economic Studies*, 42, 525–543.
- NEARY, J. P. (2003a): “Globalization and Market Structure,” *Journal of the European Economic Association*, 1, 245–271.
- (2003b): “Monopolistic Competition and International Trade Theory,” in S. Brakman and B.J. Heijdra (eds.): *The Monopolistic Competition Revolution in Retrospect*, Cambridge University Press, 159–184.
- (2007): “Cross-Border Mergers as Instruments of Comparative Advantage,” *Review of Economic Studies*, 74, 1229–1257.
- (2010): “Two and a Half Theories of Trade,” *The World Economy*, 33, 1–19.
- (2016): “International Trade in General Oligopolistic Equilibrium,” *Review of International Economics*, 24, 669–698.
- NIGAI, S. (2017): “A Tale of Two Tails: Productivity Distribution and the Gains from Trade,” *Journal of International Economics*, 104, 44–62.
- NOCKE, V. AND N. SCHUTZ (2018): “Oligopoly in International Trade: An Aggregative Games Approach,” presented to EARIE 2018, Athens.

- PARENTI, M. (2018): “Large and Small Firms in a Global Market: David vs. Goliath,” *Journal of International Economics*, 110, 103–118.
- POLLAK, R. A. (1972): “Generalized Separability,” *Econometrica*, 40, 431–453.
- SHIMOMURA, K.-I. AND J.-F. THISSE (2012): “Competition Among the Big and the Small,” *RAND Journal of Economics*, 43, 329–347.
- SIMONOVSKA, I. (2015): “Income Differences and Prices of Tradables: Insights from an Online Retailer,” *Review of Economic Studies*, 82, 1612–1656.
- STIGLITZ, J. E. (2004): “Reflections on the State of the Theory of Monopolistic Competition,” in S. Brakman and B.J. Heijdra (eds): *The Monopolistic Competition Revolution in Retrospect*, Cambridge: Cambridge University Press, 134–148.
- SUTTON, J. (1997): “Gibrat’s Legacy,” *Journal of Economic Literature*, 35, 40–59.
- TOLSTOY, L. (1878): *Anna Karenina*, translated by Rosemary Edmonds, Harmondsworth, Middlesex: Penguin Classics, 1954.
- VIVES, X. (1990): “Nash Equilibrium with Strategic Complementarities,” *Journal of Mathematical Economics*, 19, 305–321.
- WEYL, E. G. AND M. FABINGER (2013): “Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition,” *Journal of Political Economy*, 121, 528–583.
- ZHELOBODKO, E., S. KOKOVIN, M. PARENTI, AND J.-F. THISSE (2012): “Monopolistic Competition: Beyond the Constant Elasticity of Substitution,” *Econometrica*, 80, 2765–2784.