Heterogeneous Firms and Trade *

Marc J. Melitz
Harvard University, NBER and CEPR

Stephen J. Redding
Princeton University, NBER and CEPR

Preliminary Draft

August 19, 2012

Abstract

This paper reviews the new approach to international trade based on firm heterogeneity in differentiated product markets. This approach accounts for a variety of features of firm and plant data that are not well explained by traditional trade theory, such as the non-random selection of firms into export markets and within-industry reallocations of resources across firms following trade liberalization. This approach also highlights new mechanisms through which the economy is affected by trade liberalization, including endogenous increases in average industry and firm productivity.

J.E.L. CLASSIFICATION: F10, F12, F14
KEYWORDS: firm heterogeneity, international trade, productivity

*We are grateful to Keith Head, Elhanan Helpman and Thierry Mayer for helpful comments and suggestions and Cheng Chen for research assistance. The authors are responsible for any remaining limitations.

Contents

1 Introduction

Theoretical research in international trade increasingly emphasizes firm-level decisions in understanding the causes and consequences of aggregate trade. Motivated by empirical findings using micro data on plants and firms, this theoretical literature emphasizes heterogeneity in productivity, size and other characteristics even within narrowly-defined industries. This heterogeneity is systematically related to trade participation, with exporters larger and more productive than non-exporters even prior to entering export markets. Trade liberalization leads to within-industry reallocations of resources, which raise average industry productivity, as low-productivity firms exit and high-productivity firms expand to enter export markets. The increase in firm scale induced by export market entry enhances the return to complementary productivity-enhancing investments in technology adoption and innovation, with the result that trade liberalization also raises firm productivity.

This view of international trade is quite different from the perspective of traditional theories of trade. According to those traditional theories, comparative advantage drives inter-industry trade in dissimilar products, which induces reallocations of resources between industries. These in turn change relative factor demands and hence relative factor prices. In contrast, the new view of international trade emphasizes product variety, increasing returns to scale and firm heterogeneity, which generate intra-industry trade in similar products, reallocations of resources across firms within industries, and innovations in the organization of production within firms. This approach not only accounts for features of disaggregated data on plants and firms that do not have natural explanations within the confines of traditional trade theory, but also highlights new channels through which the economy is affected by trade liberalization and yields new insights for aggregate economic relationships such as the gravity equation.

The remainder of this paper is structured as follows. In Section 2, we introduce a general framework for modeling firm heterogeneity in differentiated product markets. In Section 3, we characterize the model's closed economy equilibrium and in Section 4, we examine the implications of opening to trade. In Section 6, we embed this model of firm heterogeneity within the integrated equilibrium framework of neoclassical trade theory. In Section 7, we relax the assumption of constant elasticity of substitution (CES) preferences to introduce variable mark-ups and examine the effects of market size on the selection of firms into production and exporting. In Section 8, we explore a variety of extensions, in which firm productivity is itself endogenous. Section 9 discusses factor markets and the income distributional consequences of trade liberalization. Section 10 concludes.

2 General Setup

We begin by outlining a general framework for modeling firm heterogeneity.[1] Throughout this chapter, we rely on models of monopolistic competition that emphasize product differentiation and increasing returns to scale at the level of the firm. Although this framework provides a tractable platform for analyzing a host of firm decisions in general equilibrium, it neglects strategic interactions between firms. Bernard, Eaton, Jensen and Kortum (2003) develop a heterogeneous firm framework which features head to head competition between firms, while Neary (2010) surveys the literature on oligopoly and trade in general equilibrium. In our monopolistic competition framework, all interactions between firms operate through market indices such as the mass of competing firms, and statistics of the price distribution. We begin by developing the industry equilibrium with heterogeneous firms, before embedding the sectors in general equilibrium. We start with a closed economy and then examine the implications of opening to international trade. To highlight the implications of firm heterogeneity as starkly as possible, we begin by considering a static (one-period) model, before turning to consider dynamics in a later section.[2]

Preferences

Consumer preferences are defined over the consumption of a number of sectors $j \in \{0, 1, \ldots, J\}$:

$$U = \sum_{j=0}^{J} \beta_j \log Q_j, \qquad \sum_{j=0}^{J} \beta_j = 1, \ \beta_j \geq 0. \tag{1}$$

Sector $j = 0$ is a homogeneous good produced with a unit labor requirement that is used as numeraire. In some cases, we will require that β_0 is large enough that all countries produce the good in the open economy equilibrium. Within each of the remaining $j \geq 1$ sectors, there is a continuum of horizontally differentiated varieties, and preferences are assumed to take the Constant

[1] An accompanying web appendix contains the technical derivations of results reported in the chapter.

[2] For another review of the theoretical literature on firm heterogeneity and trade see Redding (2011), while Bernard et al. (2012) survey the empirical evidence.

Elasticity of Substitution (CES) or Dixit and Stiglitz (1977) form:

$$Q_j = \left[\int_{\omega \in \Omega_j} q_j(\omega)^{(\sigma_j - 1)/\sigma_j} d\omega \right]^{\sigma_j/(\sigma_j - 1)}, \qquad \sigma_j > 1,\ j \geq 1. \tag{2}$$

This representation of consumer preferences, in which varieties enter utility symmetrically, implicitly imposes a choice of units in which to measure the quantity of each variety. There is no necessary relationship between this normalization and the units in which physical quantities of output are measured for each product in the data. Mapping physical quantities of output back to utility requires taking a stand on the relative weight of products in utility, which depends (among other things) on product quality.[3]

Using Y to denote aggregate income, the Cobb-Douglas upper tier of utility implies that consumers spend $X_j = \beta_j Y$ on sector j. The demand for each differentiated variety within sector j is given by:

$$q_j(\omega) = A_j p_j(\omega)^{-\sigma_j}, \qquad A_j = X_j P_j^{\sigma_j - 1},$$

where P_j is the price index dual to (2):

$$P_j = \left[\int_{\omega \in \Omega_j} p(\omega)^{1-\sigma_j} \right]^{1/(1-\sigma_j)}.$$

A_j is an index of market demand; it proportionally scales every firm's residual demand. This market demand index, in turn, is determined by sector spending and a statistic of the price distribution (the CES price index).

Technology

Varieties are produced using a composite factor of production L_j with unit cost w_j in sector j. This composite factor, for example, can be a Cobb-Douglas function of skilled labor (S) and unskilled labor (U): $L_j = \bar{\eta}_j S_j^{\eta_j} U_j^{1-\eta_j}$, where $\bar{\eta}_j$ is a constant such that unit cost is $w_j = w_S^{\eta_j} w_U^{1-\eta_j}$, where w_S and w_U are the skilled and unskilled wage. We index the unit cost by sector j, because even if factor prices are equalized across sectors in competitive factor markets, unit costs will still in general differ across sectors because of differences in factor intensity. The composite factor is used

[3] In some cases quality can be directly measured as for wine (see Crozet, Head, and Mayer, 2012) or inferred from input use such as in Kugler and Verhoogen (2012). Alternatively, functional form assumptions can be made about the mapping between physical and utility units as in Baldwin and Harrigan (2011) and Johnson (2012).

with the same aggregation for all productive uses within the industry, including both variable and fixed costs (incurred for overhead production as well as for entry and market access).[4] Thus, we can define a sector's aggregate employment L_j of this composite factor. In our multi-sector setting, the labor supply L_j to each sector is determined endogenously. In several instances where we wish to characterize how labor is allocated across sectors, we will return to a single homogeneous labor factor, in which case the labor supplies L_j can be summed across sectors and set equal to the country's aggregate labor endowment.

Within each industry, each firm chooses to supply a distinct horizontally-differentiated variety. Production of each variety involves a fixed production cost of f_j units of the composite input and a constant marginal cost that is inversely proportional to firm productivity φ. The total amount of the composite input required to produce q_j units of a variety is therefore:

$$l_j = f_j + \frac{q_j}{\varphi}.$$

Since all firms with the same productivity within a given sector behave symmetrically, we index firms within a sector from now onwards by φ alone. The homogeneous numeraire sector is characterized by perfect competition and is produced one for one with the composite factor, so that $w_0 = 1$.

Firm Behavior

We now focus on equilibrium in a given sector and drop the sector j subscript to avoid notational clutter. The market structure is monopolistic competition. Each firm chooses its price to maximize its profits subject to a downward-sloping residual demand curve with constant elasticity σ. From the first-order condition for profit maximization, the equilibrium price for each variety is a constant mark-up over marginal cost:

$$p(\varphi) = \frac{\sigma}{\sigma - 1}\frac{w}{\varphi},$$

which implies an equilibrium firm revenue of:

$$r(\varphi) = Ap(\varphi)^{1-\sigma} = A\left(\frac{\sigma - 1}{\sigma}\right)^{\sigma-1} w^{1-\sigma}\varphi^{\sigma-1},$$

[4] See Flam and Helpman (1987) for an analysis of non-homothetic production technologies where fixed and variable costs can have different factor intensities.

and an equilibrium firm profit of:

$$\pi(\varphi) = \frac{r(\varphi)}{\sigma} - wf = B\varphi^{\sigma-1} - wf, \qquad B = \frac{(\sigma-1)^{\sigma-1}}{\sigma^{\sigma}} w^{1-\sigma} A.$$

Since price is a constant mark-up over marginal cost, higher firm productivity is passed on fully to consumers in the form of a lower price. With elastic demand, these lower prices of more productive firms translate into higher firm revenue. Constant mark-ups and the homothetic production technology (in which the ratio of average cost to marginal cost depends solely on firm output) imply that 'variable' or 'gross' profits are a constant proportion of firm revenue. Thus, the market demand index A proportionally scales both revenues and gross profits.

Firm Performance Measures and Productivity

A key implication of the CES demand structure is that the relative outputs and revenues of firms depend solely on their relative productivities:

$$\frac{q(\varphi_1)}{q(\varphi_2)} = \left(\frac{\varphi_1}{\varphi_2}\right)^{\sigma}, \qquad \frac{r(\varphi_1)}{r(\varphi_2)} = \left(\frac{\varphi_1}{\varphi_2}\right)^{\sigma-1}, \qquad \varphi_1, \varphi_2 > 0.$$

where a higher elasticity of substitution implies greater differences in size and profitability between firms for a given difference in relative productivity.

Empirical measures of firm or plant revenue-based productivity (e.g. based on deflated sales or value-added) are monotonically related to the firm productivity draw φ. Since prices are inversely related to the firm productivity draw φ, revenue per variable input is constant across firms. Revenue-based productivity, however, varies because of the fixed production cost:

$$\frac{r(\varphi)}{l(\varphi)} = \frac{w\sigma}{\sigma-1}\left[1 - \frac{f}{l(\varphi)}\right],$$

where the labor input $l(\varphi)$ increases monotonically with φ. A higher productivity draw increases variable labor input and revenue, with the result that the fixed labor input is spread over more units of revenue.[5]

[5] In section 7, we introduce endogenous markups. This generates another channel for variations in revenue-based productivity. More productive firms set higher markups, which raises their measured revenue-based productivity relative to less productive firms.

Firm Entry and Exit

There is a competitive fringe of potential firms that can enter the sector by paying a sunk entry cost of f_E units of the composite input. Potential entrants face uncertainty about their productivity in the sector. Once the sunk entry cost is paid, a firm draws its productivity φ from a fixed distribution $g(\varphi)$, with cumulative distribution $G(\varphi)$.

After observing its productivity, a firm decides whether to exit the sector or to produce. This decisions yields a survival cutoff productivity φ^* at which a firm makes zero profits:

$$\pi(\varphi^*) = \frac{r(\varphi^*)}{\sigma} - wf = B(\varphi^*)^{\sigma-1} - wf = 0. \tag{3}$$

The relationship between profits and productivity is shown graphically in Figure 1. Firms drawing a productivity $\varphi < \varphi^*$ would make losses if they produced. Therefore these firms exit immediately, receiving $\pi(\varphi) = 0$, and hence making a loss once the sunk entry cost is taken into account. Among these active firms, a subset of them with $\pi(\varphi) > wf_E$ make positive profits net of the sunk entry cost. The remaining firms incur a loss once the sunk entry cost is taken into account. Free entry implies that in equilibrium, this expected measure of *ex-ante* profits (inclusive of the entry cost) must be equal to zero:

$$\int_0^\infty \pi(\varphi) dG(\varphi) = \int_{\varphi^*}^\infty \left[B\varphi^{\sigma-1} - wf\right] dG(\varphi) = wf_E. \tag{4}$$

Heterogeneity in firm productivity generates the systematic differences in firm employment, revenue and profits observed in micro data (see for example Bartelsman and Doms 2000). Selection into production (only firms with productivity $\varphi \geq \varphi^*$ produce) delivers the empirical regularity that exiting firms are less productive than surviving firms (see for example Davis and Haltiwanger 1992 and Dunne, Roberts and Samuelson 1989).

3 Closed Economy Equilibrium

General equilibrium can be characterized by the following variables for each sector: the survival productivity cutoff φ_j^*, the price w_j and supply L_j of the composite labor input, the mass of entrants M_{Ej}, and aggregate expenditure X_j. To determine this equilibrium vector, we use the model's recursive structure, in which the productivity cutoff φ_j^* can be determined independently of the other equilibrium variables.

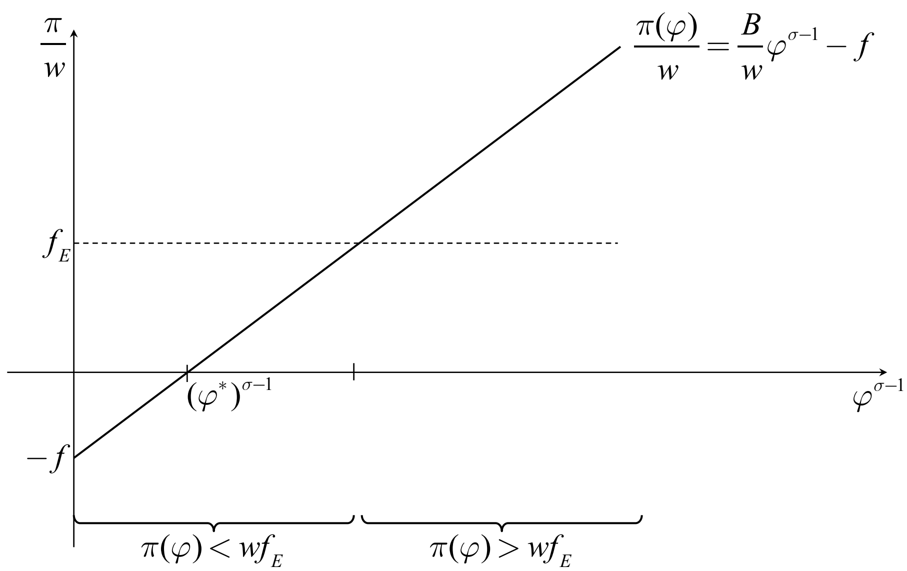

Figure 1: Closed Economy Equilibrium $\{\varphi^*, B/w\}$

Sectoral Equilibrium

Once again, we drop the sector j subscript to streamline notation, and measure all currency denominated variables (prices, profits, revenues) relative to the unit labor cost w in that sector. The zero-profit condition (3) and free entry (4) provide two equations involving only two endogenous variables: the productivity cutoff φ^* and market demand B/w. Combining these two conditions, we obtain a single equation that determines the productivity cutoff:

$$fJ(\varphi^*) = f_E, \qquad J(\varphi^*) = \int_{\varphi^*}^{\infty} \left[\left(\frac{\varphi}{\varphi^*} \right)^{\sigma-1} - 1 \right] dG(\varphi). \tag{5}$$

Since $J(.)$ is monotonically decreasing with $\lim_{\varphi^* \to 0} J(\varphi^*) = \infty$ and $\lim_{\varphi^* \to \infty} J(\varphi^*) = 0$, (5) identifies a unique equilibrium cutoff φ^*. Market demand is then $B/w = f(\varphi^*)^{1-\sigma}$. Using these two equilibrium variables, we can determine the distribution of all firm performance measures (relative to the labor cost w). Productivity φ will be distributed with cdf $G(\varphi)/[1 - G(\varphi^*)]$, and the distribution of prices, profits, revenues, output, and employment will be given by the following functions of firm productivity φ and market demand B/w:

$$\frac{p(\varphi)}{w} = \frac{\sigma}{\sigma-1}\frac{1}{\varphi}, \ \frac{\pi(\varphi)}{w} = \frac{B}{w}\varphi^{\sigma-1} - f, \ \frac{r(\varphi)}{w} = \sigma\left[\frac{\pi(\varphi)}{w} + f\right], \ q(\varphi) = \frac{r(\varphi)}{p(\varphi)}, \ l(\varphi) = \frac{q(\varphi)}{\varphi} + f.$$

Sector aggregates such as expenditures and labor supply thus have no impact on firm selection φ^*, nor on the distribution of any of the firm performance measures. Those sector aggregates will only influence the mass of firms with those characteristics. Before deriving this relationship between sector aggregates and the mass of firms, we describe some important properties of the distribution of firm performance measures.

We start by noting that the free entry condition (4) pins down the average profit (and hence the average revenue) of active firms:

$$\frac{\bar{\pi}}{w} = \frac{f_E}{1 - G(\varphi^*)}, \qquad \frac{\bar{r}}{w} = \sigma \left(\frac{\bar{\pi}}{w} + f \right).$$

Let $\tilde{\varphi}$ be the productivity of the firm earning those profits and revenues. From the free entry condition (4), we can derive $\tilde{\varphi}$ as a function of the cutoff productivity φ^*:

$$\tilde{\varphi}^{\sigma-1} = \int_{\varphi^*}^{\infty} \varphi^{\sigma-1} \frac{dG(\varphi)}{1 - G\left(\varphi^*\right)}.$$

$\tilde{\varphi}$ is a harmonic average of firm productivity φ, weighted by relative output shares $q(\varphi)/q(\tilde{\varphi})$. This productivity average $\tilde{\varphi}$ is also a reference productivity for the aggregate sector consumption index Q and the price index P in the following sense: A hypothetical monopolistic competition equilibrium with M representative firms sharing a common productivity $\tilde{\varphi}$ would induce the same consumption index $Q = M^{\sigma/(\sigma-1)} q(\tilde{\varphi})$ and price index $P = M^{1/(1-\sigma)} p(\tilde{\varphi})$ as M heterogeneous firms with the equilibrium distribution $G(\varphi)/\left[1 - G\left(\varphi^*\right)\right]$. We will also show that given the same labor supply L and expenditures X for the sector, the hypothetical equilibrium with representative firms would also feature the same mass M of active firms as in our current setup with heterogeneous firms.

In our heterogeneous firm setup, the mass M of active firms represents the portion of the mass M_E of entrants that survive. This portion depends on the survival cutoff φ^*: $M = \left[1 - G\left(\varphi^*\right)\right] M_E$. The sector's labor supply L is used both for production by the M active firms, and to cover the entry cost f_E used by all M_E entrants. Since payments to production workers must equal the difference between aggregate sector revenues R and profit Π, we can write the sector's labor market equilibrium as:

$$L = \frac{R - \Pi}{w} + M_E f_E.$$

The free entry condition ensures that aggregate profits Π exactly cover the aggregate entry cost $w M_E f_E$. Thus, aggregate sector revenue is determined by the labor supply: $R/w = L$. In a closed economy this must also be equal to the sector's expenditures X/w.

Since $L = R/w = X/w$ affects neither firm selection (the cutoff φ^*) not average firm sales $\bar{r}/w$, changes in this measure of market size must be reflected one-for-one in the mass of both active firms and entrants. This a very close parallel with Krugman (1980), where firm size is also independent of market size. In fact, a single-sector version of our model would yield the same sector aggregate variables and firm averages (for the firm with productivity $\tilde{\varphi}$) as in Krugman's (1980) model where the representative firms shared the productivity level given by $\tilde{\varphi}$. (The key distinction with our heterogeneous firms model is that the reference productivity level $\tilde{\varphi}$ is endogenously determined.) The result that market size affects neither firm selection nor the distribution of firm size is very specific to our assumption of CES preferences. In section 7, we analyze other preferences that feature a link between market size and both firm selection and the distribution of firm performance measures (size, price, markups, profit).

General Equilibrium

Now that we have characterized equilibrium in each sector j in terms of firm selection (φ_j^*), market demand (B_j/w_j), and the distribution of firm performance measures, we embed the sector in general equilibrium. The simplest way to close the model in general equilibrium is to assume a single factor of production (labor $\bar{L}$) that is mobile across sectors and indexes the size of the economy. Labor mobility ensures that the wage w is the same for all sectors j. If the homogenous numeraire good is produced, we have $w_j = w = 1$. Otherwise, we choose labor as the numeraire so that again $w_j = w = 1$.

With the zero-profit cutoff in each sector (φ_j^*) and the wage (w) already determined, the other elements of the equilibrium vector follow immediately. Aggregate income follows from $Y = w\bar{L}$ and industry revenue and expenditure follow from $R_j = X_j = \beta_j Y = \beta_j w \bar{L}$. The mass of firms in each sector is

$$M_j = \frac{R_j}{\bar{r}_j} = \frac{\beta_j \bar{L}}{\sigma \left[\frac{f_E}{1-G(\varphi^*)} + f \right]}.$$

The results on the efficiency of the market equilibrium from Dixit and Stiglitz (1977) hold in this setting with heterogeneous firms: conditional on an allocation of labor to sector j, the market allocation is constrained efficient. In other words, a social planner using the same entry technology characterized by $G_j(.)$ and f_{Ej} would choose the same mass of entrants M_{Ej} and the same distribution of quantities produced $q_j(\varphi)$ as a function of productivity, including the same productivity cutoff φ_j^* and mass of producing firms M_j with positive quantities.[6] In this multi-sector setting, the allocation of labor across sectors will not be efficient due to differences in markups across sectors (the labor allocation in high markup, low elasticity sectors will be inefficiently low). The single sector version of the model is a special case in which there are no variations in markups and the market equilibrium is therefore efficient.

4 Open Economy with Trade Costs

In the closed economy, sector aggregates such as spending X_j and labor supply L_j have no effect on firm selection (the cutoff φ_j^*) and the distribution of firm performance measures within the sector. Since opening the closed economy to free international trade is the same as increasing aggregate spending and labor supply, such a change will have no impact on those firm-level variables. Although this result for free trade provides a useful benchmark, a large empirical literature finds evidence of substantial trade costs.[7] In this section, we characterize the open economy equilibrium in the presence of trade costs, which yields sharply different predictions for the effects of trade liberalization.

The world economy consists of a number of countries indexed by $i = 1, \ldots, N$. Preferences are identical across countries and given by (1). We assume that each country is endowed with a single homogeneous factor of production (labor) that is in inelastic supply $\bar{L}_i$ and is mobile across sectors.[8] The open economy equilibrium can be referenced by a zero-profit cutoff for serving each market n from each country i in each sector j (φ_{nij}^*), a wage for each country (w_i), a mass of entrants for each country and sector (M_{Eij}), and industry expenditure for each country and sector (X_{ij}).

For much of our analysis, we assume that the additional homogeneous good (in sector $j = 0$) is produced in all countries, and we use this good as numeraire.[9] In such an incomplete specialization equilibrium, $w_i = w = 1$ for all countries i (recall that the homogenous good is produced with unit labor requirements). Combining this result with our assumption of Cobb-Douglas preferences, consumer expenditure on each sector j in each country i is determined by parameters alone: $X_{ij} = \beta_j L_i$. In some of our analysis below, we consider the case of no outside sector, in which case each

[6] See the web appendix and Dhingra and Morrow (2012) for a formal analysis of the efficiency of the equilibrium.

[7] See for example the survey by Anderson and Van Wincoop (2004).

[8] In a later section, we explore the implications of introducing multiple factors of production and Heckscher-Ohlin comparative advantage in the open economy equilibrium.

[9] The assumption that the homogeneous good is produced in all countries will be satisfied if its consumption share and the countries' labor endowments are large enough.

country's wage is determined by the equality between its income and world expenditure on its goods.

Firm Behavior

As in the closed economy, we focus on equilibrium in a given sector and drop the sector subscript. Firm heterogeneity takes the same form in each country. After paying the sunk entry cost in country i (f_{Ei}), a firm draws its productivity φ from the cumulative distribution $G_i(\varphi)$. To serve market n, firms must incur a fixed cost of f_{ni} units of labor in country i and an iceberg variable trade cost such that $\tau_{ni} > 1$ units must be shipped from country i for one unit to arrive in country n.[10] The fixed exporting cost captures "market access" costs (e.g. advertising, distribution, and conforming to foreign regulations) that do not vary with firm scale. With CES preferences, this fixed cost is needed to generate selection into export markets such that only the most productive firms export. Absent this fixed export cost all firms would export.

We denote the fixed costs of serving the domestic market by f_{ii}, which includes both "market access" costs and fixed production costs (whereas the export cost f_{ni} for $n \neq i$ incorporates only the market access cost). Thus, the combined domestic cost f_{ii} need not be lower than the export cost f_{ni} ($n \neq i$), even if the market access component for the domestic market is always lower than its export market counterparts. When incorporating the fixed production cost into the domestic cost, we are anticipating an equilibrium where all firms serve their domestic market and only a subset of more productive firms export.[11] This will imply some parameter restrictions on the fixed and per-unit trade costs. (A large empirical literature finds that only a small fraction of firms export and these exporters are systematically more productive than non-exporters; see for example Bernard and Jensen 1995, 1999). Finally, we assume lower variable trade costs for the domestic market, $\tau_{ii} \leq \tau_{ni}$, and set $\tau_{ii} = 1$ and $\tau_{ni} \geq 1$ without loss of generality.

If a firm with productivity φ supplies market n from country i, the first-order condition for profit maximization again implies that its equilibrium price is a constant mark-up over its delivered marginal cost in that destination:

$$p_{ni}(\varphi) = \frac{\sigma}{\sigma - 1} \frac{\tau_{ni}}{\varphi}.$$

[10] We focus on exporting as the mode for serving foreign markets. For reviews of the literature on Foreign Direct Investment (FDI), see Antràs and Yeaple (2012) and Helpman (2006).

[11] With zero domestic market access costs, no firm exports without serving the domestic market, because the fixed cost of production has to be incurred irrespective of whether the domestic market is served and CES preferences imply positive variable profits in the domestic market. In contrast, with positive domestic market access costs, it can be profitable in principle for firms to export but not serve the domestic market (see, for example, Lu 2011).

Revenue and profit earned from sales to that destination are:

$$r_{ni}(\varphi) = A_n p_{ni}(\varphi)^{1-\sigma}, \qquad A_n = X_n P_n^{\sigma-1},$$

$$\pi_{ni}(\varphi) = B_n \tau_{ni}^{1-\sigma} \varphi^{\sigma-1} - f_{ni}, \qquad B_n = \frac{(\sigma-1)^{\sigma-1}}{\sigma^\sigma} A_n.$$

As in the closed economy, A_n and B_n are proportional indices of market demand in country n, which are functions of sector spending X_n and the CES price index P_n. Since all firms serve the domestic market, we account for the fixed production cost in "domestic" profit ($\pi_{ii}(\varphi)$).

Firm Market Entry and Exit

The presence of fixed market access costs implies that there is a zero-profit cutoff for each pair of source country and destination market:

$$\pi_{ni}(\varphi_{ni}^*) = 0,$$

$$\frac{r_{ni}(\varphi_{ni}^*)}{\sigma} = f_{ni} \qquad \Longleftrightarrow \qquad B_n (\tau_{ni})^{1-\sigma} (\varphi_{ni}^*)^{\sigma-1} = f_{ni}, \tag{6}$$

such that firms from country i with productivity $\varphi < \varphi_{ni}^*$ do not sell in market n and receive $r_{ni}(\varphi) = \pi_{ni}(\varphi) = 0$. Total firm revenue and profit (across destinations) are $r_i(\varphi) = \sum_n r_{ni}(\varphi)$ and $\pi_i(\varphi) = \sum_n \pi_{ni}(\varphi)$. We require restrictions on parameter values that generate selection into export markets and hence $\varphi_{ii}^* \leq \varphi_{ni}^*$ for all $n \neq i$.

Just like the closed economy, the free entry condition for country i equates an entrant's ex-ante expected profits with the sunk entry cost:

$$\int_0^\infty \pi_i(\varphi) dG_i(\varphi) = \sum_n \int_{\varphi_{ni}^*}^\infty \left[B_n \tau_{ni}^{1-\sigma} \varphi^{\sigma-1} - f_{ni}\right] dG_i(\varphi) = f_{Ei}. \tag{7}$$

The zero-profit cutoff (6) and free entry conditions (7) jointly determine all the cutoffs φ_{ni}^* and market demand levels B_n. The domestic cutoffs φ_{nn}^* and market demands B_n can be solved separately by using (6) to rewrite the free entry condition (7) as

$$\sum_n f_{ni} J_i(\varphi_{ni}^*) = f_{Ei}, \tag{8}$$

where we use the same definition for $J_i(\varphi^*)$ from (5). We can then use the cutoff condition (6) again to write the cutoffs φ_{ni}^* as either a function of market demands, $\varphi_{ni}^* = (f_{ni}/B)^{1/(\sigma-1)}\,\tau_{ni}$, or as a function of the domestic cutoffs, $\varphi_{ni}^* = (f_{ni}/f_{nn})^{1/(\sigma-1)}\,\tau_{ni}\varphi_{nn}^*$. Using the former, (8) delivers N equations for the market demands B_n; while the latter delivers N equations for the domestic cutoff φ_{nn}^*.

The open economy has a recursive structure that is similar to the closed economy: The cutoffs and market demands and hence the distribution of all firm performance measures (prices, quantities, sales, profits in all destinations) are independent of the sector aggregates such as sector spending X and sector labor supply L. Thus, only the mass of firms respond to the size of the sectors. We show how the exogenous sector spending $X = \beta\bar{L}$ can be used to solve for these quantities.

Mass of Firms and Price Index

Given M_{Ei} entrants in country i, a subset $M_{ni}\,[1 - G_i(\varphi_{ni}^*)]\,M_{Ei}$ of these firms will sell to destination n. Product variety in that destination will then be given by the total mass of sellers $M_n = \sum_n M_{ni}$. The price index P_n in that destination is the CES aggregate of the prices of all these goods:

$$P_n^{1-\sigma} = \sum_i \left\{ M_{ni} \int_{\varphi_{ni}^*}^{\infty} p_{ni}(\varphi)^{1-\sigma} \frac{dG_i(\varphi)}{1 - G_i(\varphi_{ni}^*)} \right\} = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \sum_i \left\{ M_{Ei}\tau_{ni}^{1-\sigma} \int_{\varphi_{ni}^*}^{\infty} \varphi^{\sigma-1} dG_i(\varphi) \right\}. \tag{9}$$

This price index in n is also related to market demand there:

$$B_n = \frac{(\sigma-1)^{\sigma-1}}{\sigma^\sigma} X_n P_n^{-1} \iff P_n^{1-\sigma} = X_n B_n^{-1}. \tag{10}$$

Using (9) and (10), we can solve out the price index and obtain

$$\frac{X_n}{\sigma B_n} = \sum_i M_{Ei}\tau_{ni}^{1-\sigma} \int_{\varphi_{ni}^*}^{\infty} \varphi^{\sigma-1} dG_i(\varphi), \tag{11}$$

which yields a system of N equations that determines the N entry variables M_{Ei}. (Recall that we have already solved out the left-hand side of those equations.) Using (10) and (6), we can express the price index in destination n as a function of the domestic cutoff only:

$$P_n = \frac{\sigma}{\sigma-1} \left(\frac{f_{nn}\sigma}{\beta\bar{L}_n}\right)^{1/(\sigma-1)} \frac{1}{\varphi_{nn}^*}. \tag{12}$$

Welfare

This price index summarizes the contribution of each sector to overall welfare. The Cobb-Douglas aggregation of sector-level consumption into utility in (1) implies that welfare per worker in country n (with income $w_n = 1$) is:

$$\mathbb{U}_n = \prod_{j=0}^{J} P_{nj}^{-\beta_j}, \tag{13}$$

where the sectoral price index (12) depends solely on the sectoral productivity cutoff φ_{nnj}^*. Therefore, although welfare depends on both the range of varieties available for consumption and their prices (these are the components that enter into the definition of each sector's CES price index in (9)), the domestic productivity cutoffs in each sector are sufficient statistics for welfare. Changes in trade costs will lead to changes in the ranges of imported and domestically produced varieties and their prices. All of these changes have an impact on welfare but their joint impact is summarized by the change in the domestic productivity cutoff. Similarly, the impact of changes in the number of countries or their size on welfare is also summarized by the change in the domestic productivity cutoff.

Symmetric Trade and Production Costs

To provide further intuition for mechanisms in the model, we consider the special case of symmetric trade and production costs (across countries):

$$\begin{aligned} &\tau_{ni} = \tau \quad \text{and} \quad f_{ni} = f_X \quad \forall n \neq i, \\ &f_{ii} = f, \quad \text{and} \quad f_{Ei} = f_E \quad \text{and} \quad G_i(\cdot) = G(\cdot) \quad \forall i. \end{aligned}$$

The only difference across countries is country size, indexed by the aggregate (across sectors) $\bar{L}_i$. In this special symmetric case, solving the free entry conditions (8) for the market demands B_n using $\varphi_{ni}^* = (f_X/B_n)^{1/(\sigma-1)}\tau$ yields a common market demand $B_n = B$ for all countries. This, in turn, implies that all countries have the same domestic cutoff $\varphi_{ii}^* = \varphi^*$ and that there is a single export cutoff $\varphi_{ni}^* = \varphi_X^*$ for $n \neq i$. These cutoffs are the solutions to the new zero-profit cutoff conditions:

$$B(\varphi^*)^{\sigma-1} = f, \tag{14}$$

$$B\tau^{1-\sigma}(\varphi_X^*)^{\sigma-1} = f_X, \tag{15}$$

and the free entry condition then takes the following form:

$$fJ(\varphi^*) + f_X(N-1)J(\varphi_X^*) = f_E. \tag{16}$$

These three conditions (14)-(16) jointly solve for the two symmetric cutoffs φ^* and φ_X^* and the symmetric market demand B. Note that the variable trade cost τ does not enter this free entry condition. Therefore changes in τ necessarily shift the productivity cutoffs φ^* and φ_X^* in opposite directions.

Under symmetry, the domestic and exporting zero-profit cutoffs (14) and (15) imply that the exporting cutoff is a constant proportion of the domestic cutoff:

$$\varphi_X^* = \tau \left(\frac{f_X}{f} \right)^{\frac{1}{\sigma-1}} \varphi^*,$$

where selection into export markets ($\varphi_X^* > \varphi^*$) requires strictly positive fixed exporting costs and sufficiently high values of both fixed and variable trade costs: $\tau^{\sigma-1} f_X > f$.

The relationship between profits and productivity is shown graphically in Figure 2. Firms drawing a low productivity $\varphi < \varphi^*$ would make losses if they produced and hence exit immediately. Firms drawing an intermediate productivity $\varphi \in [\varphi^*, \varphi_X^*)$ serve the domestic market for which they can generate sufficient revenue to cover fixed costs ($\pi_D(\varphi) \geq 0$). Only firms drawing a high productivity $\varphi > \varphi_X^*$ can generate sufficient revenue to cover fixed costs in both the domestic and every export market. (Recall that every export market has the same level of market demand B; therefore an exporter with productivity φ earns the same export profits $\pi_X(\varphi)$ in each destination.) Export market profits (in each destination) increase less sharply with firm productivity than domestic profits as a result of variable trade costs. The slope of total firm profits $\pi(\varphi) = \pi_D(\varphi) + (N-1)\,\pi_X(\varphi)$ increases from B to $B\tau^{1-\sigma}(N-1)$ at the export cutoff φ_X^*, above which higher productivity generates profits from sales to the domestic and all export markets. While firm profits are continuous in productivity, firm revenue jumps discretely at the export cutoff due to the fixed export costs. The model therefore naturally matches empirical findings that exporters are not only more productive than non-exporters but also larger in terms of revenue and employment (e.g. Bernard and Jensen 1995, 1999).

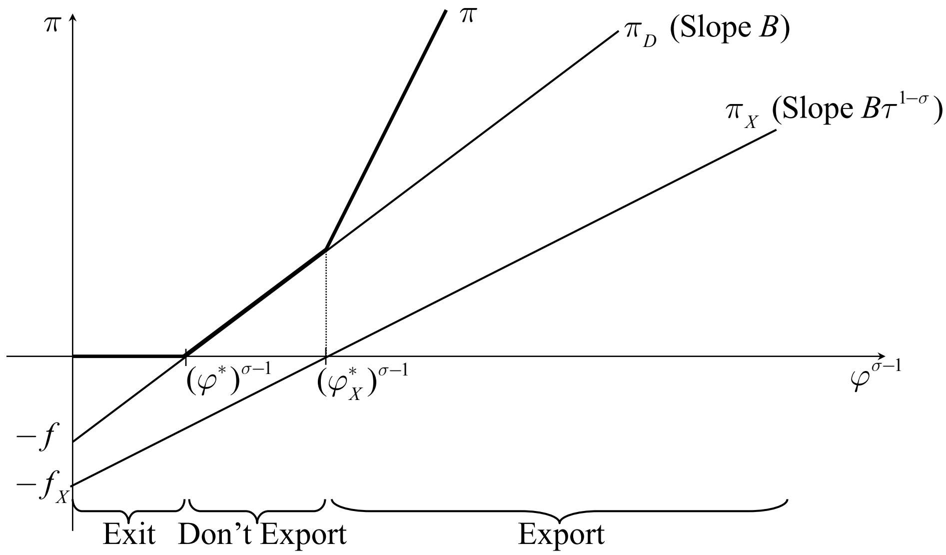

Figure 2: Open Economy Symmetric Countries

Multilateral Trade Liberalization

The impact of multilateral trade liberalization is seen most clearly in the transition between the closed and open economics with symmetric trade and production costs. Comparing the free entry conditions in the open and closed economies (16) and (5) respectively, and noting that $J(.)$ is decreasing, we see that the productivity cutoff in each sector must be strictly higher in the open economy than in the closed economy. From welfare (13), this increase in the zero-profit cutoff productivity in each sector is sufficient to establish the existence of welfare gains from trade.

The effect of opening to trade is illustrated in Figure 3. When the economy opens up to trade, the domestic market demand changes from its autarky level B^A to B (symmetric across countries). This new market demand cannot be higher than its autarky level, as this would imply that the total profit curve $\pi(\varphi)$ in the open economy is everywhere above the total profit curve $\pi^A(\varphi)$ in autarky. This would imply a rise in profits for firms of all productivities, which violates the free entry condition. Therefore, the new market demand B must be strictly below B^A. For the free entry condition to hold in both the closed and open economy equilibria, the total profit curves $\pi(\varphi)$ and $\pi^A(\varphi)$ must intersect. This implies that the combined domestic plus export market demand $B + (N-1)\,B\tau^{1-\sigma}$ must be strictly higher than the autarky demand level B^A. Therefore, the market demands must satisfy $B < B^A < B + (N-1)\,B\tau^{1-\sigma}$. The first inequality implies that all firms experience a reduction in domestic sales $r(\varphi) = \sigma B\varphi^{\sigma-1}$ (and hence a contraction in total sales for non-exporters); the second inequality implies that exporters more than make-up for their contraction in domestic sales with export sales and hence experience an increase in total sales $r(\varphi) = \sigma[B + (N-1)\, B\tau^{1-\sigma}]\varphi^{\sigma-1}$.[12]

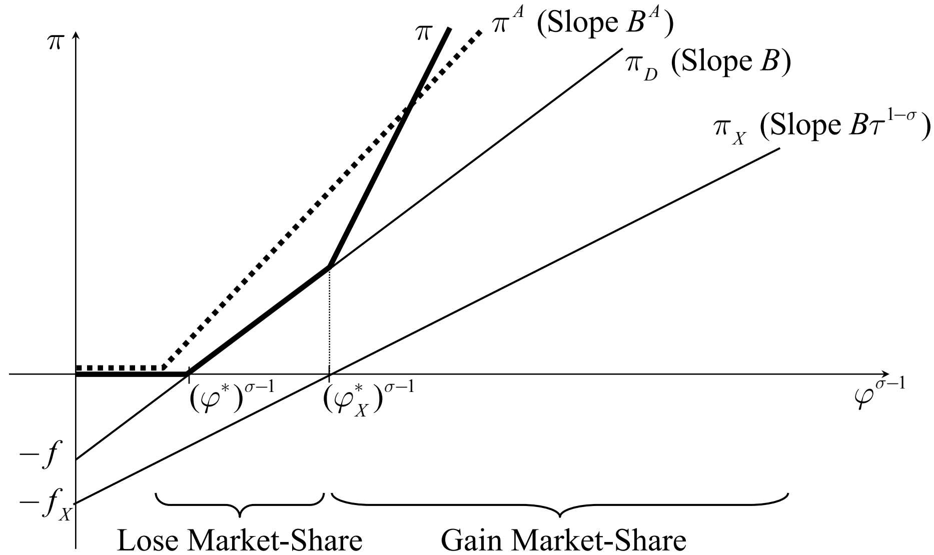

Figure 3: Open Economy Symmetric Countries

Opening to trade therefore induces a within-industry reallocation of resources between firms. The least productive firms exit with the rise in the domestic cutoff φ^*, the firms with intermediate productivity levels below the export cutoff φ_X^* contract, while the most productivity firms with productivity above the export cutoff φ_X^* expand. Each of these responses reallocates resources towards higher productivity firms generating an increase in average industry productivity.

The free entry condition implies that *ex-ante* expected profits in both the open and closed economies are equal to the entry cost f_E. *Ex-post* the average profits of surviving firms $\bar{\pi} = f_E/[1 - G(\varphi^*)]$ will be higher in the open economy due to the higher survival cutoff φ^*. (Recall that this average profit level will not depend on country size.) In both the open and closed economy, average total revenue per firm will be $\bar{r} = \sigma(\bar{\pi} + \bar{f})$. where $\bar{f}$ is the average (post-entry) fixed cost per firm. In the closed economy, $\bar{f} = f$ (the same overhead production cost paid by all firms). In the open economy $\bar{f} = f + f_X[1 - G(\varphi_X^*)]/\left[1 - G(\varphi^*)\right]$, which adds the fixed export cost weighted by the proportion of exporting firms. Thus, we see that average firm revenue $\bar{r}$ will be higher in the open economy relative to average revenue in the closed economy.

[12]Variable profits are proportional to revenues for all firms, but total profits also depend on fixed costs. Total profits move in the same direction as revenues for all firms, except for a subset of the least productive exporters. Although their total revenues increase, they experience a drop in total profit due to the additional fixed cost of exporting.

As in the closed economy, differences in country size will be reflected in the mass of entrants in a country. However, in the open economy with trade costs, the relationship between country size and entrants (and hence the mass of producing firms) will no longer be proportional. There will be a home market effect for entry, which responds more than proportionately to increases in country size: Solving the system of equations (11) under our symmetry assumptions reveals that $M_{Ei}/M_{Ei'} > \bar{L}_i/\bar{L}_{i'}$ for any two countries i and i'.[13] Differences in the mass of producing firms $M_{ii} = [1 - G(\varphi^*)] M_{Ei}$ will be proportional to entry since all countries have the same survival cutoff under our symmetry assumptions. These differences in the mass of entrants and producing firms across countries also imply disproportionate differences in labor to the differentiated sectors across countries. Recall that the free entry condition requires that aggregate sector payments to cover the entry cost, $M_{Ei}f_E$, are equal to aggregate sector profits Π_i (this property must hold for both the open and closed economies). Thus, aggregate sector labor supply L_i must be equal to aggregate sector revenue R_i:[14]

$$L_i = R_i = M_{ii}\bar{r}.$$

We see that the disproportionate response of entry and producing firms is also reflected in the disproportionate response of labor supply in larger countries. Since sector expenditures $X_i = \beta \bar{L}_i$ are proportional to country size, this implies that larger countries run a trade surplus in the differentiated good sectors.

In a single-sector version of our model, trade must be balanced, which would then lead to responses of labor supply L_i, entry M_{Ei}, and producing firms M_{ii} that are proportional to country size $\bar{L}_i$:

$$\frac{L_i}{L_{i'}} = \frac{M_{Ei}}{M_{Ei'}} = \frac{M_{ii}}{M_{i'i'}} = \frac{\bar{L}_i}{\bar{L}_{i'}}.$$

In this case, opening to trade does not affect the labor supply to the single sector, and would then induce a reduction in the mass of producing firms $M_{ii} = L_i/\bar{r}$ since average revenues are larger in the open economy. Even in this case, the response of product variety in country i, $M_i = \sum_n M_{ni}$ is ambiguous due to the availability of imported varieties. However, even if the mass of varieties available for domestic consumption falls, there are necessarily welfare gains from trade, because the

[13] This assumes that there is some available labor in the homogeneous good sector that can be moved to the differentiated good sectors. This would not be possible in the single-sector version of our model. Recall that we are assuming that differences in country size are not so large as to induce specialization away from the homogeneous sector in large countries.

[14] Here, we are assuming that all production expenses are paid to the L_i workers in country i. This includes the labor used to cover the fixed export costs f_X.

zero profit cutoff productivity rises and is a sufficient statistic for welfare in (13).[15]

The efficiency properties of the symmetric country open economy equilibrium are the same as in the closed economy: Conditional on the allocation of labor across sectors, a world social planner faced with the same entry and export technology (where the trade costs use up real resources) would choose the same distribution of quantities (as a function of firm productivity) and the same mass of producing firms as in the market equilibrium.[16] If trade costs take the form of policy interventions that do not use up real resources, national social planners can have an incentive to introduce trade policies to manipulate the terms of trade, as in Demidova and Rodriguez-Clare (2009). In a multi-sector setting with different elasticities of substitution, the allocation of labor across sectors is not efficient, providing a further potential rationale for interventionist trade policies to increase the labor allocation in high markup, low elasticity sectors.

Although for simplicity we have concentrated on opening the closed economies to trade, analogous results hold for further multilateral trade liberalization in the open economy equilibrium. This multilateral trade liberalization can take a variety of firms: (i) an increase in the number of trading partners ($N-1$), (ii) a decrease in variable trade costs (τ) and (iii) a decrease in fixed exporting costs (f_X). In each case, increased trade openness raises the zero-profit cutoff productivity and induces exit by the least productive firms, market share reallocations from less to more productive firms, and an increase in welfare. These intra-industry reallocations of resources across firms are consistent with empirical evidence from a large number of trade liberalization episodes, as examined for example in Bernard, Jensen and Schott (2006), Levinsohn (1999), Pavcnik (2002), Trefler (2004) and Tybout (2003).

Asymmetric Trade Liberalization

While the previous two sections have focused on symmetric countries, we now examine import and export liberalization between asymmetric countries. Following Demidova and Rodriguez-Clare (2011), we consider two asymmetric countries (countries 1 and 2) with a single differentiated sector and no outside sector. In this case, the relative wage between the two countries is no longer fixed and is determined by the balanced trade condition. We therefore re-introduce the wage w_1 and choose labor in country 2 as the numeraire, so $w_2 = 1$. Re-introducing wages does not change the

[15]Note the contrast with Krugman (1979), in which the opening of trade increases firm size and reduces the mass of domestically-produced varieties, but increases the mass of varieties available for domestic consumption. The underlying mechanism is also quite different: in Krugman (1979) firms are homogeneous and the rise in firm size occurs as a result of a variable elasticity of substitution.

[16]See the web appendix and Dhingra and Morrow (2012) for a formal analysis.

form of the free entry condition (8), which yields 2 conditions relating the domestic cutoff to the export cutoff for each country. We use those conditions to write the domestic cutoffs as functions of the export cutoffs: $\varphi_{22}^* = \varphi_{22}^*(\varphi_{12}^*)$ and $\varphi_{11}^* = \varphi_{11}^*(\varphi_{21}^*)$.[17] The cutoff profit conditions yield:

$$\begin{aligned} \varphi_{21}^* &= \tau_{21} \left(\frac{f_{21}}{f_{22}} \right)^{1/(\sigma-1)} (w_1)^{\sigma/(\sigma-1)} \varphi_{22}^*, \\ \varphi_{12}^* &= \tau_{12} \left(\frac{f_{12}}{f_{11}} \right)^{1/(\sigma-1)} (w_1)^{-\sigma/(\sigma-1)} \varphi_{11}^*. \end{aligned} \tag{17}$$

These conditions implicitly define the export cutoffs as functions of the wage w_1 and the domestic cutoff in the other country: $\varphi_{21}^* \equiv h_{21}(w_1, \varphi_{22}^*)$ and $\varphi_{12}^* \equiv h_{12}(w_1, \varphi_{11}^*)$. Combining all these conditions together yields a "competitiveness" condition for the export cutoff in country 1 as a function of the wage w_1:

$$\varphi_{21}^* = h_{21}\big(w_1, \varphi_{22}^*(\varphi_{12}^*)\big) = h_{21}\Big(w_1, \varphi_{22}^*\Big(h_{12}\big(w_1, \varphi_{11}^*(\varphi_{21}^*)\big)\Big)\Big) \tag{18}$$

This competitiveness condition defines an increasing relationship in (w_1, φ_{21}^*) space, as shown in Figure 4. Intuitively, a higher wage reduces a country's competitiveness and implies a higher cutoff productivity for exporting.

The "trade balance" condition is derived from labor market clearing, free entry, the zero-profit productivity cutoff conditions and the requirement that trade is balanced:

$$\begin{aligned} & M_{E1}(w_1, \varphi_{21}^*) w_1 f_{21} \left[J_1\left(\varphi_{21}^*\right) + 1 - G_1\left(\varphi_{21}^*\right) \right] \\ & = M_{E2}(w_1, \varphi_{21}^*) f_{12} \left[J_2\Big(h_{12}\big(w_1, \varphi_{11}^*\left(\varphi_{21}^*\right)\big)\Big) + 1 - G_2\Big(h_{12}\Big(h_{12}\big(w_1, \varphi_{11}^*\left(\varphi_{21}^*\right)\big)\Big)\Big) \right], \end{aligned} \tag{19}$$

which defines a decreasing relationship in (w_1, φ_{21}^*) space, as also shown in Figure 4. Intuitively, a higher productivity cutoff for exporting reduces total exports, which induces a trade deficit, and hence requires a reduction in the wage to increase competitiveness and eliminate the trade deficit. Note that since the trade balance condition (19) incorporates the competitiveness condition (18) care must be undertaken in the interpretation of these two relationships.

The effects of asymmetric trade liberalizations can be characterized most sharply for the case of a small open economy. In the monopolistically competitive environment considered here, country

[17] We continue with our notation choice that the first subscript denotes the country of consumption and the second subscript the country of production. The notation in Demidova and Rodriguez-Clare (2011) reverses the order of the subscripts.

1 is assumed to be a small open economy if (i) the zero-profit productivity cutoff in country 2 is unaffected by home variables, (ii) the mass of firms in country 2 is unaffected by home variables, (iii) total expenditure and the price index in country 2 are unaffected by home variables. Nonetheless, the export productivity cutoff in country 2 and the measure of exporters from foreign to home are endogenous and depend on trade costs.

Under these small open economy assumptions, φ_{22}^* is exogenous with respect to country 1 variables and the trade cost between the two countries, and the competitiveness condition can be evaluated using the cutoff profit condition (17). In these circumstances, a unilateral trade liberalization by country 1 (a fall in variable trade costs (τ_{12}) and/or fixed exporting costs f_{12}) leaves the competitiveness condition unchanged but shifts the trade balance condition inwards. As a result, w_1 and φ_{21}^* fall, which from the free entry condition implies a rise in φ_{11}^* and hence in country 1's welfare. Intuitively, the unilateral domestic trade liberalization reduces the price of foreign goods relative to domestic goods, and requires a fall in the domestic wage to restore the trade balance. This fall in the domestic wage increases export market profits, which induces increased entry and hence tougher selection on the domestic market.

In contrast, a unilateral reduction in variable trade costs by country 2 (a fall in τ_{21}) shifts the competitiveness condition outwards but leaves the trade balance condition unchanged. As a result, w_1 rises and φ_{21}^* again falls, which from the free entry condition implies a rise in φ_{11}^* and hence in country 1's welfare. Intuitively, the fall in foreign variable trade costs increases domestic export market profits, which induces increased entry and tougher selection on the domestic market. The domestic wage rises to restore the trade balance. Reductions in the fixed costs of exporting to country 2 (f_{21}) have more subtle effects, because they shift both the competitiveness and trade balance curves.

The key takeaway from this analysis without an outside sector is that reductions in variable trade costs on either exports or imports raise welfare. In contrast, in the presence of an outside sector, import liberalization can be welfare reducing. The reason is the home market effect, according to which production relocates to the higher trade cost country to access that market without incurring trade costs and take advantage of the lower import barriers in the liberalizing country (see for example Krugman 1980 and Venables 1987).

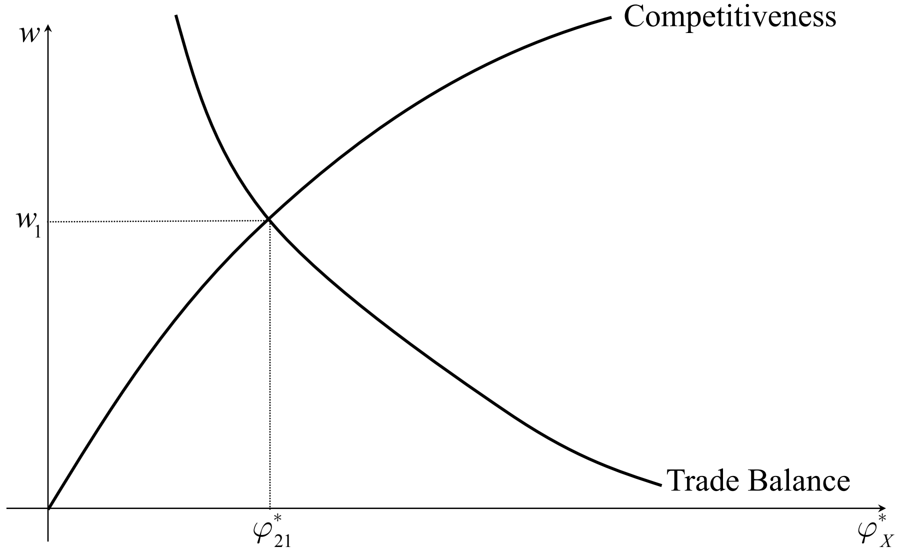

Figure 4: Competitiveness and Trade Balance Conditions

5 Quantification

To derive quantitative predictions for trade and welfare, we follow a large part of the literature in assuming a Pareto productivity distribution, as in Helpman et al. (2004), Chaney (2008) and Arkolakis et al. (2008, 2012). Besides providing a good fit to the observed firm size distribution, this assumption yields closed form solutions for the productivity cutoffs and other endogenous variables of the model. For much of the analysis, we maintain our assumption of a composite labor supply L_j for each sector j with a unit cost w_j that can vary across sectors. When we solve for factor prices, we specialize to the case of homogeneous labor with a common wage across sectors. To determine this wage, we dispense with the assumption of an outside sector, and use the equality between country income and expenditure on goods produced in that country.

Pareto Distribution

We now assume that firm productivity φ is drawn from a Pareto distribution so that

$$g(\varphi) = k\varphi_{\min}^{k}\varphi^{-(k+1)}, \qquad G(\varphi) = 1 - \left(\frac{\varphi_{\min}}{\varphi}\right)^{k},$$

where $\varphi_{\min} > 0$ is the lower bound of the support of the productivity distribution and lower values of the shape parameter k correspond to greater dispersion in productivity.[18]

A key feature of a Pareto distributed random variable is that when truncated the random variable retains a Pareto distribution with the same shape parameter k. Therefore the *ex post* distribution of firm productivity conditional on survival also has a Pareto distribution. Another key feature of a Pareto distributed random variable is that power functions of this random variable are themselves Pareto distributed. Therefore firm size and variable profits are also Pareto distributed with shape parameter $k/(\sigma-1)$, where we require $k > \sigma - 1$ for average firm size to be finite.[19]

With Pareto distributed productivity, $J(\varphi^*)$ is a simple power function of the productivity cutoff φ^*.[20] From this power function, we obtain the following closed form solutions for the zero-profit cutoff productivities in the closed economy

$$(\varphi^*)^k = \frac{\sigma-1}{k-(\sigma-1)}\frac{f}{f_E}\varphi_{\min}^k,$$

and in the symmetric country open economy equilibrium:

$$(\varphi^*)^k = \frac{\sigma-1}{k-(\sigma-1)}\left[\frac{f+(N-1)\,\tau^{-k}\left(\frac{f_X}{f}\right)^{\frac{-k}{\sigma-1}}f_X}{f_E}\right]\varphi_{\min}^k.$$

Gravity

When firm productivity and hence firm exports to any destination are distributed Pareto, we obtain some very sharp predictions for bilateral trade flows (at the aggregate sector level). Before imposing

[18]While a common shape parameter k for all countries is an important simplifying assumption, it is straightforward to accommodate cross-country differences in technology in the form of different lower bounds for the support of the productivity distribution $\varphi_{\min}$.

[19]For empirical evidence that the Pareto distribution provides a reasonable approximation to the observed distribution of firm size, see Axtell (2001). The requirement that $k > \sigma - 1$ is needed given that the support for the Pareto distribution is unbounded from above and given the assumption of a continuum of firms. If either of these conditions are relaxed (finite number of firms or a truncated Pareto distribution), then this condition need not be imposed. Empirical estimates violating this condition for some sectors therefore can be rationalized within the model by these modifications.

[20]
$$J(\varphi^*) = \frac{\sigma-1}{k-(\sigma-1)}\left(\frac{\varphi_{\min}}{\varphi^*}\right)^k$$

this distributional assumption, we can write aggregate sector exports from i to n as:

$$\begin{aligned}
X_{ni} &= M_{Ei} \int_{\varphi_{ni}^*}^{\infty} r_{ni}(\varphi) dG(\varphi) \\
&= M_{Ei} \int_{\varphi_{ni}^*}^{\infty} \left(\frac{\varphi}{\varphi_{ni}^*} \right)^{\sigma-1} \sigma w_i f_{ni} dG(\varphi) \\
&= M_{Ei} \sigma w_i f_{ni} \left[J(\varphi_{ni}^*) + 1 - G(\varphi_{ni}^*) \right]
\end{aligned}$$

Using the simple closed form derivation for $J(.)$ under Pareto productivity, we can then decompose bilateral aggregate trade into an extensive (mass of exporters) and intensive (average firm exports) margin:

$$X_{ni} = \underbrace{M_{Ei} \left(\frac{\varphi_{\min}}{\varphi_{ni}^*} \right)^k}_{\text{mass of exporters}} \underbrace{w_i f_{ni} \frac{\sigma k}{k - \sigma + 1}}_{\text{average firm exports}}. \tag{20}$$

Under this distributional assumption, we see that average firm exports are independent of variable trade costs, so that the latter reduce bilateral trade solely at the extensive margin given by the mass of exporting firms. On one hand, higher variable trade costs reduce firm-level exports for all firms; this reduces average exports per firm. On the other hand, higher variable trade costs also induce low productivity firms to exit the export market; this raises average exports per firm through a composition effect (lower productivity firms have lower exports). With a Pareto productivity distribution these two effects exactly offset one another, so as to leave average firm exports conditional on exporting independent of variable trade costs.[21] In contrast, higher fixed costs of exporting (f_{ni}) increase the exporting productivity cutoff (φ_{ni}^*), which reduces the mass of exporting firms and increases average exports conditional on exporting.

Under the assumption of Pareto productivity distribution, models of firm heterogeneity in differentiated product markets yield a gravity equation for bilateral trade flows, consistent with empirical findings from a large literature in international trade. This gravity equation holds sector by sector independently of whether or not trade is balanced. Noting that industry revenue $R_i = \sum_{n=1}^{N} X_{ni}$ and using the export productivity cutoff condition (6), bilateral exports from country i to market n in sector j (20) can be re-written as the following gravity equation:

$$X_{ni} = \frac{R_i}{\Xi_i} \left(\frac{X_n}{P_n^{1-\sigma}} \right)^{k/(\sigma-1)} \tau_{ni}^{-k} f_{ni}^{1-k/(\sigma-1)}, \tag{21}$$

[21] Note the parallel with Eaton and Kortum (2002), in which higher variable trade costs reduce bilateral exports solely through an extensive margin of the fraction of goods exported.

$$\Xi_i = \sum_n \left(\frac{X_n}{P_n^{1-\sigma}} \right)^{k/(\sigma-1)} \tau_{ni}^{-k} f_{ni}^{1-k/(\sigma-1)},$$

which takes a similar form to the standard CES gravity equation without firm heterogeneity in Anderson and Van Wincoop (2002).[22] Comparing the two expressions, a number of differences are apparent. First, variable trade costs affect aggregate trade flows through both the intensive margin (exports of a given firm) and the extensive margin (the number of exporting firms). However, the exponent on variable trade costs τ_{ni} is the Pareto shape parameter k rather than the elasticity of substitution between varieties, which reflects the invariance of average firm exports with respect to variable trade costs discussed above. Second, fixed exporting costs f_{ni} only affect aggregate trade flows through the extensive margin of the number of exporting firms, and enter with an exponent that depends on both the Pareto shape parameter and the elasticity of substitution. Third, the importer fixed effect in the standard CES formulation $(X_n/P_n^{1-\sigma})$ is amplified under firm heterogeneity by $k/(\sigma-1) > 1$, which captures the effect of market demand on the extensive margin of exporting firms. Fourth, the exporter fixed effect is the same as in the standard CES formulation (R_i/Ξ_i). This fixed effect combines an exporter's industry revenue (R_i) with its market potential (Ξ_i), where market potential is defined as in Redding and Venables (2004) as the trade cost weighted sum of demand in all markets.

A key implication of the gravity equation (21) is that bilateral trade between countries i and n depends on both bilateral trade costs $\{\tau_{ni}, f_{ni}\}$ and trade costs with all the other partners of each country ("multilateral resistance" as captured in P_n and Ξ_i). This role for multilateral resistance can be further illustrated by solving explicitly for the price indices (P_n), which depend on the mass of entrants (M_{Ei}) across countries. In general, the mass of entrants in country i will depend on both the labor supply L_i to the sector as well as the cutoffs φ_{ni}^* to all destinations n, which determine the allocation of labor between entry and production. However, under Pareto productivity, the dependence of entry on the cutoffs is eliminated and entry only depends on the labor supply to the sector (see appendix for proof):

$$M_{Ei} = \frac{(\sigma-1)}{k\sigma} \frac{L_i}{f_{Ei}}. \tag{22}$$

This allows us to write the price index in n as a function of the sector's supply of the composite

[22] For further discussion of the gravity equation literature, see the Head and Mayer (2012) chapter in this handbook.

labor input, expenditure and wage across countries:

$$P_n^{-k} = \left[\sum_v \left(L_v/f_{Ev}\right)\varphi_{\min}^k \left(\tau_{nv}w_v\right)^{-k}\left(f_{nv}\right)^{1-k/(\sigma-1)}\right]$$
$$\left(X_n\right)^{-(1-k/(\sigma-1))}\left(\tfrac{\sigma}{\sigma-1}\right)^{-k}\sigma^{-k/(\sigma-1)}\tfrac{\sigma-1}{k-\sigma+1}.$$

We can then solve out this price index from the bilateral gravity equation (21) to obtain a trade share that depends only on the prices (w_i) and supplies (L_i) of the composite labor input:

$$\lambda_{ni} \equiv \frac{X_{ni}}{X_n} = \frac{\left(L_i/f_{Ei}\right)\left(\tau_{ni}w_i\right)^{-k} f_{ni}^{1-k/(\sigma-1)}}{\sum_v \left(L_v/f_{Ev}\right)\left(\tau_{nv}w_v\right)^{-k} f_{nv}^{1-k/(\sigma-1)}}. \tag{23}$$

The trade share (23) takes the familiar gravity equation form, as in Eaton and Kortum (2002).[23] The elasticity of trade with respect to variable trade costs again depends on the Pareto shape parameter (k) and there is a unit elasticity on importer expenditure (X_n) for given sectoral labor allocations (L_i) and wages (w_i) in all countries. One key difference from Eaton and Kortum (2002) is that the trade share depends directly on the labor allocation (L_i), which reflects the presence of an endogenous measure of firm varieties compared to a fixed range of goods.

Wages and Welfare

We now turn to the general equilibrium across sectors and investigate its welfare properties. To close our model, we again assume that labor is homogeneous, with a fixed supply $\bar{L}_i$ in each country. We dispense with the assumption of an outside sector so that $j = 1, \ldots, J$ and all sectors are differentiated. Sectoral spending is given by $X_{nj} = \beta_j w_n \bar{L}_n$, and the country wages (common across sectors j) are determined by the N balanced trade conditions for each country:

$$w_i \bar{L}_i = \sum_{j=1}^{J}\sum_{n=1}^{N} \lambda_{nij}\beta_j w_n \bar{L}_n,$$

where the trade shares λ_{nij} for each sector are determined by (23).

The assumption of Pareto productivity also has some strong implications for the functional form of the welfare gains from trade, as analyzed in Arkolakis et al. (2012a). Welfare per worker (in country n) takes the same form as in (13), except that wages are no longer unitary: $\mathbb{U}_n =$

[23] This trade share (23) under Pareto productivity can also be used to show that a sufficient condition for the mass of varieties available for domestic consumption to fall following the opening of trade is $f_{ni} > f_{ii}$ for $n \neq i$, as analyzed in Baldwin and Forslid (2010).

$w_n / \left(\prod_{j=1}^{J} P_{nj}^{\beta_j} \right) = \prod_{j=1}^{J} (P_{nj}/w_n)^{-\beta_j}$. The price index for sector j in country n, in turn, can be written as a function of the domestic productivity cutoff in that sector as shown in (12):

$$P_{nj} = \frac{\sigma_j}{\sigma_j - 1} \left(\frac{f_{nnj}\sigma_j}{\beta_j w_n \bar{L}_n} \right)^{1/(\sigma_j - 1)} \frac{1}{\varphi_{nnj}^*}. \tag{24}$$

Under the assumption of Pareto productivity, we can write the domestic cutoff in each sector φ_{nnj}^* as a function of country n's domestic trade share $\lambda_{nnj} \equiv X_{nnj}/\beta_j w_n \bar{L}_n$ and the mass of entrants M_{Enj} in that sector using (20). From (22), we can write this mass of entrants in terms of the sector's labor supply L_{nj}. This yields the following expression for the domestic cutoff:

$$\left(\varphi_{nnj}^*\right)^k = \varphi_{\min}^k \frac{\sigma_j - 1}{k_j - (\sigma_j - 1)} \frac{f_{nnj}}{f_{Enj}} \frac{1}{\beta_j \bar{L}_n} \frac{L_{nj}}{\lambda_{nnj}}. \tag{25}$$

By combining (25) and (24), we obtain an expression for welfare that depends only on the endogenous ratio of labor supply to domestic trade shares L_{nj}/λ_{nnj} across sectors. This expression does not contain any per-unit or fixed trade cost measures, τ_{nij} or f_{nij} for $n \neq i$, so the ratios L_{nj}/λ_{nnj} are sufficient statistics that summarize the impact of trade costs on welfare. We can thus summarize the welfare gains from trade, measured as the welfare ratio $\hat{\mathbb{U}}_n \equiv \mathbb{U}_n^{\text{Open}}/\mathbb{U}_n^{\text{Closed}}$ (between the open and closed economies) in terms of a country's domestic trade shares $\lambda_{nnj}^{\text{Open}}$ (the domestic trade shares in the closed economy are fixed at $\lambda_{nnj}^{\text{Closed}} = 1$) and the change in sectoral labor allocations between the open and closed economies ($\hat{L}_{nj} = L_{nj}^{\text{Open}}/L_{nj}^{\text{Closed}}$):

$$\hat{\mathbb{U}}_n = \prod_{j=1}^{J} \left(\frac{\hat{\lambda}_{nnj}}{\hat{L}_{nj}} \right)^{-\frac{\beta_j}{k_j}} = \prod_{j=1}^{J} \left(\lambda_{nnj}^{\text{Open}}\right)^{-\frac{\beta_j}{k_j}} \left(\frac{L_{nj}^{\text{Open}}}{L_{nj}^{\text{Closed}}} \right)^{\frac{\beta_j}{k_j}}.$$

Also, we note that the only relevant parameters for this welfare gain calculation are the expenditure shares (β_j) and trade elasticities (k_j). Lower trade costs have a direct impact on the welfare gains by lowering the domestic trade shares $\lambda_{nnj}^{\text{Open}}$. They also have an indirect effect via the reallocation of labor across sectors. This channel operates through the welfare benefits of higher entry rates (which leads to additional product variety), and is therefore absent in models of trade where the range of consumed goods is constant. To motivate the direction of the welfare gain for this channel, we return to our scenario that adds an additional homogenous good sector $j = 0$ that is produced in every country. In that scenario (with symmetric trade and production costs), we saw that opening to trade would reallocate labor L_{nj}^{Open} to the differentiated sectors $j \geq 1$ for larger countries (larger $\bar{L}_n$). This generates distributional effects for the gains from trade, skewing those gains towards larger markets. (If we break the symmetry assumption for trade costs, then countries with better geography would also increase their relative employment in the differentiated goods sectors, skewing the gains from trade in their favor.). Balisteri et al (2011) provide a quantitative assessment of the gains from trade liberalization that accounts for these inter-sectoral labor reallocations, based on differences in country size, geography, and comparative advantage. They find that firm heterogeneity plays a key role in determining their quantitative measure of the gains from a given trade liberalization scenario.

In the special case where labor allocations across sectors do not change when opening of trade, the open economy domestic trade shares $\lambda_{nnj}^{\text{Open}}$ then provide the single sufficient statistics for the welfare gains from trade. This version of the model falls within a class of models analyzed by Arkolakis et al. (2012a). They show that when these models are calibrated to the same empirical trade shares (for the open economy equilibrium), they will all imply the same welfare gains from trade. But note that countries' trade shares with themselves are endogenous variables and have different determinants in different models. Therefore if these models are calibrated to trade shares in the open economy equilibrium they will typically have different predictions for other outcomes. More generally, different models typically have different implications for the impact of trade liberalization on the allocation of labor across sectors, in which case their predictions for the welfare gains from trade also differ.

While our discussion above concentrates on aggregate bilateral trade shares, models of firm heterogeneity in differentiated product markets provide a rationale for the prevalence of zeros in bilateral trade flows. Helpman et al. (2008) develop a multi-country version of the model in Section 2, in which the productivity distribution is a truncated Pareto. In this case, no firm exports from country i to market n if the productivity cutoff (φ_{ni}^*) lies above the upper limit of country i's productivity distribution. Estimating a structural gravity equation, they show that controlling for the non-random selection of positive trade flows and the extensive margin of exporting firms is important for estimates of the trade effects of standard trade frictions.

Structural Estimation

In addition to shedding new light on aggregate bilateral trade flows, models of firm heterogeneity in differentiated product markets also provide a natural platform for explaining a number of features of disaggregated trade data by firm and destination market. As shown in Eaton, Kortum and Kramarz

(2011), disaggregated French trade data exhibit a number of striking empirical regularities. First, the number of French firms selling to a market (relative to French market share) increases with market size according to an approximately log linear relationship. This pattern of firm export market participation exhibits an imperfect hierarchy, where firms selling to less popular markets are more likely to sell to more popular markets, but do not always do so. Second, export sales distributions are similar across markets of very different size and extent of French participation. While the upper tail of these distributions is approximately Pareto distributed, there are departures from a Pareto distribution in the lower tail, in which small export shipments are observed. Third, average sales in France are higher for firms selling to less popular foreign markets and for firms selling to more foreign markets.

To account for these features of the data, Eaton, Kortum and Kramarz (2011) use a version of the model from Section 2 with a Pareto productivity distribution and a fixed measure of potential firms as in Chaney (2008). To explain variation in firm export participation with market size under CES demand, fixed market entry costs are required. But to generate the departures from a Pareto distribution in the lower tail, these fixed market entry costs are allowed to vary endogenously with a firm's choice of the fraction of consumers within a market to serve (e), as in Arkolakis (2010). Finally, to explain imperfect hierarchies of markets, fixed market entry costs are assumed to be subject to an idiosyncratic shock for each firm ω and destination market n ($\varepsilon_{n\omega}$) as well a common shock for each source country i and destination market n (F_{ni}). Market entry costs are therefore:

$$f_{ni\omega} = \varepsilon_{n\omega} F_{ni} M\left(e\right),$$

where the function $M\left(e\right)$ determines how market entry costs vary with the fraction of consumers served (e) and takes the following form:

$$M\left(e\right) = \frac{1 - \left(1 - e\right)^{1-1/\lambda}}{1 - 1/\lambda},$$

where $\lambda > 0$ captures the increasing cost of reaching a larger fraction of consumers. Any given consumer is served with probability e, so that each consumer receives the same measure of varieties, but the particular varieties in question can vary across consumers.[24]

[24] To generate the observed departures from a Pareto distribution in the lower tail of the export sales distribution, one requires $0 < \lambda < 1$, which implies an increasing marginal cost of reaching additional consumers. An alternative potential explanation for the departures from a Pareto distribution in the lower tail is a variable elasticity of substitution. Both endogenous market entry costs and a variable elasticity of substitution provide potential explanations

To allow for idiosyncratic variation in sales conditional on entering a given export market for firms with a given productivity, demand is also subject to an idiosyncratic shock for each firm ω and destination market n, $\alpha_{n\omega}$:

$$X_{ni\omega} = \alpha_{n\omega} e_{ni\omega} X_n \left(\frac{\tau_{ni} p_{i\omega}}{P_n} \right)^{1-\sigma},$$

where X_n is total expenditure in market n and the presence of $e_{ni\omega}$ reflects the fact that only a fraction of consumers in each market are served. A firm's decision to enter a market depends on the composite shock, $\eta_{n\omega} = \alpha_{n\omega}/\varepsilon_{n\omega}$, but a firm with a given productivity can enter a market because of a low entry shock, $\varepsilon_{n\omega}$, and yet still have low sales in that market because of a low demand shock, $\alpha_{n\omega}$.

Using moments of the French trade data by firm and destination market, Eaton, Kortum and Kramarz (2011) estimate the model's five key parameters: a composite parameter including the elasticity of substitution and the Pareto shape parameter, the convexity of marketing costs, the variance of demand shocks, the variance of entry shocks, and the correlation between demand and entry shocks. These five parameters are precisely estimated and the estimated model provides a good fit to the data. Firm productivity accounts for around half of the observed variation across firms in export market participation, but explains substantially less of the variation in exports conditional on entering a market.

The estimated model is used to undertake counterfactuals, such as a 10 percent reduction in bilateral trade barriers for all French firms. In this counterfactual, total sales by French firms rise by around \$16 million, with most of this increase accounted for by a rise in sales of the top decile of firms of around \$23 million. In contrast, every other decile of firms experiences a decline in sales, with around half of the firms in the bottom decile exiting. These results suggest that even empirically reasonable changes in trade frictions can involve quantitatively large intra-industry reallocations.[25]

for empirical findings that most of the growth in trade following trade liberalization is in goods previously traded in small amounts, as in Kehoe and Ruhl (2009).

[25] See Corcos, Del Gatto and Ottaviano (2012) for a quantitative analysis of European integration using a model of firm heterogeneity and trade. See Cherkashin et al. (2010) for a quantitative analysis of the Bangladeshi apparel sector.

6 Factor Abundance and Heterogeneity

While models of firm heterogeneity in differentiated product markets emphasize within-industry reallocations, traditional trade theories instead stress between-industry reallocations. Bernard, Redding and Schott (2007) combine these two dimensions of reallocation by incorporating the model in Section 2 into the integrated equilibrium framework of neoclassical trade theory. Comparative advantage is introduced by supposing that sectors differ in their relative factor intensity and countries differ in their relative factor abundance. The production technology within each sector is homothetic such that the entry cost and the fixed and variable production costs use the two factors of production with the same intensity. The total cost of producing $q(\varphi)$ units of a variety in sector j in country i is thus:

$$\Gamma_{ij} = \left[f_{ij} + \frac{q_{ij}(\varphi)}{\varphi} \right] (w_{Si})^{\beta_j} (w_{Li})^{1-\beta_j}, \qquad 1 > \beta_1 > \beta_2 > 0$$

where w_{Si} is the skilled wage and w_{Li} is the unskilled wage.

In the special case in which fixed and variable trade costs are equal to zero, all firms export and the concept of integrated equilibrium from Dixit and Norman (1980) and Helpman and Krugman (1985) can be used to determine the set of factor allocations to the two countries for which trade in goods alone can equalize factor prices. Within this factor price equalization set, the four theorems of the Heckscher-Ohlin model – the Factor Price Equalization, Stolper-Samuelson, Rybczynski and Heckscher-Ohlin Theorems – continue to hold with firm heterogeneity.

More generally, if fixed and variable trade costs are not equal to zero, factor price equalization breaks down and, for parameter values for which there is selection into export markets, the opening of trade results in intra-industry reallocations across firms. As these intra-industry reallocations are driven by the differential impact of the opening of trade on exporters and non-exporters, they are stronger in the comparative advantage sector, where export opportunities are relatively more attractive. Although there is a decline in the relative mass of firms in the comparative disadvantage sector, as factors of production are reallocated in accordance with comparative advantage, exit by low productivity firms is strongest in the comparative advantage sector. Thus the opening of trade leads to a larger increase in the zero-profit cutoff and in average productivity in the comparative advantage sector than in the comparative disadvantage sector. This differential impact of the opening of trade across sectors according to Heckscher-Ohlin-based comparative advantage influences the effect of trade liberalization on welfare and income distribution. As the economy opens to trade, changes in product variety and aggregate productivity can more than offset the standard Stolper-Samuelson effect, so that the real wage of the scarce factor can rise.

A number of studies have further explored the relationship between within and between-industry reallocations of resources. Fan et al. (2011) and Hsieh and Ossa (2011) embed firm heterogeneity within a Ricardian model with a continuum of goods. Dan Lu (2012) incorporates firm heterogeneity into a Heckscher-Ohlin model with a continuum of goods and uses the effect of comparative advantage on the relative attractiveness of export and domestic markets to explain differences in the relative productivity of Chinese exporters and non-exporters across sectors. Burstein and Vogel (2012) provide general conditions under which changes in the factor content of trade are a sufficient condition for changes in relative factor prices. Rescheff and Harrigan (2012) explore the implications of complementarities between heterogeneous firm productivity and skills for the impact of trade liberalization on wage inequality.

7 Trade and Market Size

One limitation of the theoretical framework considered so far is its assumption of constant elasticity of substitution (CES) preferences, which imply constant mark-ups and hence that changes in aggregate demand leave the productivity cutoff for production unchanged. In this section, we extend our analysis of firm heterogeneity to the case of variable mark-ups following Melitz and Ottaviano (2008). Aggregate market conditions are summarized by the "toughness" of competition, which depends on market size in the closed economy and on both market size and trade costs in the open economy. "Tougher" competition in a market is characterized by a larger number of sellers and a lower average price of sellers, which both induce a downward shift in distribution of markups across firms. Differences in competition across markets thus feedback to influence firm location and export decisions. Markets that have more attractive fundamentals (for firms) are characterized in equilibrium by "tougher" competition, which implies that it is harder for exporters to break into these markets and harder for domestic firms to survive in these markets.

Consumer preferences are assumed to be quasi-linear between a homogeneous and differentiated sector with quadratic preferences across varieties within the differentiated sector, as in Ottaviano, Tabuchi and Thisse (2002):

$$U_i = q_{0i}^c + \alpha \int_{\omega \in \Omega_i} q_{\omega i}^c d\omega - \frac{1}{2}\gamma \int_{\omega \in \Omega_i} \left(q_{\omega i}^c\right)^2 d\omega - \frac{1}{2}\eta \left(\int_{\omega \in \Omega_i} q_{\omega i}^c d\omega\right)^2,$$

where $q^c_{\omega i}$ and q^c_{0i} denote the representative consumer's consumption of differentiated variety ω and the homogeneous good respectively; Ω_i is the set of varieties available for consumption in country i. Higher α and lower η increase demand for differentiated varieties relative to the numeraire, while higher γ implies greater love of variety, with $\gamma = 0$ corresponding to the special case of perfect substitutes.

Labor is the sole factor of production and each country i is endowed with L_i workers. Total demand for each variety is therefore $L_i q^c_{\omega i}$, where L_i indexes market size. Each country's labor endowment is assumed to be sufficiently large that it both consumes and produces the homogeneous good, which is chosen as the numeraire, so that $p_{0i} = 1$. As long as the homogeneous good is consumed, quasi-linear-quadratic preferences imply that the demand for differentiated varieties can be determined independently of income. Using the first-order conditions for utility maximization, the inverse demand curve for each differentiated variety is:

$$p_{\omega i} = \alpha - \gamma q^c_{\omega i} - \eta Q^c_i, \qquad Q^c_i = \int_{\omega \in \Omega_i} q^c_{\omega i} d\omega. \tag{26}$$

Since the marginal utility of consuming a differentiated variety is finite at zero consumption, there is a threshold price above which demand for a variety is zero, which using (26) can be expressed as:

$$p_{\omega i} \leq \frac{1}{\eta N_i + \gamma} \left(\gamma \alpha + \eta N_i \bar{p}_i \right), \tag{27}$$

where N_i is the number of consumed varieties and $\bar{p}_i$ is their average price. This threshold price decreases as the number of consumed varieties rises and as their average price falls (tougher competition). Welfare is given by the indirect utility function:

$$U_i = I^c_i + \frac{1}{2} \left(\eta + \frac{\gamma}{N_i} \right)^{-1} (\alpha - \bar{p}_i)^2 + \frac{1}{2} \frac{N_i}{\gamma} \sigma^2_{pi},$$

where I^c_i is the representative consumer's income; $\bar{p}_i$ and σ^2_{pi} are the mean and variance of prices respectively. Welfare increases when average prices fall, when the number of varieties increases (consumer love of variety) and when the variance of prices increases (as the variance of prices increases, consumers can substitute towards lower-priced varieties).

The homogeneous good is produced under conditions of perfect competition and constant returns to scale with a unit labor requirement. As long as the homogeneous good is produced, productivity in this sector pins down the wage in each country as equal to one. Differentiated varieties are produced under conditions of monopolistic competition and constant returns to scale. To enter the differentiated sector, a firm must incur a sunk entry cost of f_E units of labor, after which its unit labor requirement or cost (c) is drawn from a cumulative distribution function $G(c)$ with support on $[0, c_M]$, where this cost draw is the inverse of the productivity draw considered in Section 2. As firms with the same cost (c) behave symmetrically, firms are indexed from now on by c alone. If a firm decides to export, it faces iceberg variable costs of trade, such that $\tau_{ij} > 1$ units of a variety must be exported from country i to country j in order for one unit to arrive.

Since the marginal utility of consuming a differentiated variety is finite at zero consumption, firm exit occurs even in the absence of fixed production costs. Firms drawing a marginal cost above the threshold price (27) in the domestic market exit in equilibrium, because they cannot generate positive profits from production. In the closed economy, the zero-profit cost cutoff (c_{Di}) is a sufficient statistic that completely summarizes the competitive environment and determines firm outcomes as a function of their cost draw (c):

$$
\begin{aligned}
p_i(c) &= \frac{1}{2}\left(c_{Di} + c\right) && \text{prices} \\
\mu_i(c) &= p_i(c) - c = \frac{1}{2}\left(c_{Di} - c\right) && \text{markups} \\
r_i(c) &= \frac{L_i}{4\gamma}\left[(c_{Di})^2 - c^2\right] && \text{revenues} \\
\pi_i(c) &= \frac{L_i}{4\gamma}\left(c_{Di} - c\right)^2 && \text{profits.}
\end{aligned}
$$

Relative to other firms, more productive firms (with lower c) have lower prices ($p_i(c)$), higher mark-ups ($\mu_i(c)$), higher output and revenue ($r_i(c)$) and higher profits ($\pi_i(c)$). Firms with lower marginal cost charge higher mark-ups because their marginal cost intersects marginal revenue at a more inelastic segment of the demand curve. Since more productive firms do not fully pass on their lower marginal costs to consumers, they have higher revenue-based productivity ($r_i(l)/l_i(c)$) even in the absence of fixed production costs.

Under the assumption that productivity ($1/c$) is Pareto distributed with lower bound $1/c_M$ and shape parameter k, the closed economy cost cutoff is given by:

$$
c_{Di} = \left(\frac{\gamma\phi}{L_i}\right)^{\frac{1}{k+2}},
$$

where $\phi = 2(k+1)(k+2)c_M^k f_E$ is an (inverse) index of technology. The closed economy cost cutoff falls (higher average productivity) when varieties are closer substitutes (γ falls), when there is a better distribution of cost draws (c_M falls), when sunk costs fall (lower f_E) and in bigger markets (higher L_i). Each of these comparative statics induces an increase in the "toughness of competition" in the form of a larger number of varieties consumed (higher N_i) and lower average prices (lower $\bar{p}_i$).

The intuition for these effects of market size is as follows. As market size increases, the number of varieties consumed rises, which shifts the demand curve for each variety inwards, so that some high-cost firms exit and each surviving firm prices on a more elastic segment of its demand curve. Therefore, larger markets are characterized by lower prices, both because of higher average productivity (a lower zero-profit cost cutoff c_{Di}) and lower mark-ups for a firm with a given productivity. Larger markets are also characterized by larger firms, because of both the exit of smaller high-cost firms and an increase in the sales of a firm with a given productivity. Consumers in larger markets also enjoy higher welfare, because of both greater product variety and lower prices. While lower zero-profit cutoffs in larger markets reduce the dispersion in productivity, prices and mark-ups (by compressing the range of firm costs $[0, c_{Di}]$), they increase the dispersion of firm size in terms of both output and revenue, which is consistent with the empirical findings in Campbell and Hopenhayn (2005) and Syverson (2004).[26]

In the open economy, markets are segmented and the production technology exhibits constant returns to scale. Therefore the supplier of each differentiated variety maximizes independently the profits earned from domestic and export sales.[27] A finite marginal utility of consuming each variety at zero consumption implies that fixed exporting costs are not needed to generate selection into export markets. A firm's marginal costs may lie below the threshold price in the domestic market, but may be above the threshold price in the foreign market once variable trade costs are taken into account, in which case the firm serves only the domestic market.

In an open economy equilibrium with symmetric trade costs ($\tau_{ni} = \tau > 1$ for all $n \neq i$ and $\tau_{ii} = 1$), the zero-profit and exporting cost cutoffs are given by:

$$\begin{aligned} c_{Di} &= \left\{ \frac{\gamma \phi}{L_i \left[1 + (M-1)\,\tau^{-k}\right]} \right\}^{\frac{1}{k+2}}, && i \in \{1, \dots, M\}, \\ c_{Xni} &= \frac{c_{Dn}}{\tau}, && i, n \in \{1, \dots, M\}. \end{aligned} \tag{28}$$

[26]See Combes et al. (2012) for evidence on the contributions of agglomeration and selection towards the higher productivity of larger cities.

[27]In equilibrium, prices are such that there are no profitable arbitrage opportunities across markets, because firms absorb a portion of the trade cost difference across markets in lower export market prices (dumping).

Therefore costly trade does not completely integrate markets and market size differences affect productivity cutoffs and have qualitatively similar effects as in the closed economy. Larger markets attract more firms, which implies "tougher" competition in the presence of trade costs, and hence leads to a lower zero-profit cost cutoff and higher average productivity.

Multilateral trade liberalization (a reduction in the common value of τ) again causes intra-industry reallocation by reducing the zero-profit cost cutoff in (28), which induces low productivity firms to exit and shifts the composition of output towards more productive firms. However, the presence of variable mark-ups introduces a new "pro-competitive effect" of multilateral trade liberalization. As the zero-profit cost cutoff falls in response to trade liberalization, this reduces the mark-up charged by a firm of each productivity, so that prices fall because of both higher average productivity (a lower zero-profit cost cutoff c_{Di}) and lower mark-ups for a firm with a given productivity. This pro-competitive effect is consistent with empirical evidence from trade liberalization episodes (see for example the survey by Tybout 2003) and introduces a new mechanism through which welfare is affected by trade in addition to changes in product variety and average productivity.

Given the presence of an outside sector, the model also features a home market effect, which influences the effects of unilateral and preferential trade liberalization. In the short-run, holding the number of firms in each country fixed, all countries experience welfare gains from unilateral or preferential trade liberalization. In the long-run, once the number of firms in each country is allowed to adjust, this is no longer the case. If one country unilaterally reduces its import barriers, it can experience welfare losses, as production relocates to other countries to access these markets without trade costs and take advantage of the lower import barriers in the liberalizing country.

While we focus on the quasi-linear-quadratic demand system as a particularly tractable framework in which to examine the effects of trade in the presence of firm heterogeneity and variable mark-ups, other research has considered the Constant Absolute Risk Aversion (CARA) preferences of Behrens and Murata (2012) and the translog preferences of Feenstra (2003). Taking a different approach, Edmond, Midrigan and Xu (2012) introduce variable mark-ups into a CES demand system by considering the case of a finite number of firms.

In these models of firm heterogeneity with variable mark-ups, the impact of changes in variable trade costs on the distribution of prices depends in an important way on the productivity distribution. Under the assumption that productivity is Pareto distributed, the distribution of prices conditional on purchasing a variety is invariant to changes in variable trade costs, such that the welfare gains from trade can be expressed in terms of a country's trade share with itself and a parametric correction for variable mark-ups (see Arkolakis et al. 2012b).

8 Endogenous Productivity

Up to now, we have assumed that firm productivity is exogenously set at entry and therefore does not respond endogenously to trade liberalization. Recent research has focused on numerous extensions where firms can affect their productivity via decisions regarding the range of products produced, innovation and technology adoption, and how production is organized. These decisions, in turn, are affected by the trading environment and the firm's trade participation. This induces a complementarity between choices regarding trade and firm productivity.

In this section, we begin by developing some of these extensions regarding product scope (and the emergence of multi-product firms), innovation, and technology adoption within a static model. Firms make a one-time joint decision regarding this additional characteristic along with the production and trade decisions that we have previously analyzed. Extensions with respect to the organization of production are discussed in another chapter of this handbook. We next transition to consider some dynamic models in order to analyze the joint evolution of firm productivity and export market participation over time. In these models, firm productivity can evolve due to exogenous shocks, but also as an outcome of endogenous innovation or technology adoption decisions. These dynamic models jointly capture the complementarity between firm productivity and trade in the cross-section as well as over time: the decision to export at one point in time is linked to other decisions regarding innovation or technology adoption at another point in time.

A key feature of all the models covered in this section is that trade liberalization can raise firm-level productivity as well as generate increases in aggregate productivity via between-firm reallocations of resources (as analyzed in previous sections).

Product Scope Decision and Multi-Product Firms

One of the striking features of international trade is the extent to which it is concentrated in the hands of a relatively small number of firms supplying many products to many destinations. For example, Bernard, Jensen, Redding and Schott (2007) report using U.S. data that firms exporting more than five ten-digit products to more than five destinations account for only around 12 percent of exporters but 92 percent of export value.[28] Motivated by such evidence, a growing body of theoretical and empirical research has sought to model the implications of multi-product, multi-destination firms for understanding aggregate and disaggregate patterns of trade.[29]

The model of firm heterogeneity in differentiated product markets developed in Section 2 admits a natural generalization to incorporate multi-product firms, as explored in Bernard, Redding and Schott (2011). Suppose that the representative consumer derives utility from the consumption of a continuum of symmetric products h defined on the interval $[0,1]$:

$$U = \left[\int_0^1 C_h^{\nu} dh\right]^{\frac{1}{\nu}}, \qquad 0 < \nu < 1.$$

Within each product, a continuum of firms supply differentiated varieties of the product. Incurring the sunk entry cost f_E creates a firm brand, which can be used to supply one horizontally-differentiated variety of each of the continuum of products. Varieties are assumed to be differentiated from one another by their brand, which implies that a given brand cannot be used to supply more than one differentiated variety of each product. After incurring the sunk entry cost, a firm observes realizations of two stochastic shocks to profitability: "ability" $\varphi \in (0, \infty)$, which is common to all products and drawn from a distribution $g(\varphi)$, and "product attributes " $\lambda_h \in (0, \infty)$, which are specific to each product h and possibly to each destination market and drawn from a distribution $z(\lambda)$. A firm in each country i faces a fixed cost of supplying each market n (F_{ni}) and an additional fixed cost of supplying each product to that market (f_{ni}).

Sectoral equilibrium can be determined using a similar approach as in Section 2. There is a product attributes cutoff for each firm ability φ ($\lambda_{ni}^*(\varphi)$) above which a firm can profitably export a product from country i to market n:

$$\pi_{ni}(\varphi, \lambda_{ni}^*(\varphi)) = \frac{r_{ni}(\varphi, \lambda_{ni}^*(\varphi))}{\sigma} - f_{ni} = 0.$$

There is also a firm ability cutoff (φ_{ni}^*) above which a firm can generate enough total variable profits from exporting its range of profitable products from country i to country n to cover the fixed costs

[28] Similar findings have emerged from a wide range of countries, as summarized in Mayer and Ottaviano (2007) and World Trade Organization (2008).

[29] While early research on multi-product firms and trade, such as Ottaviano and Thisse (1999) and Allanson and Montagna (2005), modeled firms and products symmetrically, more recent research has emphasized heterogeneity both within and across firms.

of serving market n (F_n):

$$\pi_{ni}\left(\varphi_{ni}^{*}\right)=\int_{\lambda_{ni}^{*}\left(\varphi_{ni}^{*}\right)}^{\infty}\pi_{ni}\left(\varphi_{ni}^{*},\lambda\right)z\left(\lambda\right)d\lambda-F_{ni}=0.$$

Higher ability firms can generate sufficient variable profits to cover the product fixed cost at a lower value of product attributes and, therefore, supply a wider range of products to each market. For sufficiently low values of firm ability, the excess of variable profits over product fixed costs in the small range of profitable products does not cover the fixed cost of serving the market and therefore the firm does not supply the market. The lowest-ability firms exit, intermediate-ability firms serve only the domestic market and the highest ability firms export. Within exporters, products with the worst attributes are supplied only to the domestic market, while products with the best attributes are exported to the largest number of markets.

This theoretical framework features selection both within and across firms. Trade liberalization raises average industry productivity, not only through the exit of the least productive firms, but also through surviving firms dropping their least-successful products. Consistent with these predictions, U.S. firms more exposed to tariff reductions under the Canada-U.S. Free Trade Agreement reduce the number of products they produce relative to firms less exposed to these tariff reductions.[30] Selection within and across firms implies that changes in variable trade costs affect aggregate exports through the extensive margins of the number of exporting firms, the number of exported products and average exports per firm and product. The effects on average exports per firm and product are ambiguous, because higher variable trade costs reduce exports of a given firm and product but change export composition away from firms and products with small export values. Consistent with these predictions, the negative effect of distance on aggregate trade flows in the gravity equation is largely accounted for by the extensive margins of firms and products, with the effect on the intensive margin of average exports per firm and product positive but not statistically significant.

Mayer et al. (2012) introduce multi-product firms into the model of firm heterogeneity with variable mark-ups developed in Section 7. Firms face a product ladder, according to which productivity/quality declines discretely for each additional variety produced. Differences in the toughness

[30]Evidence of similar product rationalization in Canada is found by Baldwin and Gu (2009). Product adding and dropping is shown to be an important source of aggregate reallocation in Bernard, Redding and Schott (2010). Around one half of surviving U.S. firms add and/or drop products from their existing range every five years, and the contribution of these added and dropped products to aggregate output is of around the same magnitude as the contribution of firm entry and exit.

of competition across markets induces changes to both the extensive and intensive product margin within firms. Mayer et al. (2012) focus on the effects of competition on the intensive product margin. Due to the variable price elasticities, firms selling the same set of products in different markets skew their sales towards their best performing products in markets where they face tougher competition (due to the higher price elasticities in those markets). Data on French exporters across export market destinations provides strong empirical confirmation of this competitive effect.[31]

Another source of competition arises when the monopolistic competition assumption is dropped and firms internalize the effects of new products on the sales of their existing products (a cannibalization effect). Eckel and Neary (2010) develop such a model and highlight how this cannibalization effect generates an additional incentive for multi-product firms to drop their worst performing products when faced with increased competition from trade. Thus, trade liberalization generates higher firm productivity and, potentially, lower product variety.[32]

Innovation

Recent empirical work has consistently found that exporters (relative to non-exporters) are significantly more likely to innovate and adopt new technologies. For example, Verhoogen (2008) reports that Mexican exporters (plants) are more likely to be ISO 9000 certified (a proxy for the use of more advanced production techniques within the plant); and Bustos (2011) reports that Argentinian exporters (firms) spend more on new technologies (per worker).

As trade liberalization induces firms to start exporting, it is also associated with increased innovation and technology use by those new exporters. Bustos (2011) finds that the Mercosur trade liberalization agreement generated substantial increases in spending on new technologies by new exporters (and some increased spending by existing exporters). Verhoogen (2008) finds that the Mexican peso devaluation in the 1990s induced substantial increases in both plant exports and ISO 9000 certification. Lileeva and Trefler (2010) use econometric techniques to identify the effect of lower U.S. import tariffs on the innovation and technology adoption rates of new Canadian exporters. They find those tariff cuts (part of the CUSFTA trade agreement) induced higher rates of product innovation and of advanced manufacturing technologies by new Canadian exporters. Those changes, in turn, led to very large increases in labor productivity for those new exporters:

[31] See Arkolakis and Muendler (2012) and Nocke and Yeaple (2006) for other monopolistically competitive models of multi-product firms.

[32] See also Feenstra and Ma (2008) and Dhingra (2010) for other models of multi-product firms and trade featuring cannibalization effects.

over 15% between 1984 and 1996.

In the following two subsections, we describe two modeling techniques to capture the joint innovation and export decisions by heterogeneous firms. The first technique deals with a binary innovation choice (such as technology adoption) while the latter captures a continuous innovation intensity decision.

A Binary Innovation Choice: Technology Adoption

We briefly sketch how to add a binary technology adoption choice alongside the production and export decisions of heterogeneous firms. Bustos (2011) fully develops this theoretical modeling extension.[33] Every firm with productivity φ has the choice of upgrading to a new technology. This involves a tradeoff between an additional fixed cost f_I and a productivity increase to $\iota\varphi$, where the proportional productivity increase $\iota > 1$ is the same for all firms. Just like the export decision, this technology adoption choice involves a tradeoff between a fixed cost and a per-unit profit increase. Therefore technology adoption is characterized by a similar sorting according to firm productivity, such that there is a productivity cutoff φ_I^* above which all firms adopt the new technology.[34] The ranking of the export and innovation cutoffs φ_X^* and φ_I^* (assuming symmetric trade and production costs, so there is a single export cutoff) depends on the innovation parameter values f_I and ι, the trade costs f_X and τ, and the overhead production cost f. Bustos (2011) provides the conditions such that $\varphi_I^* > \varphi_X^*$, which implies that some exporters do not innovate (which is the empirically relevant case for the Argentinian data). In any event, this modeling framework implies that the most productive firms will choose to both innovate and export, while firms of lower productivity choose to do neither, and the least productive firms exit. This framework therefore captures the correlation between trade and innovation that is so prominent empirically in the cross-section of firms.

Several other firm decisions have been modeled in a similar way as involving a simple tradeoff between a fixed cost and a benefit that scales with firm size. Manova (2006) uses a similar type of tradeoff to examine the financing choice of firms and how it interacts with their export decision. Another line of work focuses on the decision to import intermediate inputs, as in Amiti and Davis (2012), Gopinath and Neiman (2012), and Goldberg et al. (2010). More generally, other research

[33] This model extension, in turn, is based on previous work by Yeaple (2005).

[34] Firm profits are log-supermodular in productivity, technology adoption, and export status, leading to this strict sorting behavior. This is a specific example of the more general case analyzed by Costinot (2009) where firms or factors can sort into multiple different activities (see the working paper version of that paper for a more detailed derivation of the firm-sorting case).

examines the choice of firm organization and how it interacts with firm productivity and export status; see Caliendo and Rossi-Hansberg (2012).

Innovation Intensity

We now turn to the modeling of innovation intensity allowing for continuous differences in the level of innovation performed by different firms. Following the seminal contributions by Aghion and Howitt (1992), Grossman and Helpman (1991) and Romer (1990), there has been a long literature analyzing how market size and globalization affect the firm innovation intensity choice. Initially, this literature did not focus on cross-sectional differences in innovation intensity across firms; more recently, Atkeson and Burstein (2010) have built a model featuring variations in innovation intensity across firms and over time in response to globalization shocks. We discuss the introduction of firm dynamics in the next section but first sketch out a static version of the innovation intensity decision used by Atkeson and Burstein (2010).

Consider a rescaling of firm productivity $\phi = \varphi^{\sigma-1}$ such that this new productivity measure ϕ is now proportional to firm size.[35] As with the case of the binary innovation choice, we assume that successful innovation increases productivity by a fixed factor $\iota > 1$ (from ϕ to $\iota\phi$). However, the probability of successful innovation is now an endogenous variable α that reflects a firm's innovation intensity choice. The cost of higher innovation intensity is determined by an exogenous convex function $c_I(\alpha) \geq 0$ and scales up proportionally with firm size and productivity ϕ, so the total cost of innovation intensity α is $\phi c_I(\alpha)$. This scaling up of innovation cost with firm size is needed in a dynamic setting to deliver the prediction of Gibrat's Law that growth rates for large firms are independent of their size.

We first examine the choice of innovation intensity in a closed economy. Consider a firm with productivity ϕ that is sufficiently high that the firm will produce even if innovation is unsuccessful. This firm will choose innovation intensity α to maximize expected profits

$$E[\pi(\phi)] = [(1-\alpha) + \alpha\iota]\, B\phi - \phi c_I(\alpha) - f,$$

where B is the same market demand parameter for the domestic economy as in previous sections.

[35] Since the rescaling involves the demand side product differentiation parameter σ, caution must be used when interpreting any comparative statics that include this parameter.

The first-order condition is given by

$$c_I'(\alpha) = (\iota - 1)\, B. \tag{29}$$

This implies that, in the closed economy, all firms (above a certain productivity threshold satisfying the no exit restriction) will choose the same innovation intensity α. In a dynamic setting, this delivers Gibrat's Law for those firms, and also generates an ergodic distribution of firm productivity (hence firm size) that is Pareto in the upper tail independently of the initial distribution of productivity upon entry.[36]

Consider now the innovation intensity choice in an open economy setting with two symmetric countries (and symmetric trade costs) and selection into export markets. The first-order condition for non-exporters will still be given by (29). However, successful innovation is more valuable to exporters because it will generate additional profits from export sales. Their first-order condition is given by:[37]

$$c_I'(\alpha) = (\iota - 1)\, B(1 + \tau^{1-\sigma}).$$

Thus, exporters will choose a higher innovation intensity than non-exporters. As with non-exporters, all large exporters (firms who will export regardless of the innovation outcome) will choose the same innovation intensity. This modeling of innovation intensity can therefore also replicate the complementarity between innovation and trade. It also offers a particularly tractable way of incorporating endogenous innovation into a dynamic model of trade and innovation, such as the one analyzed by Atkeson and Burstein (2010).

Dynamics

All the models that we have considered up to now have been static. They contrast an ex-ante period (featuring idiosyncratic firm uncertainty) with a single ex-post period where all uncertainty is lifted, firms jointly make all their decisions, and profit is earned. One can also think of this outcome as the stationary equilibrium of a dynamic model where the aggregate conditions remain constant over time. Melitz (2003) describes a simple version of such a stationary equilibrium, where firms face a single additional source of idiosyncratic uncertainty: a death shock that occurs with

[36] See the web appendix. The exact shape of the ergodic distribution is sensitive to whether Gibrat's Law holds for all productivities or only for productivities above a certain threshold. See Luttmer (2010) for a review of this literature.

[37] This is the first-order condition for firms who will export regardless of whether innovation is successful. The condition for a firm whose export decision is tied to innovation success would be different.

probability $\delta \in (0,1)$ and is independent of firm productivity φ. The key free entry and zero cutoff profit conditions that we have previously described are then very similar. In those conditions, firm profit is replaced by firm value, which is just the net present value of the non-fluctuating profits earned in every period (the death shock generates a discount factor for the value computation). As in the static version of the model, the sunk nature of the entry cost f_E is a critical component for delivering ex-post firm heterogeneity. On the other hand, the modeling of the fixed export cost as either sunk or paid in every export period does not affect the stationary equilibrium: there is no uncertainty regarding the export market so firms are indifferent between paying an overhead fixed export cost in every period or its net present value once prior to exporting for the first time. Any uncertainty regarding future export profits will break this equivalence. Sunk export costs then generate hysteresis behavior associated with export market entry and exit.[38]

The combination of sunk entry costs and uncertainty leads to option values associated with entry and exit (manifested by hysteresis). Although there is substantial empirical support for this type of behavior, the modeling of those option values in a dynamic general equilibrium model with heterogeneous firms adds some substantial complexity. As an alternative, significant gains in tractability can be achieved by analyzing dynamic versions of the model that do not feature firm option values. For example, assuming that the fixed export cost is paid per period (and not sunk) will eliminate the option value associated with export market entry/exit. The sunk entry cost f_E must be preserved to generate ex-post heterogeneity; however, if the overhead production cost f is eliminated, then firm exit is exogenously determined by the death shock δ (and not endogenously due to low productivity), and the option value associated with entry/exit is eliminated. Ghironi and Melitz (2005) make these assumptions and then embed the steady-state version of Melitz (2003) into a two country dynamic model featuring stochastic fluctuations in aggregate variables (a standard DSGE open economy model).

This type of modeling allows firm productivity to change over time due to changes in aggregate productivity, but the relative productivity of firms remains constant. Other models have incorporated sources of firm-level fluctuations such as idiosyncratic productivity shocks along side the aggregate fluctuations. Alessandria and Choi (2007) and Ruhl (2004) use variants of this type of model to analyze the growth dynamics of exporters in response to changes in trade costs. They

[38]See Baldwin (1988) for an early theoretical derivation of this hysteresis effect. Roberts & Tybout (1997) find strong evidence of such behavior for Colombian exporters. Subsequent firm-level empirical work has confirmed this effect for other countries. Das et al (2007) use the same dataset of Colombian exporters and develop estimation methods to recover the magnitude of the sunk export cost.

characterize both the firm-level responses as well as the aggregate trade response.[39] More recently, the innovation choice decision described in the previous subsection has been incorporated into these dynamic models. There is then both an endogenous (innovation) and exogenous component to the evolution of firm productivity. Aw et al. (2011) estimate this joint model using production, trade, and R&D data for Taiwanese firms. They find that endogenous productivity changes via R&D are needed to explain the joint evolution of productivity and export decisions observed in the data.

Atkeson and Burstein (2010) analyze how modelling endogenous innovations in firm productivity influences the overall welfare gains from trade liberalization. Lower trade barriers boost innovation by current and prospective exporters; however this also reduces the expected profits of new entrants (who must then compete against much larger and more productive incumbents) and thus reduces entry. Atkeson and Burstein (2010) calibrate their model to U.S. firm level data and show that these two effects are largely offsetting, so that there is no substantial effect of trade liberalization on the growth rate of welfare – even when innovation is endogenous. Burstein and Melitz (2012) use a very similar model, but focus on the transition paths following trade liberalization. They highlight how firm productivity dynamics and export market selection combine to generate long lasting adjustments to one-time changes in trade costs. Constantini and Melitz (2007) use a binary innovation choice to analyze the timing of the innovation decision relative to the export decision. They show that this relative timing is very sensitive to the timing and anticipation of trade liberalization. Productivity increases following export market need not imply a learning by exporting externality, but may rather reflect firms' joint export and innovation decisions. A more general insight is that measured firm productivity is the outcome of a number of endogenous decisions which are taken jointly with trade participation (including both exporting and importing). The contemporaneous relationship between productivity and trade participation therefore reflects complex interactions between these decisions over time and should be interpreted with caution.

Although much of the literature on firm dynamics and trade has been focused on productivity, an emerging literature considers dynamics generated by demand-side considerations. In contrast to productivity shocks, which affect firm profitability in all markets, demand shocks generate market-specific fluctuations in profitability. One strand of research emphasizes learning about uncertain demand in markets as in Albornoz et al. (2012) and Akhmetova (2012). Another line of work explores how matches between buyers and sellers evolve over time and across markets, as in Eaton

[39] Arkolakis (2011) and Irarrazabal & Opromolla (2008) analyze the stationary equilibrium in similar models. They characterize the steady-state distribution of different types of firms and cohort dynamics in that equilibrium.

et al. (2012) and Chaney (2012). These papers all seek to explain empirical patterns of firm entry and exit across export market destinations and over time. These empirical patterns include high rates of firm exit from new export destinations as well as rapid firm export growth conditional on survival in these destinations.

9 Factor Markets

A key implication of models of firm heterogeneity in differentiated product markets is that firms are unevenly affected by trade liberalization: low-productivity firms exit, intermediate-productivity domestic firms contract, and high-productivity exporting firms expand. In contrast, in the benchmark model of firm heterogeneity, workers are symmetrically affected by trade liberalization, because workers are identical and the labor market is frictionless, with the result that all workers are employed for a common wage. These labor market implications contrast with the large empirical literature that finds an employer-size wage premium (see, for example, the survey by Oi and Idson 1999) and with empirical findings of wage differences between exporters and non-exporters even after conditioning on firm size (see, in particular, Bernard and Jensen 1995, 1997).

More recent research on firm heterogeneity and trade has highlighted two sets of reasons why wages can differ across firms. One line of research assumes competitive labor markets, so that all workers with the same characteristics are paid the same wage, but wages can differ across firms because of differences in workforce composition (see for example Bustos 2007, Sampson 2012, Verhoogen 2008 and Yeaple 2005). Another line of research introduces labor market frictions, so that workers with the same characteristics can be paid different wages by different firms. One source of such labor market imperfections is search and matching frictions, which can generate variation in wages with firm revenue through bargaining over the surplus from production (see for example Davidson et al. 2008, Davidson and Matusz 2009, Cosar et al. 2011, and Helpman et al. 2010). Another source of labor market imperfections is efficiency or fair wages, which can generate similar wage variation if the wage that induces effort or is perceived to be fair varies with firm revenue (see for example Amiti and Davis 2012, Davis and Harrigan 2011, and Egger and Kreickemeier 2009).

This class of theoretical models highlights a new mechanism for trade to affect wage inequality based on wage variation across firms and the selection of firms into international trade. As shown in Helpman et al. (2010), the opening of the closed economy to trade necessarily raises within-industry wage inequality within a class of models satisfying three sufficient conditions: (a) wages and employment are power functions of productivity, (b) only some firms export and exporting raises the wage paid by a firm with a given productivity, (c) productivity is Pareto distributed. When these three conditions are satisfied, the wage and employment of firms can be expressed in terms of their productivity (φ) a term capturing whether or not a firm exports ($\Upsilon(\varphi)$) the zero-profit cutoff productivity (φ^*) and parameters:

$$l(\varphi) = \Upsilon(\varphi)^{\psi_l} l_d \left(\frac{\varphi}{\varphi^*}\right)^{\zeta_l},$$
$$w(\varphi) = \Upsilon(\varphi)^{\psi_w} w_d \left(\frac{\varphi}{\varphi^*}\right)^{\zeta_w},$$

where l_d and w_d are employment and wage of the least productive firm and:

$$\Upsilon(\varphi) = \begin{cases} \Upsilon_x > 1 & \text{for } \varphi \geq \varphi_X^* \\ 1 & \text{for } \varphi < \varphi_X^* \end{cases},$$

where Υ_x is the revenue premium from exporting for a firm of a given productivity. Using the Pareto productivity distribution, the distribution of wages across workers within the industry, $G_w(w)$, can be evaluated as:

$$G_w(w) = \begin{cases} S_{l,d} G_{w,d}(w) & \text{for} \quad w_d \leq w \leq w_d (\varphi_x/\varphi_d)^{\zeta_w}, \\ S_{l,d} & \text{for} \quad w_d (\varphi_x/\varphi_d)^{\zeta_w} \leq w \leq w_x, \\ S_{l,d} + (1 - S_{l,d}) G_{w,x}(w) & \text{for} \quad w \geq w_x, \end{cases}$$

where $w_x = w_d \Upsilon_x^{\psi_w} (\varphi_x/\varphi_d)^{\zeta_w}$ is the wage of the least productive exporter and $S_{l,d}$ is the employment share of domestic firms. The distribution of wages across workers employed by domestic firms, $G_{w,d}(w)$, is a truncated Pareto distribution:

$$G_{w,d}(w) = \frac{1 - \left(\frac{w_d}{w}\right)^{\zeta_g}}{1 - \left(\frac{w_d}{w_x}\right)^{\zeta_g}} \qquad \text{for } w_d \leq w \leq w_d (\varphi_x/\varphi_d)^{\zeta_w},$$

while the distribution of wages across workers employed by exporters, $G_{w,x}(w)$, is an un-truncated Pareto distribution:

$$G_{w,x}(w) = 1 - \left(\frac{w_x}{w}\right)^{\zeta_g} \qquad \text{for } w \geq w_x,$$

In the class of models satisfying the above three sufficient conditions, Helpman et al. (2010) show that there is strictly greater wage inequality in the open economy when only some firms export than in the closed economy, and there is the same level of wage inequality in the open economy when all firms export as in the closed economy. It follows that wage inequality is at first increasing in trade openness and later decreasing in trade openness. The intuition for these results stems from the increase in firm wages that occurs at the productivity threshold above which firms export, which is only present when some but not all firms export. When no firm exports, a small *reduction* in trade costs increases wage inequality, because it induces some firms to start exporting and raises the wages paid by these exporting firms relative to domestic firms. When all firms export, a small *rise* in trade costs increases wage inequality, because it induces some firms to stop exporting and reduces the wages paid by these domestic firms relative to exporting firms.

Helpman et al. (2012) provide evidence on the quantitative importance of this new mechanism for understanding the relationship between wage inequality and trade using Brazilian employer-employee and trade transaction data. Consistent with the class of theoretical models discussed above, wage inequality between firms within sector-occupations accounts for a substantial proportion of the level and growth of overall wage inequality, and this between-firm wage inequality remains important after controlling for observable worker characteristics. Estimating an extended version of the structural model discussed above, they find that the model has substantial explanatory power for the distribution of wages across both firms and workers. To the extent that existing empirical studies inspired by neoclassical trade theory focus on changes in relative wages between different sectors and types of workers, they abstract from an important channel through which trade liberalization can affect wage inequality.

Labor market frictions can also generate equilibrium unemployment. In this case, the opening of trade can affect the distribution of income not only through the distribution of wages across employed workers but also through changes in unemployment. Helpman and Itskhoki (2010) stress the impact of trade on unemployment through reallocations of resources across sectors while Felbermayr et al. (2011) emphasize the role of the productivity gains from trade liberalization in reducing effective search costs. Helpman & Itskhoki (2010) consider a two-country, two-sector model of international trade, in which one sector produces a homogeneous product, while the other sector produces differentiated products, and both are subject to search frictions. Differences in labor market institutions across countries and industries provide a source of comparative advantage and shape the impact of trade liberalization on aggregate unemployment.[40] Reductions in a

[40] Another setting in which cross-country differences in labor market institutions can provide a source of comparative advantage is where volatility varies across sectors, as in Cuñat & Melitz (2010).

country's labor market frictions in the differentiated sector raise its own welfare, by expanding the size of its differentiated sector and reducing its differentiated-sector price index. This expansion in the differentiated sector in one country intensifies competition in the export market faced by firms in the other country's differentiated sector. As a result, the other country's differentiated sector contracts, which reduces its welfare. In contrast, proportional reductions in labor market frictions in the differentiated sector in both countries raise welfare in each country, by expanding the size of the differentiated sector in each country.

10 Conclusions

Traditional theories of international trade emphasize the international exchange of distinct commodities. In these theories, welfare gains from trade arise from cross-country differences in the opportunity costs of producing goods. Trade liberalization induces between-industry reallocations of resources that change relative factor demands. As a result, developed countries that are abundant in skilled labor experience increased wage inequality, while developing countries that are abundant in unskilled labor experience reduced wage inequality. In the 1980s, a "New Trade Theory" introduced product differentiation, returns to scale, and imperfect competition within industries into models of trade. This theory uncovered new channels for the gains from trade: By creating a larger marketplace, trade delivered a combination of increased product variety and lower prices to consumers.

Over the past decade, this theory has been extended to incorporate within-industry reallocations driven by trade (affecting firms and workers within a sector in different ways). In this framework with firm (and potentially worker) heterogeneity, trade liberalization induces within-industry reallocations of resources from less to more productive firms, while the entry of firms into export markets stimulates innovation. The resulting increases in average industry productivity through both a change in industry composition and higher firm productivity provide the source of welfare gains from trade. To the extent that wages vary across firms, these within-industry reallocations of resources provide a new mechanism through which trade can increase wage inequality in both developed and developing countries.

This literature on firm heterogeneity provides a theoretical rationale for a number of features of aggregate and disaggregate trade data. In line with empirical evidence on firms and plants, exporters are more productive than non-exporters and trade liberalization induces the exit of the least productive suppliers and raises aggregate productivity. Consistent with empirical results from trade transactions data, the number of firms serving an export market increases with market size, and average sales in domestic markets are higher for firms selling to more foreign markets and to less popular foreign markets. At the aggregate level, these models provide a fresh perspective on long-studied relationships such as the gravity equation, highlighting the contribution of the extensive margin of exporting firms, and also accounting for the prevalence of zeros in trade flows.

While models of firm heterogeneity rationalize features of disaggregated trade data and highlight new mechanisms through which trade affects welfare, their aggregate implications for the overall welfare gains from trade are more subtle. Introducing new channels through which trade can affect welfare does not necessarily change the overall welfare gains from trade, because these new channels can lead to offsetting adjustments in existing mechanisms. An active research agenda explores the circumstances under which the welfare gains from trade can be inferred from aggregate statistics such as a country's trade share with itself. Even when these conditions hold, different mechanisms point to very different adjustments to trade liberalization, and can have quite different income distributional consequences through wage inequality and unemployment.

References

[1] Aghion P, Howitt P. 1992. A Model of Growth through Creative Destruction. *Econometrica*. 60(2): 323-51.

[2] Akhmetova Z. 2011. Firm Experimentation in New Markets. (University of New South Wales, Unpublished).

[3] Albornoz F, Calvo Pardo H, Corcos G, Ornelas E. 2012. Sequential Exporting. *Journal of International Economics*, forthcoming.

[4] Alessandria G, Choi H. 2007. Do Sunk Costs of Exporting Matter for Net Export Dynamics? *Quarterly Journal of Economics*. 122(1): 289-336.

[5] Allanson P, Montagna C. 2005. Multi-product Firms and Market Structure: An Explorative Application to the Product Lifecycle. *International Journal of Industrial Organization*. 23: 587-597.

[6] Alvarez F and Lucas R. 2007. General Equilibrium Analysis of the Eaton and Kortum Model of International Trade. *Journal of Monetary Economics*. 54: 1726-1768.

[7] Amiti M, Davis DR. 2012. Trade, Firms, and Wages: Theory and Evidence. *Review of Economic Studies*, 79(1): 1-36.

[8] Anderson J, van Wincoop E. 2002. Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*. 93(1): 170-92.

[9] Anderson, J and Van Wincoop E. 2004. Trade Costs. *Journal of Economic Literature.* 42: 691–751.

[10] Antràs P. 2003. Firms, Contracts, and Trade Structure. *Quarterly Journal of Economics*, 118(4), 1375-1418.

[11] Antràs P, Helpman E. 2004. Global Sourcing. *Journal of Political Economy.* 112(3): 552-80.

[12] Antràs P, Yeaple S. 2012. International Organization of Production. In *Handbook of International Economics*, Volume 4, ed. E Helpman, G Gopinath, K Rogoff.

[13] Arkolakis C. 2010. Market Penetration Costs and the New Consumers Margin in International Trade. *Journal of Political Economy*, 118(6): 1151-1199.

[14] Arkolakis C. 2011. A Unified Theory of Firm Selection and Growth. (Unpublished, Yale University).

[15] Arkolakis C, Klenow P, Demidova S, Rodriguez-Clare A. 2008. Endogenous Variety and The Gains from Trade. *American Economic Review.* Papers and Proceedings. 98(2): 444-50.

[16] Arkolakis C, Costinot A, Rodriguez-Clare A. 2012a. New Trade Models, Same Old Gains? *American Economic Review.* 102(1): 94-130.

[17] Arkolakis C, Costinot A, Donaldson D, Rodriguez-Clare A. 2012b. The Elusive Pro-Competitive Effects of Trade. (Unpublished, MIT).

[18] Arkolakis C, Muendler M. 2012. The Extensive Margin of Exporting Goods: A Firm-Level Analysis. (Yale University, unpublished).

[19] Atkeson A, Burstein A. 2010. Innovation, Firm Dynamics and International Trade. *Journal of Political Economy.* 118(3): 433-484.

[20] Aw BY, Roberts MJ, Xu DY. 2011. R&D Investment, Exporting, and Productivity Dynamics. *American Economic Review.* 101: 1312-1344.

[21] Axtell RL. 2001. Zipf Distribution of US Firm Sizes. *Science.* 293:1818-20.

[22] Baldwin R. 1988. Hysteresis in Import Prices: The Beachhead Effect. *American Economic Review*, 78(4): 773-785.

[23] Baldwin RE, Forslid R. 2010. Trade Liberalization with Heterogeneous Firms. *Review of Development Economics*, 14(2): 161-176.

[24] Baldwin J, Gu W. 2009. In *The Impact of Trade on Plant Scale, Production-Run Length and Diversification in Producer Dynamics: New Evidence from Micro Data*, ed. T Dunne, JB Jensen and MJ Roberts. Chicago: University of Chicago Press.

[25] Baldwin RE, Harrigan J. 2011. Zeros, Quality, and Space: Trade Theory and Trade Evidence. *American Economic Journal: Microeconomics*, 3(2): 60–88.

[26] Balistreria EJ, Hillberry RH, Rutherford TF. 2011. Structural Estimation and Solution of International Trade Models with Heterogeneous Firms. *Journal of International Economics.* 83(2): 95-108.

[27] Bartelsman EJ, Doms M. 2000. Understanding Productivity: Lessons from Longitudinal Microdata. *Journal of Economic Literature.* 38(3): 569-594.

[28] Behrens K, Murata Y. 2012. Trade, Competition and Efficiency. *Journal of International Economics.* 87(1): 1-17.

[29] Bernard AB, Eaton J, Jensen JB, Kortum SS. 2003. Plants and Productivity in International Trade. *American Economic Review.* 93(4): 1268-90.

[30] Bernard AB, Jensen JB. 1995. Exporters, Jobs, and Wages in US Manufacturing: 1976-87. *Brookings Papers on Economic Activity: Microeconomics:* 67-112.

[31] Bernard AB, Jensen JB. 1997. Exporters, Skill Upgrading and the Wage Gap. *Journal of International Economics.* 42: 3-31.

[32] Bernard AB, Jensen, JB. 1999. Exceptional Exporter Performance: Cause, Effect, or Both? *Journal of International Economics.* 47(1): 1-25.

[33] Bernard AB, Jensen JB, Redding SJ, Schott PK. 2007. Firms in International Trade. *Journal of Economic Perspectives.* 21(3): 105-30.

[34] Bernard A B, Jensen JB, Schott PK. 2006. Trade Costs, Firms, and Productivity. *Journal of Monetary Economics.* 53(5): 917-37.

[35] Bernard AB, Redding SJ, Schott PK. 2007. Comparative Advantage and Heterogeneous Firms. *Review of Economic Studies.* 74: 31-66.

[36] Bernard AB, Redding SJ, Schott PK. 2010. Multi-product Firms and Product Switching. *American Economic Review.* 100(1): 70-97.

[37] Bernard AB, Redding SJ, Schott PK. 2011. Multi-product Firms and Trade Liberalization. *Quarterly Journal of Economics*, 126(3), 1271-1318.

[38] Bernard AB, Redding SJ, Schott PK. 2012. The Empirics of Firm Heterogeneity and International Trade. *Annual Review of Economics.* In press.

[39] Burstein A, Melitz MJ. 2012. Trade Liberalization and Firm Dynamics. In *Advances in Economics and Econometrics*, ed. D Acemoglu, M Arellano, E Deckel. Cambridge: Cambridge University Press.

[40] Burstein A, Vogel J. 2011. Factor Prices and International Trade: A Unifying Perspective. (Columbia University, Unpublished).

[41] Bustos P. 2007. The Impact of Technology and Skill Upgrading: Evidence from Argentina. (Universitat Pompeu Fabra, Unpublished).

[42] Bustos P. 2011. Trade Liberalization, Exports and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinean Firms. *American Economic Review.* 101(1): 304-340.

[43] Caliendo L, Rossi-Hansberg E. 2011. The Impact of Trade on Organization and Productivity. *Quarterly Journal of Economics*, forthcoming.

[44] Campbell J, Hopenhayn H. 2005. Market Size Matters. *Journal of Industrial Economics.* 53: 1-25.

[45] Chaney T. 2008. Distorted Gravity: The Intensive and Extensive Margins of International Trade. *American Economic Review.* 98(4): 1707-1721.

[46] Chaney T. 2012. The Network Structure of International Trade. (University of Chicago, Unpublished).

[47] Cherkashin I, Demidova S, Kee HL, Krishna K. 2010. Firm Heterogeneity and Costly Trade: A New Estimation Strategy and Policy Experiments. (Penn State University, Unpublished).

[48] Combes P, Duranton G, Gobillon L, Puga D and Roux S, 2012. The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection. *Econometrica.* In press.

[49] Corcos G, Del Gatto M, Mion G, Ottaviano GIP. 2012. Productivity and Firm Selection: Quantifying the New Gains from Trade. *Economic Journal.* In press.

[50] Coçar AK, Guner N, Tybout J. 2011. Firm Dynamics, Job Turnover, and Wage Distributions in an Open Economy. (Penn State University, Unpublished).

[51] Constantini J, Melitz MJ. 2008. The Dynamics of Firm-Level Adjustment to Trade Liberalization. In *The Organization of Firms in a Global Economy*, ed. E Helpman, D Marin, T Verdier, 4:107-41. Cambridge MA: Harvard University Press.

[52] Costinot A. 2009. An Elementary Theory of Comparative Advantage. *Econometrica.* 77(4): 1165-1192.

[53] Crozet M, Head K, Mayer T. 2012. Quality sorting and trade: Firm-level evidence for French wine. *Review of Economic Studies*, 79(2): 609-644.

[54] Cuñat A, Melitz MJ. 2010. Volatility, Labor Market Flexibility and Comparative Advantage. *Journal of the European Economic Association.* 10: 225-254.

[55] Das M, Roberts MJ, Tybout J. 2007. Market Entry Costs, Producer Heterogeneity and Export Dynamics. *Econometrica.* 75(3): 837-73.

[56] Davidson C, Matusz S, Shevchenko A. 2008. Globalization and Firm-level Adjustment with Imperfect Labor Markets. *Journal of International Economics.* 75: 295-309.

[57] Davidson C, Matusz S. 2009. *International Trade with Equilibrium Unemployment.* Princeton NJ: Princeton University Press. 412 pp.

[58] Davis SJ, Haltiwanger J. 1992. Gross Job Creation, Gross Job Destruction, and Employment Reallocation. *Quarterly Journal of Economics.* 107: 819-63.

[59] Davis DR, Harrigan J. 2011. Good Jobs, Bad Jobs, and Trade Liberalization. *Journal of International Economics.* 13139.

[60] Demidova S, Rodriguez-Clare A. 2009. Trade Policy under Firm-level Heterogeneity in a Small Open Economy. *Journal of International Economics.* 78. 100-112.

[61] Demidova S, Rodriguez-Clare A. 2011. *The Simple Analytics of the Melitz Model in a Small Open Economy.* NBER Working Paper. 17521.

[62] Dixit A, Sitlitz JE. 1977. Monopolistic Competition and Optimum Product Diversity. *American Economic Review.* 67(3), 297-308.

[63] Dhingra S. 2011. Trading Away Wide Brands for Cheap Brands. (Unpublished, London School of Economics).

[64] Dhingra S, Morrow J. 2012. The Impact of Integration on Productivity and Welfare Distortions under Monopolistic Competition. (London School of Economics, Unpublished).

[65] Dixit A, Norman V. 1980. *Theory of International Trade: A Dual General Equilibrium Approach*. Cambridge UK: Cambridge University Press.

[66] Dunne T, Roberts MJ, Samuelson L. 1989. The Growth and Failure of U.S. Manufacturing Plants. *Quarterly Journal of Economics.* 104(4): 671-98.

[67] Eaton J, Eslava M, Kugler M, Tybout JR. 2010. A Search and Learning Model of Export Dynamics. (Penn State University, Unpublished).

[68] Eaton J, Kortum SS. 2002. Technology, Geography, and Trade. *Econometrica.* 70(5): 1741-79.

[69] Eaton J, Kortum SS, Kramarz F. 2011. An Anatomy of International Trade: Evidence from French Firms. *Econometrica*. 79(5), 1453-1498.

[70] Eckel C, Neary JP. 2010. Multi-product Firms and Flexible Manufacturing in the Global Economy. *Review of Economic Studies.* 77(1): 188-217.

[71] Edmund C, Midrigin V, Xu DY. 2012. Competition, Markups, and the Gains from International Trade. *NBER Working Paper*, 18041.

[72] Egger H, Kreickemeier U. 2009. Firm Heterogeneity and the Labour Market Effects of Trade Liberalization. *International Economic Review.* 50(1): 187-216.

[73] Fan H, Lai ELC, Qi H, 2011. A Model of Trade with Ricardian Comparative Advantage and Intrasectoral Firm Heterogeneity. (Unpublished, Hong Kong University of Science and Technology).

[74] Feenstra RC. 2003. A Homothetic Utility Function for Monopolistic Competition Models Without Constant Price Elasticity. *Economic Letters.* 78: 79-86.

[75] Feenstra RC, Ma H. 2008. Optimal Choice of Product Scope for Multiproduct Firms under Monopolistic Competition. In *The Organization of Firms in a Global Economy*, ed. E Helpman, D Marin, T Verdier, 6:173-99. Cambridge MA: Harvard University Press. 356 pp.

[76] Felbermayr G, Prat J, Schmerer H. 2011. Globalization and Labor Market Outcomes: Wage Bargaining, Search Frictions, and Firm Heterogeneity. *Journal of Economic Theory.* 146(1): 39-73

[77] Flam H, Helpman E. 1987. Industrial Policy Under Monopolistic Competition. *Journal of International Economics*. 22: 79-102.

[78] Ghironi F, Melitz MJ. 2005. International Trade and Macroeconomic Dynamics with Heterogeneous Firms. *Quarterly Journal of Economics*. 120: 865-915.

[79] Goldberg P, Khandelwal A, Pavcnik N, Topalova P. 2010. Imported Intermediate Inputs and Domestic Product Growth: Evidence from India. *Quarterly Journal of Economics*, forthcoming.

[80] Gopinath G, Neiman B. 2012. Trade Adjustment and Productivity in Large Crises. (Unpublished, Harvard University).

[81] Grossman GM, Helpman E. 1991. *Innovation and Growth in the Global Economy.* Cambridge MA: MIT Press. 359 pp.

[82] Harrigan J, Reshef A. 2011. *Skill-Biased Heterogeneous Firms, Trade Liberalization, and the Skill Premium.* NBER Working Paper. 17604.

[83] Head K, Mayer T. 2012. Gravity Equations. In *Handbook of International Economics*, Volume 4, ed. E Helpman, G Gopinath, K Rogoff.

[84] Helpman E. 2006. Trade, FDI, and the Organization of Firms. *Journal of Economic Literature.* XLIV: 589-630.

[85] Helpman E, Itskhoki O. 2010. Labor Market Rigidities, Trade and Unemployment. *Review of Economic Studies.* 77 (3): 1100–1137.

[86] Helpman E, Krugman PR. 1985. *Market Structure and Foreign Trade.* Cambridge MA: MIT Press. 271 pp.

[87] Helpman E, Itskhoki O, Muendler M, Redding SJ. 2012. Trade and Inequality: From Theory to Estimation. *NBER Working Paper.* 17991.

[88] Helpman E, Itskhoki O, Redding SJ. 2010. Inequality and Unemployment in a Global Economy. *Econometrica.* 78(4): 1239-83.

[89] Helpman E, Melitz MJ, Rubinstein Y. 2008. Estimating Trade Flows: Trading Partners and Trading Volumes. *Quarterly Journal of Economics.* 123(2): 441-87.

[90] Helpman E, Melitz MJ, Yeaple SR. 2004. Export Versus FDI with Heterogeneous Firms. *American Economic Review.* 94(1): 300-16.

[91] Hsieh C, Ossa R. 2011. A Global View of Productivity Growth in China and India. (University of Chicago, unpublished)

[92] Irarrazabal AA, Opromolla LD. 2008. A Theory of Entry and Exit into Exports Markets. Working Paper w200820. Banco de Portugal. Economics and Research Department.

[93] Johnson R. 2012. Trade and Prices with Heterogeneous Firms. *Journal of International Economics*, 86(1): 43-56.

[94] Kehoe T, Ruhl K. 2009. *How Important is the New Goods Margin in International Trade?* Federal Reserve Bank of Minneapolis Research Department Staff Report 324.

[95] Krugman PR. 1979. Increasing Returns, Monopolistic Competition, and International Trade. *Journal of International Economics.* 9: 469-79.

[96] Krugman PR. 1980. Scale Economies, Product Differentiation, and the Pattern of Trade. *American Economic Review.* 70: 950-59.

[97] Kugler M, Verhoogen E. 2011. The Quality-Complementarity Hypothesis: Theory and Evidence from Colombia. *Review of Economic Studies*, In press.

[98] Levinsohn J. 1999. Employment Responses to International Liberalization in Chile. *Journal of International Economics.* 47: 321-44.

[99] Lileeva A, Trefler D. 2010. Does Improved Market Access Raise Plant-Level Productivity. *Quarterly Journal of Economics.* 125(3): 1051-99.

[100] Lu, D. 2011. Exceptional Exporter Performance? Evidence from Chinese Manufacturing Firms (Princeton University, Unpublished).

[101] Luttmer, EGJ. 2010. Models of Growth and Firm Heterogeneity. *Annual Review of Economics.* 2: 547-576.

[102] Manova, K. 2012. Credit Constraints, Heterogeneous Firms, and International Trade. (Unpublished, Stanford University).

[103] Mayer T, Melitz MJ, Ottaviano G. 2012. Market Size, Competition, and the Product Mix of Exporters. *American Economic Review.* Forthcoming.

[104] Mayer T, Ottaviano GIP. 2007. *The Happy Few: New Facts on the Internationalisation of European Firms.* Bruegel – CEPR EFIM Report. Bruegel Blueprint Series.

[105] Melitz MJ. 2003. The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica.* 71: 1695-725.

[106] Melitz MJ, Ottaviano G. 2008. Market Size, Trade, and Productivity. *Review of Economic Studies.* 75(1): 295-316.

[107] Neary JP. 2010. Two and a Half Theories of Trade. *The World Economy.* 33(1): 1-19.

[108] Nocke V, Yeaple SR. 2006. Globalization and Endogenous Firm Scope. *NBER Working Paper.* 12322.

[109] Oi WY, Idson TL. 1999. Firm Size and Wages. In *Handbook of Labor Economics,* ed. O Ashenfelter, D Card, 33: 2165-214. Amsterdam: Elsevier. 2939 pp.

[110] Ottaviano GIP, Tabuchi T, Thisse JF. 2002. Agglomeration and Trade Revisited. *International Economic Review*, 43: 409–436.

[111] Ottaviano GIP, Thisse JF. 1999. Monopolistic Competition, Multiproduct Firms and Optimum Product Diversity. *CEPR Discussion Paper*, 2151.

[112] Pavcnik N. 2002. Trade Liberalization, Exit, and Productivity Improvement: Evidence from Chilean Plants. *Review of Economic Studies.* 69(1): 245-76.

[113] Redding S. 2011. Theories of Heterogeneous Firms and Trade. *Annual Review of Economics.* 3: 77-105.

[114] Redding S, Venables AJ. 2004. Economic Geography and International Inequality. *Journal of International Economics.* 62(1): 53-82.

[115] Roberts MJ, Tybout J. 1997. The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs. *American Economic Review.* 87(4): 545-64.

[116] Romer PM. 1990. Endogenous Technological Change. *Journal of Political Economy.* 98: S71-S102.

[117] Ruhl KJ. 2004. The International Elasticity Puzzle. (Unpublished, New York University).

[118] Sampson, T. 2012. Selection into Trade and Wage Inequality. (Unpublished, London School of Economics).

[119] Syverson, C. 2004. Market Structure and Productivity: A Concrete Example. *Journal of Political Economy.* 112: 1181-222.

[120] Trefler D. 2004. The Long and Short of the Canada-U.S. Free Trade Agreement. *American Economic Review.* 94: 870-95.

[121] Tybout JR. 2003. Plant- and Firm-Level Evidence on 'New' Trade Theories. In *Handbook of International Economics, ed.* E Kwan Choi and J Harrigan, 13: 388-435. Oxford: Basil-Blackwell. 435 pp.

[122] Venables A. 1987. Trade and Trade Policy with Differentiated Products: A Chamberlinian-Ricardian Model. *Economic Journal*. 97(387): 700-717.

[123] Verhoogen E. 2008. Trade, Quality Upgrading and Wage Inequality in the Mexican Manufacturing Sector. *Quarterly Journal of Economics.* 123(2): 489-530.

[124] World Trade Organization. 2008. *World Trade Report: Trade in a Globalizing World*. Geneva: World Trade Organization.

[125] Yeaple SR. 2005. A Simple Model of Firm Heterogeneity, International Trade, and Wages. *Journal of International Economics.* 65: 1–20.