Imports and productivity: firm-level evidence from Spain

(28 March 2006)

Jose C. Fariñas
(Universidad Complutense de Madrid)

Ana Martin-Marcos
(UNED)

Abstract

The objective of this paper is to explore the relationship between imports and productivity at the firm level. To organize the empirical work we rely on the model of Antràs and Helpman (2004), which predicts that high-productivity firms engage in arm's length trade (importing firms) and low-productivity firms do not outsource abroad (non-importing firms). The paper performs productivity comparisons between groups of importing and non-importing firms, applying non-parametric procedures to a sample of Spanish manufacturing firms. Our results indicate the existence of large and significant differences in productivity between importing and non-importing firms.

Keywords: Total factor productivity, imports.
JEL codes: D24; F10; M20

José Carlos Fariñas
Departamento de Economía Aplicada II
Facultad de Ciencias Económicas
Universidad Complutense
28223 Madrid
Spain
Tfno: 91 394 2642; Fax: 91 394 24 57
Mail: farinas@ccee.ucm.es

Ana Martín Marcos
UNED
C/ Senda del Rey 11
28040 Madrid
Spain
Tfno: 91 3986364
Mail: amartin@cee.uned.es

1. Introduction

According to models of industry dynamics proposed, among others, by Jovanovic (1982), Hopenhayn (1992), Ericson and Pakes (1995), the path of entry, growth and failure that characterize micro data is driven to a large extent by firm productivity differences. These models allow for industry equilibria with simultaneous flows of firm entry and exit and productivity heterogeneity at the firm level. The empirical side of this literature has been nicely reviewed by Caves (1998), Bertelsman and Doms (2000), Foster, Haltiwanger and Krizan (2001), among others.

Research on the relationship between opening to foreign markets and firm productivity is an issue that has received more attention in productivity studies using longitudinal micro-level data sets (Bartelsman and Doms, 2000). According to models of industry dynamics, patterns of entry in the export market must reflect productivity differences at the firm level. As a non-exporter must incur a sunk entry cost to enter the export market, entering exporters should have higher initial productivity relative to firms that remain outside the export market. This self-selection mechanism implies that differences between exporting and non-exporting firms precede their entry in the export market. Therefore, before entry takes place, the productivity distribution of exporters dominates the distribution of non-exporters.

The theoretical model of Melitz (2003) integrates in an open economy context the basic ingredients of industry dynamics models of firm heterogeneity, which permits to evaluate the impact of trade. In Melitz's model the assumption of sunk entry costs is crucial to predict the self-selection of most productive firms into the export market as well as the reallocation of resources within an industry following the exposure to trade. After opening to trade only the most productive firms will be able to generate the amount of revenues sufficient to amortize the sunk entry cost of exporting. Furthermore, the opportunities generated by exporting force the least productive firms to close down.

Other strand of the literature interested in the relationship between trade and firm productivity combine the intraindustry heterogeneity setting of Melitz (2003) with a property right plus incomplete contracting approach of the Grossman and Hart (1986) type. The model by Antràs and Helpman (2004) is an example of this literature, which provides a set of predictions concerning the relationship between imports and firm productivity. In this framework companies make endogenous organizational choices and this permits to address two firm decisions: the choice between integration and outsourcing and the choice between domestic and foreign outsourcing. One of the key results coming out of this model is that high-productivity firms outsource in foreign markets or arm's length trade (importing firms) and low-productivity firms do not outsource abroad (non-importing firms). The reason is that low-productivity firms that outsource at home are too far from productivity levels that make foreign sourcing profitable. Fixed organizational costs associated to search, monitoring, and communication are significantly higher in the foreign country and this requires higher levels of productivity to make this option profitable.

When integration in foreign markets is an optimal strategy, firms engage in standard vertical integration through foreign direct investment (DFI) and intrafirm trade is observed. In this case, Antràs and Helpman (2004) predicts that high productivity firms integrate through vertical FDI and low-productivity firms outsource either at home or abroad.

This paper makes a first exploration on the relationship between imports and firm productivity. We address this issue empirically, taking as reference a sample of Spanish manufacturing firms, and testing the prediction of Antràs and Helpman (2004) model that high-productivity firms engage in arm's length trade (importing firms) while low-productivity firms do not outsource from abroad (non-importing firms).

The paper uses the methodology proposed by Delgado, Farinas and Ruano (2002) and Farinas and Ruano (2005), which permits the comparison of the entire distribution of firm productivity rather than just marginal moments, typically means. In particular, we are interested in comparing the cumulative distribution function of total factor productivity for the group of importing firms and the group of non-importing firms. The paper implements a testing procedure based on the concept of stochastic dominance for ranking differences between productivity distributions.

Our results indicate the existence of large and significant differences in productivity between importing and non-importing firms. These differences are consistent with the Antràs and Helpman (2004) prediction that importing firms have higher total factor productivity than non-importing firms.

The paper is organized as follows. Section 2 describes the testing procedure that has been implemented. Section 3 presents the measurement of firm productivity. Section 4 presents the main empirical results. Finally, section 5 provides some concluding comments on directions of future work.

2. Testing procedure

The empirical implications that has to be tested can be performed comparing productivity distributions of different groups of firms. In this section we describe a procedure for testing differences between distribution functions. The procedure is based in non-parametric techniques, which has been previously used by Delgado, Fariñas and Ruano (2002); Fariñas and Ruano (2005). See both references for more details.

To perform the comparison between two productivity distributions, the procedure we use relies on the notion of first order stochastic dominance. Let define two cumulative distribution functions such that the first one corresponds to the productivity distribution of firms outsourcing abroad (importing firms), that we denote by F, and the second one corresponds to the group of firm not outsourcing abroad (non-importing firms), denoted by G. According to the predictions of Antràs and Helpman (2004) model, the productivity distribution of importing firms, F, should dominate the distribution of non-importing firms, G. Stochastic dominance of F relative to G requires two statistical conditions to be satisfied: first, both distributions are not identical, i.e. the null hypothesis H_0: $F(z)-G(z) = 0$

can be rejected; second, the sign of the difference is as expected, i.e. the null hypothesis H_0: $F(z)-G(z) \leq 0$ cannot be rejected. These two-sided and one-sided tests can be performed respectively by the following Kolmogorov-Smirnov test statistics:

$$\delta_N = \sqrt{\frac{n.m}{n+m}} \sup_{z \in Z} \left| F_n(z) - G_m(z) \right|$$

$$\eta_N = \sqrt{\frac{n.m}{n+m}} \sup_{z \in Z} \left(F_n(z) - G_m(z) \right),$$

where n and m are respectively the size of the sample of firms corresponding to the distribution F, the group of importing firms, and m is the size of the sample draw from the distribution G. The limiting distributions of both test statistics are known under independence between the sample of firms drawn from distributions F and G (see Delgado, Farinas and Ruano, 2002).

To further illustrate the comparisons between different groups of firms, we have graphed the estimates of the distribution functions for these groups. In particular, we have computed the smooth, or perturbed, sample distribution function rather than the sample distribution function itself, which provides nice smooth distribution estimates. Since the purpose here is to produce graphical representations of the differences between two groups of firms, we represent these distributions for the whole population of firms.

Next, we describe the approach that has been followed for the estimation of a univariate cumulative distribution function, F, based on a sample $Z_1,...Z_N$ of size N, where the sample is a combination composed of two random samples of small firms and large firms of sizes N_S and N_L ($N_S+N_L= N$), respectively. Given the characteristics the data-set, which we explain in the next section, we distinguish between large and small firms. For a given group of firms (say importing firms), the cumulative distribution function for the whole population, F(.), can be defined in terms of the conditional cumulative distribution functions for the two size groups, $F(.\,|\,\tau)$, where τ is a dummy variable equal to 0 for small firms and equal to 1 for large firms. F(.) can be expressed as

$$F(.) = p \times F(.\,|\,\tau = 0) + (1 - p) \times F(.\,|\,\tau = 1),$$

where p represents the probability of being a small firm in the group of importing firms. Therefore, the cumulative distribution function of the whole population of firms is a mixture of the conditional cumulative distribution functions corresponding to the two size groups of firms, where the parameter of the mixture is the probability of being a small firm in the corresponding population group. Then, the univariate cumulative distribution function for the whole population of firms can be estimated as a weighted average of some estimators of the cumulative distribution functions corresponding to both size groups. The weighted kernel distribution estimate, $\hat{F}_h$, of a univariate cumulative distribution function for the whole population of firms, F, can be expressed as:

$$\hat{F}_h(z_0) = \hat{p} \int_{-\infty}^{z_0} \left(\frac{1}{N_S} \sum_{i=1}^{N} \cdot K\left(\frac{z - Z_i}{h} \right) \right) dz + (1 - \hat{p}) \int_{-\infty}^{z_0} \left(\frac{1}{N_L} \sum_{i=1}^{N} \cdot K\left(\frac{z - Z_i}{h} \right) \right) dz \qquad [1]$$

where h is the bandwidth and $K(.)$ is the kernel function, and $\hat{p}$ represents the estimated probability of being a small firm in the considered group. Expression [1] can be rewritten as:

$$\hat{F}_h(z) = \int_{-\infty}^{z} \left(\sum_{i=1}^{N} \omega_i K\left(\frac{z - Z_i}{h} \right) \right) dz; \qquad [2]$$

where weights are given by

$$\omega_i = (1 - \tau_i) \frac{\hat{p}}{N_S} + \tau_i \frac{(1 - \hat{p})}{N_L}$$

satisfying that $\sum_{i=1}^{N} \omega_i = 1$, where τ_i a dummy variable, which is equal to 0 if firm i is small and equal to 1 otherwise[1].

3. Productivity measurement

The dataset we employ is a longitudinal survey of Spanish manufacturing firms that comes from the Encuesta sobre Estrategias Empresariales (ESEE). This dataset is collected by the Fundación Empresa Pública and sponsored by the Spanish Ministry of Industry. This database contains a longitudinal sample of firms from 1990 to 2002. Over the period 1990-2002, the panel of firms contains 21,098 observations that correspond to an average number of 3,462 firms per year.

A characteristic of the dataset is that firms participating in the survey were chosen according a selective sampling scheme. The sample of firms includes almost the universe of Spanish manufacturing firms with more than 200 employees (large firms) in 1990. In fact, the rate of participation of firms in this size category was the 67.6% of the population of large firms in 1990. Firms employing between 10 and 200 employees (small firms) were chosen according to a stratified random sampling procedure. The 3.9% of the population of firms within this size category were randomly sampled in 1990. Given the procedure used to select firms participating in the survey, both samples of small and large firms can be considered as samples that permit to estimate the distribution of any of the characteristics of the population of Spanish manufacturing firms with available information in the data set.

Firm productivity is defined by an index of total factor productivity for each firm over the period 1990-2002. The index proposed follows the framework developed by Aw, Cheng and Roberts (1997) and it is an extension of the multilateral total factor productivity index

[1] The estimation of marginal probabilities for the population of firms takes into account the sampling proportions of the data set. For the group of nonimporting firmd, the estimated probability of being smal is $\hat{p} = 0.995$ and the probability of being large is $(1-\hat{p}) = 0.005$. For the group of importing firms, the estimated probability of being small is $\hat{p} = 0.934$ and the probability of being large is $(1-\hat{p}) = 0.066$.

proposed by Caves, Christensen and Diewert (1982). The index takes as reference a hypothetical firm and measures productivity in each year relative to this reference firm. In particular, the index uses as the reference point the average firm of the industry and the size group the firm belongs to, and then chain-links the average firm for both size groups to preserve transitiveness between firms of different size groups within the same industry. A similar extension of the index can be found in Good, Nadiri and Sickless (1996).

Let each firm *f* produce a single output *y* using the set of inputs *x*, then the expression of total factor productivity for firm *f*, at time *t*, in a given industry is:

$$\begin{aligned} \ln \lambda_{ft(\tau)} = \; & \ln y_{ft} - \overline{\ln y_\tau} - \frac{1}{2}\sum_{r=1}^{R}(\varpi_{ft}^r + \overline{\varpi_\tau^r})(\ln x_{ft}^r - \overline{\ln x_\tau^r}) + \\ & + \overline{\ln y_\tau} - \overline{\ln y} - \frac{1}{2}\sum_{r=1}^{R}(\overline{\varpi_\tau^r} + \overline{\varpi^r})(\overline{\ln x_\tau^r} - \overline{\ln x^r}); \end{aligned} \quad [3]$$

where y_{ft} is the output of firm *f* at time *t*, x_{ft}^r is the quantity of input r=1…3 corresponding to firm *f* at time *t*, and ϖ_{ft}^r is the cost share of input *r*. Firms are classified in two size groups of small and large firms, $\tau = 0,1$. A bar over a variable indicates the arithmetic mean of the variable. The average value of variables with index τ, refers to a given size group of firms, otherwise the average refers to the entire sample of small and large firms.

The index $\ln \lambda_{ft}$ measures the proportional difference of total factor productivity for firm *f* at time *t* relative to a given reference firm. The reference firm varies across industries and therefore when observations of different industries are pooled, average productivity differences across industries are removed. Firms are classified in twenty manufacturing industries corresponding to the NACE classification at the two-digit level. To be more precise about the meaning of expression [3], the first line of the right hand side compares the productivity of firm *f* with the productivity of an average firm of the same size group and industry as firm *f*. Therefore, comparisons between observations corresponding to the same size group are transitive. The second set of terms measures productivity differences between the reference firm for either the size group of small or large firms, and the average firm of the entire sample of small and large firms in a given industry.

The information used to calculate the index of total factor productivity for each firm is drawn from the ESEE. The output y_{ft} is measured by the annual value of gross production of goods and services expressed in real terms using price indexes for each firm reported by the ESEE. The estimation of the index considers three inputs: labor, materials and the stock of capital. Labor inputs are measured by the number of effective hours of work per year, which is equal to normal hours plus overtime hours minus non-working hours. Material inputs are measured by the cost of intermediate inputs, including raw materials purchases, energy and fuel costs and other services paid by the firm. The value of material inputs is measured in real terms using individual price indexes on the three categories of intermediate inputs for each firm reported by the ESEE. The stock of capital is calculated according to the perpetual inventory formula for each firm:

$$k_{ft} = I_{ft} + k_{ft-1}(1 - d_{ft})\frac{P_t}{P_{t-1}},$$

where I_{ft} corresponds to the value of investment in equipment of firm f at time t, d_{ft} stands for depreciation rates and P_t is an aggregate price index for equipment investment published by the Spanish Institute of Statistics. Finally, input cost shares, w_{ft}^{r}, are defined as the fraction of the cost of each input in total input costs. Total input costs is defined by the sum of labor costs, intermediate input costs and the cost of capital. The cost of labor is measured by the sum of wages, social security contributions, and other labor costs paid by the firm. The cost of intermediate inputs is measured by the sum of costs of raw materials purchases, energy and fuel costs and other services paid by the firm. The user cost of capital is measured for each firm by the cost of long-term external debt of each firm as reported by the ESEE plus the depreciation rate, d_{ft}, minus the variation of the aggregate price index for capital goods.

4. Results

In this section we present the results on the productivity differences between the group of importing and non-importing firms. The dataset we use to test for these differences correspond to a sample of Spanish manufacturing firms over the period 1990-2002.

Figure 1 illustrates the difference between the productivity of importing and non-importing firms. The graphs show the kernel estimates of the cumulative distribution functions for both groups of firms over the period 1990-2002. These functions have been estimated using expression [2], which weights the cumulative distribution functions of both the group of small and large firms. The position of the productivity distribution of importing firms is, in all years, to the right of the distribution of non-importing firms. This position indicates that for any quantile of the distribution, importing firms have higher productivity than non-importing firms and therefore the productivity distribution of importing firms stochastically dominates the distribution of non-importing firms.

Given the assessed differences, we apply the procedure defined in section 2 to formally test the productivity differences depicted in Figure 1 between importing and non-importing firms. Given the characteristics of the data-set, the testing procedure is applied separately to the group of small and large firms for each time period. One- and two-sided tests, as described in section 2, are applied to compare the two groups of firms separately in the sample of small and large firms:

$$F_t(.\,|\,\tau=\tau_0) \quad vs \quad G_t(.\,|\,\tau=\tau_0), \quad t=1990,...2002 \quad and \quad \tau_0=0,1,$$

where τ identifies both the sample of small and large firms. F_t denotes the productivity level ($ln\ \lambda_{ft}$) distribution of importing firms and G_t denotes the distribution of non-importing firms.

Table 1 summarizes the hypothesis test statistics. First, within the sample of small firms, the null hypothesis of equality between the distributions of importing and non-importing firms can be rejected for the complete set of years. The hypothesis that the sign of the difference is as expected (higher productivity for importing firms) cannot be rejected at any level significance level. Second, for the group of large firms, the null hypothesis of equality between both distributions can be rejected at the 0.1 or lower levels, excluding years 1990, 1991 and 2002. The hypothesis that the productivity of importing firms is greater than the productivity of non-importing firms cannot be rejected at usual significance levels with the exception of year 1990. Overall, the evidence presented indicates that the productivity of importing firms is above the productivity level of non-importing firms.

5. Concluding remarks

The picture that emerges from the evidence reported in section 4 is that importing firms have higher productivity than non-importing firms. This evidence is consistent with the model of Antras and Helpman (2004) predicting that high-productivity firms outsource intermediate inputs in international markets whereas low-productivity firms acquire them at home. The preliminary evidence that is provided in this paper deserves additional research effort at least in three directions. First, the model we rely on to organize our empirical work is a framework where the firm decides about the integration within their boundaries vs. domestic/foreign outsourcing of intermediate inputs. Therefore, from a testing point of view, the group of importing firms has to include firms that engage in foreign outsourcing of intermediate inputs. Most of the firms included as importers in the empirical application reported in section 4, are firms importing intermediate inputs. However, we have to confirm this condition filtering firms that do not fulfill the criteria of being importers of intermediate inputs.

Second, according to Antras and Helpman (2004) model the prediction that links foreign/domestic outsourcing to firm productivity corresponds to component-intensive industries. Therefore, an alternative route to test the model would be to choose industries that can be characterized as component-intensive. The strategy would be to apply the testing procedure defined in section 2 to groups of importing and non-importing firms within these industries

Third, Antras and Helpman (2004) model permits to test additional predictions regarding vertical integration of firms through foreign direct investment and intrafirm trade. Therefore, the testing procedure can also be applied to the ranking of productivities of integrated firms vs. outsourcing in domestic/foreign firms.

References:

Antràs, Pol and Elhanan Helpman (2004) "Global sourcing", *Journal of Political Economy*, vol. 112, no. 3, pp 552-580.

Aw, Bee Y., Xiaomin Chen and Mark J. Roberts (1997): Firm level evidence on productivity differentials, turnover and exports in Taiwanese manufacturing", *NBER working paper* 6235.

Bartelsman, Eric J., and Mark Doms (2000) "Understanding productivity: lessons from longitudinal microdata", *Journal of Economic Literature*, 38, pp. 569-594.

Caves, Richard. (1998), "Industrial organization and new findings on the turnover and mobility of firms", *Journal of Economic Literature*, vol XXXVI, pp. 1947-1982.

Caves, Douglas W., Laurits R. Chistensen and W. Erwin Diewert (1982a): "Multilateral comparisons of output, input and productivity using superlative index numbers", *Economic Journal*, vol 92, pags. 73-86.

Delgado, Miguel, Jose C. Fariñas and Sonia Ruano (2002), "Firm productivity and export markets: a non-parametric approach", *Journal of International Economics,* 57, pp. 397-422.

Ericson, Richard and Ariel Pakes (1995), ``Markov-perfect industry dynamics: A framework for empirical work'', *Review of Economic Studies*, 62, pp. 53-82.

Fariñas, Jose C. and Sonia Ruano (2005) "Firm productivity, heterogeneity, sunk costs and market selection", *International Journal of Industrial Organization*, 23 pp. 505-534

Foster, Lucia, John Haltiwanger and C.J. Krizan (2001), "Aggregate productivity growth: lessons from microeconomic evidence" in *New developments in productivity analysis*, editors: E. Dean, M. Harper and Ch. Hulten, Chicago University Press.

Good, David H., M. Ishaq Nadiri and Robin C. Sickles (1996): "Index numbers and factor demand approaches to the estimation of productivity", *NBER Working Paper* 5790.

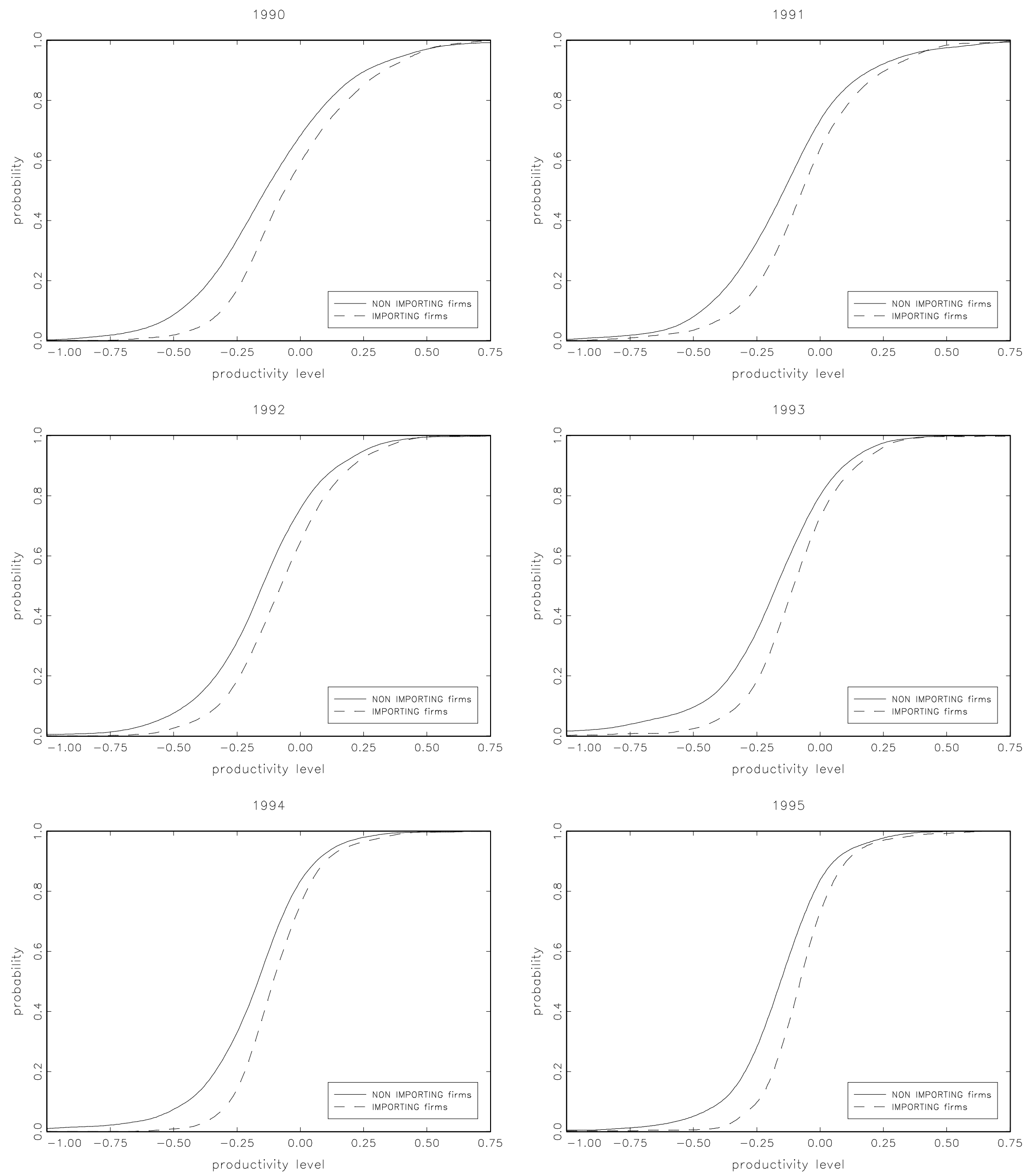
Grossman, Sanford and Oliver Hart (1986) "The costs and benefits of ownership: a theory of vertical and lateral integration", *Journal of Political Economy*, vol. 94 (August), pp. 691-719.

Hopenhayn, Hugo (1992), "Entry, Exit, and firm dynamics in long run equilibrium"', *Econometrica*, 60, September, pp.1127-1150.

Jovanovic, Boyan (1982), ``Selection and the evolution of industry'', *Econometrica*, 50, pp. 649-670.

Melitz, Marc J. (2003), “The impact of trade on intra-industry reallocations and aggregate industry productivity”, *Econometrica*, vol 71, (November), pp. 1695-1725.

Figure 1
Productivity differences of importing firms versus non-importing firms
(Smooth sample distribution function)



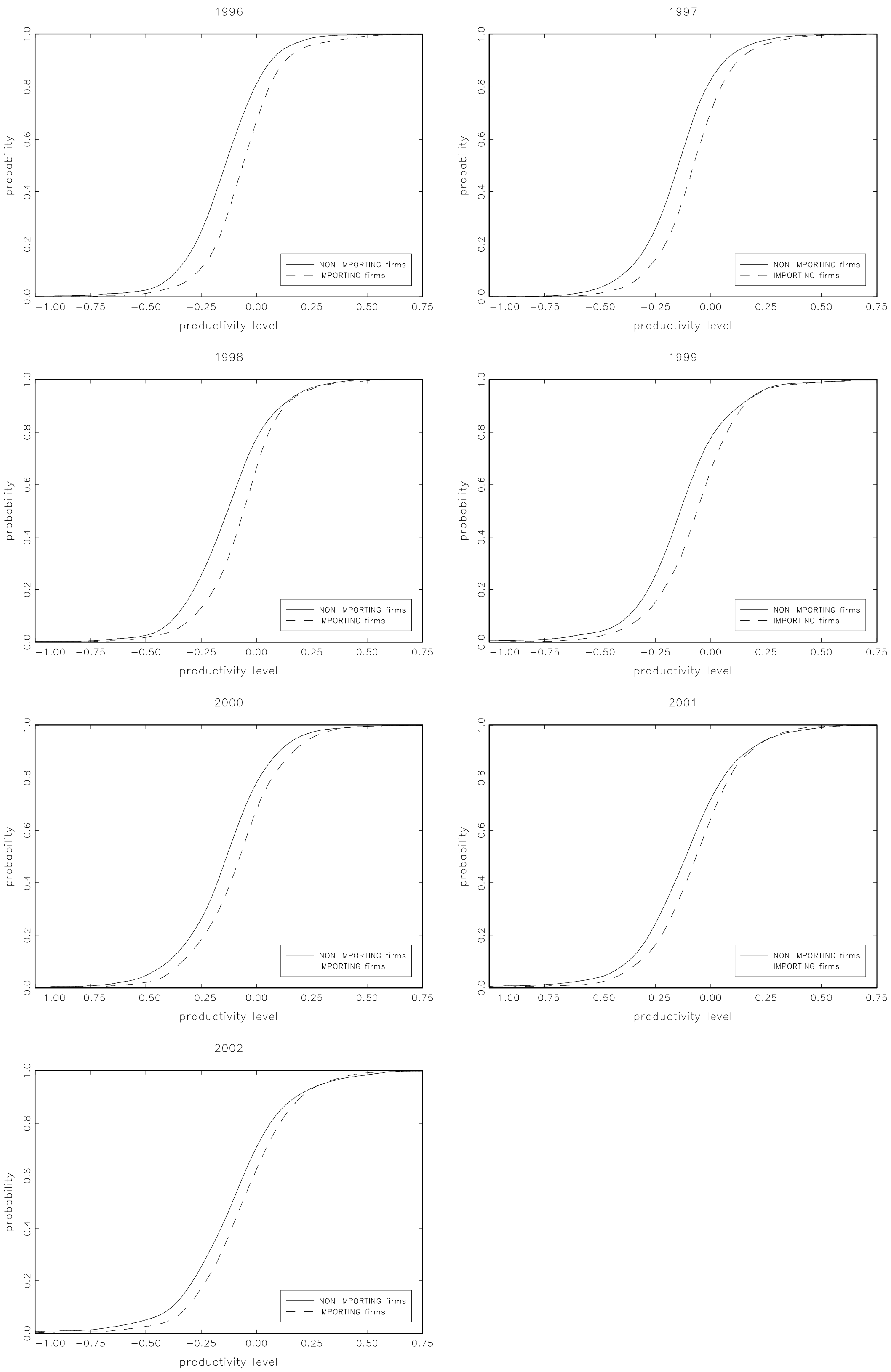

1996
1997
1998
1999
2000
2001
2002
probability
productivity level
NON IMPORTING firms
IMPORTING firms
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75
0.0 0.2 0.4 0.6 0.8 1.0

Table 1
Productivity level differences between importing firms and non-importing firms: hypotheses test statistics

	Small importing firms vs. small non-importing firms						Large importing firms vs. large non-importing firms					
	Number of observations		Equity of distributions		Differences favorable to importing firms		Number of observations		Equity of distributions		Differences favorable to importing firms	
Year	Importi ng	Non - importing	Statistics	P-value	Statistics	P-value	Importing	Non- importing	Statistics	P-value	Statistics	P-value
1990	287	543	2.621	0.000	0.151	0.955	409	45	1.034	0.236	1.034	0.118
1991	331	691	2.428	0.000	0.211	0.915	509	65	0.896	0.398	0.075	0.989
1992	392	750	2.555	0.000	0.048	0.995	552	69	1.291	0.071	0.426	0.696
1993	390	706	3.019	0.000	0.031	0.998	498	63	1.673	0.007	0.427	0.695
1994	420	649	3.396	0.000	0.049	0.995	551	54	1.282	0.075	0.107	0.977
1995	412	574	3.669	0.000	0.000	1.000	521	49	1.198	0.113	0.11	0.976
1996	444	592	3.543	0.000	0.027	0.999	501	41	1.851	0.002	0.202	0.922
1997	592	644	3.463	0.000	0.027	0.999	497	36	1.639	0.009	0.068	0.991
1998	567	595	3.368	0.000	0.131	0.966	485	29	1.632	0.010	0.062	0.992
1999	604	594	3.592	0.000	0.191	0.929	463	26	1.234	0.095	0.661	0.418
2000	595	590	2.629	0.000	0.06	0.993	530	34	1.545	0.017	0.26	0.874
2001	538	538	1.799	0.003	0.213	0.913	483	30	1.208	0.108	0.458	0.658
2002	539	509	1.756	0.004	0.228	0.901	442	26	0.46	0.984	0.258	0.875