Tech on the ROC:
A New Way of Looking at Exporting Firms

Stefano Costa*, Federico Sallusti*, Claudio Vicarelli*, Davide Zurlo*

Abstract

Policies aimed at increasing firm participation in international markets have been playing an increasing role. Using a new approach to estimate export threshold for manufacturing firms, and considering the technology adoption, this paper analyses the potential mismatch between the conditions required for a firm to become exporter and the pattern of technology in the industry. The export threshold – which is estimated on the basis of the ROC methodology – is the minimum combination of productivity and "economic size" (a broad measure of firm size composed of employment, age, turnover and capital intensity) that firms need to achieve in order to access international markets. In turn, the technology prevailing in each industry is expressed in terms of the relative weights of productivity and size corresponding to a (firm-level) technology level higher than the average within the industry. The interaction between this "technology line" and the export threshold allows for deriving a new firm-based taxonomy that can be useful to study exporting and non-exporting firms in the light of their position with respect to the technology prevailing in the given industry, allowing to have a more efficient selection of policy targets (e.g. intensive or extensive margins).

JEL code: F14, L60, L11, O14

Keywords ROC analysis, export threshold, technology adoption, extensive margin of exports

***** Italian National Insitute of Statistics. We are grateful to Matteo Bugamelli, Luca De Benedictis, Sergio De Nardis, Andrea Linarello, Fabrizio Traù and the participants in the 25/9/2019 Seminar of ISTAT Thematic Research Laboratories for their helpful observations. The opinions expressed in this work are those of the authors and do not involve the responsibility of the National Institute of Statistics.

1. Introduction

Export activity is important for firm competitiveness and, more in general, for the economic growth of countries. As a consequence, policies aimed at increasing firm participation in international markets, both in terms of intensive and extensive margins, have been playing an increasing role. This in turn highlights the importance of being able to detect the firm-level determinants of export, i.e. the minimum requirements firms have to bear to become an exporter.

In a previous paper (Costa *et al.*, 2019), we applied the Receiver Operating Characteristics (ROC) analysis to develop a new methodology for the estimation of the "export threshold", i.e. the combination of productivity and "economic size" (a broad measure of firm size composed of employment, age, turnover and capital intensity) corresponding to the transition from non-exporter to exporter status. This enables us to position each firm according to its distance from the threshold.

In this paper, we enrich that analysis by explicitly taking into account firms technological level. In particular, we analyse the potential mismatch between the conditions required for a firm to become exporter and the pattern of technology adoption in the industry. We do so by assessing the positioning of firms across the export threshold and a "technology line". This permits to design a map of the business system that is particularly useful from a policy-making point of view, as it allows for more targeted policies aimed at boosting firm participation to foreign markets.

The possibility of a mismatch between the sorting of firms in terms of export premia and technological intensity has been widely studied in recent literature related to firm heterogeneity. Among the most influential works, Bustos (2011) developed a model with heterogeneous firms where exporting firms can upgrade their technology after entering foreign markets, so that productivity differences identify a sorting of firms in three groups: high-technology exporters, low-technology exporters and low-technology domestics. It follows that some exporting firms (i.e. new exporters but also firms that were already exporting) are not more technology intensive than non-exporters, even though they can upgrade their technology faster once they enter the export markets and/or variable trade costs fall (i.e. due to a fall in tariffs). In a similar vein, using data on the Canadian business system, Lileeva and Trefler (2010) show that heterogeneity in firms' investing choices can affect the productivity-export relationship *via* technology adoption. Analogous results are found for U.S. (Bernard *et al.*, 2003), Spain (Delgado *et al*., 2002), and Germany (Bertschek *et al.*, 2016).

The empirical literature has largely analyzed the relationship between export, productivity and size (see Wagner, 2012 and ISGEP, 2008 for detailed surveys). The existence of some export thresholds characterizes all the theoretical works on firm heterogeneity originating from the seminal paper of Melitz (2003), where only firms above a minimum productivity level are able to sell abroad (Melitz and Ottaviano, 2008; Chaney, 2008; Bernard *et al*., 2011). However, from the empirical point of view, several works have showed that in many countries, firm productivity distributions of exporters and non-exporters may overlap, implying that enterprises might not export even though their productivity levels would enable them to (i.e. are above the productivity threshold).[1] Moreover, others (Schröder and Sørensen, 2012; Geishecker *et al*., 2017) have shown that the mismatch between Melitz's theory and empirical evidence is actually linked to the definition of productivity: empirical works are forced to use average cost-based productivity measures, while theoretical ones rank firms according to their marginal productivity.

Also firm size is relevant to explain the ability to export, because it loosens the constraint represented by sunk costs. Indeed, empirical studies did find a direct relationship between export and size: exporters tend

[1] See, among others, Castellani and Zanfei (2007) for the Italian case and Schröder and Sørensen (2012) for a survey.

to be larger than non-exporters (Bernard and Jensen, 1995; Wagner, 2007; Máñez-Castillejo *et al.*, 2010). This raises important questions about the sources of export premia and, more specifically, whether, and to what extent, such sources could be size-related. Internal sources include managerial talent, quality of inputs, information technology, R&D, learning by doing, and innovation (Syverson, 2011): small and large firms could differ in terms of access to these sources (Leung *et al*., 2008). External factors such as regulation and access to financing could also be responsible for heterogeneity between small and large firms (Tybout, 2000).

The rest of the paper is organized as follows. Section 2 presents a description of the dataset and empirical strategy. Section 3 illustrates the ROC methodology for the estimate of the export threshold. Section 4 introduces the technology line. Section 5 show the new taxonomy obtained from the interaction between the export threshold and the technology line. Section 6 concludes.

2. Data

The main statistical source of this work is the business register "Frame-Sbs" for 2016. Released by ISTAT since 2011, it annually provides administrative-based information on the structure (e.g. number of employees, business sector, location, age, belonging to a group) and the main Profit and Loss Account variables (e.g. value of production, turnover, value added, labour cost) for the whole population of about 4.4 million of Italian firms.

This database is then integrated with other information drawn from Custom Trade Statistics, a census dataset reporting, for each Italian firm, the values of imports, exports, and trade balance with both EU (intra-EU trade) and non-EU operators (extra-EU trade).

In order to focus on relevant business units, some restrictions are imposed to the dataset. In particular, in the light of the extremely fragmented structure of the Italian business system – where in 2016 the firm average size was less than 4 workers, and the enterprises with just one worker accounted for over 50% of total firms and 12% of total employment – we exclude units which do not have "economic relevance" for the analysis of export strategies. Consequently, we consider firms that have positive value added, no less than 1 employee, and positive consumption of fixed capital. Moreover, we only retain firms operating in manufacturing (excluding Tobacco, Refined petroleum products, Maintenance and repair, and Other manufacturing),[2] which in 2016 accounted for 83% the total Italian exports. The final dataset includes 208,627 firms, accounting for about 54% of manufacturing firms, 85% of workforce, 93% of value added, 84% of exports. Table 1 reports industry composition and main information about the strata of analysis.

[2] The exclusion of Tobacco and Refined petroleum products is connected with the peculiar characteristics of these activities (regulation and monopoly). Maintenance and repair has been excluded because of its high content of services. Other manufacturing has been excluded because it includes miscellaneous activities (see NACE Rev. 2 Classification).

Table 1. The sample: Industry classification and firms' characteristics

Industry	Nace Rev.2 code included	Number of firms	Share of firms	Share of value added	Share of employees	Share of exports
Food and beverage	10, 11	39,356	18.9	12.1	12.9	7.9
Textile	13	8,274	4.0	2.8	3.4	2.6
Wearing apparel	14	11,957	5.7	3.3	4.8	4.1
Leather	15	8,634	4.1	3.3	4.0	5.1
Wood	16	15,410	7.4	1.7	2.8	0.5
Paper and print	17, 18	12,927	6.2	4.4	4.7	2.3
Chemicals and pharmaceutics	20, 21	3,679	1.8	9.6	5.2	13.3
Rubber and plastic	22	7,732	3.7	5.6	5.4	5.0
Non metalic minerals	23	11,766	5.6	4.3	4.6	2.8
Metals	24, 25	46,319	22.2	16.5	18.6	13.6
Electronics	26, 27	9,082	4.4	7.8	7.3	7.8
Machinery	28	18,429	8.8	16.3	14.5	20.4
Automotive	29, 30	3,269	1.6	9.5	8.1	12.0
Furniture	31	11,793	5.7	2.8	3.9	2.5
Total		208,627	100.0	100.0	100.0	100.0

Source: Authors' calculation on ISTAT data.

3. ROC methodology and export threshold

3.1. The basics of the ROC analysis

Following the methodology developed in our previous work (Costa et al., 2019), we estimate the export threshold on the basis of the joint application of the Receiver Operating Characteristics (ROC hereinafter) analysis and Youden's (1950) *J* index. This permits the identification of a cut-off point over an independent variable in a logit model (in our case: a combination of productivity and economic size), so as to efficiently cluster observations with respect to a dependent binomial variable (in our case: the exporter status).

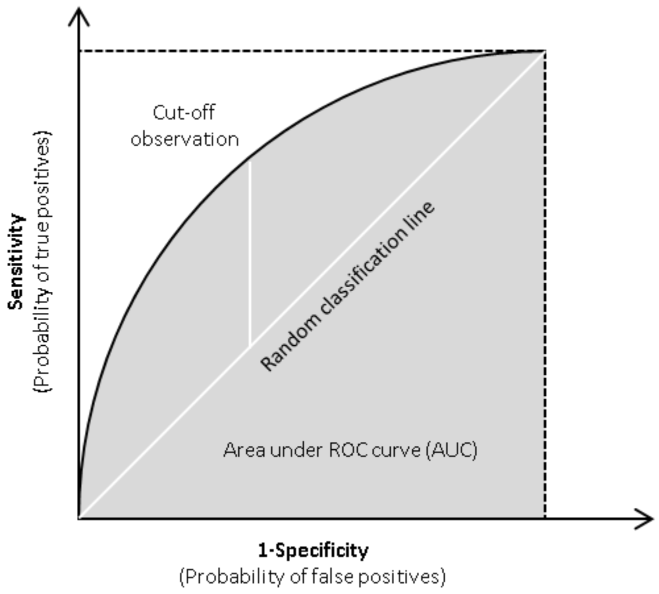
The application of the ROC analysis is quite new in Economics. To the best of our knowledge, so far this methodology has been used to test the accuracy of business cycle classification made by the Business Cycle Dating Committee of the National Bureau of Economic Research (Berge and Jorda, 2011) and in credit risk literature (Khandani *et al*., 2010). However, it has been widely adopted in medicine (Lusted, 1960), and it is now a common standard of evaluation of medical and psychological tests (Pepe, 2003). Furthermore, ROC methodology is used in machine learning (Majnik and Bosnić, 2013), and natural science (Warnock and Peck, 2010).

According to Fawcett (2005), classification models (or classifiers) can give four possible outcomes: True positive (TP), False positive (FP), True negative (TN), False negative (FN).

The validity of a classifier can be measured based on two main metrics: Sensitivity and Specificity. Sensitivity represents the probability of detecting true positives. Specificity is the probability of detecting true negatives. This latter is usually considered in its reciprocal expression (1–Specificity), which measures the probability of false positives.

Once a classifier is applied, the ROC curve displays the position of each observation in the space of Sensitivity and 1–Specificity (Figure 1), showing the tradeoff between the probability of detecting true positives or false positives across all possible cut-off points (Kumar and Indrayan, 2011).

Figure 1. The ROC curve



The area under the ROC curve (AUC, grey portion in Figure 1) provides a measure of the extent to which the clustering obtained by the model is more efficient than a pure random classification (the 45° line). In this respect, the AUC criterion is largely used to measure the goodness of fit of logit models, and to define the relative relevance of a set of variables in determining the overall logistic distribution of probability.

In order to single out along the ROC curve the observation that most efficiently discriminates between positives and negatives ($\widehat{\text{Cut}}$), the following equation is used:

$$\widehat{Cut} = h * Sensitivity - (1-h) * (1 - Specificity) \qquad [1]$$

where h and $(1-h)$ represent the relative weights to manage the trade-off between true and false positives. By setting up $h = 0.5$, we opt for a "neutral" selection between the two outcomes.[3] In doing so, Equation [1] turns out to be equal to Youden's (1950) J index:

$$(Sensitivity + Specificity - 1) \qquad [2]$$

Youden's J – which identifies the observation that maximizes equation [2] and, consequently, the vertical distance between ROC curve and the 45° line (see Figure 1) – is the most commonly used criterion for

[3] Values of $h > 0.5$ (i.e., finding true positives is more relevant than avoiding false positives) would correspond to a "liberal" selection, which assigns positive classification even in the presence of weak evidence. Conversely, setting up $h < 0.5$ (i.e., detecting true positives is less relevant than avoiding false positives) would correspond to a "conservative" selection, which assigns positive classifications only in presence of strong evidence.

detecting optimal cut-offs.[4] Moreover, the J index – implying a "neutral" choice between false positives and negatives – is all the more suitable for our purposes because we have no *a priori* bias in dealing with the trade-off.[5]

3.2. Definition of the "export threshold"

As in our previous work (Costa *et al.*, 2019) in order to apply the ROC analysis to the identification of the export threshold, we firstly estimate the probability to export of the i-th firm in the h-th industry based on the following logit model:

$$Prob\ (Export = 1|X)_{h,i} = \Lambda(\alpha X)_{h,i} \qquad [3]$$

where Λ is the cumulative distribution of the logistic function, α is the estimated parameter, and X is the covariate.

Once estimates have been obtained, we use Youden's J to identify the cut-off observation in the h-th industry, thus also determining the value of the covariate representing the threshold:

$$X_h^e = X_{h,c} \qquad [4]$$

where c is the cut-off firm.

Using this threshold, each firm can be classified as exporter or non-exporter according to its laying above or under this threshold.

In particular, we use a composite model (Z-model, where $\mathrm{X}_h^e = \mathrm{Z}_h^e$), where the export threshold is defined over a combination (Z_h^e) of productivity and economic size.[6]

The Z indicator is derived from a three-step procedure. In the first step, for each industry, the economic size is defined, using factor analysis over a set of four variables: number of workers; turnover; consumption of fixed capital; age (in terms of number of months from the date of inclusion in the Italian Business Register). For each firm in a given industry, economic size is thus obtained from the linear combination of the four variables as resulting from the first (rotated) autovector.

In the second step, the following logit model is estimated for the h-th industry:

[4] Beside the *J* index, two other criteria are used to find optimal threshold point along a ROC curve: *a)* the minimization of the distance from the (0,1) point; *b)* the cost minimization, which considers several types of costs, e.g. for correct and false classification, for further investigation etc., but it is rarely used due to its assessment difficulty.

[5] Actually, the best cut-off depends on whether one needs to maximize sensitivity at the expense of 1-specificity or vice versa. This often happens in medicine. The first case leads to a test that is maximal sensitive (i.e. correctly identifying diseased people at the expense of a high number of false positives). The second case generates a test that is better at ruling out the disease. The Youden's *J* maximizes both.

[6] In Costa et al. (2019), we tested two alternative models: a pure sales model (S-model, where $\mathrm{X} = \mathrm{Sales}$), in which the export threshold is defined over the value of firms' turnover, and a pure productivity model (π-model, where $\mathrm{X} = \mathrm{Productivity}$), in which the export threshold is defined over the value of labour productivity (value added-per-worker). Both S-model and π-model have been proved to be consistent with Melitz's theory (Geishecker *et al.*, 2017). Fitting tests showed that the Z-model outperforms the other two.

$$Prob\ (Exporter_i = 1|S_i, \pi_i, G_i, I_i) = \Lambda(\ \alpha_1 S_i + \ \alpha_2 \pi_i + \alpha_3 G_i + \alpha_4 I_i\ ) \qquad [5]$$

where Λ is the cumulative distribution of the logistic function, α_j are estimated parameters, S is the economic size, π is the productivity (expressed in terms of value added-per-worker and standardized), G is a set of dummy variables indicating the location of firms,[7] and I is a set of dummy variables related to NACE 4-digit levels of economic activity.

In the third step, the estimated coefficients of productivity and economic size from equation [5] are used to obtain, for each industry, the composite indicator $Z^e_{h,i}$ for the i-th firm. In particular, estimated parameters are used as weights, while variables are taken at individual level:

$$Z^e_{h,i} = \ \hat{\alpha}_{1,h} S_{h,i} + \ \hat{\alpha}_{2,h} \pi_{h,i} \qquad [6]$$

where $Z^e_{h,i}$ is the covariate to be used in equation [3].

Following the ROC methodology, the optimal cut-off observation c (the “export threshold firm”) is identified. Finally, substituting the productivity and economic size of c in Equation [6], we obtain the export threshold as:

$$Z^e_{h,c} = \ \hat{\alpha}_{1,h} S_{h,c} + \ \hat{\alpha}_{2,h} \pi_{h,c} \qquad [7]$$

In the rest of the paper we refer to $Z^e_{h,c}$ as Z^e.

3.3. Fitting tests of ROC estimates

Three types of test are carried out. First, we apply the usual Area Under Curve (AUC) test to compare the model based on the composite indicator Z with an alternative, strictly “Melitz-compliant” pure productivity model (π-model), in which the export threshold is defined over the labour productivity, measured in terms of value added per worker ($X = \pi$ in Equation [3]).

Results are reported in Table 2: both π- and Z-model show a high goodness of fit (never below 70% for the π-model, always over 75% for the Z-model). However, the Z-model significantly outperforms the pure productivity one for all strata.

[7] We refer to five geographical areas: North-West, North-East, Centre, South, Islands.

Table 2. Area under ROC curve (AUC): comparison between π-model and Z-model

Industry	AUCs		Z-model - π-model				
	Z-model	π-model	Difference estimate	Standard error	Lower bound	Upper bound	P-value
Food and beverage	0.865	0.849	0.017	0.002	0.014	0.020	0.000
Textile	0.824	0.767	0.058	0.004	0.050	0.065	0.000
Wearing apparel	0.777	0.730	0.047	0.005	0.037	0.056	0.000
Leather	0.756	0.698	0.058	0.005	0.048	0.067	0.000
Wood	0.831	0.753	0.078	0.005	0.069	0.087	0.000
Paper and print	0.843	0.785	0.058	0.003	0.051	0.064	0.000
Chemicals and pharmac	0.787	0.741	0.046	0.008	0.030	0.063	0.000
Rubber and plastic	0.818	0.742	0.076	0.005	0.066	0.085	0.000
Non metalic minerals	0.769	0.732	0.037	0.004	0.030	0.044	0.000
Metals	0.850	0.772	0.079	0.002	0.074	0.083	0.000
Electronics	0.786	0.718	0.068	0.005	0.058	0.079	0.000
Machinery	0.778	0.700	0.078	0.004	0.070	0.085	0.000
Automotive	0.790	0.724	0.066	0.008	0.050	0.083	0.000
Furniture	0.833	0.734	0.099	0.004	0.091	0.108	0.000

Source: Authors' calculation on ISTAT data.

Second, we consider the capability of the cut-off identified by the J index of classifying firms as exporters and non-exporters in terms of Precision and Accuracy. In particular, Precision measures the share of true positives over the total number of observations the model classifies as positives (i.e. the percentage of firms correctly classified as exporters):

$$Precision = \frac{TP}{TP+FP} \qquad [8]$$

In turn, Accuracy measures the share of true positive and negative outcomes of the model (i.e. the proportion of firms correctly classified as exporters and non-exporters) over the total number of observations:

$$Accuracy = \frac{TP+TN}{Total\ observations} \qquad [9]$$

On such bases, we assess the capability of our model in detecting the bulk of Italian exporters by calculating the weight of true positive observations in terms of total exports.

The results of these latter tests are reported in Table 3. Our model shows a high capability of correctly clustering exporters: in 9 out of 14 industries, the Precision (column 2) is over 60% (over 80% in four industries). With regard to correct and wrong classifications (columns 3 to 5), the model shows a good performance in detecting true positives (i.e. in correctly classifying exporters), so discharging clustering errors on false negatives (i.e. firms that the model classifies as non-exporters despite they actually sell abroad some of their products). The last column confirms that our clustering method grasps an extremely large share of total exports (over 97% in all industries), suggesting that false negatives (which largely bear the bias of the model) are negligible exporters.

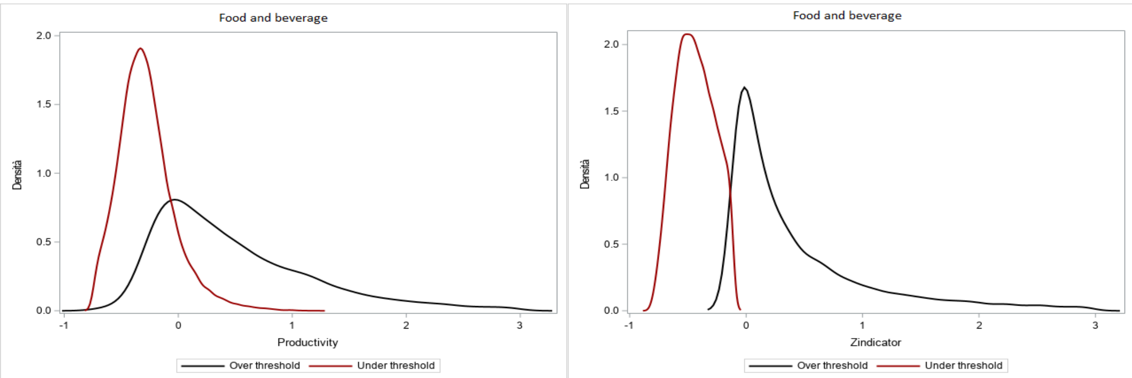
Table 3. Fitting tests of the ROC estimates

Industry	Precision	Accuracy (correct clustering)	Share of false positives	Share of false negatives	Share of export for true positives
Food and beverage	49.1	80.5	15.8	3.7	99.5
Textile	62.7	75.4	17.4	7.2	98.7
Wearing apparel	59.3	71.7	19.3	9.0	97.5
Leather	73.7	74.5	11.5	13.9	97.6
Wood	34.0	76.0	21.3	2.8	97.9
Paper and print	57.4	78.9	15.4	5.7	99.3
Chemicals and pharmaceutics	84.3	71.0	11.2	17.8	99.3
Rubber and plastic	84.2	73.9	9.1	17.0	97.9
Non metalic minerals	54.8	74.0	18.1	7.9	98.4
Metals	59.4	80.0	14.1	5.9	98.7
Electronics	83.8	72.2	8.7	19.0	97.3
Machinery	85.5	69.3	9.1	21.6	96.6
Automotive	74.0	71.3	15.3	13.5	99.3
Furniture	67.9	79.9	11.9	8.2	97.5

Source: Authors' calculation on Istat data

Third, another way of looking at how the Z-model outperforms the pure productivity model concerns the distribution of exporting and non-exporting firms according to their values of productivity and Z. As Figure 2 clearly shows, once we take into account the Z indicator – i.e. once we move from considering just productivity as in Melitz (2003) to considering a combination of productivity and economic size – in all industries the distributions overlap substantially shrinks to a very limited area.

Figure 2. Labour productivity (left) and Z indicator (right) for firms over and under the export threshold[8]



Source: Authors' calculations on Istat data

4. The "Technology line"

In this section, we use productivity and economic size to estimate the firms' positioning within the industry in terms of technology intensity. In fact, the minimum combination between these two variables which is necessary to become an exporter might not be consistent with the adoption of an advanced technology. As mentioned in Section 1, the literature has shown that an exporting firm can display the same (low) level of technology of a non-exporting firm (Bustos 2011, Bertschek *et al.*, 2016). It follows that technology may be

[8] We included in the text only Food and beverage. Figures for all industries in Appendix A.

a not-crucial element in determining the exporting status of a firm, and a mismatch between the conditions required to export and those ensuring a high level of technology may emerge.

In order to shed further lights on such mismatch, we set up a two-step procedure. First, we estimate the relative weights of economic size and productivity corresponding to a "high" level of technology, here defined as a higher than average level within the industry. Second, to assess the technology level of firms laying over and under the export threshold, we define a "technology line", which is the locus of all the combinations of economic size and productivity which guarantee the same probability of lying above the average technology as the one of the export threshold firm (i.e. the first exporter). The possibility of measuring each firm's distance from the export threshold and the technology line provides new insights on the possible mismatch between the technological levels of exporters and non-exporters, and has some important consequences for policy-making purposes.

More in details, in the first step, we build a firm-level proxy of technology using the same measure of technology as in Bustos (2011), which includes spending on computers and software, payments for technology transfers and patents, spending on R&D.[9] Successively, for each industry, a logit model of the probability for a firm to have a technology level higher than the industry average value is estimated, using the same covariates and controls as in Equation [5]:

$$Prob\ (Technology\ level_i > Average \mid S_i, \pi_i, G_i, I_i) = \Lambda(\ \alpha_1 S_i +\ \alpha_2 \pi_i + \alpha_3 G_i + \alpha_4 I_i) \qquad [10]$$

From this estimate, we use the relative weights of economic size (S) and productivity (π) to calculate, for each firm in the h-th industry, the following composite indicator:

$$Z^t_{h,i} =\ \hat{\alpha}_{1,h} S_{h,i} +\ \hat{\alpha}_{2,h} \pi_{h,i} \qquad [11]$$

In the second step, among the bundle of parallel lines represented by Equation [11], we identify the "technology line" in a plane, with $x - axis = S$ and $y - axis = \pi$, as the line passing through the values of economic size and productivity of the export threshold firm c ($S_{h,c}$ and $\pi_{h,c}$, respectively):

$$Z^t_{h,c} =\ \hat{\alpha}_{1,h} S_{h,c} +\ \hat{\alpha}_{2,h} \pi_{h,c} \qquad [12]$$

where $Z^t_{h,c}$ (hereinafter: Z^t) is the minimum combination of productivity and economic size which corresponds to a level of technology equal to the export threshold firms' (hereinafter: the "benchmark").

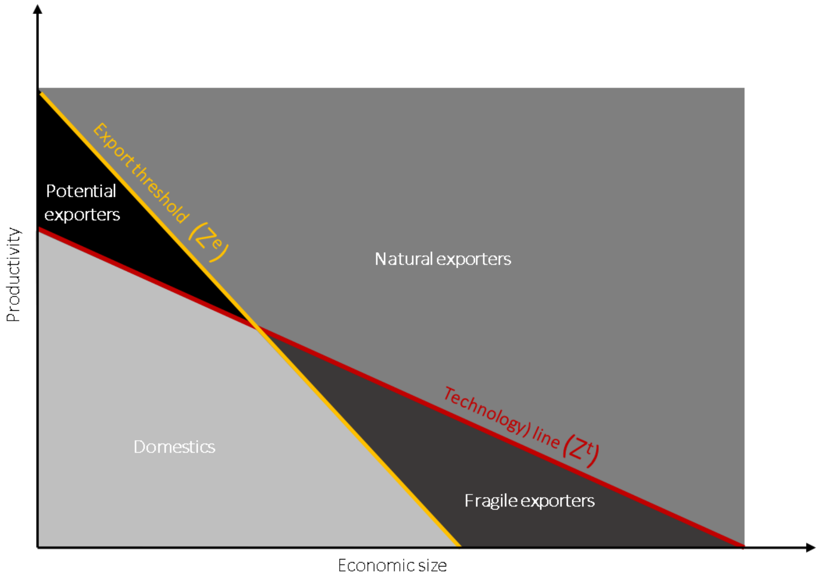
5. Mapping the business system: a new taxonomy of firms

On the basis of the positioning of firms with respect to Z^e and Z^t it is possible to derive a four-class taxonomy which qualifies the comparison between exporting and non-exporting firms in the light of their

[9] These informations are taken from administrative sources and included in the aforementioned business register "Frame-Sbs". We summarize them through a factor analysis in a synthetic indicator. Then we build a dichotomy variable to be used as a dependent variable in Equation [10], which takes value 1 when firm expenditure on technology is higher than industry average, 0 otherwise.

technology within the industry. In fact, the space defined by the interaction of Z^e and Z^t ideally defines four areas as depicted in Figure 3:

Figure 3. The taxonomy of firms export orientation

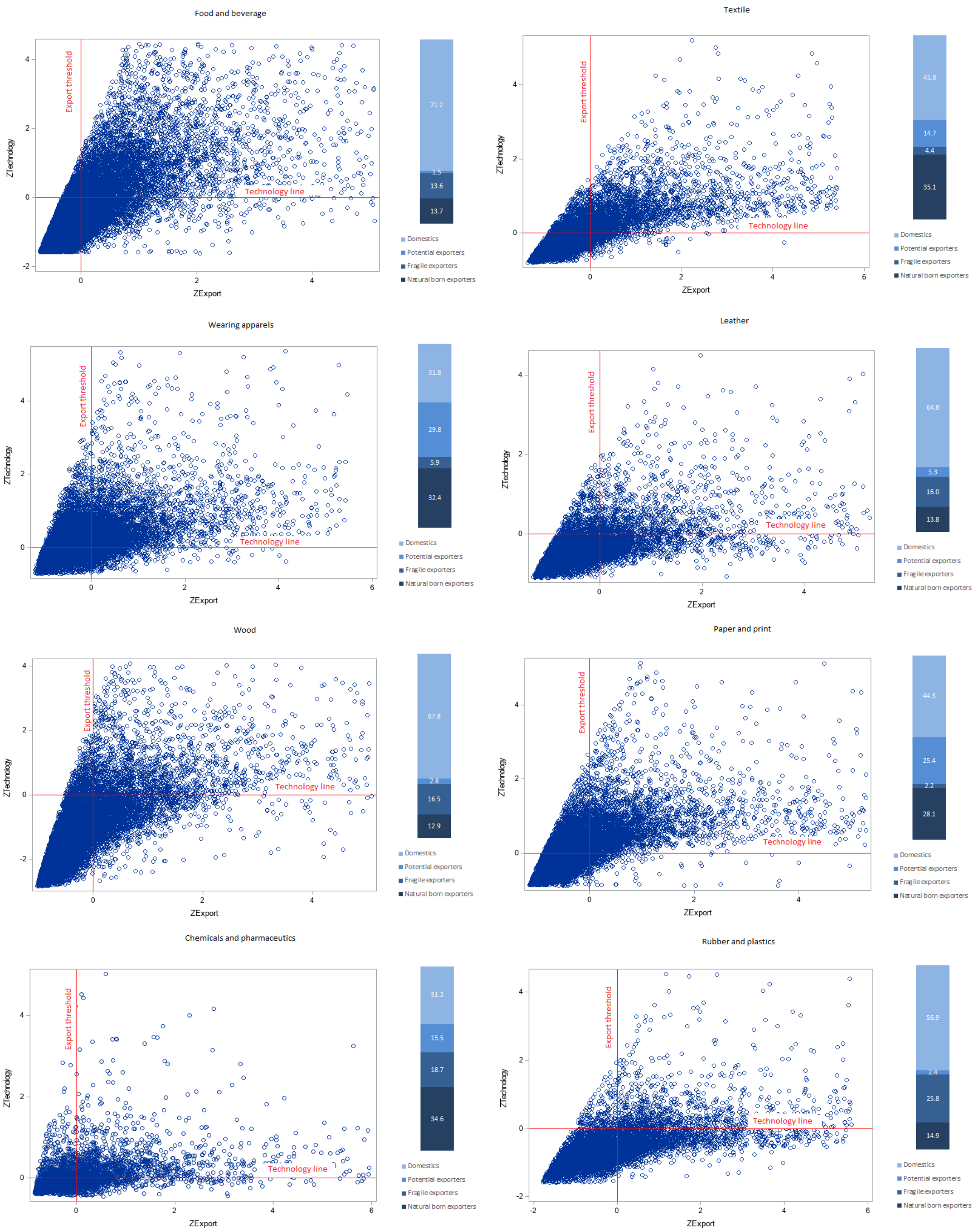


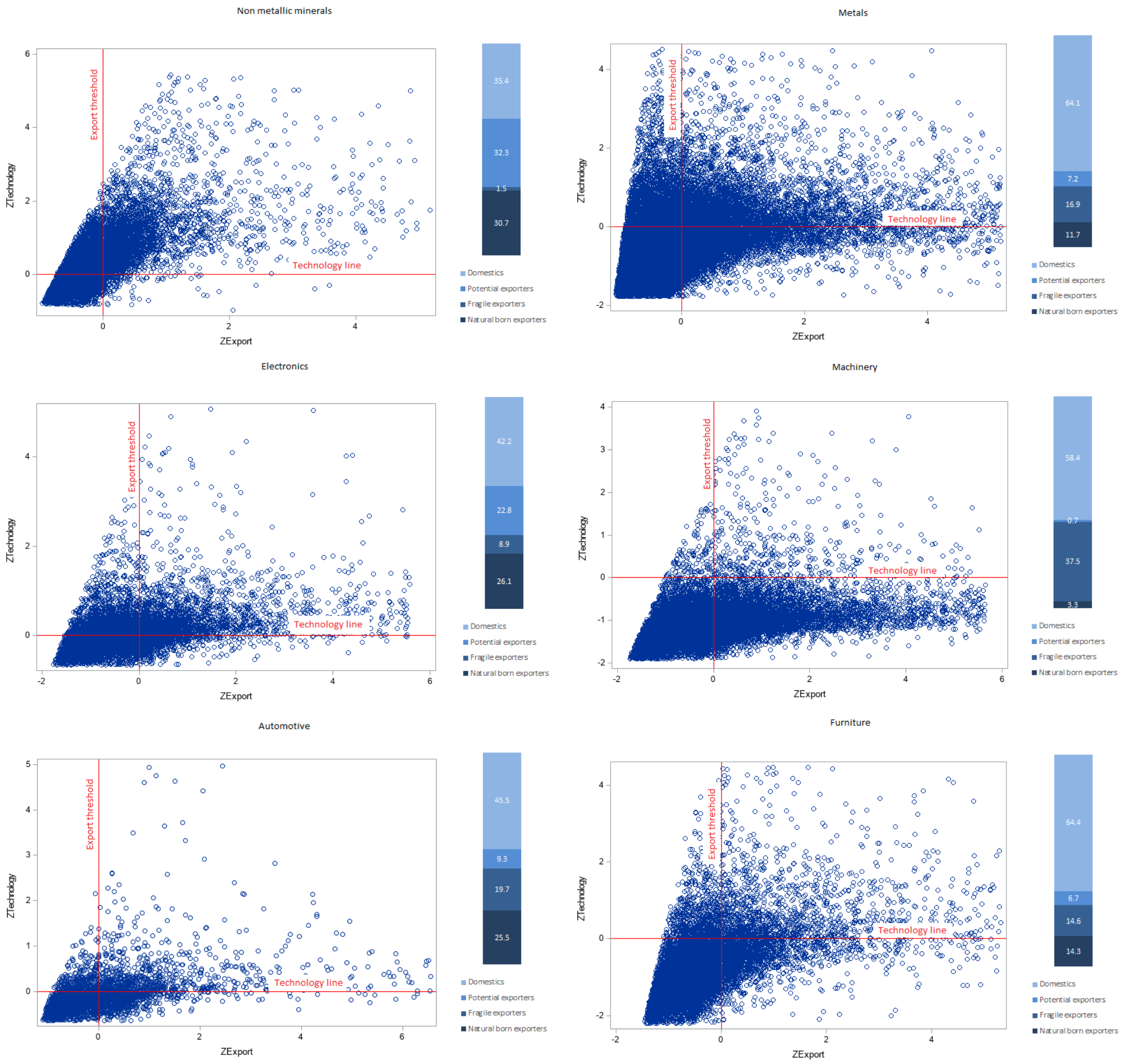
- "*Natural exporters*": firms with Z_i^e > Z^e and Z_i^t > Z^t, i.e. with a combination of productivity and economic size higher than both the export threshold and the technology line. These units are productive and/or "large" enough to produce efficiently and export.
- "*Fragile exporters*": firms with Z_i^e > Z^e and Z_i^t < Z^t. These units are thus classified as exporters notwithstanding their combination of productivity and size corresponds to a technology level lower than the benchmark.
- "*Potential exporters*": firms with Z_i^e < Z^e and Z_i^t > Z^t. These units have levels of productivity and size consistent with an over-the-benchmark efficiency, but insufficient to export.
- "*Domestics*": firms with Z_i^e < Z^e and Z_i^t < Z^t. These units have low levels of technology and do not reach the minimum combinations of productivity and economic size required to export.

Fragile and Potential exporters are the two classes where a mismatch between export activity and technology levels (Z^e and Z^t) emerges.

The distribution of firms in the four classes are plotted in Figure 4 according to their respective values of the Z^e and Z^t. The noticeable heterogeneity among the exporters (Fragile and Naturals) and the non-exporters (Domestics and Potentials) clearly emerges. Moreover, in all industries, the class of Domestics tends to outnumber the others with the exceptions of Machinery and Chemical and pharmaceutics, i.e. the industries with the highest percentages of exporting firms and especially of "Natural exporters" (but these latter are numerous also in Textile and Rubber and plastic).

Figure 4. Mapping business system: interaction between export threshold and technology line





Source: Authors' calculation on Istat data.

Moving from this taxonomy, Table 4 reports some descriptive evidence about the different classes by industry. Italian manufacturing comes out as a polarized system: in almost every industry, while Domestics account for the majority of firms, Natural exporters largely dominate in terms of share of value added and turnover.

From analytical and policy-making points of view, however, the most interesting groups are those of Fragile and Potential exporters. The formers, which lay above the export threshold despite a size-productivity combination under the technology line, are numerous especially in Machinery (where they account for over one third of the total), Rubber and plastics, Automotive, Chemicals and pharmaceutics. This might be related to factors such as the participation in GVC and/or intra-group trade. In this respect, the last column of Table 4 reports, for each class, the share of firms belonging to a multinational group. However, the incidence among Fragile exporters is generally low, ranging from 4.9% in Wood and 27.3% in Chemicals and pharmaceutics. This implies that the class of the taxonomy is only partially affected by this aspect.

Table 4. Characteristics of firms by typology and industry

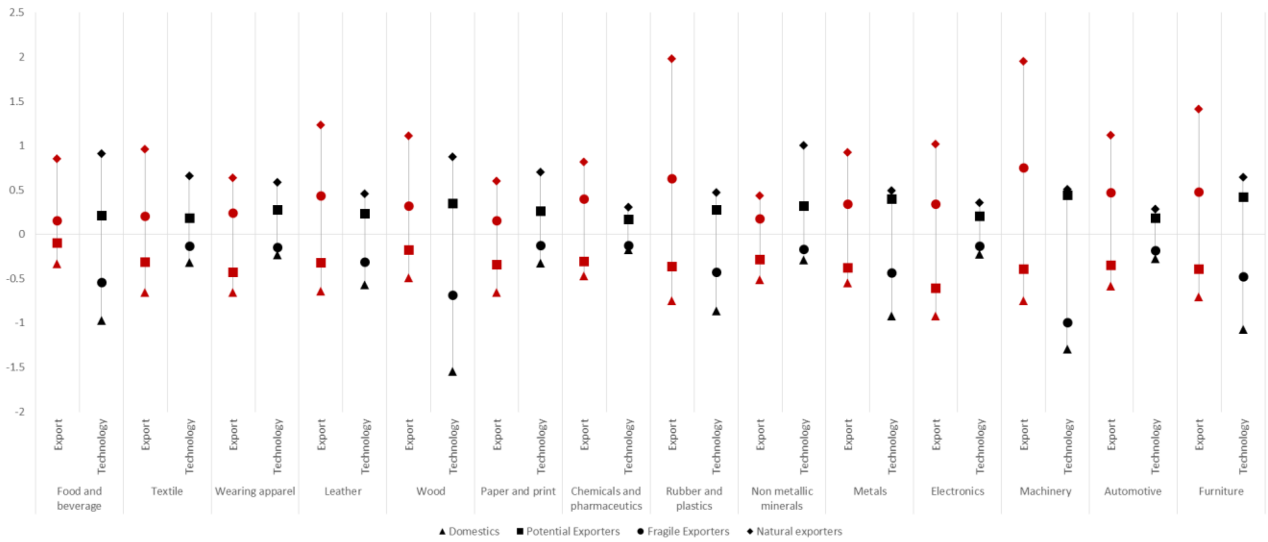
Industry	Taxonomy	Firms (shares of total industry)	Value added (Shares of total industry)	Turnover (Shares of total industry)	Employment (Mean, Workers)	Labor productivity (Mean, thousand euro)	Export/turnover ratio (Mean %)	Shares of firms belonging to MNE
Food and beverage	Domestics	71.2	8.8	5.2	4.0	19.7		0.8
	Potential Exporters	1.5	0.4	0.2	3.0	52.2		1.4
	Fragile Exporters	13.6	19.2	18.2	21.2	42.7	11.1	6.0
	Natural exporters	13.7	71.7	76.3	32.5	103.4	19.4	13.8
Textile	Domestics	45.8	5.0	4.0	4.0	18.9		1.0
	Potential Exporters	14.7	3.5	2.8	4.3	39.3		0.9
	Fragile Exporters	4.4	1.8	1.7	13.5	21.3	18.9	7.5
	Natural exporters	35.1	89.7	91.5	27.8	64.4	28.5	11.9
Wearing apparel	Domestics	31.8	3.4	2.8	4.7	13.1		0.2
	Potential Exporters	29.8	7.0	5.8	4.6	29.6		0.4
	Fragile Exporters	5.9	2.0	2.1	14.8	13.6	25.7	5.2
	Natural exporters	32.4	87.5	89.3	26.8	58.7	32.9	8.9
Leather	Domestics	64.8	11.3	7.7	5.9	23.5		1.0
	Potential Exporters	5.3	1.5	1.2	3.8	57.0		1.3
	Fragile Exporters	16.0	20.7	19.3	30.5	33.9	38.2	7.8
	Natural exporters	13.8	66.6	71.8	39.7	97.0	43.2	16.1
Wood	Domestics	67.8	14.4	11.8	2.4	20.7		1.3
	Potential Exporters	2.8	1.1	0.8	1.7	50.7		0.2
	Fragile Exporters	16.5	21.4	19.7	9.9	30.2	14.8	4.9
	Natural exporters	12.9	63.1	67.6	18.0	62.3	17.3	8.0
Paper and print	Domestics	44.3	3.6	3.0	3.1	18.8		1.4
	Potential Exporters	25.4	5.8	4.6	4.1	39.2		1.4
	Fragile Exporters	2.2	4.5	6.0	31.4	45.0	6.8	10.3
	Natural exporters	28.1	86.0	86.3	29.1	74.0	11.3	10.7
Chemicals and pharmaceutics	Domestics	31.2	1.0	1.1	5.5	33.2		2.6
	Potential Exporters	15.5	1.7	1.7	7.6	80.3		10.0
	Fragile Exporters	18.7	42.6	45.8	112.1	110.8	25.3	27.3
	Natural exporters	34.6	54.6	51.4	58.7	146.4	31.1	43.2
Rubber and plastic	Domestics	56.9	8.8	7.7	6.2	37.9		2.5
	Potential Exporters	2.4	0.5	0.5	3.9	88.6		2.7
	Fragile Exporters	25.8	26.3	25.3	30.0	51.2	21.7	15.5
	Natural exporters	14.9	64.3	66.5	70.2	93.1	32.0	31.8
Non metalic minerals	Domestics	35.4	2.0	2.0	3.1	14.3		1.5
	Potential Exporters	32.3	6.3	5.6	4.5	33.2		1.1
	Fragile Exporters	1.5	17.8	19.2	123.9	71.2	27.8	23.6
	Natural exporters	30.7	73.8	73.2	24.9	73.9	24.7	12.4
Metals	Domestics	64.1	12.2	8.6	4.4	32.3		1.3
	Potential Exporters	7.2	3.3	2.1	4.5	74.8		2.3
	Fragile Exporters	16.9	24.8	23.3	26.0	41.9	19.4	10.1
	Natural exporters	11.7	59.8	66.0	42.6	89.0	27.4	17.3
Electronics	Domestics	42.2	4.1	3.3	5.9	29.4		1.9
	Potential Exporters	22.8	5.4	4.4	7.0	60.4		6.8
	Fragile Exporters	8.9	4.1	4.8	26.8	30.7	25.6	14.2
	Natural exporters	26.1	86.4	87.5	71.4	82.5	36.1	32.1
Machinery	Domestics	58.4	9.9	8.8	6.7	47.2		3.4
	Potential Exporters	0.7	0.2	0.2	2.6	163.4		1.5
	Fragile Exporters	37.5	63.6	63.1	44.8	70.1	41.1	24.0
	Natural exporters	3.3	26.3	27.9	112.3	129.5	51.3	40.2
Automotive	Domestics	45.5	2.0	1.1	8.2	31.8		1.7
	Potential Exporters	9.3	0.9	0.6	9.2	66.2		7.3
	Fragile Exporters	19.7	4.4	3.2	37.7	35.7	27.1	14.7
	Natural exporters	25.5	92.7	95.0	254.9	86.4	39.5	37.2
Furniture	Domestics	64.4	11.0	8.5	3.6	23.9		1.5
	Potential Exporters	6.7	2.4	1.9	3.4	53.0		2.0
	Fragile Exporters	14.6	18.4	18.3	19.5	32.5	24.6	7.9
	Natural exporters	14.3	68.2	71.3	34.4	69.2	32.9	12.1

Source: Authors' calculation on Istat data.

As for Potential exporters, which show combinations of productivity and economic size corresponding to a relatively high technology but do not export, they are relatively numerous (with shares ranging from 22.8 to over 32%) in Non metallic minerals, Wearing apparels, Electronics, Paper and print. This class represents the target that policy measures aiming at increasing the number of exporting firms (i.e. to stimulate domestic units to cross the export threshold) should actually focus on, taking into account their peculiarities and heterogeneity. In this vein, an important size-related aspect emerges, because all Potential exporters, in all industries, are small-sized enterprises, counting less than 50 workers. In other terms, this class of the taxonomy includes generally small (possibly undersized) firms which nonetheless have a significant economic size – possibly due to relatively high levels of turnover and/or capital intensity, or because they are characterized by a long-lasting activity – and show technology levels comparable to those of Natural exporters.[10]

Even more interestingly, in virtually each industry Potential exporters are substantially smaller and more productive than Fragile exporters. On the one hand, this suggests that on average, in order to have the Potential exporters cross the export threshold, a recovery in size appears more adequate than an increase in productivity. On the other hand, in order for Fragile exporters to become Natural exporters, an increase in productivity appear to be more necessary than a recovery in size.

Figure 5. Distance from the export threshold and technology line by industry and taxonomy (median values)



Source: Authors' calculation on Istat data.

However, the extent to which Potential exporters (Fragile exporters) may reach the export threshold (technology line) by recovering size (productivity) also depends on the initial positioning of firms with respect to the export threshold (technology line) itself. In this context, our approach allows for measuring the distance of each firm from both: the median value of the distance for the four classes is displayed in Figure 5. In all industries, with the exception of food and beverage and wood, where the share of exporters is lower, the distribution of firms across the export threshold (red marks) appears to be more dispersed with respect to the one referring to the technology line (black marks). In other terms, the path of technology adoption appears to be more concentrated than the capability to export, confirming again the fact that the exporter status does not necessarily entail high level of technology as in Bustos (2011).

[10] There are a number of possible reasons for this. For example, in terms of the model by Lileeva and Trefler (2010), such firms may be domestic units which have invested in technology and are expected to be shifting to exporter status (in our terms: crossing the export threshold). Moreover, they may also be units belonging to enterprises groups in which specific branches are in charge of the export activity of the entire group. Furthermore, our Potential exporters may include suppliers of other exporting firms; in this case a possible high-technology, exporting buyer could stimulate its component or intermediate goods suppliers to adopt an advanced technology, so that the (generally small-sized) suppliers would end up crossing the technology line without reaching the export threshold.

6. Conclusions

In this paper, we analyse the potential mismatch between the conditions required for a firm to become exporter and the pattern of technology adoption. In particular, we provide a methodology that allows to cluster business units according to their export orientation and technology, so that it becomes possible to distinguish what firms are able to export despite their relatively (within the sector) low technology, and, even more importantly, what firms do not export notwithstanding their high level of technology within their industry.

To do so, we firstly use our ROC-based methodology to estimate, for each industry, the export threshold, defined as the firm-level minimum combination of productivity and economic size corresponding to the transition from the non-exporter to exporter status. Successively, we introduce the technology line, i.e. the combination of productivity and economic size which corresponds to a technology level higher than the average of the industry. The presence of a compensation between productivity and economic size in exporting, in fact, may imply the possibility of a mismatch between export- and technology- related combinations.

The interaction between the export threshold and the technology line permits to derive a taxonomy that classifies firms in terms of the conditions needed to export and to adopt a high level of technology. This classification is especially important from a policy-making point of view, because it allows for a new breakdown of exporters and non-exporters. On the one hand, Fragile exporters are low-tech exporting firms, while, on the other hand, Potential exporters are high-tech non exporting firms. The formers are comparable to the particular set of exporting firms that have not yet adopted the higher technology, as pointed out by the literature. The latters are instead a new class of firms identified by our approach, allowing to stress the existence of a specific group of non-exporting enterprises which are more likely to become exporters. This portion of the manufacturing sector would be the ideal target for policy measures aimed at increasing the participation of firms in international markets. The possibility of singling it out within the universe of non exporting firms allows to design more precise policies, so increasing their effectiveness and, eventually, reducing their costs for Governments.

References

Berge, T.J. and Ò. Jorda (2011), Evaluating the Classification of Economic Activity into Recessions and Expansions, *American Economic Journal: Macroeconomics* 3: 246–277.

Bernard, A.B. and J.B. Jensen (1995), Exporters, Jobs and Wages in U.S. Manufacturing: 1976–1987. *Brookings Papers on Economic Activity, Microeconomics*, *1*, 67–119.

Bernard, A.B., J. Eaton, J.B. Jensen and S. Kortum (2003), "Plants and Productivity in International Trade," *American Economic Review*, 93, 1268–1290.

Bernard, A.B., S.J. Redding and P.K. Schott (2011), Multi-product Firms and Trade Liberalization. *Quarterly Journal of Economics*, *126*(3), 1271–1318. https ://doi.org/10.1093/qje/qjr021

Bertschek I., J. Hogrefe and F. Rasel (2016), Trade and Technology: New Evidence on the Productivity Sorting of Firms, *Review of World Economics* 151 (1): 53–72.

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

Castellani, D. and A. Zanfei (2007). Internationalisation, Innovation and Productivity: How Do Firms Differ in Italy, *World Economy* 30: 156-176.

Chaney, T. (2008), Distorted gravity: The Intensive and Extensive Margins of International Trade. *American Economic Review*, *98*(4), 1707–1721. https ://doi.org/10.1257/aer.98.4.1707

Costa S., F. Sallusti, C. Vicarelli and D. Zurlo (2019), Over the ROC Methodology: Productivity, Economic Size and Firms' Export Thresholds, *Review of International Economics*, 27: 955-980. DOI: 10.1111/roie.12405.

Delgado M.A., J.C. Fariñas and S. Ruano (2002), Firm Productivity and Export Markets: a Non-parametric Approach, *Journal of International Economics*, 57, 397–422.

Fawcett, T. (2005), An Introduction to ROC Analysis, *Pattern Recognition Letters*, 27, 861-874. https://doi.org/10.1016/j.patrec.2005.10.010.

Geishecker, I., P.J.H. Schröder and A. Sørensen (2017), Explaining the Size Differences of Exporter Premia: Theory and Evidence, *Review of World Economics*, 1*53* (2), 327-351.

ISGEP – International Study Group on Exports and Productivity (2008), Understanding cross-country differences in export premia: Comparable evidence for 14 countries. *Review of World Economics*, *144*(4), 596–635.

Khandani, A.E., J.K. Adlar and W.Lo. Andrew (2010), Consumer Credit-Risk Models via Machine-Learning Algorithms. *Journal of Banking and Finance*, 34(11): 2767–87.

Kumar, R., and A. Indrayan (2011), Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatrics*, *48*(4), 277–287. https ://doi.org/10.1007/s13312-011-0055-4.

Leung, D., C. Meh and Y. Terajima (2008), Productivity in Canada: Does Firm Size Matter? *Bank of Canada Review*, Autumn, 7–16.

Lileeva, A. and D. Trefler (2010), Improved Access to Foreign Markets Raises Plant-level Productivity... For Some Plants. *Quarterly Journal of Economics*, 125(3), 1051–1099.

Lusted, L.B. (1960), Logical Analysis in Roentgen Diagnosis: Memorial Fund Lecture. *Radiology*, 74(2): 178–93.

Majnik, M. and Z. Bosnić (2013), ROC Analysis of Classifiers in Machine Learning: A Survey. *Intelligent Data Analysis*, 17: 531-558. doi: 10.3233/IDA-130592

Máñez-Castillejo J.A., M.E. Rochina-Barrachina and J.A. Sanchis-Llopis (2010), Does Firm Size Affect Self-selection and Learning-by-exporting?, *The World Economy*, 33(3), 315–346.

Melitz, M.J. (2003), The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71, 1695-1725. doi: 10.1111/1468-0262.00467.

Melitz, M.J. and G.I.P. Ottaviano (2008), Market Size, Trade and Productivity. *Review of Economic Studies*, 75, 295-316. doi.org/10.1111/j.1467-937X.2007.00463.x.

Pepe, M.S. (2003), *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford, UK: Oxford University Press.

Schröder P.J.H. and A. Sorensen (2012), Second Thoughts on the Exporter Productivity Premium. *Canadian Journal of Economics*, *45*(4), 1310–1331. https ://doi.org/10.1111/j.1540-5982.2012.01742.x

Syverson, C. (2011), What Determines Productivity? *Journal of Economic Literature*, *49*(2), 326–365. https://doi.org/10.1257/jel.49.2.326.

Tybout, J. (2000), Manufacturing Firms in Developing Countries: How Well Do They Do and Why? *Journal of Economic Literature*, *38*(1), 11–44. https://doi.org/10.1257/jel.38.1.11.

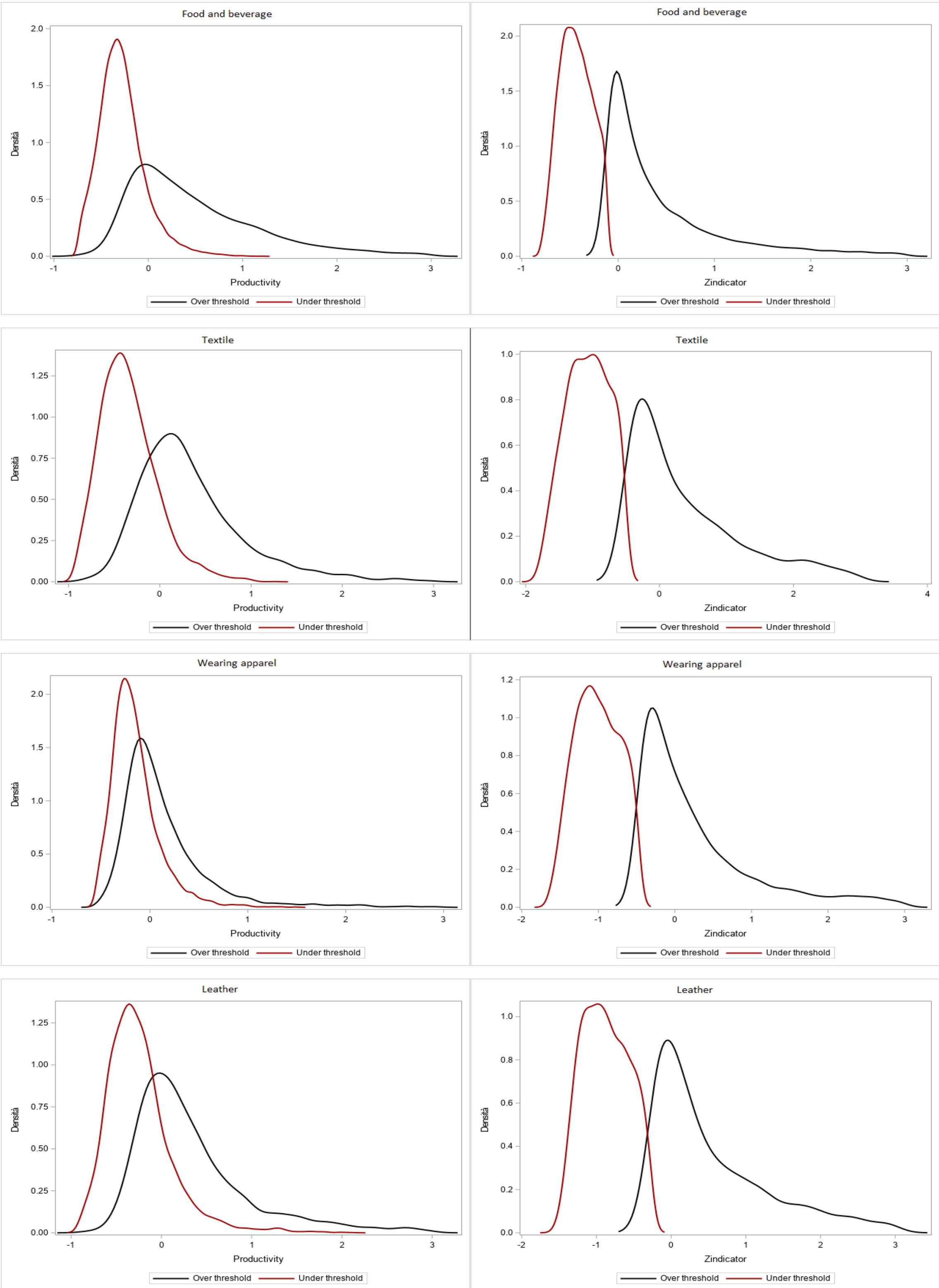
Wagner, J. (2007), Exports and productivity: A survey of the evidence from firm level data. *World Economy*, *30*(1), 60–82. https ://doi.org/10.1111/j.1467-9701.2007.00872.x.

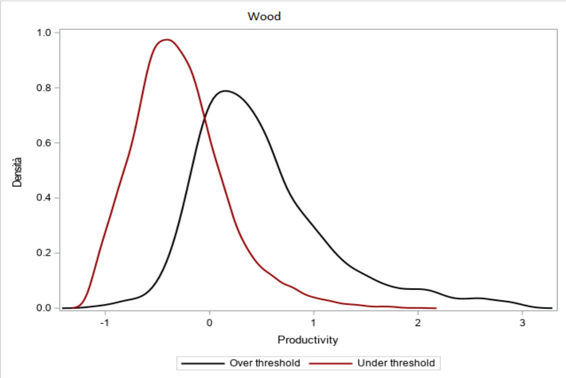
Wagner, J. (2012), International trade and firm performance: A survey of empirical studies since 2006. *Review of World Economics*, *148*(2), 235–267. https ://doi.org/10.1007/s10290-011-0116-8.

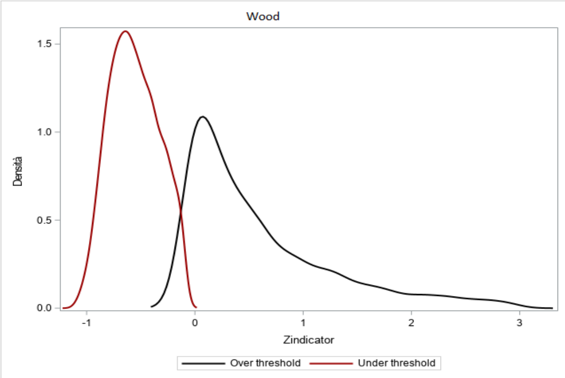
Warnock, D.G. and C. Peck (2010), A Roadmap for Biomarker Qualification. *Nature Biotechnology*, 28, 444–445. doi: 10.1038/nbt0510-444.

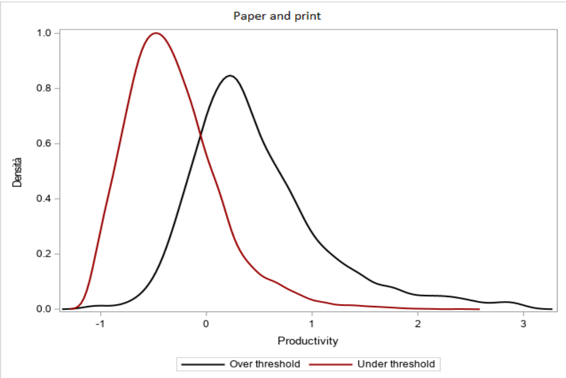
Youden, W.J. (1950), Index for Rating Diagnostic Tests. *Cancer*. 3: 32–35. Doi: 10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.

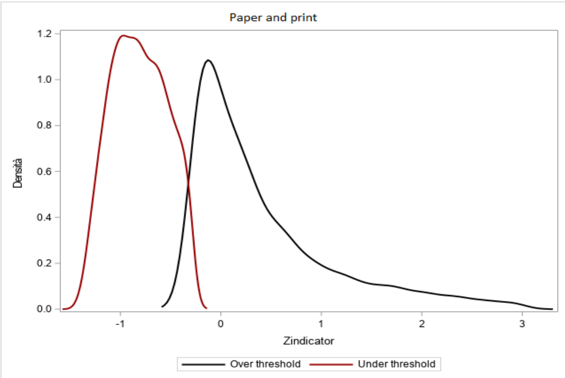
Appendix A. Distribution of labour productivity (Left) and Z indicator (Right) for firms over and under the export threshold (Kernel density)

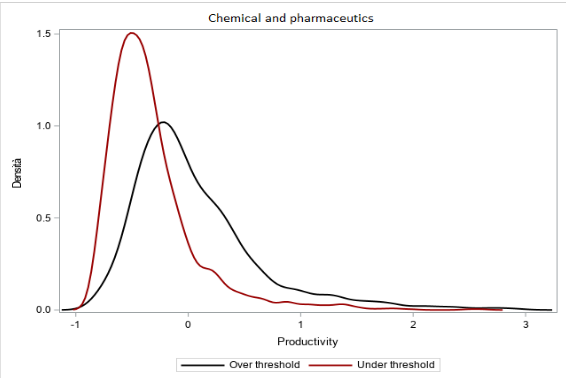


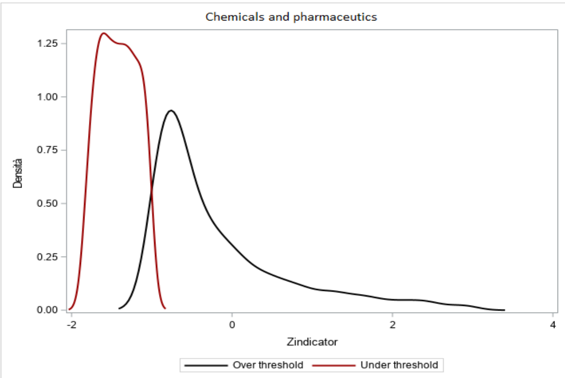

Wood
Densità
Productivity
Over threshold
Under threshold

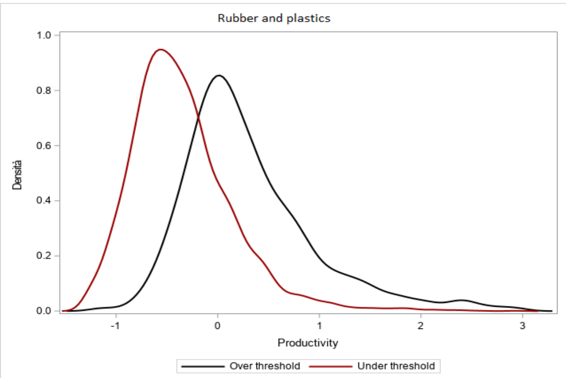

Wood
Densità
Zindicator
Over threshold
Under threshold

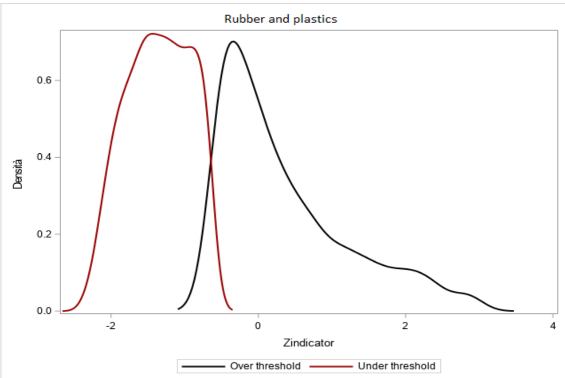

Paper and print
Densità
Productivity
Over threshold
Under threshold

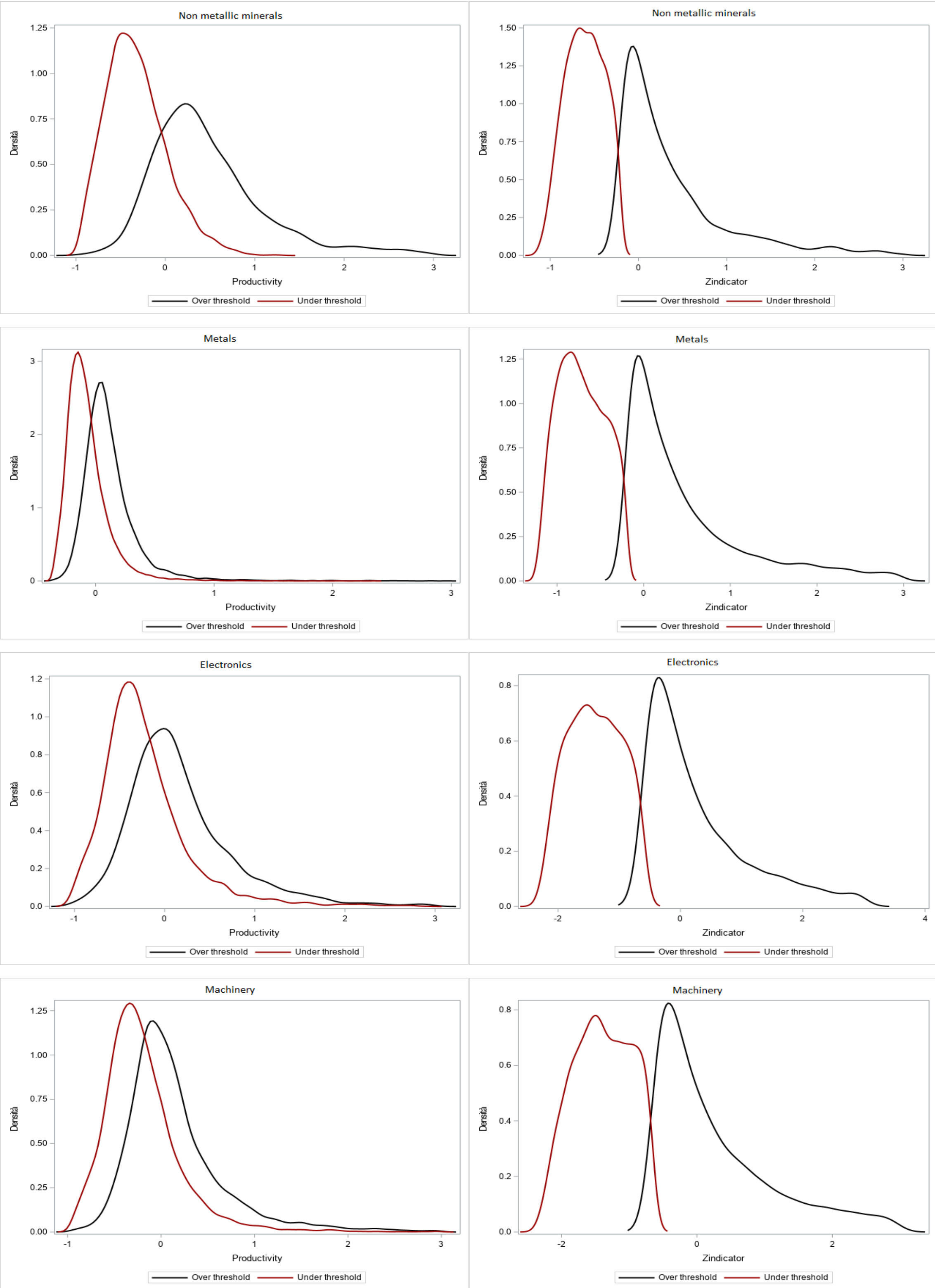

Paper and print
Densità
Zindicator
Over threshold
Under threshold

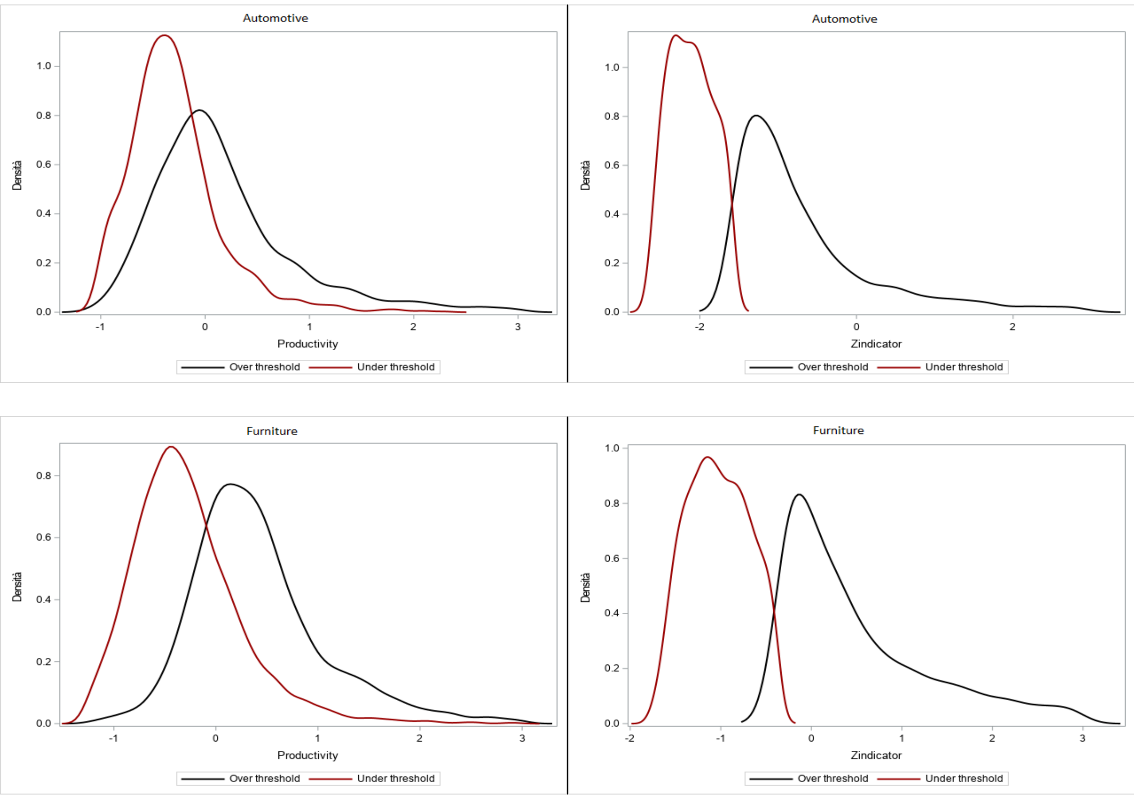

Chemical and pharmaceutics
Densità
Productivity
Over threshold
Under threshold


Chemicals and pharmaceutics
Densità
Zindicator
Over threshold
Under threshold

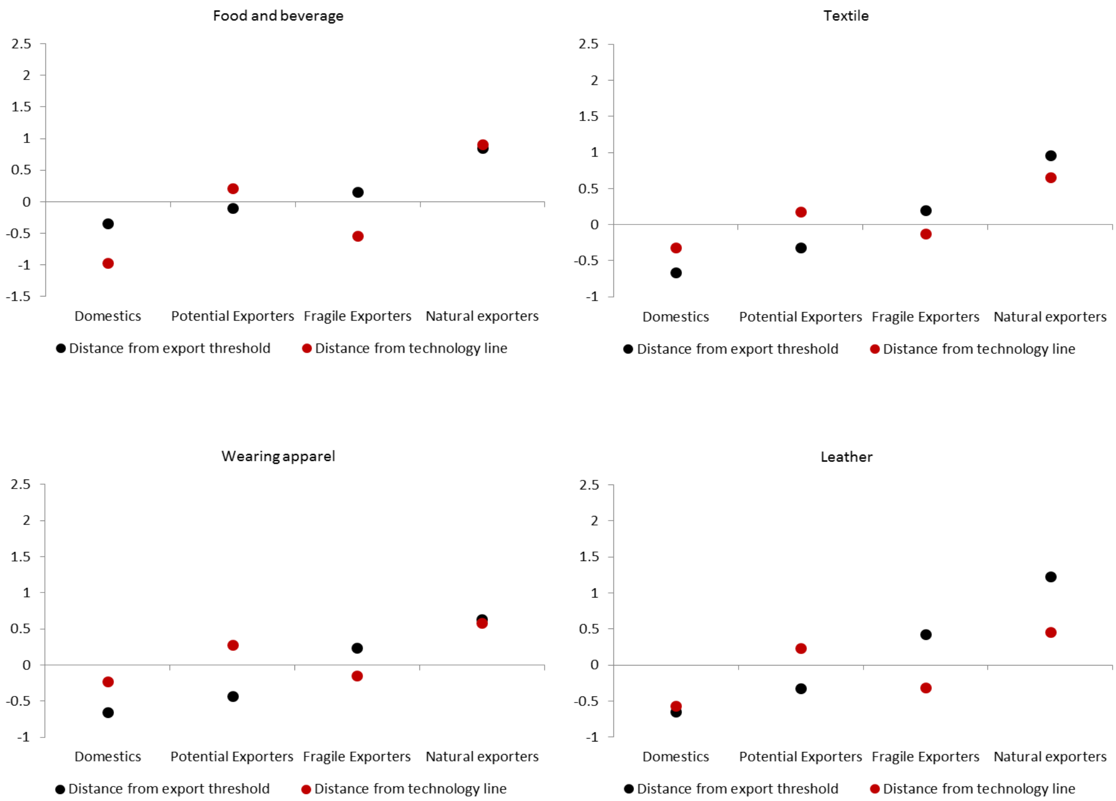

Rubber and plastics
Densità
Productivity
Over threshold
Under threshold


Rubber and plastics
Densità
Zindicator
Over threshold
Under threshold

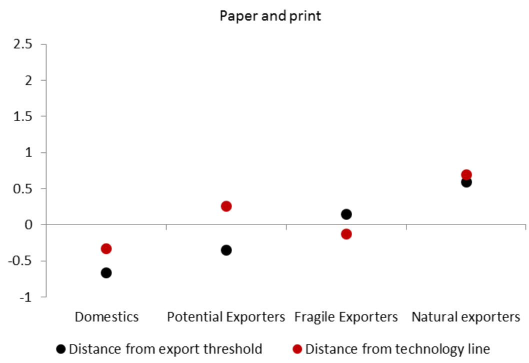

Non metallic minerals
Densità
Productivity
Over threshold
Under threshold
Non metallic minerals
Densità
Zindicator
Over threshold
Under threshold
Metals
Densità
Productivity
Over threshold
Under threshold
Metals
Densità
Zindicator
Over threshold
Under threshold
Electronics
Densità
Productivity
Over threshold
Under threshold
Electronics
Densità
Zindicator
Over threshold
Under threshold
Machinery
Densità
Productivity
Over threshold
Under threshold
Machinery
Densità
Zindicator
Over threshold
Under threshold

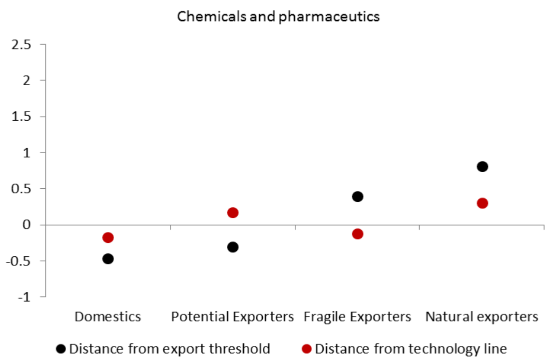


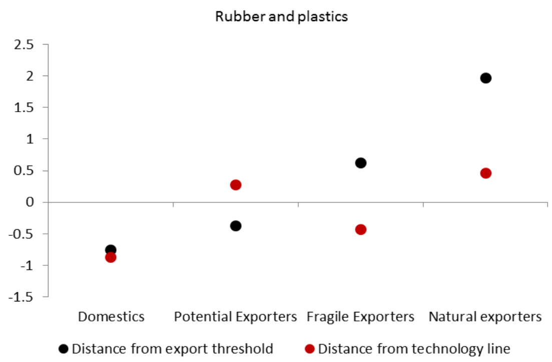
Appendix B. Distance from export threshold and technology line, by industry and taxonomy (median values)

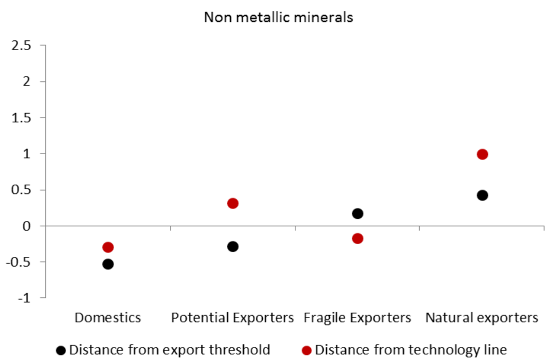


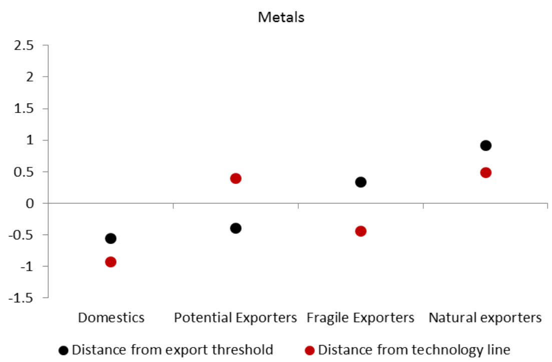

Wood
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
Domestics
Potential Exporters
Fragile Exporters
Natural exporters
Distance from export threshold
Distance from technology line

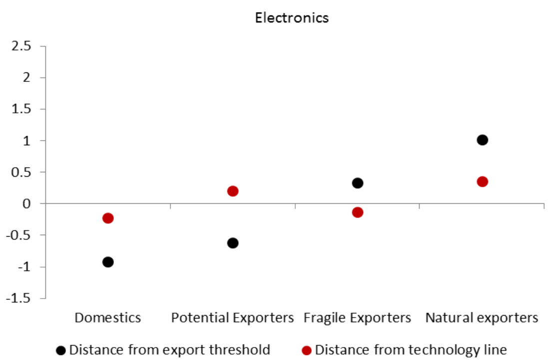

Paper and print
2.5
2
1.5
1
0.5
0
-0.5
-1
Domestics
Potential Exporters
Fragile Exporters
Natural exporters
Distance from export threshold
Distance from technology line

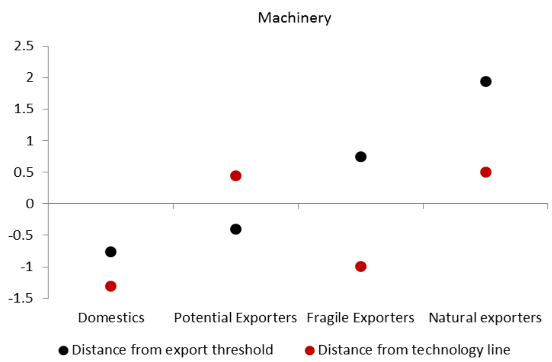

Chemicals and pharmaceutics
2.5
2
1.5
1
0.5
0
-0.5
-1
Domestics
Potential Exporters
Fragile Exporters
Natural exporters
Distance from export threshold
Distance from technology line

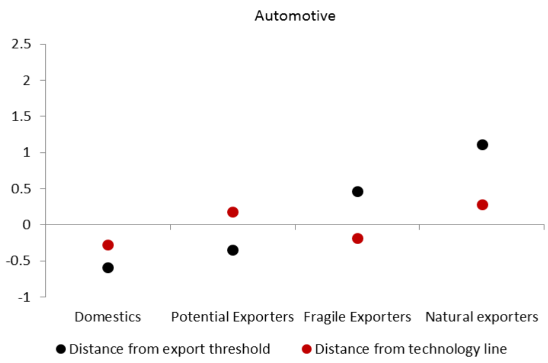

Rubber and plastics
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
Domestics
Potential Exporters
Fragile Exporters
Natural exporters
Distance from export threshold
Distance from technology line

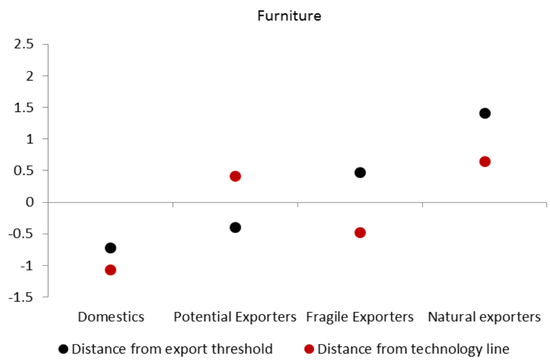

Non metallic minerals
2.5
2
1.5
1
0.5
0
-0.5
-1
Domestics
Potential Exporters
Fragile Exporters
Natural exporters
Distance from export threshold
Distance from technology line


Metals
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
Domestics
Potential Exporters
Fragile Exporters
Natural exporters
Distance from export threshold
Distance from technology line


Electronics
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
Domestics
Potential Exporters
Fragile Exporters
Natural exporters
Distance from export threshold
Distance from technology line


Machinery
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
Domestics
Potential Exporters
Fragile Exporters
Natural exporters
Distance from export threshold
Distance from technology line


Automotive
2.5
2
1.5
1
0.5
0
-0.5
-1
Domestics
Potential Exporters
Fragile Exporters
Natural exporters
Distance from export threshold
Distance from technology line


Furniture
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
Domestics
Potential Exporters
Fragile Exporters
Natural exporters
Distance from export threshold
Distance from technology line