Trade Integration, Firm Selection and the Costs of Non-Europe.*

Massimo Del Gatto
University of Cagliari
and
CRENOS

Giordano Mion
CORE, Université Catholique de Louvain
and
FNRS

Gianmarco I.P. Ottaviano
University of Bologna
FEEM
and
CEPR

March 25, 2006

Abstract

In models with heterogeneous firms trade integration has a positive impact on aggregate productivity through the selection of the best firms as import competition drives the least productive ones out of the market. To quantify the impact of firm selection, we calibrate and simulate a multi-country multi-sector model with monopolistic competition and variable markups using firm-level data and aggregate trade figures on a panel of 11 EU countries. We find that EU trade has a sizeable impact on aggregate productivity. In 2000 the introduction of prohibitive trade barriers would have caused an average productivity loss of roughly 13 per cent, whereas a reduction of intra-EU trade costs by 5 per cent would have generated a productivity gain of roughly 2 per cent. Productivity losses and gains, however, vary a lot across countries and sectors depending on market accessibility and trade costs. We provide evidence that our results are robust to alternative distance and productivity measures.

Keywords: European integration, firm-level data, firm selection, gains from trade, total factor productivity
J.E.L. Classification: F12, R13.

*Massimo Del Gatto, Department of Economics, University of Cagliari, Italy. Email: mdelagatto@unica.it. Giordano Mion, CORE, Université Catholique de Louvain, Belgium. Email: mion@core.ucl.ac.be. Gianmarco I.P. Ottaviano, Department of Economics, University of Bologna, Italy. Email: ottavian@economia.unibo.it. We thank Stefano Breschi, Francesco Caselli, Armando Dominioni, Jonathan Eaton, Eric Strobl and Silvana Tenreyro for helpful comments and discussions as well as seminar participants at Bocconi University Milan, London School of Economics and Université Catholique de Louvain. We are indebted with ZEW Mannheim for access to their MIP dataset. The second author gratefully acknowledges financial support from MIUR. Errors are ours.

1 Introduction

In recent models with heterogeneous firms international trade integration has a positive impact on aggregate productivity through the selection of the best firms (Bernard et al., 2003; Melitz, 2003). The reason is a combination of import competition and export market access. On the one hand, as lower trade costs allow foreign producers to target the domestic markets, the operating profits of domestic firms in those markets shrink whatever their productivities. On the other hand, some domestic firms gain access to foreign markets and get additional profits from their foreign ventures. These are the firms that are productive enough to cope with the additional costs of foreign activity (such as those due to transportation and remaining administrative duties or institutional and cultural barriers). The result is the partition of the initially active domestic firms in three groups. As they start making losses in their home markets without gaining access to foreign markets, the least productive firms are forced to exit. On the contrary, as they are able to compensate lost profits on home sales with new profits on foreign sales, the most productive firms survive and expand their market shares. Finally, firms with intermediate levels of productivity also survive but, not being productive enough to access foreign markets, are relegated to home sales only and their market shares fall. Since international trade integration eliminates the least productive firms, average productivity grows through the reallocation of productive resources from less to more efficient producers.

This mechanism finds empirical support in firm-level analyses that have tried to identify the direction of causation hidden in the positive correlation between the export status of a firm and its productivity (called 'exceptional exporter performance' by Bernard and Jensen, 1999). This is a crucial issue for trade policy. Causation going from export status to firm performance would reveal the existence of 'learning by exporting' and therefore call for export promotion. However, apart from peculiar cases, most of the evidence supports reverse causation in the form of 'selection into export status': firms that already perform better have a stronger propensity to export than other firms (Tybout, 2002). Selection comes with two additional effects that are consistent with the theoretical argument discussed above. First, exposure to trade forces the least productive firms to shut down (Bernard and Jensen, 1999; Aw, Chung and Roberts, 2000; Clerides, Lach and Tybout, 1998). Second, trade liberalization leads to market share reallocations towards the most productive firms (Pavcnik, 2002; Bernard, Jensen and Schott 2003). On both counts, aggregate average productivity improves.

The empirical relevance of the selection effect motivates additional efforts towards quantifying its contribution in terms of gains from trade. This line of research is heralded by Tybout (2002) and pursued by Bernard et al. (2003) through the simulation of counterfactual scenarios. Those authors start with noting that the differences in measured productivity (e.g. value added per worker) across firms can be generated only by theoretical models assuming imperfect competition. Two options are then available. Differences can be derived from constant markups (i.e. Dixit-Stiglitz monopolistic competition) as well as fixed costs of entry and of exporting. This is the option chosen by Melitz (2003). Alternatively, those differences can be obtained from variable markups even without any fixed cost. This is the option chosen by Bernard et al. (2003), who propose a model obtained by introducing Bertrand competition in the probabilistic Ricardian framework developed by Eaton and Kortum (2002). They then calibrate the parameters of their model on U.S. data. In particular, they use aggregate production and trade data among the 47 leading U.S. export destinations (including the U.S. itself) to calibrate the parameters governing geographic barriers, aggregate technology differences, and differences in input costs. U.S. plant level data are used, instead, to calibrate the parameters that relate to the heterogeneity of goods in production and consumption. Finally, the calibrated model is used to assess the impacts of various counterfactual scenarios. In the focal one, Bernard et al. (2003) report a 4.7 percent increase in the average productivity of U.S. firms resulting from a 5 percent reduction in world trade barriers.[1]

The counterfactual analysis by Bernard et al. (2003) reveals the empirical tractability of their model. Their treatment of market structure has, however, some theoretical shortcoming. Specifically, a certain good in a certain country is always supplied by one firm only. This is the lowest-cost supplier of that good to that country. Indeed, under Bertrand competition, all other firms are kept out through limit pricing: the lowest-cost producer quotes a price that matches the second-lowest cost of supplying that good to that country. Accordingly, to derive the price and markup quoted for each good in each country, one needs to know which firms are the lowest and the second-lowest cost suppliers from each potential source country. This problem becomes tractable if one is ready to assume that in each country the lowest and the second-lowest costs are realizations of random variables drawn from probability distributions. In this respect, Bernard et al. (2003) show that the Fréchet family yields tractable distributions for prices and markups along with simple expressions for bilateral trade shares. Differently, building on Dixit-Stiglitz monopolistic competition, the model by Melitz (2003) allows many producers with different costs to simultaneously serve the same market, which is consistent even with arm-chair evidence. This model, however, is analytically tractable only if countries are identical and all bilateral trade barriers are the same. This clearly undermines empirical calibration and makes the model a fairly unpromising tool to deal with counterfactual analysis. Moreover, its implication of constant and equal markups across all firms, no matter where they produce and sell, flies in the face of empirical evidence showing that markups do vary across firms and markets.

The aim of the present paper is to supplement the existing literature in two respects. First, we perform a counterfactual analysis in the case of monopolistic competition to match what has been achieved by Bernard et al. (2003) with Bertrand competition. Second, in so doing we propose a multi-sector empirical implementation of the model by Melitz and Ottaviano (2005), who obtain variable markups in a highly tractable asymmetric multi-country framework with monopolistic competition, thus overcoming some of the theoretical limitations of both the Bertrand and Dixit-Stiglitz models.

Our dataset covers aggregate and firm-level evidence on a panel of 11 EU countries. To the best of our knowledge, this is the first analysis that uses comparable individual panel data across countries to investigate how trade integration affects aggregate productivity in the presence of firm selection. In particular, we use the calibrated model to explore two scenarios. In the first scenario, we assess the productivity losses that would be associated with autarky ('costs of non-Europe').[2] We find that, if in 2000 trade had been inhibited altogether, average productivity would have dropped by rougly 13 percent. A similar thought experiment is performed by Eaton and Kortum (2002) in their probabilistic Ricardian framework applied to a sample of 19 OECD countries in 1990. For our 11 countries they calculate that the average productivity loss associated with autarky is equal to roughly 4%. The difference may be explained by the fact that they do not have imperfect competition, their base year pre-dates ours by a decade, and OECD countries are generally less integrated than EU ones. The second scenario is designed to assess the productivity gains from

[1] Recent evidence on the existence of causation from trade to aggregate income and productivity is provided by Frankel and Rose (2002), who find per capita income to be positively affected by the formation of currency unions, thanks to their positive impact on trade, and by Alcalà and Ciccone (2004), who report strong support for a positive causal effect of trade on labor productivity. With respect to our analysis, Alcalà and Ciccone (2004) provide the interesting insight that, at the aggregate level, such a positive causation mainly acts through total factor productivity.

[2] The expression 'cost of non-Europe' was introduced to refer to the economic cost of failing to complete the common market. This is the subject of a landmark study by the European Commission, the Cecchini report, presented in March 1988.

further integration ('gains from (freer) trade'). For the sake of comparison we follow Bernard et al. (2003) and study the impact of a 5 per cent reduction in trade barriers. We find that in 2000 such reduction would have caused an average productivity increase of roughly 2 per cent. This is smaller than the increase of roughly 5 per cent obtained by Bernard et al. (2003) for the US. Although the two outcomes may appear to differ considerably, it is worth noting that US productivity is higher than the average European productivity. Indeed, when focusing on the most productive countries in our European sample, namely Germany and Denmark, we get very similar effects (respectively, 4.6 and 4.4 per cent). More generally, in both scenarios we find that productivity gains vary a lot across countries and sectors depending on market accessibility and trade costs.[3]

The rest of the paper is organized in six sections. Section 2 presents the theoretical model. Section 3 derives its equilibrium properties, then designs the calibration and simulation strategy. Section 4 describes the dataset. Section 5 calibrates the model. Section 6 simulates alternative integration scenarios while Section 7 provides evidence that our results are robust to alternative distance and productivity measures. Section 8 concludes.

2 The model

Consider an economy with M countries, indexed by $l = 1, ..., M$. Country l hosts L^l consumers, each supplying one unit of labor.

2.1 Preferences and Demand

Preferences are defined over S horizontally differentiated goods, indexed by $s = 1, ..., S$, and a homogenous good chosen as numeraire. Each differentiated good s is composed by a continuum of varieties indexed by $i \in \Omega_s$. All consumers share the same preferences in all countries and the same income in each country. The utility function for a representative consumer in country l is given by

$$U^l = d_0^l + \sum_{s=1}^{S} \left\{ \alpha_s \int_{i\in\Omega_s} d_s^l(i)di - \frac{1}{2}\gamma_s \int_{i\in\Omega_s} \left[d_s^l(i)\right]^2 di - \frac{1}{2}\eta_s \left(\int_{i\in\Omega_s} d_s^l(i)di \right)^2 \right\}, \qquad (1)$$

where d_0^l and $d_s^l(i)$ represent the individual consumption levels of the numeraire good and variety i of good s. The demand parameters α_s, η_s, and γ_s are all positive. For each differentiated good s, the parameters α_s and η_s index the substitution pattern between its varieties and the numeraire: increases in α_s and decreases in η_s both shift out the demand for the differentiated varieties relative to the numeraire. The parameter γ_s indexes the degree of product differentiation between the varieties of good s. In the limit when $\gamma_s = 0$, consumers only care about their total

[3] Quantifying the gains from trade associated with the integration process has also been of interest in other analytical contexts. In particular, it is worth mentioning the analysis by Smith and Venables (1988), who simulate the effects of a reduction of intra-EU trade costs in a model with imperfect competition and homogeneous firms, as well as the vast literature of Computable General Equilibrium (CGE). Concerning the former, our gains from trade, obtained under the assumption that firms differ in terms of productivity, are (see section 6.2) significantly bigger than those achieved by Smith and Venables without firm heterogeneity but still in an imperfectly competitive environment. This circumstance can be easily interpreted as supporting the idea that the selection effect provides an important contribution to increasing average productivity when markets become more integrated. As for the latter, it has to be noticed that, differently from the Computable General Equilibrium (CGE) approach, our analysis is not conceived as an ex ante or ex post evaluation of the overall effects of a given policy, but rather as an exercise of comparative statics carried out under the ceteris paribus assumption. In particular, as our "gains from trade" stem only from the selection effect, our analysis is better understood as an attempt to capture the implications of endogenous productivity, which is usually taken as given by the CGE literature (Haaland and Norman,1992).

consumption level over all varieties of that good, $D_s^l = \int_{i \in \Omega_s} d_s^l(i) di$. Such varieties are then perfect substitutes. The degree of product differentiation increases with γ_s as consumers give increasing weight to the distribution of consumption levels across varieties.

The marginal utilities for all goods are bounded, and a consumer may thus not have positive demand for any particular good. We assume that consumers have positive demands for the numeraire good ($d_0^l > 0$). The inverse demand by country l for each variety i of good s is then given by

$$p_s^l(i) = \alpha_s - \gamma_s d_s^l(i) - \eta_s D_s^l, \tag{2}$$

whenever demand $d_s^l(i) > 0$. Let $\widetilde{\Omega}_s^l \subset \Omega_s^l$ be the subset of varieties of good s consumed in country l. Then (2) can be inverted to yield the linear market demand system for these varieties:

$$q_s^l(i) \equiv L^l d_s^l(i) = \frac{\alpha_s L^l}{\eta_s N_s^l + \gamma_s} - \frac{L^l}{\gamma_s} p_s^l(i) + \frac{\eta_s N_s^l}{\eta_s N_s^l + \gamma_s} \frac{L^l}{\gamma_s} \bar{p}_s^l, \ \forall i \in \widetilde{\Omega}_s^l, \tag{3}$$

where N_s^l is the measure of consumed varieties in $\widetilde{\Omega}_s^l$ and $\bar{p}_s^l = \left(1/N_s^l\right) \int_{i \in \widetilde{\Omega}_s^l} p_s^l(i) di$ is their average price. The set $\widetilde{\Omega}_s^l$ is the largest subset of Ω_s^l that satisfies

$$p_s^l(i) \leq \frac{1}{\eta_s N_s^l + \gamma_s} \left(\gamma_s \alpha_s + \eta_s N_s^l \bar{p}_s^l\right) \equiv p_s^l. \tag{4}$$

Any price above α_s must violate this condition since the marginal utility in (2) is bounded above by α_s; hence $\bar{p}_s^l \leq \alpha_s$ (the inequality must be strict when there is any price heterogeneity). For a given level of product differentiation γ_s, lower average prices $\bar{p}_s^l$ or a larger number of competing varieties N_s^l induce an increase in the price elasticity of demand and decrease the price bound (or choke price) defined in (4). On both counts, lower $\bar{p}_s^l$ or a larger N_s^l generate a 'tougher' competitive environment.

2.2 Production and Firm Behavior

Labor is the only factor of production and is inelastically supplied in a competitive market. The numeraire good is produced under constant returns to scale at unit cost. Its market is also competitive and perfectly integrated among countries. These assumptions imply a unit wage in all countries as long as the numeraire good is produced in all countries, which is henceforth assumed. Entry in each differentiated product sector is modeled as a research and development (R&D) process with uncertain outcome. Specifically, each entrant has to invent its own variety and a corresponding production process by making an irreversible investment of $f_{E,s}$ units of numeraire. In so doing a prospective entrant knows it is going to find a new variety for sure and that production will occur under constant returns to scale. It does know, however, the marginal cost of production c as this will be randomly determined only after the investment in R&D has been sunk. In each country l and sector s uncertainty is modeled as a draw from a common and known distribution $G_s^l(c)$, with support $[0, c_{M,s}^l]$, which varies across sectors and countries. This allows us to introduce (probabilistic) 'comparative advantage' stemming from technological differences that affect the distribution of firm-level productivity draws. For example, if $\left(c_{M,s}^l / c_{M,r}^l\right) < \left(c_{M,s}^h / c_{M,r}^h\right)$, countries l and h are said to have comparative advantages in sectors s and r respectively: relative to entrants in h (l), entrants in l (h) have a 'better chance' of getting lower cost draws in sector s than in sector r.[4]

Since the entry cost $f_{E,s}$ is sunk, only firms that can cover their marginal cost survive and produce. All other entrants exit without even starting production. Surviving firms maximize

[4]See Section 3.1 for additional details.

their profits facing the demand function (3). Given the continuum of competitors, a firm takes the average price level $\bar{p}_s^l$ and numbers of firm N_s^l as given. This is the essence of monopolistic competition. Moreover, national markets are segmented, although firms can produce in one market and sell in the other, incurring a per-unit trade cost. The overall cost of a delivered unit with cost c from country h to country l is $\tau_s^{hl}c$ with $\tau_s^{lh} > 1$, where $(\tau_s^{hl} - 1)c$ is the frictional trade cost. We interpret such cost in a wide sense as stemming from all distance-related barriers. For this reason, even within countries, trade may not be costless and we allow for $\tau_s^{ll} \geq 1$.

Let $p_s^{lh}(c)$ and $q_s^{lh}(c)$ represent the levels of the profit maximizing price and quantity sold for a firm in sector s producing in country l with cost c and selling to country h. Since national markets are segmented and firms produce under constant returns to scale, they independently maximize the profits earned from sales to different countries. Let $\pi_s^{lh}(c) = \left[p_s^{lh}(c) - \tau_s^{lh}c\right] q_s^{lh}(c)$ denote the maximized value of these profits as a function of the firm's marginal cost c. Then the profit maximizing prices and output levels must satisfy: $q_s^{lh}(c) = \left(L^h/\gamma_s\right)\left[p_s^{lh}(c) - \tau_s^{lh}c\right]$. Only firms earning non-negative profits in a market will choose to sell in that market. This leads to similar cost cutoff rules for firms selling in the various markets. Let c_s^{lh} denote the upper bound cost for firms producing in country l and selling to country h. This cutoff must then satisfy:

$$c_s^{lh} = \sup\left\{c : \pi_s^{lh}(c) > 0\right\} = \frac{p_s^h}{\tau_s^{lh}} \tag{5}$$

This implies $\tau_s^{hl}c_s^{hl} = \tau_s^{kl}c_s^{kl}$, e.g. higher trade barriers from h to l make it harder for exporters from h to break even relative to their competitors from k. The cutoffs summarize all the effects of market conditions relevant for firm performance. In particular, the optimal prices and output levels can be written as:

$$p_s^{lh}(c) = \frac{\tau_s^{lh}}{2}(c_s^{lh} + c), \quad q_s^{lh}(c) = \frac{L^h}{2\gamma_s}\tau_s^{lh}\left(c_s^{lh} - c\right) \tag{6}$$

which yield the following maximized profit levels:

$$\pi_s^{lh}(c) = \frac{L^h}{4\gamma_s}\left(\tau_s^{lh}\right)^2\left(c_s^{lh} - c\right)^2. \tag{7}$$

Finally, entry is unrestricted in all countries. Firms choose a production location prior to entry and sink the corresponding entry cost. Free entry of firms in country l implies zero expected profits in equilibrium, hence:

$$\sum_{h=1}^{M}\left[\int_0^{c_s^{lh}} \pi_s^{lh}(c)dG_s^l(c)\right] = f_{E,s}. \tag{8}$$

which, together with (5), determines the cost cutoffs c_s^{lh}. These cutoff, in turn, determine the numbers of sellers. Indeed, since $\tau_s^{lh}c_s^{lh} = p_s^h$, (4) implies:

$$\tau_s^{lh}c_s^{lh} = \frac{1}{\eta_s N_s^h + \gamma_s}\left(\gamma_s\alpha_s + \eta_s N_s^h\bar{p}_s^h\right).$$

This yields the zero cutoff profit condition:

$$N_s^h = \frac{2\gamma_s}{\eta_s}\frac{\alpha_s - \tau_s^{lh}c_s^{lh}}{\tau_s^{lh}c_s^{lh} - \bar{c}_s^h}, \tag{9}$$

where $\bar{c}_s^h = \sum_{l=1}^{M}\left\{\left[\int_0^{c_s^{lh}} cdG_s^l(c)\right]/G_s^l(c_s^{lh})\right\}$ is the average cost of surviving firms.

3 Equilibrium

We are now ready to determine the equilibrium distribution of firms across countries and the associated trade flows.

3.1 Parametrization of Technology

All the results derived in the previous section hold for any distribution of cost draws $G_s^l(c)$. However, to implement the model empirically, we use a specific parametrization for the distribution whose empirical relevance will then be tested. In particular, we assume that in sector s and country l productivity draws $1/c$ follow a Pareto distribution with lower productivity bound $1/c_{M,s}^l$ and shape parameter $k_s \geq 1$. This implies a distribution of cost draws c given by

$$G_s^l(c) = \left(\frac{c}{c_{M,s}^l}\right)^{k_s}, \ c \in [0, c_{M,s}^l]. \tag{10}$$

The shape parameter k_s indexes the dispersion of cost draws in sector s and it is the same in all countries. When $k_s = 1$, the cost distribution is uniform on $[0, c_{M,s}^l]$. As k_s increases, the relative number of high cost firms increases, and the cost distribution is more concentrated at these higher cost levels. As k_s goes to infinity, the distribution becomes degenerate at $c_{M,s}^l$. Any truncation of the cost distribution from above at $c_s^{lh} < c_{M,s}^l$ retains the same distribution function and shape parameter k_s. The productivity distribution of firms producing in l and selling to h is therefore also Pareto with shape k_s, and the truncated cost distribution is given by $G_s^{lh}(c) = \left(c/c_s^{lh}\right)^{k_s}, \ c \in [0, c_s^{lh}]$.

3.2 Production and Entry

Let $\rho_s^{lh} \equiv \left(\tau_s^{lh}\right)^{-k_s} \in (0, 1]$ measure the 'freeness' of trade for exports from l to h, which allows us to define the following trade freeness matrix for sector s:

$$P_s \equiv \begin{pmatrix} \rho_s^{11} & \rho_s^{12} & \cdots & \rho_s^{1M} \\ \rho_s^{21} & \rho_s^{22} & \cdots & \rho_s^{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_s^{M1} & \rho_s^{M2} & \cdots & \rho_s^{MM} \end{pmatrix}.$$

Given our parametrization, the free entry condition (8) in country l can be rewritten as:

$$\sum_{h=1}^{M} \rho_s^{lh} L^h \left(c_s^{hh}\right)^{k_s+2} = \frac{2\gamma_s(k_s+1)(k_s+2)f_{E,s}}{\psi_s^l} \qquad l = 1, ..., M,$$

where $\psi_s^l = \left(c_{M,s}^l\right)^{-k}$ is an index of absolute advantage in sector s. This yields a system of M equations that can be solved for the M equilibrium domestic cutoffs in sector s using Cramer's rule:

$$c_s^{hh} = \left(\frac{2(k_s+1)(k_s+2)f_{E,s}\gamma_s}{|P_s|} \frac{\sum_{l=1}^{M} \left|C_s^{lh}\right| / \psi_s^l}{L^h}\right)^{\frac{1}{k_s+2}} \qquad h = 1, ..., M, \tag{11}$$

where $|P_s|$ is the determinant of the trade freeness matrix and $\left|C_s^{lh}\right|$ is the cofactor of its ρ_s^{lh} element. Cross-country differences in cutoffs arise from three sources: own country size (L^h), as well as a combination of market access and comparative advantage ($\sum_{l=1}^{M} \left|C_s^{lh}\right| / \psi_s^l$). Countries benefiting from a larger local market, a better distribution of productivity draws, and better market accessibility have lower cutoffs.

The mass of sellers N_s^l in sector s and country l (including domestic producers in l and exporters to l) is still given by (9). With a positive mass of entrants $N_{E,s}^l$ in all countries, there are $G_s^l(c_s^{ll})N_{E,s}^l$ domestic producers and $\sum_{h\neq l} G_s^l(c_s^{hl})N_{E,s}^h$ exporters selling in l since c_s^{hl} is the export cutoff from h to l. This implies:

$$\sum_{h=1}^{M} \rho_s^{hl}\psi_s^h N_{E,s}^h = \frac{N_s^l}{(c_s^{ll})^{k_s}}.$$

The latter provides a system of M linear equations that can be solved for the number of entrants in the M countries using Cramer's rule:[5]

$$N_{E,s}^l = \frac{2\left(k_s+1\right)\gamma_s}{\eta_s\left|P_s\right|\psi_s^l}\sum_{h=1}^{M}\frac{\left(\alpha_s - c_s^{hh}\right)\left|C_s^{lh}\right|}{\left(c_s^{hh}\right)^{k_s+1}} \tag{12}$$

Given $N_{E,s}^l$ entrants in country l, $N_{E,s}^l G_s^l(c_s^{ll})$ firms survive and produce for the local market. Among the latter, $N_{E,s}^l G_s^l(c_s^{lh})$ export to country h.

3.3 Trade Flows

Our model yields a gravity equation for aggregate bilateral trade flows. In sector s an exporter from l to h with cost c generates export sales $r_s^{lh}(c) = p_s^{lh}(c)q_s^{lh}(c)$ where (see (5) and (6))

$$\begin{aligned} p_s^{lh}(c) &= \frac{\tau_s^{lh}}{2}\left(c_s^{lh}+c\right) = \frac{1}{2}\left(c_s^{hh}+\tau_s^{lh}c\right), \\ q_s^{lh}(c) &= \frac{L^h\tau_s^{lh}}{2\gamma_s}\left(c_s^{lh}-c\right) = \frac{L^h}{2\gamma_s}\left(c_s^{hh}-\tau_s^{lh}c\right). \end{aligned}$$

Aggregating these export sales $r_s^{lh}(c)$ over all exporters from l to h (with cost $c \leq c_s^{lh}$) yields the aggregate bilateral exports in sector s from l to h:[6]

$$\begin{aligned} EXP_s^{lh} &= N_{E,s}^l \int_0^{c_s^{lh}} r_s^{lh}(c)dG_s^l(c) \\ &= N_{E,s}^l\frac{L^h}{4\gamma_s}\int_0^{c_s^{hh}/\tau_s^{lh}}\left[\left(c_s^{hh}\right)^2 - \left(\tau_s^{lh}c\right)^2\right]dG_s^l(c) \\ &= \frac{1}{2\gamma_s\left(k_s+2\right)}N_{E,s}^l\psi_s^l L^h\left(c_s^{hh}\right)^{k_s+2}\rho_s^{lh}. \end{aligned} \tag{13}$$

This is a gravity equation in so far as it determines bilateral exports as a (log-linear) function of bilateral trade barriers and country characteristics. In particular, it reflects the combined effects of country size, technology (comparative advantage), and geography (accessibility) on both the extensive (number of traded goods) and intensive (amount traded per good) margins of trade flows.[7] It highlights how a lower cutoff c_s^{ll} dampens exports by making it harder for potential exporters to break into that market.

[5] We use the properties that relate the freeness matrix P and its transpose in terms of determinants and cofactors.

[6] The integration measure $G_s^l(c_s^{lh})$ represents the proportion of entrants $N_{E,s}^l$ in l that export to h.

[7] See Eaton and Kortum (2002) and Helpman, Melitz and Rubinstein (2004) for similar results derived from different models.

3.4 From Theory to Simulation

How large are the productivity gains from EU integration due to the impact of freer trade on competition and selection? Our model can be used to address this question by building on equations (11) and (13). In so doing, we proceed in two stages. We start with calibrating the model. Then we use the calibrated model to investigate the effects of different integration scenarios.

In the calibration stage, we first use trade and geographical data for the year 2000 to recover the sectoral freeness of trade $\rho_s^{lh} = \left(\tau_s^{lh}\right)^{-k_s}$ from the gravity equation (13). This allows us to get the freeness matrix P_s and to compute its determinant and co-factors that appear in equation (11). We then use a database on manufacturing firms belonging to 11 EU countries to estimate individual total factor productivities (TFP) for the year 2000. From such productivities we recover two additional elements of equation (11): the shape parameter of the underlying Pareto distribution (k_s) and the M endogenous domestic cut-offs (c_s^{hh}) by sector. Using the computed values of P_s, k_s and c_s^{hh} together with data on population L^h, we finally solve (11) to obtain the index of absolute advantage ψ_s^l up to a sector specific constant (related to $f_{E,s}$ and γ_s).

In the simulation stage, we run a counterfactual analysis on the calibrated model. In particular, we simulate the changes in productivity induced by different trade costs by recomputing c_s^{hh} for alternative freeness matrices P_s. Two scenarios are considered. One in which international trade costs are prohibitive ($\rho_s^{lh} = 0$ for $l \neq h$) and one in which international trade costs (τ_s^{lh} for $l \neq h$) are reduced by 5 per cent. The first scenario provides us with an assessment of the 'costs of non-Europe'. The second scenario gives us a measure of the 'gains from (freer) trade', that is, the gains from further integration.

4 Data

In our empirical analysis we take advantage of different datasets. For the productivity estimations we extensively use the Amadeus database provided by the Bureau Van Dijk. This dataset gives (harmonized) yearly balance-sheet information on the biggest 250,000 European firms for the period 1994-2003. To the best of our knowledge, it is the only dataset that provides comparable individual figures for a relatively large group of countries. In particular, Amadeus provides information on value added, fixed assets (capital), sales, and the cost of materials (intermediates consumption) in thousands of euros, as well as on the number of employees. We focus on manufacturing firms in western Europe for the year 2000. We choose that year because of the quality of the data and the fact that no major economic change took place. We consider only those countries for which a reasonable data coverage exists. We eliminate missing values and extreme observations, defined as having either a capital/employees or value added/employees ratio which is out of the range identified by the 1st and the 99th percentile. This leaves us with a sample of 22,120 firms across 11 countries as listed in Table 1.[8]

As one can see in the table, data coverage for Germany, which is the biggest EU economy, is rather poor. This is the reason why we complement our Amadeus data with information coming from the MIP database on German firms provided by the Zentrum für Europäische Wirtschaftsforschung (ZEW). The MIP database has relatively smaller firms than Amadeus. However, the productivity of German firms in the two samples is not much different and both samples reveal that Germany is the most productive country. The MIP contains information on value added, employment and input consumption. The capital variable is reconstructed by using the book value

[8]Sample statistics suggest that observations are missing at random within each country so that our sample seems to be representative.

Table 1: Data coverage across countries for the year 2000: Amadeus only.

Country initials	Country	Frequency	Percent
BE	Belgium	1557	7.04
DE	Germany	385	1.74
DK	Denmark	309	1.40
ES	Spain	2730	12.34
FI	Finland	529	2.39
FR	France	3956	17.88
GB	Great Britain	4514	20.41
IT	Italy	5735	25.93
NL	Netherlands	861	3.89
PT	Portugal	156	0.71
SE	Sweden	1388	6.27
Total		22120	100

of capital in 1998, adding investments at the end of the period and applying the relevant deflators. After eliminating missing as well as extreme observations, the MIP database provides us with roughly 700 additional firms. Although our results are virtually the same when we use the Amadeus data only, the actual sample we rely on for productivity estimations contains those additional firms. Descriptive statistics of the main variables in the combined Amadeus-MIP database are given in Table 2.

Table 2: Descriptive statistics of the OLS estimation dataset (Amadeus and MIP).

	N. firms	Mean	St. dev.	Min	Max
Sales	22801	146008.9	1739573	2	162000000
Value added	22801	47083.5	511309	18	44500000
Capital	22801	72865.69	937859.3	8	89100000
Intermed. consumpt.	22801	57920.58	428622.7	1	26900000
Employees	22801	667.84	6027.05	1	449594

Note: All variables except Employees are in thousand of euros.

As a benchmark we estimate firm-level productivity by means of simple OLS in the year 2000. However, one may wonder to what extent our results are robust to considering more accurate estimators of productivity. To address this concern, in Section 7.1 we also implement the approach by Levinsohn and Petrin (2003), which uses intermediates consumption to control for unobservables.[9] As panel information is needed in this case, we build another dataset using years 1998, 1999 and 2000. When eliminating missing (as well as extreme) observations for intermediates consumption, the combined Amadeus-MIP sample is reduced to approximately 14,500 firms per year (for a total of 42,663 observations in the three years). This reduction is mainly due to the fact that there is no information available on firms' input consumption for Great Britain and Denmark. This explains why, as detailed in Table 3, the panel for the Levinsohn-Petrin procedure covers 9 countries only. Descriptive statistics of the main variables used from this panel dataset are given in Table 4.

[9] The alternative approach by Olley and Pakes (1996), which uses investment to control for unobservables, can not be applied to the Amadeus database because this does not contain any information on firms' investments.

Table 3: Data coverage across countries for the Levinsohn and Petrin estimation (Amadeus and MIP): years 1998-1999-2000.

Country initials	Country	Frequency	Percent
BE	Belgium	4107	9.63
DE	Germany	2202	5.16
ES	Spain	6603	15.48
FI	Finland	1290	3.02
FR	France	9951	23.32
IT	Italy	15399	36.09
NL	Netherlands	987	2.31
PT	Portugal	294	0.69
SE	Sweden	1830	4.29
Total		42663	100.00

Table 4: Descriptive statistics of the Levinsohn and Petrin estimation dataset (Amadeus and MIP).

	N. firms	Mean	St. dev.	Min	Max
Sales	42663	92577.97	699496.4	1	65700000
Value added	42663	26198.83	232179.1	12	19700000
Capital	42663	35942.36	385963.2	4	35700000
Intermed. consumpt.	42663	47006.06	308579.9	1	26900000
Employees	42663	385.53	2142.79	1	142881

Note: All variables except Employees are in thousand of euros.

Turning to the industry disaggregation, we work with a 18 sectors breakdown of manufacturing activities, which derives from merging the information contained in the Amadeus and MIP databases – organized by 2-digits Nace rev. 1 and thus leading to 23 manufacturing sectors – with that contained in the database we use to compute trade costs (see below) – organized by 3-digits ISIC rev 2 and thus leading to 26 manufacturing industries. The resulting industry disaggregation is detailed in Table 5.

The data we use to compute trade costs are provided by the Centre d'Etude Prospectives et d'Informations Internationales (CEPII). The main dataset, used in Mayer and Zignago (2005), involves trade and production figures in a compatible ISIC 3-digit classification for a large set of countries over the 1976-2001 period.[10] These data allow us to recover both the internal (EXP_s^{ll}) and the external (EXP_s^{lh} with $l \neq h$) flows of goods. To estimate the freeness of trade ρ_s^{lh} from the gravity equation (13), we complement trade and production data with geographical variables, in particular bilateral distances and common language indicators, which are also provided by CEPII. For both geographical variables several alternative measures are available.[11]. To recover the bilateral trade costs for our 11 countries in 2000, we consider trade among 15 European countries (our 11 countries plus Austria, Greece, Ireland and Norway) in the years from 1999 to 2001. We use a larger number of countries and three years to obtain more accurate measures. Table 6 shows descriptive statistics of the trade and geographical variables. The data are organized by flows and

[10] For details, see http://www.cepii.fr/anglaisgraph/bdd/TradeProd.htm.
[11] For details, see http://www.cepii.fr/anglaisgraph/bdd/distances.htm.

Table 5: Sectoral disaggregation

Industry code	Industry description
1	Food beverages and tobacco
2	Textiles
3	Wearing apparel except footwear
4	Leather products and footwear
5	Wood products except furniture
6	Paper products
7	Printing and Publishing
8	Petroleum and coal
9	Chemicals
10	Rubber and plastic
11	Other non-metallic mineral products
12	Metallic products
13	Fabricated metal products
14	Machinery except electrical
15	Electric machinery
16	Professional and scientific equipment
17	Transport equipment
18	Other manufacturing

the number of observations (12,150) is given by the number of origins ($M = 15$), times the number of destinations ($M = 15$), times the number of sectors ($S = 18$), times the number of years (which are 3).

Table 6: Descriptive statistics of the trade and geographical variables used.

	N. observ.	Mean	St. dev.	Min	Max
EXP_s^{lh}	12150	746444.9	4308752	0	113000000
Common language	12150	0.06	0.24	0	1
dist	12150	1320.49	767.00	68.44	3362.98
distcap	12150	1333.98	760.74	68.44	3362.98
distw	12150	1344.83	732.85	66.78	3383.27
distwces	12150	1288.80	757.22	13.18	3364.83
Population	11	32159.82	28060.35	5176.209	82211.51

Note: EXP_s^{lh} is in thousand of US dollars while all distance variables are in km. The common language variable is a dummy while population is in thousand of inhabitants.

The variable EXP_s^{lh} in the table corresponds to trade flows (both internal and external) in thousands of US dollars. The common language variable is a dummy indicating whether a couple of countries share a common official language.[12] The distance variables are in km and cover simple measures (*dist* and *distcap*) and weighted measures (*distw* and *distwces*). Simple (geodesic)

[12] In the CEPII database, there are two alternative common language indicators based on different definitions. One indicator considers that two countries share a common language as long as at least 20 per cent of the two populations speaks that language. The other one is similar, but the threshold is now between 9 per cent and 20 per cent. We experimented both indicators getting similar results.

distances are calculated following the great circle formula, which uses the latitudes and longitudes of the most important cities/agglomerations (in terms of population) for the *dist* variable and the geographic coordinates of the capital cities for the *distcap* variable. Both variables incorporate internal distances that are allowed to be non-zero. As in Head and Mayer (2002, 2004), the internal distance d^{ll} of country l is calculated from its area as $d^{ll} = (2/3)\sqrt{area^l/\pi}$. This formula models the average distance between a producer and a consumer on a stylized geography where all producers are centrally located and the consumers uniformly distributed across a disk-shaped region.

By contrast, weighted distances use city-level data on distances and the geographic distribution of population (in 2004) inside each nation. The basic idea is to calculate the distance between two countries as the weighted average bilateral distance between their biggest cities with the corresponding weights determined by the shares of those cities in the overall national populations. This procedure can be used in a totally consistent way for both internal and external distances. Specifically, the general formula developed by Head and Mayer (2002) calculates the distance between country l and h as:

$$d^{lh} = \left(\sum_{p \in l} \sum_{r \in h} (pop^p/pop^l)(pop^r/pop^h) \ (d^{pr})^\theta \right)^{1/\theta} \tag{14}$$

where pop^p (pop^r) designates the population of agglomeration p (r) belonging to country l (h). The parameter θ measures the sensitivity of trade flows to bilateral distance d^{pr}. For the *distw* variable, θ is set equal to 1. The *distwces* calculation sets it equal to -1, which corresponds to the standard distance coefficient estimated from gravity equations. Our benchmark distance variable is *distw*. We will show, however, in Section 7.3 that our results are robust to alternative measures of distance.

For some robustness checks, we will use value added per hour worked in US dollars for our 11 countries by NACE rev.1 industries in the year 2000 as well as the corresponding total amount of hours worked for each country-sector cell. These data are available from the Groningen Growth and Development Centre (GGDC).[13] Finally, data on population come from the New Cronos database provided by EUROSTAT. They refer to the year 2000 in thousands of inhabitants.

5 Calibration

To gauge the impact of trade openness on domestic productivity, we need to recover the parameters of the model and in particular those of equation (11).

5.1 Trade costs

The starting point of our estimation strategy is the gravity equation (13), which will allow us to get the freeness of trade matrix P_s, whose generic element is $\rho_s^{lh} \equiv \left(\tau_s^{lh}\right)^{-k_s}$. From equation (13), one can easily see that the only term that depends on both l and h is ρ_s^{lh}. In fact, the other terms either depend on the origin country only ($N_{E,s}^l \psi_s^l$), or on the destination country only ($L^h \left(c_s^{hh}\right)^{k_s+2}$), or they are constant ($1/[2\gamma_s (k_s+2)]$). Therefore, as in Head and Mayer (2004) and Redding and Venables (2004), one can isolate the effects of these latter terms by means of dummies for origin

[13] Respectively, ICOP Database 1997 Benchmark (http://www.ggdc.net) and 60-Industry Database October 2004 (http://www.ggdc.net).

(EX_l) and destination (IM_h) countries.[14] As to the freeness of trade, we follow Head and Mayer (2004), assuming that $\rho_s^{lh} = \exp(\beta^h + \lambda\ Lang^{lh})\ (d^{lh})^{\delta_s}$ if $l \neq h$, and $\rho_s^{lh} = (d^{lh})^{\delta_s}$ if $l = h$, where d^{lh} is distance between l and h, β^h is a coefficient capturing the fall in trade due to crossing country h border, and $Lang^{lh}$ is a dummy variable that takes value one if l and h share a common language. In other words, as standard in the gravity literature, trade costs are a power function of distance while crossing a border or not sharing the same language impose additional costs. Taking the log of equation (13) we thus get the following regression:

$$\ln(EXP_s^{lh}) = EX_l + IM_h + \delta_s\ \ln(d^{lh}) + \beta^h\ Border^{lh} + \lambda\ Lang^{lh} Border_{lh} + \epsilon_{lh} \qquad (15)$$

where $Border^{lh}$ is a dummy variable that takes value one if $l \neq h$ ('border effect'). In estimating 15, we use data from years 1999, 2000, and 2001 to run a single regression in which we also put year and industry dummies. The coefficient on distances is industry specific while the border effect is country specific. We do not consider country-industry specific border effects because they impose too many parameters and their estimation would be inaccurate. It is important to stress that the specification used to estimate ρ_s^{lh} gives country-industry-sector specific transportation costs and that in general $\rho_s^{lh} \neq \rho_s^{hl}$ due to border effects. Moreover, ρ^{ll} is always less than one due to internal distances.

A visual impression of how openness to international trade varies across countries is given by Figure 1, which shows the average (across sectors and origins) of the estimated ρ_s^{lh} excluding ρ_s^{ll}, i.e. $\bar{\rho}^h = \sum_s \sum_l \hat{\rho}_s^{lh}/((M-1)S)$, for $l \neq h$. The figure reveals that centrality matters with the Netherlands and Belgium being the most open countries. Germany, Denmark and Great Britain are also relatively open, while peripheral Italy, Spain, Norway and Greece are rather closed. The sectoral variation of trade costs (which is essentially due to the δ_s) is shown in Table 7, which reports the estimated distance coefficients. The values are in line with previous findings for Europe by Head and Mayer (2004). In particular, 'Printing and publishing' as well as 'Petroleum and coal' are the least tradable goods while 'Textiles as well as leather products and footwear' are characterized by the smallest trade costs.

5.2 Total factor productivity

After calculating ρ_s^{lh}, we still have to recover the shape parameter of the underlying Pareto distribution of productivity (k_s), and the M endogenous domestic cut-off c_s^{hh} for each sector. For this we need to estimate the distributions of firm-level productivities for all sectors and countries. As a benchmark, we will rely on simple OLS estimations based on the regression

$$\ln(VA_i) = const + a \ln(CAP_i) + b \ln(EMPL_i) + \varepsilon_i \qquad (16)$$

where VA_i is value added, CAP_i is capital (fixed assets), $EMPL_i$ is the number of employees of firm i and the sector/country indices have been dropped to alleviate the notation. The estimated productivity of firm i is thus $\hat{Prod}_{i,OLS} = \exp(\hat{const} + \hat{\varepsilon}_i)$. As robustness check, in Section 7.1 we will show that a more sophisticated estimation methodology of firm productivity, based on Levinsohn and Petrin (2003), yields very similar results.

OLS estimations of productivity are carried out separately for each of the 18 manufacturing industries considered. However, we do not make separate estimations by country assuming *de facto*

[14] This 'fixed effect' approach does not suffer from the specification problems of standard gravity equations discussed by Anderson and Van Wincoop (2003). In particular, these authors show that fixed effects regressions generate parameter estimates that are very similar to those obtained using their multilateral resistance terms.

Table 7: Sectoral trade elasticities with respect to distance

Industry code	Industry description	δ_s
1	Food beverages and tobacco	-1.7434
2	Textiles	-1.0527
3	Wearing apparel except footwear	-1.3452
4	Leather products and footwear	-1.1064
5	Wood products except furniture	-2.0185
6	Paper products	-1.4278
7	Printing and Publishing	-2.4556
8	Petroleum and coal	-2.4164
9	Chemicals	-1.3820
10	Rubber and plastic	-1.6271
11	Other non-metallic mineral products	-1.7603
12	Metallic products	-1.4470
13	Fabricated metal products	-1.7210
14	Machinery except electrical	-1.4861
15	Electric machinery	-1.1236
16	Professional and scientific equipment	-1.5079
17	Transport equipment	-1.4588
18	Other manufacturing	-1.7206
Average		-1.6000

that countries have the same technology up to a scale factor (Hicks neutral factor augmenting technology). The evidence provided by Trefler (1995) supports this assumption for all countries in the sample except Portugal.[15] Note that our assumption does not imply that all countries use the same capital/labor ratio. If Germany is more capital abundant than Italy, then capital would be relatively cheaper in Germany with firms having a higher capital/labor ratio. The fact that we do not carry out separate estimations by country has strong empirical advantages. First, it allows us to have a better measure of productivity as for some countries there are very few country-sector observations. Second, it avoids the following problem. If we found the sum of the coefficients a and b to differ between two countries, then our estimated average productivity would turn out to be higher in the country with the lower sum simply because this would map into a higher value of the constant.[16]

Table 8 shows average (across firms) OLS productivity by country, as well as per capita income in PPS.[17] As one can see, the two measures are closely related, with the correlation being 0.61. The table shows that our OLS estimates of productivity are generally in line with aggregate figures. A notable exception is Germany, whose omission increases the correlation between productivity and GDP to 0.88. The reason is that both the Amadeus and the MIP databases have a strong bias towards West German firms, which are known to be much more productive than East German ones. However, our results on the aggregate gains from trade are not much sensitive to the exclusion of Germany, so we decided to keep it in the analysis. Figure 2 reports all estimated productivities by

[15] In unreported estimations, where we exclude Portugal, we find qualitatively similar results to the ones we report when Portugal is included in the sample.

[16] The estimated values of $a + b$ are all between 0.9 and 1.0 and never significantly different from 1. They are thus in line with constant returns to scale.

[17] Per capita income in PPS should be a better measure of 'physical' productivity because it deflates nominal values by country specific price indices.

sector and country.

Table 8: Productivity across countries: OLS estimations.

Country initials	Country	OLS Productivity	Per capita GDP in PPS
BE	Belgium	43.22	104.30
DE	Germany	63.63	101.60
DK	Denmark	50.26	114.56
ES	Spain	32.61	83.78
FI	Finland	37.17	102.59
FR	France	40.22	103.13
GB	Great Britain	38.95	102.14
IT	Italy	40.30	99.35
NL	Netherlands	42.25	108.53
PT	Portugal	24.24	73.08
SE	Sweden	34.44	106.91
Average		40.06	100.00

Now that we have $\hat{Prod}_{i,OLS}$ for all firms in the year 2000, we can use them to estimate the shape parameter k_s of the Pareto distribution and the cutoffs c_s^{hh}.[18] To estimate the former, it is useful to recall the following property. Consider a random variable X (our productivity) with observed cumulative distribution $F(X)$. If the variable is distributed as a Pareto with shape parameter k_s, then the OLS estimate of the slope parameter in the regression of $\ln(1 - F(X))$ on $\ln(X)$ plus a constant is a consistent estimator of $-k_s$ and the corresponding R^2 is close to one.[19] Table 9 shows the estimated k_s and the R^2 of our regressions by sector. For all sectors the R^2 is far above 0.8, which shows that the Pareto is a fairly good approximation of the underlying productivity distributions, and the average k_s is estimated to be close to 2. Large values of k_s characterize sectors in which the productivity distribution is skewed towards relatively small and inefficient firms firms ('Leather products and footwear', 'Wood products except furniture', 'Rubber and plastic', 'Fabricated metal products', 'Machinery except electrical'). Small values of k_s are associated, instead, with an even distribution of firms across all productivity levels and sizes ('Wearing apparel except footwear', 'Petroleum and coal', 'Chemicals', 'Professional and scientific equipment').[20]

Turning to the cut-offs c_s^{hh}, these are country and sector specific. In the model, they represent the highest cost (or equivalently the reverse of the lowest productivity) of active domestic firms. The maximum likelihood estimator of the cut-off for a Pareto distribution is the minimum observed value. However, this is probably a rather unreliable method to implement with micro data because of extreme observations. Consequently, we prefer to use a moment estimator based on the formula of the mean of a Pareto. Specifically, if X (our productivity) is distributed as a Pareto with shape parameter k_s and cut-off x then its mean is $E(X) = x\ k_s/(k_s - 1)$. Using the country-sector average productivities and the previously estimated shape parameters, such formula can be inverted to recover all the productivity cut-offs, which are simply scaled average productivities. Finally, in order to pass from productivity cutoffs to cost cutoffs, which are those needed in equation (11), we

[18] One can reasonably argue that considering only one year may be problematic whenever the business cycles do not coincide across countries. To address such concern, we have also estimated 7-year average individual productivities. The corresponding results are very similar to the 1-year estimates, so we do not report them.

[19] See Norman, Kotz and Balakrishnan (1994).

[20] We have also tried alternative estimation techniques for k_s, such as the inverse of the standard error of $\ln(X)$. Overall results are very similar.

Table 9: Sectoral k_s and the R^2 from the regression method

Industry code	Industry description	k_s	R^2
1	Food beverages and tobacco	2.004	0.898
2	Textiles	2.248	0.872
3	Wearing apparel except footwear	1.804	0.904
4	Leather products and footwear	2.345	0.893
5	Wood products except furniture	2.454	0.871
6	Paper products	1.966	0.827
7	Printing and Publishing	1.988	0.898
8	Petroleum and coal	1.604	0.862
9	Chemicals	1.811	0.848
10	Rubber and plastic	2.372	0.868
11	Other non-metallic mineral products	2.156	0.826
12	Metallic products	2.206	0.848
13	Fabricated metal products	2.450	0.875
14	Machinery except electrical	2.346	0.898
15	Electric machinery	1.930	0.881
16	Professional and scientific equipment	1.844	0.856
17	Transport equipment	2.062	0.861
18	Other manufacturing	2.128	0.900
Average		2.095	0.871

simply need to raise the productivity cut-offs to the power of minus one.[21]

6 Simulation

We start with taking a logarithmic transformation of equation (11):

$$\ln(c_s^{hh}) = \ln(a_s) + \frac{1}{k_s+2}\left[\ln(b_s) + \ln\left(\frac{\sum_{l=1}^{M}\left|C_s^{lh}\right|}{|P_s|}\frac{1}{\psi_s^l}\right)\right] \tag{17}$$

where $a_s = f_{E,s}\ \gamma_s$, and $b_s = 2(k_s+1)(k_s+2)/L^h$. So far, we have computed ρ_s^{lh} (so we are able to evaluate both $\left|C_s^{lh}\right|$ and $|P_s|$), k_s and c_s^{hh}. We also have L^h, which is simply the population of county h. Therefore, we know everything about (17) except a_s, which is industry specific, and ψ_s^l, which is country and industry specific. However, a_s is a log-linear term that simply cancels out when comparing (as we do) productivity scenarios obtained with different trade costs. For each sector s we can thus set $a_s = 1$ without loss of generality and then solve the resulting non-linear system of 11 equations (17) in the 11 unknown ψ_s^l's (one per country). With the ψ_s^l's we can finally simulate the model and evaluate the changes in average productivity induced by changes in the freeness of trade. This amounts to computing the new equilibrium c_s^{hh}'s and comparing them with the initial ones.[22] In particular, we consider the two following scenarios:

[21] Two sectors (number 4 and 8) in Denmark have missing c_s^{hh} because we do not have observations (firms) for them. This explains why, in the Tables where we show the results of our simulations, average gains across sectors and countries do not perfectly match.

[22] If one were interested in ranking countries according to their average absolute advantage ψ_s^l, this could be obtained by solving the system of equations (17) and removing the sectoral scaling term $a_s = f_{E,s}\ \gamma_s$ from the solutions. Specifically, sectoral scaling could be removed by taking the logarithm of the solutions and regress the

1. A situation in which international trade barriers are prohibitive (i.e. $\rho_s^{lh} = 0$ for $l \neq h$). This provides an assessment of the overall 'costs of non-Europe' as measured by foregone productivity were EU countries to become autarkic.

2. A situation in which international trade costs (τ_s^{lh} for $l \neq h$) are reduced by 5 per cent. This provides an assessment of the 'gains from (freer) trade'.

6.1 Costs of non-Europe

Table 10 shows the average 'costs of non-Europe' by country. These are measured as minus percentage changes in mean productivity (i.e. percentage changes in the cutoff c_s^{hh}) with respect to the estimate in the year 2000.

Table 10: Costs of non-Europe by country: OLS estimations.

Country initials	Country	c_s^{hh} with trade	c_s^{hh} with no trade	% Cost of non-Europe	$\bar{\rho}^h$
BE	Belgium	0.049	0.059	16.72	0.00028
DE	Germany	0.037	0.048	23.29	0.00005
DK	Denmark	0.043	0.055	22.29	0.00009
ES	Spain	0.058	0.066	11.63	0.00001
FI	Finland	0.058	0.067	13.34	0.00002
FR	France	0.053	0.061	13.05	0.00002
GB	Great Britain	0.060	0.061	3.31	0.00005
IT	Italy	0.058	0.061	6.66	0.00001
NL	Netherlands	0.051	0.057	12.03	0.00026
PT	Portugal	0.090	0.093	4.62	0.00002
SE	Sweden	0.068	0.076	11.43	0.00004
Average		0.057	0.064	12.58	0.00008

Based on our simulations, autarky reduces average productivity by 12.58%. This is a sizeable number and suggests that the selection effect is an important channel through which the benefits of international trade materialize. This number, however, varies considerably across countries. Germany, with 23.29%, is the country that loses the most, closely followed by Denmark (22.29%). At the other extreme, there is Great Britain with a mere 3.31% followed by Portugal (4.62%) and Italy (6.66%). According to our economic model, such losses are positively correlated (0.26) with the openness to trade (see $\bar{\rho}^h$ in the last column), so central countries lose more when turning to autarky. Such correlation is not perfect because of comparative advantage and different trade costs across sectors. For example, the big losses of Germany and Denmark are essentially due to their high underlying productivities (low c_s^{hh} with no trade) whose beneficial effect is magnified by international trade.

It is interesting to compare our results with those obtained by Eaton and Kortum (2002) for the same thought experiment in a probabilistic Ricardian framework. Those authors calculate the fall in productivity (measured as GDP per worker) due to autarky for a sample of 19 OECD

results on a complete set of industry dummies. The country average of the regression residuals across sectors could then be used as a measure of the average absolute advantage. When we follow such procedure, we find that Finland, Sweden and Denmark are the top three countries in terms of average absolute advantage. Interestingly enough, these three countries are ranked first, third and fourth respectively (the US come second) in terms of the Global Competitiveness Index calculated by the World Economic Forum (2005).

countries in 1990. For our 11 European countries the average decrease in productivity is 3.95%, much smaller than our 12.58%. The discrepancy may be due to the different sample of interacting countries. In particular, OECD countries are generally less integrated than EU ones. Moreover, Eaton and Kortum (2002) simulate a perfectly competitive model and their base year pre-dates ours by a crucial decade for European integration.

Table 11: Costs of non-Europe by industry: OLS estimations.

Industry code	Industry description	c_s^{hh} with trade	c_s^{hh} with no trade	δ_s	% Cost of non-Europe
1	Food beverages and tobacco	0.096	0.104	-1.743	8.95
2	Textiles	0.033	0.040	-1.053	18.45
3	Wearing apparel except footwear	0.042	0.049	-1.345	16.66
4	Leather products and footwear	0.043	0.054	-1.106	24.17
5	Wood products except furniture	0.043	0.046	-2.019	6.51
6	Paper products	0.076	0.092	-1.428	17.37
7	Printing and Publishing	0.039	0.041	-2.456	5.92
8	Petroleum and coal	0.145	0.157	-2.416	9.51
9	Chemicals	0.059	0.069	-1.382	13.96
10	Rubber and plastic	0.050	0.056	-1.627	10.00
11	Other non-metallic mineral products	0.071	0.079	-1.760	10.55
12	Metallic products	0.059	0.066	-1.447	10.16
13	Fabricated metal products	0.039	0.042	-1.721	8.75
14	Machinery except electrical	0.036	0.040	-1.486	11.23
15	Electric machinery	0.055	0.063	-1.124	14.49
16	Professional and scientific equipment	0.042	0.049	-1.508	14.71
17	Transport equipment	0.064	0.072	-1.459	12.33
18	Other manufacturing	0.038	0.042	-1.721	11.69
Average		0.057	0.064	-1.600	12.52

Turning to sectoral disaggregation, Table 11 shows the average costs of non-Europe by industry and compares them with the elasticity of trade flows to distance (δ_s), as obtained from the gravity equation. In the table losses are positively correlated with the elasticity (0.78): the smaller the absolute value of the elasticity (i.e. the more a sector is open), the larger the corresponding loss from autarky.

6.2 Gains from (freer) trade

The effects of a 5 per cent reduction in international trade barriers are reported in Table 12, which shows the percentage changes in average productivity ('gains from (freer) trade') by country and compares them with the 'costs of non-Europe'. The table shows that the 'gains from (freer) trade' are highly correlated with the 'costs of non-Europe' (0.90), Portugal being the main exception. As before, both accessibility and comparative advantage positively affect the productivity increase due to a more competitive environment. The overall gains from trade are 2.13%. This is a sizeable number, although smaller than the 4.7% productivity increase obtained by Bernard et al. (2003) for the US when considering the same percentage fall in trade costs. Besides differences in the underlying models, the fact that the US are a very productive country may help explain such discrepancy. Indeed, the two most productive countries in our sample, Denmark and Germany, experience comparable gains, respectively 4.4% and 4.6%. Furthermore, the fact that we analyze trade among a smaller set of countries and that we do not consider intermediate goods, whose price reduction is the main driver of the gains in Bernard et al. (2003), may also explain why we get smaller impacts.

It is also interesting to compare our findings with those in Smith and Venables (1988), who simulate the effects of a reduction of intra-EU trade costs in a CGE model with increasing returns to scale, segmented markets and product differentiation. Firms are identical within countries but they are allowed to differ in size and product lines between countries. Market structure is alternatively modeled as Cournot or Bertrand competition with free or restricted entry. Thus, while our model

Table 12: Gains from (freer) trade by country: OLS estimations.

Country initials	Country	% Gains from trade	% Costs of non-Europe
BE	Belgium	2.86	16.72
DE	Germany	4.60	23.29
DK	Denmark	4.40	22.29
ES	Spain	2.11	11.63
FI	Finland	1.81	13.34
FR	France	1.74	13.05
GB	Great Britain	0.23	3.31
IT	Italy	0.66	6.66
NL	Netherlands	1.24	12.03
PT	Portugal	2.02	4.62
SE	Sweden	1.81	11.43
Average		2.13	12.58

stresses the impact of trade on market share reallocations ('selection effect'), their model focuses instead on scale economy exploitation ('scale effect'). In Smith and Venables (1988) a decline in trade costs makes competition fiercer, decreases prices, and expands sales. Due to increasing returns to scale, average costs fall, especially with free entry. However, as firms are identical within countries, no market share reallocations take place towards more productive firms. Though simulations are run for many industries, only for 'Domestic electrical equipment' reported data allow for a reasonable comparison with our analysis. In the closest specification to our model, Smith and Venables (1988) estimate that a 8% reduction in trade costs yields a 0.76% drop in average production costs. This is smaller than the 1.94% decrease in average costs we find for 'Electric Machinery' as response to a 5% reduction in trade costs (see Table 13). We interpret this difference as capturing the relative importance of the scale and selection effects. Indeed, as argued by Tybout and Westbrook (1996), the neglect of firm heterogeneity implies that scale effects may be even overstated in CGE models such as Smith and Venables (1988). On the one hand, exporting plants are typically the largest in their industry, so they are not likely to exhibit much potential for further scale economy exploitation. On the other hand, large plants also account for most of the production in any industry, so foregone economies of scale due to downscaling in import-competing sectors are also likely to be minor.

As to the sectoral dimension, Table 13 reports the simulated 'gains from (freer) trade' by industry as percentage changes in average productivity with respect to the year 2000. These are compared with the corresponding 'costs of non-Europe'. As expected, the correlation between the two effects is high (0.91). In particular, the finding that industries with small distance elasticity gain more from trade liberalization is confirmed.[23]

7 Robustness checks

In this section we explore how sensitive our results are to alternative measures of productivity and distance as well as to alternative ways of recovering the shape parameter k_s.

[23] In unreported simulations we have also considered alternative changes in trade costs, obtaining results that nearly perfectly correlate with the ones we report. For instance, a 5% increase in international trade costs yields a 1.57% reduction in average productivity, with correlations with the 'costs of non-Europe' and the 'gains from (freer) trade' as high as 0.97 and 0.98 respectively.

Table 13: Gains from (freer) trade by industry: OLS estimations.

Industry code	Industry description	% Gains from trade	% Cost s of non-Europe
1	Food beverages and tobacco	1.12	8.95
2	Textiles	3.16	18.45
3	Wearing apparel except footwear	2.52	16.66
4	Leather products and footwear	6.40	24.17
5	Wood products except furniture	0.97	6.51
6	Paper products	4.65	17.37
7	Printing and Publishing	0.71	5.92
8	Petroleum and coal	1.62	9.51
9	Chemicals	1.70	13.96
10	Rubber and plastic	1.65	10.00
11	Other non-metallic mineral products	1.72	10.55
12	Metallic products	1.23	10.16
13	Fabricated metal products	1.30	8.75
14	Machinery except electrical	1.75	11.23
15	Electric machinery	1.94	14.49
16	Professional and scientific equipment	2.41	14.71
17	Transport equipment	1.59	12.33
18	Other manufacturing	1.89	11.69
Average		2.13	12.52

7.1 Individual productivity

In section 5.2 we estimated individual productivity as the exponential of the residual of a simple OLS regression of (the log of) value added on (the log of) capital and labor. All the results obtained so far are based on that estimate. We now repeat the analysis relying on a different estimation procedure. The aim is to assess the robustness of our findings when one takes into account the possible presence of a simultaneity bias in the OLS estimate. The main idea is that a firm hires capital and labor after having observed the realization of its random TFP. The larger the realization, the larger the quantities of inputs hired. As long as the realization is unobservable to the econometrician, the regressors in (16) are thus correlated with the residual of the OLS regression.

The bias can be removed by identifying an observable proxy variable and introducing it as an additional regressor in (16). The proxy is such that, according to economic theory, it can be expected to respond to the TFP realization observed only by the firm. Accordingly, the residual of the new regression is free from any correlation with the inputs due to the asymmetric observability of realized TFP. This approach, originally proposed by Olley and Pakes (1996) using investment as a proxy, has been recently extended by Levinshon and Petrin (2003) (henceforth LP) using intermediate inputs instead. Data availability forces us to choose the latter proxy since for our sample of Amadeus firms there is no information available on investment. Moreover, as information on the cost of materials is not available for Denmark and Great Britain, also the LP procedure can be implemented on a sample of 9 countries only. That is why we preferred to rely on OLS in our benchmark analysis.[24]

The LP estimates for the year 2000 are reported in Table 14, which shows the average productivity by country.[25] Although absolute levels are quite different, the correlation between the OLS and LP estimates *within* a sector is around 0.9, so the two estimates differ only for a scaling factor, which is innocuous for our counterfactual analysis. Still, a problem seems to appear for Germany and Sweden, which are respectively too much and too little productive in the LP estimates with respect to the OLS ones.

Tables 15, 16 and 17 show the results of our counterfactuals when LP rather than OLS produc-

[24] There is indeed another point that led us to prefer OLS: the returns to scale estimated by LP are in some cases significantly smaller than one. This is probably due to the bad measurement of input consumption, whose definitions may not be homogeneous across countries.

[25] The relevant capital, intermediates, and value-added deflators are used.

Table 14: Productivity across countries: LP estimations.

Country initials	Country	LP Productivity	OLS Productivity
BE	Belgium	128.58	43.22
DE	Germany	233.03	63.63
ES	Spain	115.14	32.61
FI	Finland	118.52	37.17
FR	France	121.96	40.22
IT	Italy	117.92	40.30
NL	Netherlands	168.42	42.25
PT	Portugal	113.93	24.24
SE	Sweden	80.75	34.44
Average		133.14	39.79

Table 15: Costs of non-Europe by country: LP estimations.

Country initials	Country	% Costs of non-Europe LP	% Costs of non-Europe OLS
BE	Belgium	11.30	16.72
DE	Germany	22.73	23.29
ES	Spain	9.51	11.63
FI	Finland	15.19	13.34
FR	France	10.78	13.05
IT	Italy	5.52	6.66
NL	Netherlands	12.19	12.03
PT	Portugal	5.35	4.76
SE	Sweden	6.09	11.43
Average		10.96	12.54

tivities are used to calibrate the model. The two former tables report the 'costs of non-Europe' for countries and sectors respectively. The latter displays the 'gains from (freer) trade'. To ease comparison, we also report the OLS results for the same countries and sectors.[26] The tables show that results are largely consistent between the LP and OLS simulations. On average, both the 'costs of non-Europe' and the 'gains from (freer) trade' decrease, but this is not unexpected because there are missing countries in the LP sample and thus the potential benefits from trade are smaller.

7.2 Aggregate productivity

As shown in Table 1, data coverage in the Amadeus dataset is not very satisfactory for some countries, namely Denmark, Germany, and Portugal. In the case of Germany we have complemented our data with additional firm level information provided by ZEW (see Section 4). Nevertheless, even though we eventually succeed in having a good sample of German firms, an issue of representativeness due to small sample size still remains for the other countries. More generally, one could also

[26]Four sectors (number 4, 8, 16 and 18) for Portugal have missing c_s^{hh} because we do not have observations (firms) for them. This is the reason why, when comparing LP with OLS gains, OLS values for Portugal are slightly different from those reported in the previous section. Furthermore, considering that two countries are lost when computing LP productivities, OLS sectoral gains are also different from those previously reported.

Table 16: Costs of non-Europe by industry: LP estimations.

Industry code	Industry description	% Cost s of non-Europe LP	% Cost s of non-Europe OLS
1	Food beverages and tobacco	7.36	7.98
2	Textiles	14.84	18.27
3	Wearing apparel except footwear	16.90	18.12
4	Leather products and footwear	29.41	29.27
5	Wood products except furniture	5.84	6.87
6	Paper products	16.51	18.32
7	Printing and Publishing	4.98	6.23
8	Petroleum and coal	10.51	11.63
9	Chemicals	12.21	13.40
10	Rubber and plastic	6.45	10.10
11	Other non-metallic mineral products	8.51	9.31
12	Metallic products	9.05	10.33
13	Fabricated metal products	6.77	9.06
14	Machinery except electrical	12.95	10.75
15	Electric machinery	10.60	13.28
16	Professional and scientific equipment	10.74	12.70
17	Transport equipment	9.45	12.16
18	Other manufacturing	8.43	13.31
Average		11.20	12.84

Table 17: Gains from (freer) trade: LP estimations.

Country initials	Country	% Gains from trade LP	% Gains from trade OLS
BE	Belgium	1.28	2.86
DE	Germany	5.19	4.60
ES	Spain	1.56	2.11
FI	Finland	2.61	1.81
FR	France	1.20	1.74
IT	Italy	0.50	0.66
NL	Netherlands	1.24	1.24
PT	Portugal	0.12	2.46
SE	Sweden	0.89	1.81
Average		1.62	2.14

argue that, since the Amadeus coverage is generally biased towards large firms, representativeness is potentially an issue for all our countries.

To address these concerns, we repeat our simulation exercise relying on the country-sector productivities provided by GGDC (see Section 4). Under the Pareto assumption, such aggregates allow us to recover alternative measures of the domestic cutoffs c_s^{hh} by adequately scaling up the corresponding average productivities. The GGCD dataset contains the (producer) price adjusted value added per hour worked for each of the 18x11 country-sector pairs in US dollars for the year 2000.[27] It represents the most accurate and comparable existing record of international productivities. Compared with our individual estimates, it has the advantage of being based on a very large firm coverage. Moreover, its productivities are deflated by industry-specific purchasing power parities, which accounts for the fact that prices vary across countries and thus gives a more reliable measure of 'physical' productivities. GGDC productivities, however, do not take into account capital intensity as we instead do when estimating individual TFP. Moreover, for one sector ('Petroleum

[27] Specifically, we have combined the ICOP Industrial Database (New Benchmarks), which reports the productivities, with the 60-Industry Database (October 2004), which reports the total number of hours worked. Data originally follow the Nace rev.1 classification. We have converted them into our 18-industry classification by weighting the disaggregated productivities by the total number of hours worked in each Nace industry.

and coal') no productivity data is available and we have to restrict our analysis on 17 industries.

Table 18: Costs of non-Europe by country: AP estimations.

Country initials	Country	% Costs of non-Europe AP	% Costs of non-Europe OLS
BE	Belgium	26.79	17.14
DE	Germany	8.67	24.06
DK	Denmark	28.94	22.29
ES	Spain	10.43	10.31
FI	Finland	22.88	12.80
FR	France	25.17	13.20
GB	Great Britain	5.81	3.40
IT	Italy	12.39	6.98
NL	Netherlands	19.72	12.52
PT	Portugal	3.30	4.89
SE	Sweden	23.89	12.02
Average		17.09	12.69

Table 19: Gains from (freer) trade: AP estimations.

Country initials	Country	% Gains from trade AP	% Gains from trade OLS
BE	Belgium	11.00	2.98
DE	Germany	0.86	4.81
DK	Denmark	5.54	4.40
ES	Spain	1.48	1.69
FI	Finland	7.53	1.71
FR	France	14.29	1.79
GB	Great Britain	0.59	0.23
IT	Italy	2.10	0.69
NL	Netherlands	3.82	1.30
PT	Portugal	0.55	2.14
SE	Sweden	9.15	1.91
Average		5.17	2.15

Tables 18 and 19 show the simulations based on GGDC data together with (comparable) simulations based on individual data.[28] In the former table, the average 'costs of non-Europe' increase from 12.69% to 17.09%. The correlation between the results based on the two datasets is 0.90 (0.53) if Germany is excluded (included). Hence, an issue of representativeness may exist for Germany only. Indeed, in the GGDC database, where the coverage is as good for the former East Germany as for West Germany, German firms are much less productive with respect to other countries. Once this is controlled for, the overall 'costs of non-Europe' and much more their distribution across countries seem to be fairly robust. If anything, our benchmark results may underestimate the overall selection effects. Similar conclusions can be reached for the 'gains from (freer) trade' by inspecting Table 19.

[28]When simulating the model calibrated on GGDC data, we use the k_s's from the OLS-based individual productivity distributions.

7.3 Trade costs

So far we have relied on a specific bilateral distance metric (*distw*). We now check to what extent our results depend on such metric. First of all, we simulate the model again for each of the alternative metrics (*distswces*, *dist* and *distcap*) discussed in Section 4. While all of them account for internal distances their own way, there is still no general consensus on how to deal with them and even on whether they should be included at all. For this reason, we also present the results under the rather unrealistic assumption of irrelevant internal distances ($\rho_s^{ll} = 1$).[29] Finally, as shown by Santos Silva and Tenreyro (2006), the presence of heteroschedasticity and truncation in log-linearized models (such as a gravity equation) may induce a systematic bias in OLS estimation. Thus, we also estimate trade costs following the Poisson Pseudo-Maximum Likelihood (PPML) method they propose.

Table 20: Costs of non-Europe by country with different distance measures and estimation techniques.

Country initials	Country	% non-Europe	% non-Europe	% non-Europe	% non-Europe	% non-Europe	% non-Europe
BE	Belgium	16.72	24.23	24.66	23.71	39.94	23.91
DE	Germany	23.29	21.81	26.83	18.63	21.48	26.27
DK	Denmark	22.29	29.30	22.31	23.14	39.62	30.50
ES	Spain	11.63	8.83	8.10	9.46	7.91	13.80
FI	Finland	13.34	12.32	21.58	21.94	19.23	21.77
FR	France	13.05	12.73	12.87	14.79	9.70	15.36
GB	Great Britain	3.31	2.99	4.64	5.21	5.09	5.48
IT	Italy	6.66	7.35	4.03	4.39	5.76	9.86
NL	Netherlands	12.03	16.58	15.39	15.84	30.52	18.12
PT	Portugal	4.62	2.04	4.47	4.45	8.04	7.83
SE	Sweden	11.43	10.52	12.06	13.01	12.92	16.70
Average		12.58	13.52	14.27	14.05	18.20	17.24
Distance measure		distw	distwces	dist	distcap	no internal dist.	distw
Estimation method		OLS	OLS	OLS	OLS	OLS	PPML

Tables 20 and 21 show that, no matter how internal distances are measured, results are essentially unchanged (Columns 3 to 6). When internal distances are, instead, neglected (Column 7), both the 'costs of non-Europe' (18.20%) and the 'gains of (freer) trade' (5.12%) are considerably larger than in the baseline case. This is not surprising because omitting internal distances amounts to setting border effects to zero, thus increasing the observed international openness. Consequently, going back to autarky becomes more costly and a given reduction in trade costs corresponds to a more open environment. Moreover, setting $\rho_s^{ll} = 1$ underestimates (overestimates) the degree of internal freeness of trade in small (large) countries, thus increasing (decreasing) the relative importance of international trade for competition and selection. This is why for small countries like Belgium and the Netherlands the cost of non-Europe is bigger when all ρ_s^{ll} are set to 1. Similar arguments apply to PPML-based simulations. A constant outcome of such a methodology is to provide smaller estimates of the distance elasticity (δ_s) with respect to OLS. Therefore, the world

[29] With no internal distances, the freeness of trade ρ_s^{lh} is still computed on the basis of equation 15 with international distances based on *distw*. However, only observations for which $l \neq h$ are considered and borders effects are no more identifiable. Moreover, to make ρ_s^{lh} economically comparable to the assumed $\rho_s^{ll} = 1$, we have to re-scale the unit of measurement of distance so that, once trade costs are computed with internal distances, the average ρ_s^{ll} is still equal to one for each sector. In this respect, it is important to point out that the value of ρ_s^{lh} has no 'absolute' interpretation because of the arbitrariness in the choice of the distance unit. More precisely, the unit of measurement of distance (which affects the scale of both ρ_s^{ll} and ρ_s^{lh}) is immaterial for the simulations as long as internal distances are considered because it simply has a multiplicative effect on all elements of the matrix P_s. What matters, instead, is the ratio between ρ_s^{ll} and ρ_s^{lh}, which determines the 'relative' degree of international openness and so the contribution of foreign competition to selection.

Table 21: Gains from (freer) trade by country with different distance measures and estimation techniques

Country initials	Country	% trade gains	% trade gains	% trade gains	% trade gains	% trade gains	% trade gains
BE	Belgium	2.86	6.87	5.65	8.95	33.63	4.08
DE	Germany	4.60	4.09	4.85	2.64	1.82	5.54
DK	Denmark	4.40	11.27	4.34	4.54	13.88	7.44
ES	Spain	2.11	1.53	1.14	1.51	1.24	2.56
FI	Finland	1.81	1.83	4.52	4.55	3.65	5.17
FR	France	1.74	1.90	1.28	1.46	0.82	1.92
GB	Great Britain	0.23	0.21	0.31	0.33	0.21	0.31
IT	Italy	0.66	0.82	0.37	0.41	0.51	0.96
NL	Netherlands	1.24	2.06	1.67	1.37	-1.70	3.61
PT	Portugal	2.02	0.32	1.67	1.50	0.70	0.73
SE	Sweden	1.81	1.73	2.46	2.82	1.54	1.54
Average		2.13	2.97	2.57	2.73	5.12	3.08
Distance measure		distw	distwces	dist	distcap	no internal dist.	distw
Estimation method		OLS	OLS	OLS	OLS	OLS	PPML

appears to be much more open to trade and autarky is more costly. As in the previous section, if anything, our benchmark results may underestimate the overall effects.

7.4 Shape parameter

The final issue we address is the calibration of the shape parameter of the underlying Pareto distribution of productivity draws (k_s). As argued by Bernard et al. (2003), individual TFP productivity $\hat{Prod}_i$ is certainly measured with an error. Even if such error were uncorrelated with the 'true' productivity $Prod_i$, it should nonetheless deflate the value of k_s.[30] In other words, measurement errors may make us observe too much TFP variability across firms so that our k_s (whose average in OLS estimations is 2.095) may be too low. Bernard et al. (2003) propose to solve that problem by reconstructing the parameters of the productivity distribution from aggregate data. In particular, they recover the shape parameter of their Fréchet distribution by matching the productivity and size advantage of exporters between simulated and actual US data. Such solution comes at the cost of imposing more structure. Indeed, one has to believe that not only firm productivity follows a certain distribution but also all other assumptions (on demand, market structure, etc.), which are needed to obtain the average productivity of exporters from the theoretical model, hold. While we consider our approach as more reliable, we find it useful to check how sensitive our results to the potential mismeasurement of k_s.

Unfortunately we do not have reliable export data in the Amadeus database, so we can not exploit the size and the productivity advantage of exporters as Bernard et al. (2003) do. Therefore, we simply check how our benchmark results change when we adopt their value $k_s = 3.6$, which is the same for all sectors since their analysis has no sectoral disaggregation. It is important to remark that we can use their estimate because the Fréchet and the Pareto distributions are closely related. Indeed, in their model, where only the lowest cost supplier is active in any particular country, if all potential suppliers draw their costs from an unobservable Pareto distribution with shape parameter k_s, then the corresponding observable extreme value distribution is precisely a Fréchet distribution with shape parameter k_s.

[30] The variance of a Pareto distribution is a decreasing function of k_s. In particular, if X is distributed as a Pareto with shape parameter k_s, the standard error of $\ln(X)$ is $1/k_s$. A simple way to recover k_s is thus to consider the standard error of $\ln(\hat{Prod}_i) = \hat{const} + \hat{\varepsilon}_i$. Clearly, as long as there is an uncorrelated measurement error in the estimates of $const$ and ε_i, the standard error of the variable $\hat{const} + \hat{\varepsilon}_i$ is greater than that of $const + \varepsilon_i$, and so k_s is underestimated.

There is, however, a different reason why our calibration of k_s may instead be too high. In our model more (physically) productive firms set lower prices, so they do not entirely translate their higher productivity into higher values of sales per worker. Therefore, our productivity estimation, which, as standard, is based on the value of sales or value added, may underestimate (overestimate), the TFP of more (less) productive firms, thus reducing the observed variance of productivity and overstating k_s. This potential bias is studied by Foster, Haltiwanger and Syverson (2005), who present the only TFP research we know that relies on firm-level physical quantities. They find that at plant level physical productivities are inversely correlated with prices so that the standard error of (the ln of) revenue-based TFP is lower than the one of the output-based TFP, as predicted by our model. However, they also find that the correlation between revenue-based and output-based TFP's is not negligible (0.64) and the bias in terms of standard errors of (the ln of) productivities, which is all that matters for our k_s, is rather small (20%).[31]

Table 22: Costs of non-Europe by country with different k_s: OLS estimations.

Country initials	Country	% Costs of non-Europe	% Costs of non-Europe	% Costs of non-Europe
BE	Belgium	17.68	16.29	16.72
DE	Germany	22.89	26.76	23.29
DK	Denmark	23.20	21.20	22.29
ES	Spain	11.56	13.94	11.63
FI	Finland	14.15	13.22	13.34
FR	France	13.60	13.77	13.05
GB	Great Britain	3.59	4.39	3.31
IT	Italy	6.87	8.31	6.66
NL	Netherlands	12.64	12.58	12.03
PT	Portugal	4.97	5.27	4.62
SE	Sweden	12.73	10.35	11.43
Average		13.08	13.28	12.58
k_s		our k_s's reduced by 20%	$k_s = 3.6$ for all s	our k_s's

Tables 22 and 23 respectively compare our benchmark results on the 'costs of non-Europe' and the 'gains from (freer) trade' with those obtained by either reducing all our k_s's by 20% in the wake of Foster, Haltiwanger and Syverson (2005) or attributing $k_s = 3.6$ to all sectors in the wake of Bernard et al. (2003). This amounts to increasing our average k_s by 70% from its OLS value of 2.1. In general, one should not expect the productivity gains from selection to change monotonically with k_s as two opposing effects operate. On the one hand, for given trade freeness, as k_s grows the impact of selection on productivity becomes stronger because the cost distribution becomes more skewed towards less productive firms. On the other hand, for a given cost distribution, as k_s grows trade freeness $\rho_s^{lh} = \left(\tau_s^{lh}\right)^{-k_s}$ falls, which weakens firm selection. Non-monotonicity characterizes Table 22 as both reducing and increasing k_s slightly raise the average 'costs of non-Europe', although this is not the case for all countries. Table 23 shows, instead, that the average 'gains from (freer) trade' change monotonically with k_s, although this is not the case for all countries. Overall, our simulations suggest that correcting k_s the 20% upward bias changes our results only slightly, whereas imposing $k_s = 3.6$ increase substantially only the 'gains from (freer) trade'. Once more, if anything,

[31] The problem of the unobserved price bias is highlighted by Klette and Griliches (1996). Melitz (2000) proposes a correction based on a simple CES Dixit-Stigliz model with constant markups. Unfortunately the complexity of our model with variable markups prevents us from applying any simple comparable correction.

Table 23: Gains from (freer) trade with different k_s: OLS estimations.

Country initials	Country	% Gains from trade	% Gains from trade	% Gains from trade
BE	Belgium	2.24	3.06	2.86
DE	Germany	2.99	13.24	4.60
DK	Denmark	3.28	15.78	4.40
ES	Spain	1.40	5.93	2.11
FI	Finland	1.40	5.63	1.81
FR	France	1.31	5.02	1.74
GB	Great Britain	0.19	-0.71	0.23
IT	Italy	0.51	1.14	0.66
NL	Netherlands	0.96	4.15	1.24
PT	Portugal	1.03	-0.90	2.02
SE	Sweden	1.34	0.19	1.81
Average		1.51	4.78	2.13
k_s		our k_s's reduced by 20%	$k_s = 3.6$ for all s	our k_s's

our benchmark results may underestimate the overall selection effect.

8 Conclusion

We have calibrated a multi-country multi-sector model with heterogeneous firms, monopolistic competition and variable markups on firm-level data and trade figures for a panel of 11 EU countries. When simulating different integration scenarios, we have found that in the year 2000 an increase of trade barriers to prohibitive levels would have caused an average productivity loss of roughly 13 per cent. On the other hand, a 5 per cent reduction in trade costs, would have raised average productivity by roughly 2 per cent. These estimates have been shown to be fairly robust to alternative measures of productivity and trade costs. Indeed, the robustness checks overall suggest that, if anything, those numbers may actually underestimate the overall selection effects. This shows that the Darwinian selection of the best firms is an important effect of trade liberalization.

....

References

[1] Alcalà F. and A. Ciccone (2004) Trade and Productivity, *Quarterly Journal of Economics* 119, 613-646.

[2] Anderson J.E., E. van Wincoop (2003) Gravity with Gravitas: A Solution to the Border Puzzle, *American Economic Review* 93(1), 170-192.

[3] Aw B., S. Chung and M. Roberts (2000) Productivity and Turnover in the Export Market: Micro-level Evidence from the Republic of Korea and Taiwan (China), *World Bank Economic Review* 14, 65-90.

[4] Bernard A., J. Eaton, J. Jensen and S. Kortum (2003) Plants and Productivity in International Trade, *American Economic Review* 93, 1268-1290.

[5] Bernard A. and B. Jensen (1999) Exceptional Exporter Performance: Cause, Effect, or Both?, *Journal of International Economics* 47, 1-25.

[6] Bernard A, J. Jensen and P. Schott (2003) Falling Trade Costs, Heterogeneous Firms, and Industry Dynamics, NBER Working Paper No.9639.

[7] Clerides S., S. Lach and J. Tybout (1998) Is Learning by Exporting Important? Micro-dynamic Evidence from Colombia, Mexico, and Morocco, *Quarterly Journal of Economics* 113, 903-947.

[8] Eaton J. and S. Kortum (2002) Technology, Geography, and Trade, *Econometrica* 70, 1741-1779.

[9] Foster L., J. Haltiwanger and C. Syverson (2005) Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?, mimeo, University of Chicago. *http://home.uchicago.edu/~syverson/selection2.pdf*

[10] Frankel J. and Rose A. (2002) An Estimate of the Effect of Common Currencies on Trade and Income, *Quarterly Journal of Economics* 117, 437-466.

[11] Haaland J. and V. Norman (1992) Global Production Effects of European Integration. In: Winters L. (ed.) *Trade Flows and Trade Policy After 1992*, Cambridge University Press.

[12] Head K. and T. Mayer (2002) Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade, CEPII Working Paper No.2002-01.

[13] Head K. and T. Mayer (2004) Market Potential and the Location of Japanese Firms in the European Union, *Review of Economics and Statistics* 86, 959-972.

[14] Helpman E., M. Melitz and Y. Rubinstein (2004) Trading Partners and Trading Volumes, mimeo, Harvard University.

[15] Klette T. and Z. Griliches (1996) The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous, *Journal of Applied Econometrics* 11, 343-361.

[16] Levinsohn J. and A. Petrin (2003), Estimating Production Functions Using Inputs to Control for Unobservables, *Review of Economic Studies* 70, 317-342.

[17] Mayer T. and S. Zignago (2005) Market Access in Global and Regional Trade, CEPII Working Paper No.2005-02.

[18] Melitz M. (2000) Estimating Firm-Level Productivity in Differentiated Product Industries, mimeo, Harvard University. *http://post.economics.harvard.edu/faculty/melitz/papers/prodest.pdf*

[19] Melitz M. (2003) The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity, *Econometrica* 71, 1695-1725.

[20] Melitz M. and G. Ottaviano (2005) Market size, trade and productivity, NBER Working Paper No.11393.

[21] Norman L., S. Kotz and N. Balakrishnan (1994) *Continuous Univariate Distributions*, Volume 1, 2nd Edition, Wiley.

[22] Olley S. and A. Pakes (1996) The Dynamics of Productivity in the Telecommunications Equipment Industry, *Econometrica* 64, 1263-1297.

[23] Pavcnik N. (2002) Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants, *Review of Economic Studies* 69, 245-276.

[24] Redding S. and A. Venables (2004) Economic Geography and International Inequality, *Journal of International Economics* 62, 53-82.

[25] Santos Silva J. and S. Tenreyro (2006) The Log of Gravity, *Review of Economics and Statistics*, forthcoming.

[26] Smith A. and A. Venables (1988) Completing the Internal Market in the European Community. Some industry simulations, *European Economic Review* 32, 1501-1525.

[27] Trefler D. (1995) The Case of Missing Trade and Other Misteries, *American Economic Review* 85, 1029-1046.

[28] Tybout J. (2002) Plant and Firm-Level Evidence on New Trade Theories. In: Harrigan J. (ed.) *Handbook of International Economics*, Vol. 38, Basil-Blackwell.

[29] Tybout J. and M. Westbrook (1996) Scale Economies as a Source of Efficiency Gains. In: Roberts M. and Tybout J. (eds.) *Industrial Evolution in Developing Countries*, Oxford University Press.

[30] World Economic Forum (2005), *Global Competitiveness Report 2005-2006*, Palgrave Macmillan.

Accessibility in 2000.

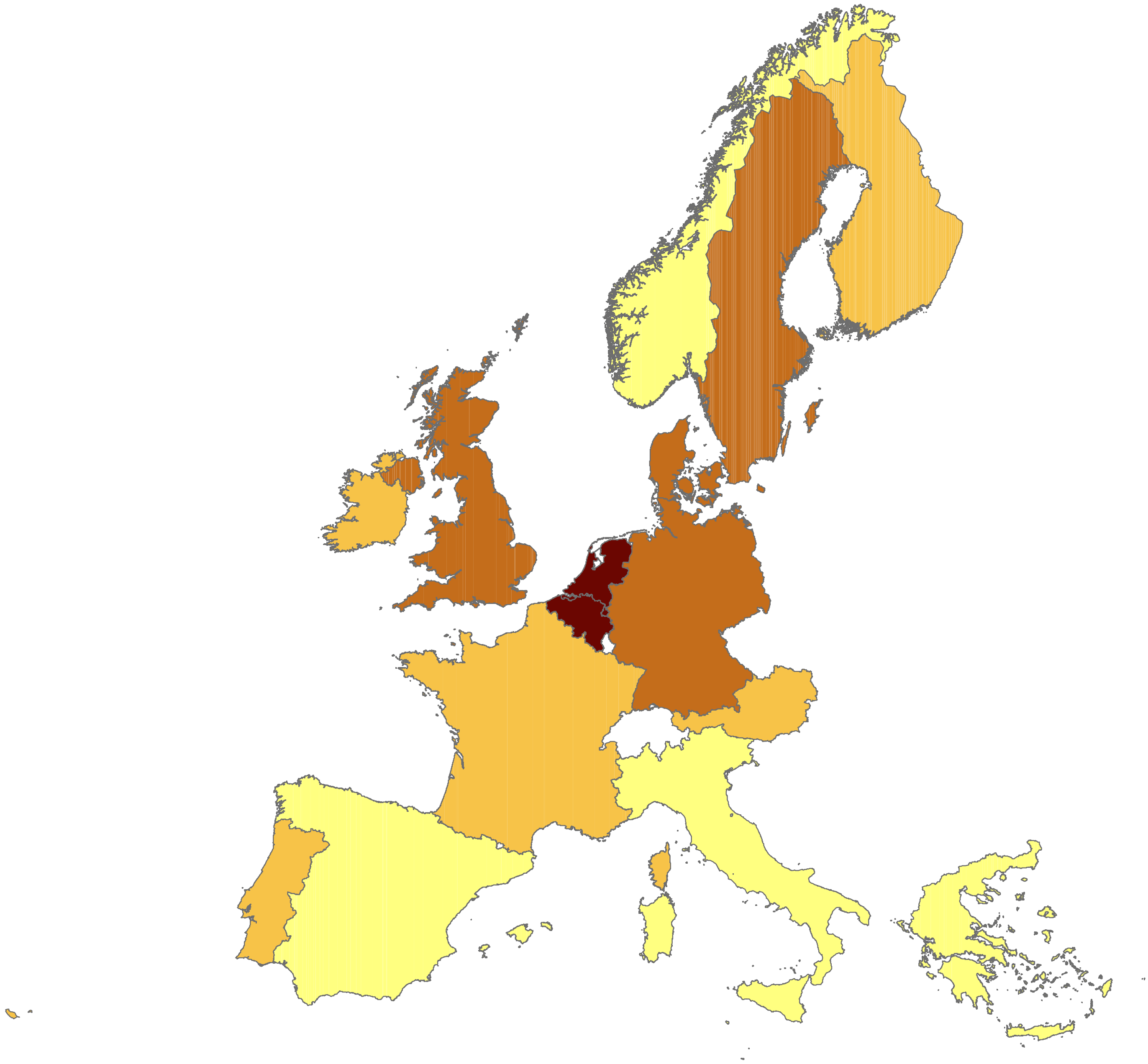

Figure 1 - Accessibility