Internationalization choices: an ordered probit analysis at industry-level

Filomena Pietrovito - Alberto Pozzolo - Luca Salvatici

Abstract

Consistently with the heterogeneous firms theory, this paper analyzes the link between heterogeneity within sectors, in terms of firm size and the average level of productivity, and internationalization choices, namely trade and foreign direct investments (FDI). We explain differences across sectors and countries in the choices to serve foreign markets through exports or FDI using trade and mergers and acquisitions (M&A) data for a sample of countries and industries between 1994 and 2004. This is done by performing an ordered probit analysis. The results confirm that industries with higher productivity levels and with a distribution of firms shifted toward large firms are more likely to internationalize through both trade and FDI.

Keywords: internationalization choices, exports, foreign direct investment, ordered probit

JEL Classification: C25, D21, F10, F14, F20, F23

F. Pietrovito (corresponding author)
Università degli Studi del Molise, Dipartimento di Economia, Gestione, Società e Istituzioni
II Edificio Polifunzionale
Via F. de Sanctis, snc
86100 Campobasso
e-mail: filomena.pietrovito@unimol.it
Tel. +39 0874 404335
Fax: +39 0874 98043

A. Pozzolo
Università degli Studi del Molise, Centro Studi Luca D'Agliano, MoFiR and CASMEF
Università degli Studi del Molise, Dipartimento di Economia, Gestione, Società e Istituzioni
II Edificio Polifunzionale
Via F. de Sanctis, snc
86100 Campobasso

L. Salvatici
Università degli Studi Roma Tre and International Food Policy Research Institute
Università degli Studi Roma Tre, Dipartimento di Economia
Via Silvio D'Amico, 77
00145 Roma

1. Introduction

Firms in a given sector can serve foreign consumers through two main channels: (*i*) producing at home and then exporting and (*ii*) setting or acquiring foreign establishments to produce abroad. The international economics and business research literature has shown that the choice of the foreign entry mode hinges on characteristics of products, firms, sectors and countries (Barba Navaretti and Venables 2004; Slangen et al. 2011).

While there is a long tradition of studies on the factors underlying specific patterns of foreign expansion through trade or foreign direct investments (FDI),[1] analyses focusing on different forms of internationalization are relatively more recent. In the traditional proximity-concentration trade-off literature, a well-accepted result is that FDI become more convenient than exports as both the size of the foreign market and the costs of exporting increase, and less convenient as the costs of setting up foreign production grow (Brainard 1993, 1997; Yeaple 2003).[2] As pointed out by Brainard (1993), firms can be expected to invest abroad when the gains from avoiding transport and tariff costs outbalance the costs of maintaining capacity in multiple markets.

This literature does not predict which firms in each sector become international (Head and Ries 2003). More recent contributions, starting from the seminal paper by Melitz (2003), address this issue taking into account the role of heterogeneity in firm-level productivity. Building on this theoretical framework, Helpman et al. (2004) expand the set of internationalization choices by including FDI in addition to trade. This new setting typically leads to a ranking in terms of productivity and size: multinational firms outperform exporters, which in turn outperform domestic firms. The empirical analyses searching for a validation of the theoretical framework of Melitz (2003) and Helpman et al. (2004) in most cases focus on single countries, the few for which firm-level data on export and FDI are

[1] For a recent survey of this literature see Greenaway and Kneller (2007).
[2] Another strand of the literature focuses on the distinction between horizontal and vertical FDI (see, e.g., Carr et al. 2001; Conconi et al. 2013). However, this issue is out of the scope of our analysis.

available. Moreover, they do not disentangle the impact of different aspects of firm performance, such as productivity and size, since firm size is typically assumed to depend on the level of productivity, that in turn follows a Pareto distribution as in Helpman et al. (2004).

In this paper, we analyse the relationship between country-sector characteristics and internationalization strategies (i.e., export and FDI) distinguishing the impact of productivity and size. Moreover, we do so for a large sample of countries, including 25 origin countries, 91 destination countries between 1994 and 2004. The shortcoming of our data-set is that we only have information on average productivity at the level of 57 manufacturing industries, not at the firm-level.

Disentangling the effect of productivity from that of firm size, we test two separate hypotheses on the relationship between industry heterogeneity and internationalization: (*i*) that sectors featuring higher productivity levels are more likely to internationalize through both exports and FDIs; and (*ii*) that sectors featuring a size distribution shifted toward large firms are also likely to internationalize through both exports and FDIs. To this end, we use an ordered probit model to assess the impact of several covariates at sector level on a discrete foreign expansion index ranging from 0 to 2 according to whether: sectors serve uniquely the domestic market, export only, export and perform FDI as well.[3] Results confirm that more productive sectors and sectors with a distribution of firms shifted toward large firms are more prone to foreign expansion, through both trade and FDI.

Our analysis contributes to the literature explaining the nature of the internationalization processes, along three dimensions. *First*, we expand the span of variables considering separately the productivity level and the distribution of firms by size in each sector. This allows a better characterization of sectoral heterogeneity since we consider both intensity and dispersion, and this marks a departure from previous contributions which have either focused on the former or on the latter type of variables.

[3] Whereas in a linear regression, a sector with an index equal to 2 would be twice as internationalized as one with an index equal to 1, in the ordered probit model, no such presumption of cardinality is made: a value of 2 simply indicates more internationalization than a value of 1.

Second, we use bilateral flows of trade and FDI at sector level for a large sample including both developed and least-developed countries. This allows us to simultaneously measure the impact on the internationalization index of several country-level and sector-level factors, alongside with productivity and the distribution of firms by size, taking into account potential heterogeneities within as well as across countries.[4] Moreover, our dataset allows the estimation of the effects of average productivity and firms size controlling for all country and sector invariant unobserved characteristics. This reduces the risk of the possible reverse causality problems, that would be present if countries and/or sectors that are intrinsically more internationalized turned also out to be the most productive and/or those with firms of larger size.[5] It also reduces the risk of possible endogeneity problems if an omitted country or sector characteristic caused firms to be at the same more productive and/or larger, and more international. *Third*, from a methodological point of view, we analyze the complexity of the internationalization process in a multinomial framework. Adopting the view that the internationalization process is complex and cumulative, since it is based on accumulating experience and higher commitment, we use an ordered probit model to analyze the determinants of different internationalization involvements (domestic – i.e., no internationalization at all, only exports and both exports and FDI).

Our approach allows us to verify the robustness of results obtained in the literature by using firm-level data. *First*, we verify whether the impact of firm size and productivity translates in an aggregate significant effect at the sector level. *Second*, by using sector data we are able to verify this effects for a large sample of countries, both of origin and destination.[6]

[4] To disentangle potential differences between groups of countries, we also provide evidence on the patterns of internationalization depending on the level of country development.

[5] As highlighted by several studies (Bernard and Jensen 1995; Bernard and Jensen 2004; Castellani and Zanfei 2007), not only firms self-select into internationalization modes, with more productive and larger firms becoming more involved in international activities, but their level of productivity and size could also be influenced by internationalization involvement.

[6] On the other hand, even if the results of this paper cannot be attributed directly to firms, moving from the country to the industry perspective allows us to draw some indirect implications for single firms. Indeed, we analyze the sector-level implications of firm-specific characteristics, such as firm size according to the international business literature, controlling

The paper is organized as follows. Section 2 briefly discusses the theoretical and empirical background and the hypotheses to be tested. Section 3 presents the empirical model. Section 4 describes the data used in the analysis. The main results of the analysis are presented in Section 5, along with a number of robustness checks. Section 6 draws some conclusions.

2. Previous literature and testable hypotheses

Two related aspects of the role of firms heterogeneity in the choice of the mode of internationalization have been analyzed in the literature: the impact on the value of existing exports or investments to the same destination(s) – the *intensive margin* – and the impact on the number of export items or foreign countries where firms export or set up a foreign subsidiary – the *extensive margin*.[7]

The most influential theoretical model to study the choice between internationalization through trade or FDI is that of Helpman et al. (2004). The analytical framework focuses on the *intensive margin* and builds on the seminal paper by Melitz (2003), where monopolistically competitive firms draw different levels of productivity from an exogenous distribution and find internationalization profitable only if they are productive enough to reach the scale that is necessary to sustain the fixed costs of exporting. A key feature of this model is that firm productivity maps exactly into firm size, and therefore exporting firms are at the same time more productive and larger. Making the additional assumption that the fixed costs of setting up a foreign subsidiary are higher than those of exporting, Helpman et al. (2004) show that a higher within-industry heterogeneity in firm sales is associated with a higher incidence of sales by foreign affiliates relative to exports, because greater dispersion implies a larger share of firms with a sufficiently high level of productivity to find it profitable to invest abroad. Using data on exports and foreign subsidiaries' sales of US manufacturing firms in 30 countries and 52 industries, they find that a

for the impact of location- and country-specific advantages.

[7] There is a large literature comparing the performances of domestic, export-oriented or multinationals firms. However, since it is not explicitly aimed at explaining internationalization choices we will not mention it here.

wider dispersion of firms size (and therefore of productivity) within each sector is associated with a larger incidence of foreign affiliates' sales relative to exports.[8] Additional empirical evidence, surveyed by Bernard et al. (2007) and Greenaway and Kneller (2007), confirms the theoretical hypothesis that firms self-select into internationalization strategies depending on their productivity level and size.[9]

The literature analysing the *extensive margin* of trade and FDI is mostly empirical. In the real world, the choice of a firm to enter or not a given foreign country ranges from "no internationalisation" to all possible combinations of the available set of internationalisation modes. These options are analysed estimating the pattern of internationalization conditional on several firms, industry and country characteristics, by means of either non-ordered or ordered multiple choice models.

In the framework of non-ordered models (bivariate probit, multinomial logit and probit), choices are typically exhaustive and mutually exclusive, and each firm is assumed to choose the strategy that maximises its profit function. Several contributions in the literature use non-ordered models to analyse internationalization choices in specific countries. To estimate the productivity effects on the probability of investing abroad or exporting, Oberhofer and Pfaffermayr (2012) use a bivariate probit model that allows for both modes: the number of employees as a measure of firm size and productivity of companies increase the probability of both strategies, but the effect is larger for the probability of investing abroad.[10]

Concerning the multinomial logit approach, Bougheas and Görg (2008) estimate the probability that Irish firms choose one of the modes of internationalization, conditional on a number of plant characteristics (including productivity). They find that (*i*) exporters are more productive than non-

[8] Similarly, Oldenski (2010) extends the analysis of Helpman et al. (2004) showing that greater firm-level heterogeneity in firm size significantly increases FDI relative to exports also in service industries.

[9] A partly contrasting result is that of Todo (2011) who, allowing firm heterogeneity in unobserved characteristics by estimating a multinomial logit model with random intercepts and random coefficients (a mixed logit model), finds a small economic impact of productivity on the probability that a firm exports or invests abroad for Japanese firms.

[10] Similarly, Kimura and Kiyota (2006), by adopting a probit model with random effects, find that the most productive firms are those that engage both in FDI and in export.

exporter and (*ii*) exporting firms that also invest abroad are more productive than firms that only export. Using the same methodology, Benfratello and Razzolini (2009) confirm the same ranking of productivity for a sample of 4,000 Italian firms.

Since the multinomial logit model is subject to the constraint of the independence of irrelevant alternatives, some papers in the related literature estimate a multinomial probit model. For example, Engel et al. (2009) analyse the relationship between firm-specific characteristics and the entry and exit pattern in foreign markets for the two main modes of internationalization, namely export and FDI. For a sample of French firms, they find that high productivity firms have a significantly higher propensity to invest abroad than low productivity firms. In such models, choices are exhaustive and mutually exclusive and the firm chooses only the alternative that maximises the profit function.

Unfortunately, multiple-choice models become cumbersome for a large number of internationalisation forms because the different forms can be combined and each combination defines a choice. For this reason, Calia and Ferrante (2010) use a multivariate probit model to estimate the relevant associations between different internationalisation patterns and variables describing firm characteristics. In particular, they study Italian firms considering a wide range of internationalisation forms, including offshoring of production and outsourcing of services abroad, as well as non-equity forms, such as commercial penetration operations and agreements, in addition to the exports and FDI modes. Regarding productivity, their results suggest that it affects the choice to stay domestic or to have international activities, but not the choice among different internationalization modes.

To the best of our knowledge, in the framework of ordered models only Basile et al. (2003), focusing on Italian manufacturing firms, and Demirbas et al. (2013), considering Indian firms, adopt an ordered probit to investigate the determinants of foreign expansion through exports and investment. Basile et al. (2003) postulate that a higher internationalization level implies a greater cumulative commitment to foreign markets and a better firm's position in those markets: their results suggest that firm size, the

relationships with other firms, innovation and geographic location are very important determinants of variations in the foreign expansion index across firms in different points in time. Demirbas et al. (2013) empirical findings suggest that firm characteristics evolve through time, and that there is a ladder of quality where some firms evolve towards exporting and FDI.

Following the seminal paper by Melitz (2003), in most of the literature cited above, firm size is a function of its productivity, and therefore the impacts of the two characteristics are not separately evaluated since "a more productive firm will be bigger (larger output and revenues), charge a lower price, and earn higher profits than a less productive firm" (Melitz 2003, 1700).

Although it is true that a firm's superiority in one dimension may be associated with enhanced performance in others, it should not be taken for granted that more efficient firms are the largest ones. Mrazova and Neary (2012) have recently provided a general characterization of which firms will select alternative ways of serving a market showing that if and only if firms' maximum profits are 'supermodular' in production and market-access costs, more efficient firms will select into the activity with lower market-access costs. As a consequence, there could be cases with firms less productive overall, but relatively more productive at higher levels of output.

On the contrary, in our empirical framework we choose to disentangle the effect of firms productivity from that of firm size. As far as the former is concerned, we focus on the differences across sectors/countries in terms of average productivity formulating the following hypothesis:

Hypothesis 1: *a higher average level of productivity in a given sector of a given country is associated with a higher probability of exporting and carrying out FDI.*

Concerning firms size, we focus on the differences across sectors in terms of dispersion formulating the following hypothesis:

Hypothesis 2: *a firm distribution shifted toward large firms in a given sector of a given country is associated with higher probability of exporting and carrying out FDI.*

It is worth noting that in the model of Helpman et al. (2004) it would be impossible to shift the firm size distribution to the right holding productivity constant since productivity and domestic sales are summarized in a single variable (namely, firms size dispersion) under the assumption that both follow the same (Pareto) distribution. Even if our analysis is not directly comparable with the Helpman et al. (2004), the two hypotheses are closely related to their results since they state that sectors with higher productivity levels and firm distribution shifted toward large firms, cumulate different and more demanding forms of internationalization to enlarge their involvement. As a matter of fact, when firms are able to assume higher risks associated with international activities, they enter the international market with forms requiring higher experience, investments and commitment (Conconi et al. 2013).[11] Finally, much of the applied work on export and FDI choices has been based on micro-level data allowing to explain firms behaviour taking into account inter-firm variability. Aggregate data average out this variability but the loss of information does not imply that the findings of studies based on them are irrelevant. Data based on inter-industry and inter-country variability, as a matter of fact, allow to confirm whether the micro relationships are confirmed for a large set of sectors and countries.

3. Empirical methodology

To test the two hypotheses put forward in the previous section, we design a set of regression models. These models are based on the estimation of an ordered discrete choice model to evaluate how country and sector characteristics affect the likelihood of different international involvement. In general, in a *J*-choice ordered probit model *y* is an ordered response where the values we assign to each outcome

[11] Conconi et al. (2013) provide an interesting methodology to analyze the internationalization process. They identify the year in which firms start investing in a foreign market and show that firms follow the type of gradual internationalization process in the sense that they export to a foreign market before investing there. However, with data aggregated by sector it is not straightforward to identify the precise year in which firms start investing in a foreign country.

represent a specific order along a continuum, but not the magnitude of difference between the options. In our specification, y is an indicator of international involvement at sector level ranging between zero and 2, with: $y = 0$ for sectors that are not internationalized at all ("domestic"), $y = 1$ for sectors that internationalize only through trade ("export") and $y = 2$ for sectors that have both trade and FDI ("export and FDI"). The fact that 2 indicates a higher international involvement than 1 (and 0) conveys useful information, even though the index itself has only an ordinal meaning.

For such an ordinal dependent variable, using multinomial probit or logit would not be efficient because these models would mis-specify the data-generating process in assuming that there is no order in the different categories that the dependent variable can take. OLS regression estimation would also be inappropriate, since it would consider the difference in the dependent variable between a 0 and a 1 as equivalent to the difference between a 1 and a 2. Greene (2008) summarizes the previous remarks pointing out that when "the outcome is discrete, the multinomial logit or probit model would fail to account for the ordinal nature of the dependent variable. Ordinary regression analysis would err in the opposite direction, however" (Greene 2008, 831). Moreover, this strategy reflects the intuition that "in the face of uncertainty, exporting allows to experiment in foreign market at a lower fixed cost: if a firm discovers that it can earn large enough profits by serving foreign consumers, it establishes production facilities to avoid paying the trade costs" (Conconi et al. 2013, 1).

The ordered probit model for y can be derived from a latent or unobserved continuous variable, y^*, related to a set of explanatory variables according to a standard linear model:

$$y^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_K x_K + \varepsilon \qquad (1)$$

where, $x_{1,\ldots,K}$ are the explanatory variables, which may include sector and country characteristics influencing the probability of different internationalization involvements, $\beta_{1\ldots k}$ are the associated parameters, and ε is a random error term drawn from a standardized normal distribution. Although y^* is unobserved, y is observed and related to y^* by the following relationship:

$$
\begin{aligned}
&y = 0 \quad \text{if} \quad y^* \le \alpha_1 \\
&y = 1 \quad \text{if} \quad \alpha_1 < y^* \le \alpha_2 \\
&y = 2 \quad \text{if} \quad y^* > \alpha_2
\end{aligned}
\tag{2}
$$

where, $\alpha_1 < \alpha_2$ are the unobserved *cut points* identifying the boundaries between the different levels of international involvement. Therefore, given the standard normal assumption for the error term, we can derive each response probability of observing a sector as being "domestic" (i.e., the dependent variable *y* taking the value of 0) as:

$$
\begin{aligned}
\Pr[y = 0] &= \Pr[y^* \le \alpha_1] \\
&= \Pr[\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_K x_K + \varepsilon \le \alpha_1] \\
&= \Pr[\varepsilon \le \alpha_1 - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_K x_K)] \\
&= \Phi(\alpha_1 - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_K x_K)) \\
&= \Phi(\alpha_1 - \mathbf{x\beta})
\end{aligned}
\tag{3}
$$

where $\mathbf{\Phi}(.)$ is the standard normal distribution function. Similarly, we can obtain the probability of y = 1 and y = 2 in the following way:

$$
\begin{aligned}
&\Pr[y = 1] = \Pr[\alpha_1 < y^* \le \alpha_2] = \Phi(\alpha_2 - \mathbf{x\beta}) - \Phi(\alpha_1 - \mathbf{x\beta}) \\
&\Pr[y = 2] = \Pr[y^* > \alpha_2] = 1 - \Phi(\alpha_2 - \mathbf{x\beta})
\end{aligned}
\tag{4}
$$

The β parameters together with the threshold levels on the latent variable that characterize the transition from one observed categorical response to the next (cut points α) can be obtained through the maximum likelihood estimation.

In our empirical setting, the main specification adopted in the empirical analysis is the following:

$$
\begin{aligned}
y_{ij}^h = {} & \beta_0 + \beta_1 TFP_i^h + \beta_2 Number\_of\_large\_firms_i^h + \beta_3 \mathbf{Z}_i^h + \beta_4 \mathbf{T}_{ij} + \beta_5 \mathbf{V}_{ij}^h \\
& + \beta_6 \mathbf{DU}_i + \beta_7 \mathbf{DU}_j + \beta_8 \mathbf{DU}^h + \varepsilon_{ij}^h
\end{aligned}
\tag{5}
$$

Where (excluding indices) *y* is the ordered dependent variable that takes the value 0 for sectors of country *i* not exporting to country *j*, the value 1 for sectors of country *i* presenting exports but not FDI

to country *j*, and the value 2 for sectors of country *i* featuring both exports and FDI;[12] *TFP* is the average productivity level in sector *h* in country *i*; *Number_of_large_firms* is the number of firms of sector *h* in country *i* in the 10th decile of the world firms distribution of total sales in sector *h*; Z is a set of control variables for sector *h* of country *i* (i.e., capital and technological intensity); T is the set of control variables describing the bilateral relationship between countries *i* and *j* (e.g., distance, islands, common language and common religion); X is the set of control variables describing the bilateral relationship between countries *i* and *j* in a given sector *h* (i.e., tariffs, number of common partners in trade or FDI); and DUs are three sets of dummies controlling for unobserved common characteristics at the level of the origin country *i*, the destination country *j*, and the sector *h*.

We control for country and industry invariant characteristics by introducing fixed effects for origin countries, destination countries and sector of economic activity. In addition, we include a set of control variables that are based on characteristics specific of each industry in each country, on country pairs characteristics, and on country pairs/industry characteristics. This specification allows to control for potential effects of country and sector specific characteristics that might contemporaneously enhance the international activity as well as the average productivity and the distribution of firms.

According to the two main assumptions presented in the previous section, we expect the estimated coefficients of our key independent variables to be positive and statistically significant, after controlling for other industry and country characteristics. In general, increasing one of the independent variables, while holding coefficients and cut points constant, is equivalent to shifting the distribution to the right. Accordingly, finding a positive coefficient for an independent variable implies that the change of the probability of being a "domestic" sector (Pr[y = 0]) moves in the opposite direction with respect to β_1 and β_2, while the change of the probability mass of being an "export and FDI" sector (Pr[y = 2]) move in the same direction. However, what happens to the middle category (Pr[y = 1]) is ambiguous, because

[12] The very few cases of sectors that have FDI but no trade are dropped from the sample.

the probability mass moving from "domestic" to the "export" can be either larger or smaller than that of moving from "export" to "export and FDI".

4. Data and sample[13]

4.1 The dependent variable

To construct the dependent variable for the ordered model, we need data on both exports and FDI. The main statistical source of data on exports is the database UN Comtrade, managed by the statistical division of the United Nations, that reports data on the bilateral flows in several industrial sectors. In particular, it contains annual international trade statistics, detailed by commodity and partner country, for a very large set of countries. Commodities are classified according to different recognized classifications, such as the standard international trade classification (SITC) and the harmonized commodity description and coding system (HS). We use the international standard industry classification (ISIC), Revision 3, at 4-digits level to be able to concord data on export with other data used in the empirical analysis.

Much less information is available on FDI, especially at the bilateral and sector levels. To overcome these shortcomings, we use information on Mergers and Acquisitions (M&A) as a proxy for FDI. While this is a limitation of our analysis, we believe that it is unlikely to affect the qualitative results, because cross-border M&A are by and large the most widely used mode of operating a foreign firm (Herger et al. 2008). However, we cannot rule out the possibility of a selection bias due to the fact that a firm's decision to use M&A versus greenfield investment may be related to its productivity or size. While this is a limitation of our analysis, since larger and more productive firms are more likely to perform greenfield investment, we believe that the possible bias actually make the confirmation of our hypotheses more difficult and it is unlikely to affect the qualitative results.

[13] Variables used in the analysis and their sources are listed in Appendix 1.

Data on M&A are sourced from SDC Platinum Global Mergers and Acquisitions, a database provided by Thomson financial securities data that records all deals involving a change in ownership of at least 5% of total equity and exceeding 1 million US dollar. The Thomson dataset allows to analyze M&A for a large range of countries and years. It records two related aspects of cross-border acquisitions: the number of acquisitions and their value.[14] A common choice in the literature on M&A is to consider disclosed and complete deals for which the value of the transaction is available. This choice allows us to construct a more reliable database. The database also contains information on target and acquirer profiles, such as industry classification, based on the primary activity and location, that are used in our empirical analysis. In particular, we identify cross-border deals in manufacturing standard industry classification (SIC) codes at 4-digits level.[15]

Using these information, we build an indicator variable at the country and sector levels. This indicator is constructed in such a way that higher values correspond to greater involvement of sector in international activities. This variable (y) distinguishes between sectors that are not internationalized at all (with a value of zero), sectors that internationalize only through trade (with a value of one) and those that have both trade and FDI (with a value of two).

4.2 Key independent variables

Our two key explanatory variables are measures of productivity and firm size. The average industry TFP is calculated under the assumption of constant returns to scale Cobb-Douglas production function as:

[14] The main sources of information of data on M&A are financial newspapers and specialized agencies like Bloomberg and Reuters. It should be kept in mind that until the mid-1980s Thomson focused very much on M&A for the USA only, and it is only for about the last 20 years that (systematic) M&A data gathering took place for other countries (Brakman et al. 2005).

[15] Domestic M&A, i.e., acquisitions with acquirer and target located in the same country, could still provide access to foreign markets if the target firm is active abroad or if the acquirer is controlled by a foreign firm. However, in the former case we do not know what are the foreign markets (possibly) involved, while in the latter case we have no information about foreign controls: as a consequence, we exclude domestic M&A from our sample.

$$TFP_i^h = \frac{Y_i^h}{(K_i^h)^{\alpha}(L_i^h)^{1-\alpha}} \quad (6)$$

Where, Y_i^h is value added in sector *h* of country *i*; K_i^h and L_i^h are the stock of capital and the number of employees in sector *h* of country *i*, respectively; and α, the capital share, is assumed to be 1/3.[16]

Total factor productivity at the national sector level is calculated from data on investment and labour from UNIDO (Indstat4, 2008 version), where each sector's capital stock is estimated by the inventory method (Bernanke and Gurkaynak 2002; Isaksson 2009). In particular: (*i*) for each country, we calculate the sector's share of investment using flow information for the first five years of data available; (*ii*) we use investment shares to allocate each country's total capital, sourced from the UNIDO's World Productivity Database, across sectors; (*iii*) we use the estimates of the country and sector specific initial stock of capital obtained as described above as the starting point to apply the inventory method, i.e., adding each year's value of real term investment and applying a sector specific rate of depreciation to account for obsolescence.

The use of TFP as a measure of productivity implies that Germany is excluded from the analysis since data on aggregate capital are not available from the UNIDO's World Productivity Database. However, Germany is included in the sample used for the robustness check based on an alternative measure of productivity, namely the ratio between value added and number of employees in a given sector. Data on labour productivity are drawn from UNIDO (Indstat4, 2008 version).

To measure the distribution of firms by size in a sector, we first calculate the deciles of the world distribution of firms by total sales in each sector and then we count the number of firms that each

[16] While countries might use different capital-labour ratios, in absence of detailed information, we follow Bernanke and Gurkaynak (2002) in calculating TFP growth rates under the assumption that labour share is the same and fixed for both developed and developing countries. More generally, lack of data prevents us from using more refined approaches for the estimation of total factor productivity (e.g., Levinsohn and Petrin 2003).

country has in the 10th decile of the world/sector distribution.[17] This indicator proxies for the incidence in each country and sector of those firms that are large enough to overcome the higher fixed costs of expanding abroad through FDI rather than exports (Helpman et al. 2004).

As additional measures of large firms in a sector we use alternatively: (*i*) the number of firms in the 9th and 10th decile of the world distribution of firms by size, (*ii*) the number of firms in the 7th decile and higher of the world distribution of firms by size and (*iii*) the dispersion of the distribution of sales within sectors – that allows to compare our results with those obtained by Helpman et al. (2004).

Data on firm's sales are drawn from the Worldscope database including financial statements of about 29,000 companies listed in developed and emerging markets, representing approximately 95% of the global market capitalization. Since we focus on large firms, excluding non-listed companies is unlikely to introduce a relevant bias in our measure of each sector's ability to internationalize. Data are classified according to the SIC classification at 4-digits level.

4.3 Control variables

To limit the potential for omitted-variable bias, we add to the main variables of interest three sets of controls, that are based on the vast literature focusing on trade and on M&A. First, we control for some relevant sector characteristics in the country of origin. Second, we control for a set of characteristics of the bilateral relationship between each couple of countries. Finally, we include some sector characteristics that are specific to each pair of countries.

4.3.1 Sector-level variables for the country of origin.

Helpman et al. (2004) show that capital intensity is a useful predictor of a larger incidence of exports relative to FDI while the opposite is true as far as technological intensity is concerned. Accordingly, we

[17] Considering the world rather than the national distribution(s) we avoid the risk of a country-specific definition of "large firms". On the other hand, the total number of firms in each sector may be influenced by technological peculiarities, such as the existence of economies of scale. To account for this issue, it is possible either to use the share rather than the absolute number of large firms or, as we do in this paper, account for all sector-specific features through the use of sector dummies.

use the ratio between capital and number of employees for each country and sector from UNIDO to construct a measure of capital intensity, and the number of utility patents granted by the US Patent Office, provided by the national bureau of economic research (NBER), also at the country and sector level, as a measure of technological intensity.[18]

4.3.2 Bilateral country-level variables.

The empirical literature has identified a large set of variables that influence foreign market entry modes, though the magnitudes and even the signs of the impact on either trade or FDI are not always consistent (see, for example, Blonigen 2005; Disdier and Head 2008; Helpman et al. 2008; Herger et al. 2008; Oldenski 2010; Slangen and Beugelsdijk 2010; Wang et al. 2010; Slangen et al. 2011). Distance directly increases transaction costs because of the transportation costs of shipping products, the cost of acquiring information about other economies, and the cost of finding a partner and contracting at a distance. Similarly, the number of islands in each country pair, common language and common religion are expected to affect bilateral relationships, both through trade and investment. Our data on bilateral characteristics are drawn from the dataset provided by the *centre d'etudes prospectives et d'informations internationales* (CEPII).[19]

4.3.3 Bilateral country- and sector-level variables.

We consider two bilateral sector-level variables. First, bilateral trade tariffs, that we expect to favor FDI, according to the well-known "tariff jumping" effect pointed out in the literature (Brainard 1997; Carr et al. 2001; Markusen and Maskus 2002; Yeaple 2003; Helpman et al. 2004). To make data

[18] Since the original data on patents are classified according to the US Patent Classification, we combined them with other information adopting the correspondence scheme between the US Patent Classification and the International Patent Classification and between the latter and the ISIC3 provided by Johnson (2002).

[19] The CEPII follows the great circle formula and uses latitudes and longitudes of the most important cities (in terms of population) to calculate the average of distances between city pairs. Data on distances are available at: http://www.cepii.fr/anglaisgraph/bdd/distances.htm. We also adopted distances between capitals as an alternative measure and the results remain unchanged.

comparable to other data used in the analysis, we aggregate HS 6-digits level data on tariffs from TRAINS to the 4-digits ISIC classification through simple averages. Second, building on the results of Chaney (2011) – who show that the existing contacts of a firm can be used to find new ones – we include in our specification two "network indexes" calculated as the number of common partners in trade and in M&A of each couple of countries (Francois 2010). We expect that a higher number of common partners in exports (or in M&A) between two countries increases the probability of exporting (or doing M&A) between those same countries. Data on the number of common partners is built from our information on trade and FDI.

4.4 Sample summary statistics

Matching our different sources, we construct an original database that associates bilateral trade and FDI flows at the sector level in a common classification, for a sample of developed as well as developing countries. Industries including finance and utilities are excluded, along with wholesale and retail trade, because of the non-tradable nature of these activities. We also exclude agriculture and primary sectors (i.e., mining and oil and gas extraction) due to the lack of data on productivity. As a result, we focus on manufacturing sectors (i.e., sectors with an ISIC code between 1511 and 3720).

Since our measures of M&A and sales are available in the SIC classification, we mapped SIC codes into ISIC codes, both at 4-digits level, using the concordances produced by Statistics Canada, as in Brakman et al. (2005).[20] To take into account that at the 4-digits level of disaggregation we have a large number of empty cells, both in exports and in M&A, we aggregate data available at 3 digits of ISIC classification. Matching the different sources yields a dataset including 25 origin countries and 91 destination countries, covering 57 manufacturing industries at the 3 digits ISIC level from 1994 to

[20] Concordance tables are available from: http://www.macalester.edu/research/economics. Appendix 2 lists concordances between SIC and ISIC classifications at the 3-digits.

2004.[21] Appendix 3 and 4 list countries and sectors included in our analysis. Table 1 presents the descriptive statistics for the variables used in the estimations, showing substantial variation in all our key variables.[22] TFP shows a high variability around the average, and sectors presenting (on average) the highest TFP are: Refined petroleum products, Tobacco products, Motor vehicles and Automobiles. Labour productivity presents an even larger variability; sectors featuring the highest labor productivity are: Tobacco products, Refined petroleum products and Man-made filament tow or staple fibers.

The average number of firms in the 10th decile of the world distribution of firms by total sales is 2 and shows a high within sample variability, with values ranging from 0 to 52. The number of patents, reflecting the level of technological development, shows an average value of 17 and a high variability since it ranges between 0 and 1,465.

Among bilateral characteristics, tariffs show a high variability, with values ranging between 0 and 58% and an average level of 12%. The average number of common partners in trade is 58, with values ranging between 0 and 117, whereas the average number of common partners in FDI is much lower and the range narrower (between 0 and 30). This difference highlights that the two "networks" are quite different and the former is much larger than the latter.

Appendix 5 reports the summary statistics (means and standard deviations) for all variables in our data set, distinguishing among "domestic", "export" and "export and FDI" sectors. The first category, grouping 5,917 observations, includes country pairs-sectors not involved in an international relationship at all; the second, by far the most numerous (62,758 observations), includes country pairs-sectors that are involved in exports only; the third category, featuring 4,229 observations, includes country pairs-sectors that are involved in both exports and FDI. The distribution of the key explanatory

[21] We average data over the ten years period to get a cross-section framework. Even though one could expect the number of observations to be equal to 129,675, our regressions are estimated on a more limited number of observations since not all commodities are traded between each country pair.

[22] Descriptive statistics are computed on the largest sample, i.e. the one including Germany in the group of domestic country.

variables in the three samples implies that: the higher the internationalization involvement of sectors, the higher the level of productivity and the presence of large firms, independently of the measure adopted. This suggests, as expected, that sectors that are involved both in trade and in investment are the most productive and show the highest incidence of large firms. Sectors that are only active in exports represent 86% of our sample, while domestic and exporter and investor sectors represent, respectively, 8% and 6% of the total.

Table 2 reports simple correlations among the variables used in the empirical model. TFP and labour productivity levels are positively correlated with the dependent variable: higher levels of productivity in a given sector determine higher internationalization and higher probability of both trade and investment. Further, the correlation between the ordered dependent variable, distinguishing internationalization, and the number of large firms is positive, suggesting that having firm distribution by size shifted towards large firms favours both trade and FDI.

Even though summary statistics and bilateral correlations are suggestive, they do not control for potentially confounding factors. For this reason, in what follows we perform a more refined econometric analysis.

5. Results

5.1 Estimations on the whole sample

The first step of our empirical analysis consists in estimating the ordered probit model described in equation (5) on the whole sample that includes the 67,975 cases. This approach allows us to analyse the impact of our variables of interest, along with other controls, on the probability and the degree of different internationalization involvement.[23]

[23] All estimations reported include three sets of dummies controlling for the domestic country, the foreign country and the sector-specific fixed effects, as described in Section 3.

The threshold parameters – *cut1* and *cut2* – define the boundaries between different categories of internationalization. Both estimated threshold values are statistically significant and their coefficients are different from 1, implying that the ordinal categories are not equally spaced (Basile et al. 2003). In unreported analyses, available upon request, we also verified that cut-off points are also statistically different from each other, meaning that observed internationalization categories do not overlap.

Results in column (1) of Table 3 show that the average level of TFP has a positive impact on foreign expansion, with a coefficient of 1.191 that is statistically significant at the 99% level. Our finding is therefore consistent with the theoretical hypothesis that more productive sectors internationalize. Since the sign of the coefficient can only tell us about how an independent variable affects the probability of the end categories (Greene 2008 ; Wooldridge 2010), to get a sense of the magnitude of this impact, we also calculated the changes in the predicted probability of each category for a variation of average sector TFP. As it is customary in the literature (Rajan and Zingales 1998; Benfratello and Razzolini 2009)[24], we take as reference levels the 25th and the 75th percentiles of the world distribution.[25] Table 3 shows that such variation in sector TFP would determine a decrease of 8.9% in the probability that a sector is not internationalized (Column 2), an increase of 2.1% in the probability that its foreign expansion takes place only through exports (Column 3) and an increase of 6.8% in the probability that both exports and FDI are present (Column 5).

Similar results are obtained for our second key explanatory variable: a shift of the distribution of firms by size to the right increases the probability for a sector to internationalize. Even controlling for other covariates, as well as industry and country dummies, the positive impact of the number of large firms in a sector is confirmed. The positive impact of this variable, statistically significant at the 99% level, is

[24] Benfratello and Razzolini (2009) adopt as reference level the 10th and 90th percentile. Adopting this range in the estimates does not change our main results.

[25] Qualitatively similar results are obtained using different variations, for example from 20th to 80th or from 30th to 70th percentile.

consistent with our second hypothesis: when the distribution of firms in a given sector-country is shifted towards large firms, it is more likely that domestic sectors begin to explore foreign markets, via exports and/or foreign investment. Looking at the economic impact, an increase in the number of large firms from the 25th to the 75th percentile reduces the probability mass of being "domestic" sector by 3.9% and favours internationalization through export only (by increasing the probability by 1.6%) and through both exports and investment (by increasing the probability by 2.3%). Interestingly, this impact is much lower than that of productivity.

Even if the percentage changes look similar, it is worth recalling that the average predicted probabilities are quite different since 86% of the country-sectors in the database register some exports whereas only 6% of the observations are characterized by both exports and FDI. Accordingly, the relative impact on the share of observations using both internationalization strategies (i.e., exports and FDI) is much larger (38%) than that of using only exports (2%). Overall, our results provide support to the hypotheses stated in Section 2 that sectors characterized by a high level of productivity and by a higher incidence of large firms are more likely to be able to afford the higher fixed costs required to serve foreign consumers.

Concerning other control variables, the level of capital intensity and the innovation activity also favour internationalization of sectors. In particular, an increase in capital and technological intensity makes sectors to leave the "domestic" category in favour of the "export" and, even more, "export and FDI" categories.[26] Regarding country-level bilateral characteristics, a first group of control variables (i.e., distance, and tariffs) presents a negative and statistically significant impact on the probability of foreign expansion. These results provide evidence that such factors induce firms to remain at home, instead of internationalize through export and/or FDI. It may be surprising that distance has a larger

[26] These results are consistent with those of Helpman et al. (2004) as far as capital intensity is concerned, not in the case of innovation activity. It should be noted, though, that we differ from them in terms of the variable used to proxy for innovation: the number of patents rather than R&D expenses.

negative impact on “export and FDI” than on “export” alone but, even if we control for bilateral features concerning language or religion, the geographical distance is likely to be positively related to other variables increasing the cost of investing abroad. In the same vein, the restraining impact of tariffs is larger when both internationalization modalities are taken into account. This may suggest the absence of “tariff-jumping” FDI, although it should be recalled that our “2” category lumps together both exports and FDIs.

The opposite is true for a second group of bilateral characteristics (i.e. islands, common language and common religion), showing a positive impact on internationalization choices and especially on the export and FDI mode. Finally, the coefficients associated with the number of common partners in trade or FDI confirms the relevance of the network effects. Apparently, firms in sectors with a higher number of foreign contacts are more likely to enter an additional market, and sectors benefit from the contacts of their contacts. In other words, if a firm *k* has a contact in country *j'* which itself has a contact in country *j*, then firm *k* is more likely to enter country *j*. However, the FDI network has an impact almost three times larger than the trade one.

5.2 Does the level of country development matter?

Up to now, we have estimated the ordered probit on the whole sample of observations, and we have found that sectors characterized by a high level of productivity and by a higher incidence of large firms are more likely to be able to afford the higher fixed costs required to serve foreign consumers. However, since our sample includes several origin and destination countries with different levels of development, it is of interest to analyze the behavior of sectors in the internationalization process in different groups of countries. As a matter of fact, non-traditional destination countries of FDI play an increasingly important role, and this raises the question of whether the determinants of FDI differ systematically between developed and developing countries.

For this reason, in Tables 4-6 we present the findings obtained considering different samples of countries. In particular, we concentrate on developed countries as source of FDI and we first estimate the internationalization strategies for the sample of developed countries towards all destination countries and then we split the destination sample into developed and developing countries.[27]

Restricting the sample of origin countries to developed countries does not change the overall picture in terms of our hypotheses (Table 4). The impact of productivity level and the distribution of firms by sales is lower for "export" and "export and FDI" modes of internationalization than that of the whole sample. In particular, looking at the economic impact, a variation of TFP from the 25th to the 75th percentile would determine a decrease of 0.3% of the probability for a sector to remain at home and by 0.2% in the probability of foreign expansion through export only. In turn, this is offset by an increase of 0.5% in the probability that internationalization takes place through both export and FDI. On the other hand, an increase in the number of large firms from the 25th to the 75th percentile reduces the probability mass of being a "domestic" sector by 1.1% and that to internationalize through export only by 0.7%, while it increases the probability of having both exports and investment by 2%. This impact is much higher than that of productivity. However, it must be noticed that in this sample, the difference between the *cut points* identifying the boundaries between internationalization through export and FDI is not statistically significant.

The sign and the significance of the other coefficients remain by and large unchanged, with only few exceptions.[28] Concentrating on the subsample of developed countries, the positive effect of tariffs is

[27] In addition, firms in developing countries face difficulties in expanding in their countries as a result of a less developed institutional environment. However, multinational firms in developing countries may be successful in other countries, despite these disadvantages in their country of origin. In other estimates, not reported but available on request, we have found that sectors from developing countries are more likely to internationalize the higher the level of TFP. On the other hand, the distribution of firms by size is not relevant for them.

[28] It should be recalled that to compare the point estimates across samples and groups within samples, it is necessary to assume that the unobserved heterogeneity is the same across the compared samples or groups (Mood 2010).

consistent with a 'tariff-jumping' strategy that may be explained by the fact that FDI among developed countries are expected to be more horizontal in nature (Blonigen 2002).

Considering the same sample of developed countries as origins, and distinguishing between destination countries, we found some interesting results. As reported in Table 5, the internationalization process of developed countries towards other developed countries is not affected by productivity. The coefficient of TFP is indeed negative, but it is not statistically significant. On the other hand, the distribution of firms by size is still a determinant of the choice of serving foreign markets.

Table 6 presents results for sectors in developed countries internationalizing towards developing markets. Results confirm that international sectors are both more productive and have a higher presence of large firms. Also in this case, the coefficients of the other control variables remain by and large unchanged. The only relevant exception is the coefficient of tariffs, that becomes statistically insignificant. A possible explanation is that tariffs imposed by developing countries are often used not only to protect imports from other countries, but also to finance public balances. Moreover, imports demand curve for products of high quality produced by advanced economies is already inelastic despite tariff barriers.[29]

5.3 Robustness checks: different measures of the number of large firms and labour productivity

In our empirical specification we have employed the number of firms in the 10th decile as an indicator of the distribution of firms by size. However, this could be considered as an *ad-hoc* choice. For this reason, Table 7 reports several robustness checks aimed at verifying that our results do not depend on the specific threshold adopted. In particular, we use three different thresholds to define large firms in a sector: the number of firms in the 9th and 10th decile of the distribution, those in the 7th decile and

[29] In unreported regressions, available on request, we obtain similar results for the group of developing countries as destination of international expansion from all other countries (i.e. both developed and developing).

higher, and a measure of the dispersion of sales in a sector. The last measure, that is constructed as described in Section IV-B, is similar to that adopted by Helpman et al. (2004).

Reassuringly, the coefficients of all different measures of the incidence of large firms confirm the positive impact on the internationalization index and are in line with the first hypothesis. Our main results on the productivity level are confirmed also. Moreover, as already mentioned in Section IV-B, the use of TFP as a measure of productivity implies Germany to be excluded from the analysis sample. For this reason, in the robustness checks reported in Table 8, we include this country and adopt as an alternative measure, the average level of labour productivity. Also in this case, the main results are confirmed. Compared to the impact of TFP reported in Table 3, productivity has a higher coefficient and a higher impact on the probability of internationalize with both exports and investment. The impact of remaining coefficients is almost unchanged.

6. Conclusions

The literature studying firms' choice between exporting at arms' length and serving foreign market through FDI, traditionally modeled as a proximity-concentration trade-off (Brainard 1993, 1997), has been enriched by more recent empirical contributions taking into account heterogeneity in firm productivity (Yeaple 2003; Helpman et al. 2004; Oldenski 2010). In particular, Helpman et al. (2004), building on the theoretical framework of Melitz (2003), demonstrate that a wider dispersion of firms size – reflecting high productivity level – within each sector is associated with a higher incidence of foreign affiliates' sales relative to exports. However, in the existing literature, the distinction between the impact of firm size and firm productivity in the modes of exports is often blurred. In addition, while generating important insights, these studies have generally focused on single-country analysis.

In this paper we analyse the relationship between country and sector characteristics and different internationalization strategies (i.e. export and the FDI), paying special attention to the role of both productivity and the distribution of firms by size in a sector. In particular, we make the hypotheses that

a higher productivity level and a distribution of firms by size shifted toward large firms are associated with a higher level of foreign expansion along the extensive margin, with a stronger effect on FDI than on exports.

From a methodological point of view, we adopt an ordered-choice model. More importantly, we go beyond country studies to look for general patterns. In point of fact, we enlarge previous empirical analyses using a large dataset including 25 domestic countries, 91 foreign countries and 57 manufacturing industries between 1994 and 2004.

The results obtained from the whole sample of countries are qualitatively consistent with the hypotheses that more productive sectors internationalize and that sectors characterized by a distribution of firms by size shifted toward large firms are more likely to be able to afford the higher fixed costs required to serve foreign consumers. These results are also consistent with the theoretical model of Helpman et al. (2004) suggesting that larger and more productive firms should be more likely to internationalize through foreign investment.

Provided that multinational enterprises have increasingly considered developing countries as profitable investment locations, we obtain that for sector in developed countries internationalizing toward these countries both productivity and distribution of firms by size matter. On the other hand, sectors in developed countries going to other developed countries do not need to be more productive, but they need to be large. Finally, our results are also robust to different measures of the number of large firms and the productivity in a sector.

Acknowledgment
We would like to acknowledge the comments of participants at the Italian Trade Study Group conference held in Catania, 15-16 June 2012, and at the European Trade Study Group conference held in Leuven, 13-15 September 2012. In particular, we wish to thank Maurizio Zanardi and Filippo Vergara Caffarelli for helpful comments on a previous version of this paper. A special thanks to the Electronic Resources Area of Bocconi University's library for providing the access to the Worldscope Database.

References

Barba Navaretti G, Venables AJ (2004) Multinational firms in the world economy. Princeton, Princeton University Press.

Basile R, Giunta A, Nugent J (2003) Foreign expansion by Italian manufacturing firms in the nineties: an ordered probit analysis. Rev Ind Organ 23: 1-24

Benfratello L, Razzolini T (2009) Firms' productivity and internationalisation choices: evidence for a large sample of manufacturing firms. In: Piscitello L, Santangelo G (eds) Multinationals and local competitiveness. Milano, Franco Angeli.

Bernanke BS, Gurkaynak RS (2002) Is growth exogenous? Taking Mankiw, Romer and Weil seriously. NBER Macroecon Annual 16: 62-70

Bernard AB, Jensen J B (1995) Exporters, jobs, and wages in US manufacturing, 1976-1987. Brook Pap Econ Act, Microecon 67-119

Bernard AB, Jensen JB (2004) Why some firms export? Rev Econ Stat 86: 561-569

Bernard AB, Jensen JB, Redding S, Schott PK (2007) Firms in international trade. J of Econ Perspect 21: 105-30

Blonigen B (2002) Tariff-jumping antidumping duties. J Int Econ 57: 31-49

Blonigen B (2005) A review of the empirical literature on FDI determinants. Atl Econ J 33: 383-403

Brakman S, Garretsen H, Van Marrewjk C (2005) Cross-border mergers and acquisitions: on revealed comparative advantage and merger waves. CESifo Working Paper n. 1602. http://wwwSSRNcom/abstract=870389. Accessed 3 March 2010

Brainard SL (1993) A simple theory of multinational corporations and trade with a trade-off between proximity and concentration. NBER Working Paper n. 4269. http://wwwnberorg/papers/w4269. Accessed 3 March 2010

Brainard SL (1997) An empirical assessment of the proximity-concentration trade-off between multinational sales and trade. Am Econ Rev 87: 520-44

Bougheas S, Gorg H (2008) Organizational forms for global engagement of firms. University of Nottingham, GEP Research Paper n. 2008/33. http://ssrncom/abstract=1448554. Accessed 27 February 2012

Calia P, Ferrante MR (2010) How do firms combine different internationalization modes? A multivariate probit approach. Paper presented at the ETSG 2010 Annual Conference, 9-11 Septembe 2010, Lausanne, Switzerland. http://wwwetsgorg/ETSG2010/ETSG2010Programmehtml. Accessed 14 February 2012

Carr DL, Markusen JR, Maskus KE (2001) Estimating the knowledge capital model of the multinational enterprises. Am Econ Rev 91: 693-708

Castellani D, Zanfei A (2007) Internationalisation, innovation and productivity: how do firms differ in Italy. World Econ 30: 156-176

Chaney T (2011) The network structure of international trade. NBER Working Paper n. 16753. http://wwwnberorg/papers/w16753. Accessed 14 February 2012.

Conconi P, Sapir A, Zanardi M (2013) The internationalization process of firms: from exports to FDI. Centre for Economic Policy Research (CEPR) Discussion Paper n. 9332. http://164.15.27.46/ecare/personal/conconi$/web/internationalization.pdf. Accessed 27 June 2012.

Demirbas D, Patnaik I, Shah A (2013) Graduating to globalisation: a study of Southern multinationals. Indian Growth Dev Rev (forthcoming).

Disdier AC, Head K (2008) The puzzling persistence of the distance effect on bilateral trade. Rev Econ Stat 90: 37-48

Engel D, Procher V, Schmidt CM (2009) Foreign market dynamics and the symmetric role of firm-specific characteristics - evidence for French firms. http://www2.druid.dk/conferences/viewpaper.php?id=5415&cf=32. Accessed 30 March 2013.

Francois J (2010) Who trades with whom. Paper presented at the 12th European Trade Study Group, Lausanne 9-11 September, 2010.

Greene WH (2008) Econometric Analysis. Upper Saddle River, NJ, Prentice-Hall.

Greenaway J, Kneller J (2007) Firm heterogeneity, exporting and foreign direct investment. Econ J 117: F134-F161

Head K, Ries J (2003) Heterogeneity and the FDI versus export decisions of Japanese manufacturers. J Jpn Int Econ 17: 448-467

Helpman E, Melitz MJ, Yeaple SR (2004) Export versus FDI with heterogeneous firms. Am Econ Rev 94: 300-16

Helpman E, Melitz MJ, Rubistein Y (2008) Estimating trade flows: trading partners and trading volumes. Q J Econ 73: 441-487

Herger N, Kostoggiannis C, McCorriston S (2008) Cross-border acquisitions in the global food sector. Eur Rev Agric Econ, 35: 563-587

Isaksson A (2009) The UNIDO world productivity database: an overview. Inter Product Monitor 18: 38-50

Johnson DKN (2002) The OECD technology concordance (OTC): patents by industry of manufacture and sector of use OECD Science, Technology and Industry. Working Papers n. 2002/5. http://wwwoecd-ilibraryorg/science-and-technology/the-oecd-technology-concordance-otc_521138670407. Accessed 14 February 2012.

Kimura, F, & Kiyota, K (2006) Exports, FDI, and productivity: dynamic evidence from Japanese firms. Rev World Econ/Weltwirtschaftliches Archiv 142: 695-719

Levinsohn J, Petrin A (2003), Estimating production functions using inputs to control for unobservables. Rev Econ Stud 70: 317-41

Markusen JR, Maskus K (2002) Discriminate among alternative theories of the multinational enterprise. Rev Intl Econ 10: 694-707

Melitz MJ (2003) The impact of trade on intra-industry reallocation and aggregate industry productivity. Econom 71: 1695-1725

Mood C (2010) Logistic regression: why we cannot do what we think we can do, and what we can do about it. Europ Sociolog Rev, 26: 67-82

Mrázová M, Neary J (2013). Selection effects with heterogeneous firms. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2210258. Accessed 13 June 2013.

Oberhofer H, Pfaffermayr M (2012) FDI versus exports: multiple host countries and empirical evidence. World Econ 35: 316-330 doi: 101111/j1467-9701201101403x

Oldenski L (2010) Export versus FDI: a task based approach. Working Paper, Georgetown University, 2010 http://www9georgetownedu/faculty/lo36/Oldenski_XvsFDI_Nov2010pdf. Accessed 14 February 2010.

Rajan RG, Zingales L (1998) Financial dependence and growth. Am Econ Rev 88: 559-586

Slangen AHL, Beugelsdijk S (2010) The impact of institutional hazards on foreign multinational activity: A contingency perspective. J Int Bus Stud 41: 980-995. doi:101057/jibs20101

Slangen AHL, Beugelsdijk S, Hennart JMA (2011) The impact of cultural distance on bilateral arm's length exports: an international business perspective. Manag Intl Rev 51: 875-896

Todo Y (2011) Quantitative evaluation of the determinants of exports and FDI: firm-level evidence from Japan. World Econ 34: 355-381. doi: 101111/j1467-9701201101331x

Wang C, Wei Y, Liu X (2010) Determinants of bilateral trade flows in OECD countries: evidence from gravity panel data models. World Econ 33: 894-915. doi: 101111/j1467-9701200901245x

Wooldridge JM (2010) Econometric analysis of cross section and panel data. Second Edition. MIT Press.

Yeaple S (2003) The role of skill endowments in the structure of US outward foreign investments. Rev Econ and Stat 85: 726-734

Table 1. Summary statistics (whole sample)

Variable	*Mean*	*Median*	*St. dev.*	*Min*	*25th*	*75th*	*Max*	*Obs.*
TFP	206.043	179.537	185.047	9.590	118.977	240.535	2,448.199	67,975
Labour productivity	606.414	514.074	689.905	11.125	282.181	704.687	13,135.190	72,904
Num. of large firms (9th decile)	2.230	0.364	5.662	0	0	1.600	52.818	72,904
Num. of large firms (10th decile)	2.353	0	5.999	0	0	1.714	51.727	72,904
Num. of large firms (4th quintile)	4.193	1	9.733	0	0.200	3.182	83.143	72,904
Num. of large firms (5th quintile)	4.583	0.909	11.498	0	0	3.091	104.546	72,904
Sales dispersion	1.499	1.223	1.084	0.030	0.791	1.951	7.840	72,825
Capital intensity	1.681	1.664	0.178	1.309	1.565	1.774	2.468	67,975
Patents	17.236	0.008	86.534	0	0	2.682	1,465.436	72,904
Distance	8,322	8,224	4,237	215	5,519	10,470	19,772	72,904
Islands	0	0	1	0	0	1	2	72,904
Common language	0	0	0	0	0	0	1	72,904
Common religion	0	0	0	0	0	0	1	72,904
Tariffs	0.117	0.093	0.108	0	0.034	0.172	0.582	72,904
Common partners in trade	57.931	57	36.866	0	24	92	117	72,904
Common partners in FDI	0.399	0	1.440	0	0	0	30	72,904

Variables description and sources are provided in Appendix 1. Summary statistics are computed after excluding influential outliers. 25th and 75th refer to the percentiles of the world distribution. All descriptive statistics are computed on variables in levels.

Table 2. Correlation matrix

	Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1)	*y*	1																
(2)	*TFP*	0.087*	1															
(3)	*Labour productivity*	0.128*	0.775*	1														
(4)	*Num. of large firms (9thdecile)*	0.212*	0.119*	0.233*	1													
(5)	*Num. of large firms (10thdecile)*	0.221*	0.131*	0.250*	0.944*	1												
(6)	*Num. of large firms (4thquintile^{j}*	0.202*	0.097*	0.202*	0.951*	0.901*	1											
(7)	*Num. of large firms (5thquintile)*	0.220*	0.127*	0.245*	0.985*	0.987*	0.938*	1										
(8)	*Sales dispersion*	0.225*	0.097*	0.211*	0.627*	0.643*	0.636*	0.644*	1									
(9)	*Capital intensity*	0.058*	0.307*	0.275*	-0.158*	-0.140*	-0.180*	-0.151*	-0.195*	1								
(10)	*Patents*	0.131*	0.056*	0.140*	0.417*	0.437*	0.366*	0.433*	0.349*	-0.108*	1							
(11)	*Distance*	-0.132*	-0.034*	-0.009*	0.051*	0.046*	0.058*	0.049*	0.066*	-0.093*	0.017*	1						
(12)	*Islands*	0.003	0.143*	0.107*	0.123*	0.135*	0.114*	0.114*	0.030*	0.095*	-0.001*	0.064*	1					
(13)	*Common language*	0.090*	-0.011*	-0.007	0.042*	0.033*	0.042*	0.038*	0.019*	-0.071*	0.059*	-0.088*	0.127*	1				
(14)	*Common religion*	-0.046*	-0.043*	-0.058*	-0.080*	-0.083*	-0.093*	-0.082*	-0.163*	-0.013*	-0.028*	-0.066*	-0.055*	0.110*	1			
(15)	*Tariffs*	-0.025*	0.020*	0.024*	-0.019*	-0.013*	-0.028*	-0.016*	-0.032*	-0.018*	0.011*	-0.020*	-0.058*	0.020*	-0.138*	1		
(16)	*Common partners in trade*	0.208*	-0.076*	-0.048*	0.031*	0.028*	0.036*	0.030*	0.037*	-0.122*	0.020*	-0.043*	-0.009*	-0.063*	-0.004	-0.222*	1	
(17)	*Common partners in FDI*	0.329*	0.028*	0.059*	0.219*	0.223*	0.217*	0.224*	0.203*	-0.019*	0.113*	-0.041*	0.073*	0.083*	0.014*	-0.171*	0.337*	1

Variable description and sources are provided in Appendix 1. Correlations are computed after excluding influential outliers. * indicates significance at the 1% level. Correlations are computed on variables in levels.

Table 3. Ordered probit on the whole sample

	Coefficients		***Changes in predicted probability***		
			Domestic	Export	Export and FDI
	(1)		(2)	(3)	(4)
TFP[a]	1.191	***	-0.089	0.021	0.068
	(0.028)				
Num. of large firms (10th decile)[b]	0.425	***	-0.039	0.016	0.023
	(0.015)				
Capital intensity[a]	4.914	***	-0.069	0.026	0.043
	(0.127)				
Patents[b]	0.154	***	-0.015	0.006	0.009
	(0.009)				
Distance[a]	-0.459	***	0.034	-0.014	-0.020
	(0.012)				
Islands	0.912	***	-0.127	-0.093	0.220
	(0.035)				
Common language	0.417	***	-0.036	0.003	0.033
	(0.028)				
Common religion	0.291	***	-0.027	0.005	0.022
	(0.037)				
Tariffs[b]	-0.335	**	0.004	-0.001	-0.003
	(0.133)				
Common partners in trade[b]	0.352	***	-0.040	0.009	0.031
	(0.017)				
Common partners in FDI[b]	0.904	***	-0.090	0.030	0.060
	(0.020)				
cut1	5.424	***			
	(0.215)				
cut2	9.895	***			
	(0.222)				
Average predicted probability			0.087	0.855	0.058
Observations			67,975		

Variables description and sources are provided in Appendix 1. Column (1) reports coefficients of estimations. Columns (2)-(4) report changes in predicted probability for continuous variables varying from 25th to 75th of the world distribution, for discrete variables or dummies varying from the minimum to the maximum value and marginal effects for "*Common partners in FDI*". *cut1* and *cut2* indicate thresholds between one category and the next. Standard errors robust to heterosckedasticity are reported in parentheses. Standard errors for cut points and changes in predicted probabilities are calculated with the delta method. Changes in predicted probabilities are significant at the 1% level. ***, ** and * indicate statistical significance at the 1%, 5% and 10% level, respectively.

[a] This variable is included as ln(variable).

[b] This variable is included as ln(1+variable).

Table 4. Developed countries as origin vs. all other countries

	Coefficients		***Changes in predicted probability***		
			Domestic	Export	Export and FDI
	(1)		(4)	(3)	(4)
TFP[a]	0.119	**	-0.003	-0.002	0.005
	(0.060)				
Num. of large firms (10th decile)[b]	0.190	***	-0.011	-0.007	0.018
	(0.019)				
Capital intensity[a]	-1.263	***	0.006	0.006	-0.012
	(0.240)				
Patents[b]	0.046	***	-0.004	-0.002	0.006
	(0.012)				
Distance[a]	-0.523	***	0.017	0.009	-0.026
	(0.019)				
Islands	1.157	***	-0.080	-0.214	0.294
	(0.055)				
Common language	0.458	***	-0.015	-0.026	0.042
	(0.038)				
Common religion	0.373	***	-0.014	-0.018	0.033
	(0.070)				
Tariffs[b]	0.061		-0.000	-0.001	0.001
	(0.198)				
Common partners in trade[b]	0.238	***	-0.012	-0.014	0.025
	(0.023)				
Common partners in FDI[b]	0.508	***	-0.023	-0.017	0.040
	(0.024)				
cut1	-5.338	***			
	(0.417)				
cut2	-0.072				
	(0.417)				
Average predicted probability			0.026	0.898	0.076
Observations			47,563		

Variables description and sources are provided in Appendix 1. Column (1) reports coefficient of estimations. Columns (2)-(4) report changes in predicted probability for continuous variables varying from 25th to 75th of the world distribution, for discrete variables or dummies varying from the minimum to the maximum value and marginal effects for "*Common partners in FDI*". *cut1* and *cut2* indicate thresholds between one category and the next. Standard errors robust to heterosckedasticity are reported in parentheses. Standard errors for cut points and changes in predicted probabilities are calculated with the delta method. Changes in predicted probabilities are significant at the 1% level. ***, ** and * indicate statistical significance at the 1%, 5% and 10% level, respectively.

[a] This variable is included as ln(variable).

[b] This variable is included as ln(1+variable).

Table 5. Developed countries as origin vs. other developed countries

	Coefficients		***Changes in predicted probability***		
			Domestic	Export	Export and FDI
	(1)		(2)	(3)	(4)
TFP[a]	-0.110		0.001	0.007	-0.007
	(0.106)				
Num. of large firms (10th decile)[b]	0.194	***	-0.004	-0.033	0.037
	(0.033)				
Capital intensity[a]	-0.453		0.001	0.007	-0.008
	(0.426)				
Patents[b]	0.068	***	-0.002	-0.015	0.018
	(0.020)				
Distance[a]	-0.504	***	0.010	0.062	-0.072
	(0.031)				
Islands	1.626	***	-0.100	-0.385	0.485
	(0.108)				
Common language	0.619	***	-0.007	-0.090	0.097
	(0.067)				
Common religion	-0.363		0.007	0.039	-0.046
	(0.252)				
Tariffs[b]	1.164	**	-0.001	-0.006	0.007
	(0.510)				
Common partners in trade[b]	0.370	***	-0.002	-0.017	0.019
	(0.070)				
Common partners in FDI[b]	0.447	***	-0.006	-0.062	0.068
	(0.035)				
cut1	-8.054	***			
	(0.781)				
cut2	-2.666	***			
	(0.781)				
Average predicted probability			0.010	0.812	0.178
Observations			12,646		

Variables description and sources are provided in Appendix 1. Column (1) reports coefficient of estimations. Columns (2)-(4) report changes in predicted probability for continuous variables varying from 25th to 75th of the world distribution, for discrete variables or dummies varying from the minimum to the maximum value and marginal effects for "*Common partners in FDI*". *cut1* and *cut2* indicate thresholds between one category and the next. Standard errors robust to heterosckedasticity are reported in parentheses. Standard errors for cut points and changes in predicted probabilities are calculated with the delta method. Changes in predicted probabilities are significant at the 1% level. ***, ** and * indicate statistical significance at the 1%, 5% and 10% level, respectively.

[a] This variable is included as ln(variable).

[b] This variable is included as ln(1+variable).

Table 6. Developed countries as origins vs. developing countries

	Coefficients		***Changes in predicted probability***		
			Domestic	Export	Export and FDI
	(1)		(2)	(3)	(4)
TFP[a]	0.261	***	-0.006	0.0003	0.006
	(0.069)				
Num. of large firms (10th decile)[b]	0.208	***	-0.014	0.001	0.012
	(0.023)				
Capital intensity[a]	-1.485	***	0.009	0.000	-0.009
	(0.283)				
Patents[b]	0.036	**	-0.003	0.000	0.003
	(0.016)				
Distance[a]	-0.585	***	0.022	-0.003	-0.019
	(0.027)				
Islands	0.957	***	-0.072	-0.125	0.197
	(0.065)				
Common language	0.364	***	-0.016	-0.007	0.023
	(0.048)				
Common religion	0.312	***	-0.015	-0.004	0.019
	(0.074)				
Tariffs[b]	-0.108		0.001	0.000	-0.001
	(0.233)				
Common partners in trade[b]	0.170	***	-0.012	-0.001	0.013
	(0.026)				
Common partners in FDI[b]	0.411	***	-0.022	0.0004	0.022
	(0.047)				
cut1	-5.800	***			
	(0.508)				
cut2	-0.534				
	(0.505)				
Average predicted probability			0.032	0.929	0.039
Observations			34,917		

Variables description and sources are provided in Appendix 1. Column (1) reports coefficient of estimations. Columns (2)-(4) report changes in predicted probability for continuous variables varying from 25th to 75th of the world distribution, for discrete variables or dummies varying from the minimum to the maximum value and marginal effects for "*Common partners in FDI*". *cut1* and *cut2* indicate thresholds between one category and the next. Standard errors robust to heterosckedasticity are reported in parentheses. Standard errors for cut points and changes in predicted probabilities are calculated with the delta method. Changes in predicted probabilities are significant at the 1% level. ***, ** and * indicate statistical significance at the 1%, 5% and 10% level, respectively.

[a] This variable is included as ln(variable).

[b] This variable is included as ln(1+variable).

Table 7 Robustness checks: different measures of the incidence of large firms

	9th and 10th decile					**4th and 5th quintile**					**Sales dispersion**				
	Coefficients		***Changes in predicted probability***			***Coefficients***		***Changes in predicted probability***			***Coefficients***		***Changes in predicted probability***		
			Domestic	Export	Export and FDI			Domestic	Export	Export and FDI			Domestic	Export	Export and FDI
	(1)		(2)	(3)	(4)	(5)		(6)	(7)	(8)	(9)		(10)	(11)	(12)
TFP[a]	1.204	***	-0.088	0.021	0.067	1.179	***	-0.085	0.020	0.065	1.197	***	-0.090	0.021	0.069
	(0.028)					(0.028)					(0.027)				
Num. of large firms (9th decile)[b]	0.466	***	-0.044	0.019	0.025										
	(0.019)														
Num. of large firms (10th decile)[b]	0.203	***	-0.018	0.006	0.011										
	(0.018)														
Num. of large firms (4th quintile)[b]						0.093	***	-0.011	0.004	0.007					
						(0.016)									
Num. of large firms (5th quintile)[b]						0.533	***	-0.076	0.035	0.041					
						(0.015)									
Sales dispersion[a]											0.209	***	-0.018	0.005	0.013
											(0.013)				
Capital intensity[a]	5.023	***	-0.070	0.026	0.044	5.175	***	-0.072	0.026	0.046	5.050	***	-0.072	0.025	0.046
	(0.129)					(0.131)					(0.128)				
Patents[b]	0.110	***	-0.011	0.004	0.007	0.106	***	-0.010	0.004	0.006	0.195	***	-0.020	0.008	0.012
	(0.010)					(0.010)					(0.009)				
Distance[a]	-0.464	***	0.034	-0.013	-0.020	-0.467	***	0.034	-0.013	-0.020	-0.458	***	0.034	-0.013	-0.021
	(0.012)					(0.012)					(0.012)				
Islands	0.850	***	-0.118	-0.076	0.194	0.806	***	-0.112	-0.065	0.177	0.923	***	-0.130	-0.099	0.228
	(0.035)					(0.035)					(0.035)				
Common language	0.421	***	-0.035	0.002	0.033	0.425	***	-0.035	0.002	0.033	0.403	***	-0.035	0.002	0.033
	(0.028)					(0.028)					(0.027)				
Common religion	0.343	***	-0.031	0.005	0.026	0.379	***	-0.033	0.005	0.028	0.306	***	-0.028	0.005	0.023
	(0.037)					(0.037)					(0.037)				
Tariffs	-0.311	**	0.004	-0.001	-0.003	-0.318	**	0.004	-0.001	-0.003	-0.368	***	0.005	-0.001	-0.003
	(0.135)					(0.136)					(0.133)				
Common partners in trade[b]	0.353	***	-0.039	0.008	0.031	0.350	***	-0.038	0.008	0.030	0.340	***	-0.039	0.008	0.031
	(0.017)					(0.017)					(0.017)				
Common partners in FDI[b]	0.866	***	-0.085	0.028	0.057	0.842	***	-0.082	0.026	0.055	0.937	***	-0.094	0.030	0.064
	(0.020)					(0.020)					(0.020)				
cut1	5.559	***				5.570	***				5.399	***			
	(0.217)					(0.217)					(0.213)				
cut2	10.091	***				10.145	***				9.836	***			
	(0.225)					(0.225)					(0.220)				
Average predicted probability			0.087	0.856	0.058			0.087	0.855	0.058			0.086	0.856	0.058
Observations			67,975					67,975					67,896		

Variables description and sources are provided in Appendix 1. Columns (1),(5) and (9) report coefficient of estimations. Columns (2)-(4), (6)-(8) and (10)-(12) report changes in predicted probability for continuous variables or dummies varying from 25th to 75th of the world distribution, for discrete variables varying from the minimum to the maximum value and marginal effects for "*Common partners in FDI*". *cut1* and *cut2* indicate thresholds between one category and the next. Standard errors robust to heterosckedasticity are reported in parentheses. Standard errors for cut points and changes in predicted probabilities are calculated with the delta method. Changes in predicted probabilities are significant at the 1% level. ***, ** and * indicate statistical significance at the 1%, 5% and 10% level, respectively.

[a] This variable is included as ln(variable).

[b] This variable is included as ln(1+variable).

Table 8

Robustness checks: labour productivity

	Coefficients		***Changes in predicted probability***		
			Domestic	Export	Export and FDI
	(1)		(2)	(3)	(4)
Labour productivity[a]	1.280	***	-0.120	0.020	0.100
	(0.024)				
Num. of large firms (10th decile)[b]	0.324	***	-0.030	0.009	0.020
	(0.014)				
Patents[b]	0.128	***	-0.015	0.005	0.011
	(0.009)				
Distance[a]	-0.452	***	0.028	-0.009	-0.018
	(0.012)				
Islands	0.614	***	-0.085	-0.042	0.127
	(0.031)				
Common language	0.464	***	-0.036	-0.002	0.038
	(0.028)				
Common religion	0.325	***	-0.028	0.003	0.025
	(0.036)				
Tariffs[b]	-0.439	***	0.005	-0.001	-0.004
	(0.133)				
Common partners in trade[b]	0.331	***	-0.035	0.005	0.030
	(0.017)				
Common partners in FDI[b]	0.876	***	-0.081	0.021	0.060
	(0.020)				
cut1	4.360	***			
	(0.202)				
cut2	8.912	***			
	(0.209)				
Average predicted probability			0.081	0.860	0.059
Observations			72,904		

Variables description and sources are provided in Appendix 1. Column (1) reports coefficient of estimations. Columns (2)-(4) report changes in predicted probability for continuous variables varying from 25th to 75th of the world distribution, for discrete variables or dummies varying from the minimum to the maximum value and marginal effects for "*Common partners in FDI*". *cut1* and *cut2* indicate thresholds between one category and the next. Standard errors robust to heterosckedasticity are reported in parentheses. Standard errors for cut points and changes in predicted probabilities are calculated with the delta method. Changes in predicted probabilities are significant at the 1% level.. ***, ** and * indicate statistical significance at the 1%, 5% and 10% level, respectively.

[a] This variable is included as ln(variable).

[b] This variable is included as ln(1+variable).

Appendix 1

Variables description and sources

Definition	Description and Source
	Dependent variable
y	Categorical variable taking the value of zero if sector *h* in country *i* neither exports nor invests in country *j*, the value of 1 if sector *h* in country *i* only exports in country *j* and the value of 2 if sector *h* in country *i* both exports and invests in country *j*. *Source*: UN Comtrade for exports and SDC Platinum for FDI
	Key independent variables
TFP[a]	Average level of total factor productivity in sector *h* in country *i*. *Source*: UNIDO (Indstat4, 2008 version)
Labour productivity[a]	Ratio between value added and number of employees in sector *h* in country *i*. *Source*: UNIDO (Indstat4, 2008 version)
Num. of large firms (*9th decile*)[b]	Number of firms in country *i* in the 9th decile of the world distribution of firm sales in a given sector *h*. *Source:* Worldscope Database
Num. of large firms (*10th decile*) [b]	Number of firms in country *i* in the 10th decile of the world distribution of firm sales in a given sector *h*. *Source:* Worldscope Database
Num. of large firms (*4th quintile*) [b]	Number of firms in country *i* in the 4th quintile of the world distribution of firm sales in a given sector *h*. *Source:* Worldscope Database
Num. of large firms (*5th quintile*) [b]	Number of firms in country *i* in the 5th quintile of the world distribution of firm sales in a given sector *h*. *Source:* Worldscope Database
Sales dispersion[a]	Standard deviation of the world distribution of the size of firms, measured by total sales, in a given sector *h*. *Source:* Worldscope Database
	Sector-level variables for the country of origin
Capital intensity[a]	Ratio between capital and number of employees in sector *h* in country *i*. *Source*: UNIDO (Indstat4, 2008 version)
Patents [b]	Number of patents produced in a country *i* and in a given sector *h* and granted by the US Patent Office. *Source*: NBER
	Bilateral country-level variables
Distance [a]	Average distance between countries *i* and *j* calculated through the great circle formula that uses latitudes and longitudes of the most important cities (in terms of population). *Source*: CEPII http://www.cepii.fr/anglaisgraph/bdd/distances.htm
Islands	Number of countries that are islands in the pair of countries *i* and *j*. *Source*: CEPII http://www.cepii.fr/anglaisgraph/bdd/distances.htm

Appendix 1
(continued)

Common religion	Dummy variable equal to 1 if country *i* and *j* share the same religion. *Source*: CEPII http://www.cepii.fr/anglaisgraph/bdd/distances.htm
	Bilateral country- and sector-level variables
Tariffs[b]	Tariffs applied from country *j* to country *i* in sector *h*. *Source*: TRAINS
Common partners in trade[b]	Number of partners in trade common to country *i* and *j* in sector *h*. *Source*: UN Comtrade
Common partners in FDI[b]	Number of partners in FDI common to country *i* and *j* in sector *h*. *Source*: SDC Platinum

[a] This variable is included in the estimations as ln(variable).
[b] This variable is included in the estimations as ln(1+variable).

Appendix 2

Concordances between SIC and ISIC classification at 3-digits

SIC	Description	ISIC code			
101	Iron Ores	131			
102	Copper Ores	132			
103	Lead and Zinc Ores	132			
104	Gold and Silver Ores	132			
106	Ferroalloy Ores, Exc Vanadium	132			
109	Miscellaneous Metal Ores	120	132		
122	Bituminous Coal and Lignite Mining	101	102		
123	Anthracite Mining	101			
124	Coal Mining Services	101			
131	Crude Petroleum and Natural Gas	101	111		
138	Oil and Gas Field Services	112			
141	Dimension Stone	141	142		
142	Crushed and Broken Stone	141	142		
144	Sand and Gravel	141			
145	Clay, Ceramic, & Refractory Minerals	141	142		
147	Chemical and Fertilizer Minerals	142			
149	Miscellaneous Nonmetallic Minerals	103	141	142	
201	Meat Products	151	154		
202	Dairy Products	152	154		
203	Preserved Fruits and Vegetables	151	153	154	
204	Grain Mill Products	151	153	154	
205	Bakery Products	154			
206	Sugar and Confectionery Products	154			
207	Fats and Oils	151			
208	Beverages	11	151	154	155
209	Misc. Foods and Kindred Products	151	152	154	
211	Cigarettes	160			
212	Cigars	160			
213	Chewing and Smoking Tobacco	160			
214	Tobacco Stemming and Redrying	160			

Appendix 2
(continued)

221	Broadwoven Fabric Mills, Cotton	171						
222	Broadwoven Fabric Mills, Manmade	171						
223	Broadwoven Fabric Mills, Wool	171	172					
224	Narrow Fabric Mills	172						
225	Knitting Mills	172	173	181				
226	Textile Finishing, Exc Wool	171	172					
227	Carpets and Rugs	172						
228	Yarn and Thread Mills	171						
229	Misc. Textile Goods	171	172	372				
231	Men's and Boys' Suits and Coats	181						
232	Men's and Boys' Furnishings	181						
233	Women's, Misses', and Juniors' Outerwear	181	182					
234	Women's and Children's Undergarments	181						
235	Hats, Caps and Millinery	181						
236	Girl's and Children's Outwear	181						
238	Misc. Apparel and Accessories	181	252					
239	Misc. Fabricated Textile Products	172	191	181	252	343	369	
241	Logging	20	201					
242	Sawmills and Planing Mills	201	202	361				
243	Millwork, Plywood and Structural Members	201	202	361				
244	Wooden Containers	202						
245	Wood Buildings and Mobile Homes	202	342					
249	Misc. Wood Products	201	202	252	289	361		
251	Household Furniture	361						
252	Office Furniture	361						
253	Public Building & Related Furniture	361						
254	Partitions and Fixtures	202	281	361				
259	Misc. Furniture and Fixtures	172	202	252	289	331	359	361
261	Pulp Mills	210						
262	Mills, Exc Building Paper	210	269					
263	Paperboard Mills	210						

Appendix 2
(continued)

265	Paperboard Containers and Boxes	210							
267	Misc. Converted Paper Products	172	210	252	289	369			
271	Newspapers	221							
272	Periodicals	221							
273	Books	221	222						
274	Miscellaneous Publishing	221	369						
275	Commercial Printing	210	222						
276	Manifold Business Forms	222							
277	Greeting Cards	221	222						
278	Blankbooks and Bookbinding	222							
279	Printing Trade Services	222							
281	Industrial Inorganic Chemicals	233	241	242	272				
282	Plastics Materials and Synthetic	241	243						
283	Drugs	242							
284	Soaps, Cleaners and Toilet Goods	242							
285	Paints and Allied Products	242							
286	Industrial Organic Chemicals	155	233	241	242				
287	Agricultural Chemicals	241	242						
289	Misc. Chemical Products	241	242	269	369				
291	Petroleum Refining	232							
295	Asphalt Paving and Roofing Materials	269							
299	Misc. Petroleum and Coal Products	101	102	202	232	242			
301	Tires and Inner Tubes	251							
302	Rubber and Plastic Footwear	192							
305	Hose & Belting & Gaskets & Packing	172	191	202	251	252	269	289	291
306	Fabricated Rubber Products, Nec	192	251	292	314	351	369	372	
308	Misc. Plastics Products, Nec	192	241	252	314	342	343	351	
311	Leather Tanning and Finishing	191							
313	Footware Cut Stock	192	289						
314	Footwear, Exc Rubber	192							
315	Leather Gloves and Mittens	181							

Appendix 2
(continued)

316	Luggage	191									
317	Handbags and Personal Leather Goods	191									
319	Leather Goods, Nec	181	191	192	369						
321	Flat Glass	261									
322	Glass and Glassware, Pressed or Blown	261	332								
323	Products of Purchased Glass	261									
324	Cement, Hydraulic	269									
325	Structural Clay Products	269									
326	Pottery and Related Products	269									
327	Concrete, Gypsum and Plaster Products	269									
328	Cut Stone and Stone Products	269									
331	Blast Furnace and Basic Steel Products	231	271	272	289	359					
332	Iron and Steel Foundries	271	273								
333	Primary Nonferrous Metals	272									
334	Secondary Nonferrous Metals	272	371								
335	Nonferrous Rolling and Drawing	272	313								
336	Nonferrous Foundries (Castings)	273									
339	Misc. Primary Metal Products	271	272	289							
341	Metal Cans and Shipping Containers	289									
342	Cutlery, Hand Tools and Hardware	289	291	292	331	343	361	369			
343	Plumbing and Heating, Exc Electric	281	289	291	293						
344	Fabricated Structural Metal Products	271	281	289	291	292	293	300	342	351	
345	Screw Machine Products, Bolts, etc.	289									
346	Metal Forgings and Stampings	289	343								
347	Metal Services, Nec	289									
348	Ordnance and Accessories, Nec	172	292	242							
349	Misc. Fabricated Metal Products	251	271	272	289	291	319	333	359	361	369
351	Engines and Turbines	291	311	341	343						
352	Farm and Garden Machinery	281	289	291	292	359					
353	Construction and Related Machinery	289	291	292	342	352	359				
354	Metalworking Machinery	289	292	311	331						

Appendix 2 (continued)

355	Special Industry Machinery	281	291	292	319	331	332												
356	General Industrial Machinery	172	291	293	331														
357	Computer and Office Equipment	300	323	333															
358	Refrigeration and Service Machinery	291	292	293	359														
359	Industrial Machinery, Nec	289	291	343	353	369													
361	Electric Distribution Equipment	311	312																
362	Electrical Industrial Apparatus	269	311	312	319	321													
363	Household Appliances	292	293	369															
364	Electric Lighting and Wiring Equipment	172	312	315	319														
365	Household Audio & Video Equipment	221	223	323															
366	Communications Equipment	311	322	323	331	353													
367	Electronic Components and Accessories	311	312	313	319	321	323												
369	Misc. Electrical Equipment & Supplies	242	292	293	313	314	315	319	322	353									
371	Motor Vehicles and Equipment	292	319	341	342	343													
372	Aircraft and Parts	291	353																
373	Ship and Boat Building and Repairing	351																	
374	Railroad Equipment	291	341	352															
375	Motorcycles, Bicycles and Parts	359																	
376	Guided Missiles, Space Vehicles, Parts	292	353																
379	Misc. Transportation Equipment	292	341	342	343	359													
381	Search and Navigation Equipment	331																	
382	Measuring and Controlling Devices	291	311	319	331	332	333	361											
384	Medical Instruments & Supplies	172	181	191	242	252	269	289	292	331	359	369							
385	Ophthalmic Goods	331	332																
386	Photographic Equipment and Supplies	242	300	323	331	332													
387	Watches, Clocks, Watchcases & Parts	333																	
391	Jewelry, Silverware and Plated Ware	289	333	369															
393	Musical Instruments	369																	
394	Toys and Sporting Goods	369																	
395	Pens, Pencils, Office, & Art Supplies	172	210	242	291	331	361	369											
396	Costume Jewelry and Notions	191	289	292	333	369													
399	Misc. Manufacturing	171	181	182	202	221	222	252	289	291	292	293	315	319	331	332	361	369	

Appendix 3

List of foreign and domestic countries and number of observations

Foreing country	Observations	Domestic country	Observations
Algeria	783	Argentina	2,007
Argentina	906	Austria	2,636
Australia	953	Belgium	2,568
Austria	587	Colombia	347
Bangladesh	798	Denmark	2,839
Belgium	587	Finland	2,603
Belize	447	France	982
Bermuda	297	India	3,733
Bolivia	834	Indonesia	3,587
Brazil	999	Italy	3,613
Burkina Faso	563	Japan	4,635
Cambodia	548	Korea	4,482
Canada	970	Malaysia	3,956
Chad	338	Mexico	2,453
Chile	935	Morocco	686
China	998	Netherlands	3,180
Colombia	933	Norway	2,883
Costa Rica	801	Peru	621
Cyprus	729	Philippines	1,306
Czech Republic	934	Portugal	1,624
Denmark	587	Singapore	3,860
Dominican Republic	735	Sweden	2,951
Ecuador	855	Thailand	1,716
Egypt	760	United Kingdom	4,121
Ethiopia	648	United States	4,586
Finland	587	**Total**	**67,975**
France	582		
Gabon	664		
Germany	587		
Ghana	714		
Greece	587		
Guatemala	790		
Guyana	529		
Honduras	716		
Hong Kong	677		
Hungary	865		
Iceland	831		
India	807		
Indonesia	870		
Iran	617		
Ireland	587		
Israel	605		
Italy	586		

Appendix 3
(continued)

Jamaica	772
Japan	969
Jordan	745
Korea	879
Madagascar	600
Malaysia	856
Mauritius	654
Mexico	882
Morocco	782
Mozambique	714
Nepal	607
Netherlands	587
New Zealand	928
Nicaragua	747
Nigeria	772
Norway	892
Oman	709
Pakistan	738
Panama	670
Paraguay	846
Peru	837
Philippines	919
Poland	868
Portugal	587
Romania	806
Russian Federation	938
Rwanda	389
Saudi Arabia	940
Singapore	826
South Africa	938
Spain	587
Sri Lanka	849
Sweden	587
Switzerland	949
Taiwan	953
Tanzania	742
Thailand	872
Trinidad and Tobago	767
Tunisia	740
Turkey	851
United Arab Emirates	576

Appendix 3
(continued)

United Kingdom	587
United States	940
Uruguay	836
Venezuela	851
Vietnam	833
Zambia	576
Zimbabwe	756
Total	**67,975**

Appendix 4

List of sectors and number of observations

Isic	Description	Obs.
151	Production, processing and preservation of meat, fish, fruit, vegetables, oils and fats	1,766
152	Manufacture of dairy products	760
153	Manufacture of grain mill products, starches and starch products, and prepared animal feeds	1,216
154	Manufacture of other food products	1,612
155	Manufacture of beverages	1,212
160	Manufacture of tobacco products	270
171	Spinning, weaving and finishing of textiles	1,289
172	Manufacture of other textiles	1,539
173	Manufacture of knitted and crocheted fabrics and articles	731
181	Manufacture of wearing apparel, except fur apparel	1,166
182	Dressing and dyeing of fur; manufacture of articles of fur	226
191	Tanning and dressing of leather; manufacture of luggage, handbags, saddlery and harness	1,085
192	Manufacture of footwear	1,064
201	Sawmilling and planing of wood	652
202	Manufacture of products of wood, cork, straw and plaiting materials	1,139
210	Manufacture of paper and paper products	1,644
221	Publishing	1,274
222	Printing and service activities related to printing	1,123
231	Manufacture of coke oven products	156
232	Manufacture of refined petroleum products	862
233	Processing of nuclear fuel	201
241	Manufacture of basic chemicals	1627
242	Manufacture of other chemical products	1640
243	Manufacture of man-made fibres	897
251	Manufacture of rubber products	1,115
252	Manufacture of plastics products	1,646
261	Manufacture of glass and glass products	878
269	Manufacture of non-metallic mineral products n.e.c.	1,731
271	Manufacture of basic iron and steel	1,255
272	Manufacture of basic precious and non-ferrous metals	1,301
281	Manufacture of structural metal products, tanks, reservoirs and steam generators	1,364
289	Manufacture of other fabricated metal products; metal working service activities	1,719
291	Manufacture of general purpose machinery	1,667
292	Manufacture of special purpose machinery	1,655
293	Manufacture of domestic appliances n.e.c.	1,508
300	Manufacture of office, accounting and computing machinery	1,259
311	Manufacture of electric motors, generators and transformers	1,408
312	Manufacture of electricity distribution and control apparatus	1,292
313	Manufacture of insulated wire and cable	1,278
314	Manufacture of accumulators, primary cells and primary batteries	1,164
315	Manufacture of electric lamps and lighting equipment	1,348
319	Manufacture of other electrical equipment n.e.c.	1,484
321	Manufacture of electronic valves and tubes and other electronic components	1,223

Appendix 4
(continued)

322	Manufacture of television and radio transmitters and apparatus for line telephony and line telegraphy	1,211
323	Manufacture of television and radio receivers, sound or video recording or reproducing apparatus, and associated goods	1,278
331	Manufacture of medical appliances and instruments and appliances for measuring, checking, testing, navigating and other purposes, except optical instruments	1,409
332	Manufacture of optical instruments and photographic equipment	1,297
333	Manufacture of watches and clocks	911
341	Manufacture of motor vehicles	1,066
342	Manufacture of bodies (coachwork) for motor vehicles; manufacture of trailers and semi-trailers	1,220
343	Manufacture of parts and accessories for motor vehicles and their engines	1,524
351	Building and repairing of ships and boats	927
352	Manufacture of railway and tramway locomotives and rolling stock	457
353	Manufacture of aircraft and spacecraft	941
359	Manufacture of transport equipment n.e.c.	1,218
361	Manufacture of furniture	1,381
369	Manufacturing n.e.c.	1,689

Appendix 5

Summary statistics by international involvement

	Domestic		Export		Export and FDI	
Variable	***Mean***	***St. dev.***	***Mean***	***St. dev.***	***Mean***	***St. dev.***
TFP	159.973	204.127	208.325	184.885	242.277	138.110
Labour productivity	375.653	662.184	611.656	684.608	851.495	708.312
Num. of large firms (9^{th}decile)	0.399	1.756	2.057	5.221	7.362	10.603
Num. of large firms (10^{th}decile)	0.371	1.851	2.153	5.491	8.089	11.369
Num. of large firms (4^{th}quintile)	1.217	3.229	3.903	9.015	12.663	18.083
Num. of large firms (5^{th}quintile)	0.770	3.529	4.210	10.550	15.450	21.670
Sales dispersion	1.061	0.667	1.476	1.047	2.452	1.469
Capital intensity	1.624	0.197	1.688	0.177	1.662	0.146
Patents	2.029	19.977	15.195	78.790	68.804	182.669
Distance	10,069	4,383	8,236	4,185	7,166	4,116
Islands	0	1	0	1	0	1
Common language	0	0	0	0	0	0
Common religion	0	0	0	0	0	0
Tariffs	0.105	0.112	0.121	0.107	0.083	0.097
Common partners in trade	51.811	35.840	55.774	36.149	98.495	21.785
Common partners in FDI	0.082	0.354	0.254	0.841	2.992	4.229
Observations	5,917		62,758		4,229	

Variables description and sources are provided in Appendix 1. Summary statistics are computed, after excluding influential outliers, on three groups of sectors depending on internationalization involvement: “domestic”, “export” and “export and FDI”. All descriptive statistics are computed on variables in levels.