Stefano Costa
Federico Sallusti
Claudio Vicarelli
Davide Zurlo

Istat

Where Frontiers Interact: Mapping Firms Participation in International Markets

Abstract

This paper provides new insights on Italian manufacturing firms participation in international markets in 2016. In particular, we are able to position each firm with respect to an "export threshold" and a "technology line". The former – 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 line, estimated starting from a stochastic frontier model, is the combination of economic size and productivity that is representative of firm's industry.

The interaction between the technology line and export threshold permits: (1) to define a classification of firms ("Natural exporters", "Locals", "Smarts", "Potential exporters") which is particularly relevant from a policy-making point of view; (2) to compare, for each industry, the relative importance of firms' economic size and productivity in determining export threshold and technology line, in order to have a measure of the "export friendliness" of industries.

JEL code: F14, L60, L11

Keywords ROC analysis, export threshold, intensive and extensive margin of exports

1. Introduction

Export activity is important for firm competitiveness and, more in general, for a country's economic growth. 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 became an exporter.

In a previous paper, we applied the Receiver Operating Curve (ROC) analysis to develop a new methodology for the estimation of the "export threshold", i.e. the combination of productivity and "economic size" (defined over a set of firm-level size-related variables) corresponding to the transition from non-exporter to exporter status (Costa *et al.*, 2019). In this paper, we enrich that analysis by introducing a new "frontier", which identifies the representative technology of each industry on the basis of a combination of productivity and "economic size" (a broad measure of firm size composed of employment, age, turnover and capital intensity). The interaction between the export threshold and the "technology line", along with the position of firms with respect to these two frontiers, determine a new taxonomy of firms that allow to better analyse firms export orientation and (especially) export potential.

On such bases, the main contribution of this paper is to provide a map of business system that is especially useful from a policy-making point of view, as it allows for more precisely targeting policies to boost firm participation to foreign markets.

For our scope, the main strand of literature is that related to firm heterogeneity, which points out the role of firms structural characteristics (e.g. size, location, business sector, exporting status), strategies (e.g. different forms of innovation, inter-firms relationships) and performance (e.g. revenues, profitability, productivity, innovation) in determining firms competiveness. In particular, we are interested in the relationship between firm productivity, (economic) size and ability to export.

The existence of an "export threshold" characterizes all the theoretical works focused on firm heterogeneity starting from the seminal paper of Melitz (2003), where only firms above a minimum productivity level are able to sell both domestically and abroad (Melitz and Ottaviano, 2008; Chaney, 2008; Bernard *et al*., 2011). However, from the empirical point of view, several works showed that in many countries, firms' productivity distributions between exporters and non-exporters may overlap[1], implying that firms might not export even though their productivity is above the threshold. Moreover, other works have shown that the mismatch between Melitz's theory and empirical evidence is only apparent, being mainly 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 (Schröder and Sørensen, 2012; Geishecker *et al*., 2017).

Firm size, a proxy of the ability to afford sunk costs of exporting, may be relevant to explain a preference for, or an aversion to, exporting. Empirical studies did find a direct relationship between export and size: exporters tend to be larger than non-exporters (Bernard and Jensen, 1995; Wagner, 2007). 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 regulations and access to

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

financing could also be responsible for productivity differentials between small and large firms (Tybout, 2000).

In the empirical literature, causal relationship between productivity, size and export activity has been largely analyzed (see Wagner, 2012 and ISGEP, 2008 for detailed surveys), especially the "self-selection" versus learning-by-exporting hypothesis. In the first one, a firm should reach a minimum level (a "threshold") of productivity before starting to export; in the second one, knowledge flowing from international buyers and competitors helps improve the post-entry performance of exporters. However, the learning-by-exporting effect may also be related to the size of firms. Focusing on Spanish manufacturing, Máñez-Castillejo *et al.* (2010) demonstrate the existence of a process of self-selection into exporting among small firms but not among large firms, while the learning-by-exporting effect is significant independent from size.

The rest of the paper is organized as follows. Section 2 presents a description of the dataset and empirical strategy. Section 3 illustrates, for each industry, the main results obtained from the interaction between the export threshold and the technology line, describing the new taxonomy of firms export propensity. Section 4 summarizes and provides conclusions.

2. The data

Our main statistical source is the firm-level dataset "Frame-Sbs" for 2016. Released by ISTAT since 2011, it annually provides administrativebased information on the structure (number of employees, business sector, location, age) and the main Profit and Loss Account variables (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-type 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 our 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 persons employed, and the enterprises with just one person employed accounted for over 50% of total firms and 12% of total employment –we choose to focus on units with "economic relevance" for the analysis of export strategies. Consequently, we imposed a number of restrictions. (1) We only focus on firms operating in manufacturing (excluding Tobacco, Refined petroleum products, Maintenance and repair, Other manufacturing), which in 2016 accounted for 83% the total Italian exports. (2) In order to consider economic relevance, we select only firms that have positive value added, no less than 1 employee, and positive consumption of fixed capital. (3) Finally, we rule out irregular and one-off exporting firms, and consider only the "stable exporters", namely those exporting on a regular basis over the three-year period 2014-2016.[2]

In doing so, the final dataset, referring to 2016, 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] There is no universally accepted definition of "stable exporter", except that, for a firm to be defined as such, it has to be exporting on a regular basis over a specified (more than a year) period. We preferred the 2014-2016 time span also because it is more homogeneous from a business cycle point of view, as it fully covers the Italian post-recession period.

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 (hereinafter, ROC) analysis and Youden's (1950) *J* index. This permits the identification of a cut-off point over an independent variable in a logit model, 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 (NBER; Berge and Jorda 2011) and in the 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 Bosnic, 2013), and natural science (Warnock and Peck, 2010).

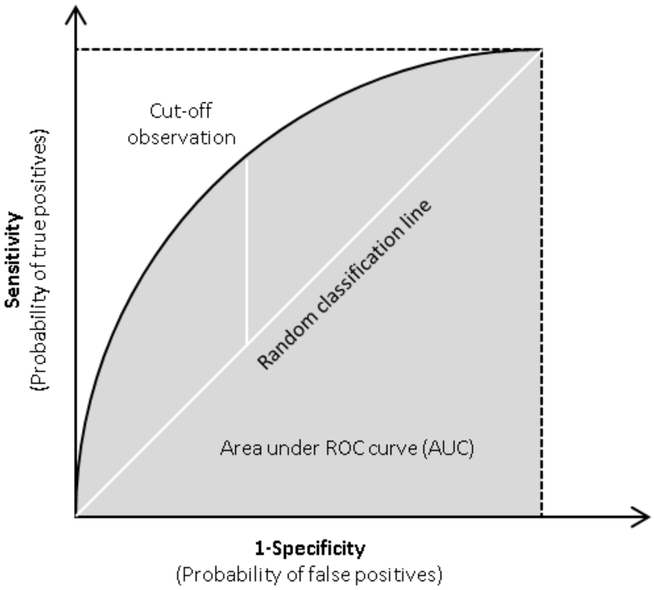
According to Fawcett (2005), classification models (or classifiers) can give four possible outcomes:

1. True positives (TP): positive observations are correctly classified as positive by the model;
2. False negatives (FN): positive observations are erroneously classified as negative by the model;
3. False positives (FP): negative observations are erroneously classified as positive by the model;
4. True negatives (TN): negative observations are correctly classified as negative by the model.

The validity of a classifier can be measured based on two main metrics: Sensitivity and Specificity. Sensitivity represents the probability of detecting true positive cases. Specificity is the probability of detecting true negative cases. 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 & 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 given 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 ($\widehat{\mathrm{Cut}}$), the following equation is to be maximized:

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

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:

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

[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.

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 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"

Like 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 firm h in industry i based on the following logit model:

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

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 i-th industry, thus also determining the combination of the covariate representing the threshold:

$$\overline{\mathrm{X}}_{\mathrm{i}} = \mathrm{X}_{\mathrm{c,i}} \qquad [4]$$

where c is the cut-off firm. Using this cut-off, it is possible to classify firms as exporters or non-exporters according to their being over or under this threshold.

In particular, we use a composite model (Z-model, where $\overline{\mathrm{X}} = \mathrm{Z}^e$), in which the export threshold is defined over a combination (Z^e) of productivity and economic size (which in turn synthesises four size-related variables).[6]

The composite indicator Z is derived from a three steps procedure. In the first step, for each industry, the "economic size" indicator 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 the i-th industry, economic size is thus given by 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 each 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., and 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.

$$\text{Prob}\,(\text{Export} = 1|\text{S}, \pi, \text{G}, \text{I}) = \Lambda(\,\alpha_1\text{E} + \,\alpha_2\pi + \alpha_3\text{G} + \alpha_4\text{I}\,) \qquad [5]$$

where Λ is the cumulative distribution of the logistic function, α_j are estimated parameters, E is the "economic size" of firms, π is their productivity (in terms of value added-per-worker), G is a set of dummy variables indicating the location of firms,[7] and I is a set of dummy variables related to NACE 2-digit levels of economic activity.

In the third step, the estimated coefficients of productivity and "economic size" from equation [6] are used to obtain the composite indicator $Z_{h,i}$ for each firm h in the i-th industry. In particular, estimated parameters are used as weights, while variables are taken at individual level:

$$\text{Z}_{h,i} = \,\widehat{\alpha}_{1,\text{i}}E_{h,i} + \,\widehat{\alpha}_{2,\text{i}}\pi_{h,i} \qquad [6]$$

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

3.3. Fitting tests of ROC estimates

Three types of test have been carried out on the results. Firstly, 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]).

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

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

Source: Authors' calculation on ISTAT data.

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

Results of the test are reported in Table 2. On the basis of the AUC, 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), with statistically significant differences in favour of the Z-model.

Secondly, we consider the capability of the cut-offs identified by the *J* index in 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. firms correctly classified as exporters):

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad [7]$$

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

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{Total observations}} \quad [8]$$

Finally, in order to assess to what extent this characteristic could result in a distorted selection, we calculated the weight of true positive observations in term of total exports.

Table 3. Fitting tests of the ROC estimates

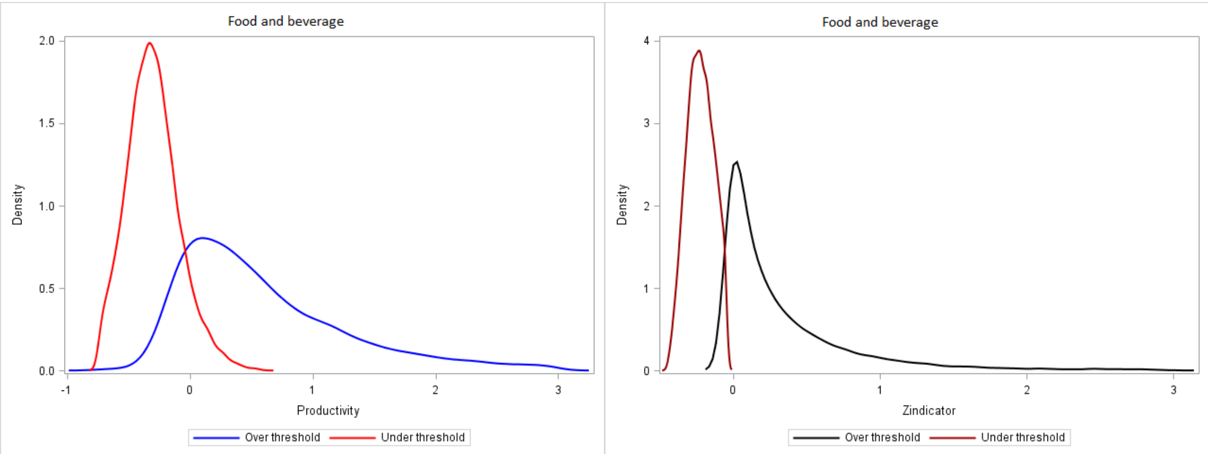
Industry	Precision	Accuracy (correct clustering)	Share of false positives	Share of false negatives	Share of export for true positives
Food and beverage	50.4	81.6	15.0	3.4	99.6
Textile	63.7	75.8	16.7	7.5	98.6
Wearing apparel	60.5	72.1	18.4	9.6	97.3
Leather	66.8	72.1	16.0	11.9	98.2
Wood	38.4	79.3	17.5	3.2	97.3
Paper and print	55.3	77.9	16.7	5.4	99.4
Chemicals and pharmaceutics	84.3	70.9	11.1	17.9	99.3
Rubber and plastic	81.3	74.0	11.1	14.8	98.3
Non metalic minerals	55.6	74.4	17.5	8.0	98.3
Metals	59.2	79.9	14.2	5.9	98.7
Electronics	83.9	72.4	8.7	18.9	97.3
Machinery	81.6	69.9	12.0	18.1	97.5
Automotive	74.0	71.3	15.2	13.5	99.3
Furniture	66.0	79.2	13.0	7.8	97.7

Source: Authors' calculation on Istat data

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), 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 in all industries (always over 97%), suggesting that false negatives are negligible exporters.

Another way to look at how the Z-model outperforms the pure productivity model concerns the distribution of exporting and non-exporting firms (i.e. units above and below the export threshold, respectively) according to their values of productivity and Z. As Figure 2 clearly shows, once we take into account the Z indicator – that is, once we move from considering just the labour 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 and Z indicator for firms over and under the export threshold[8]



Source: Authors' calculations on Istat data

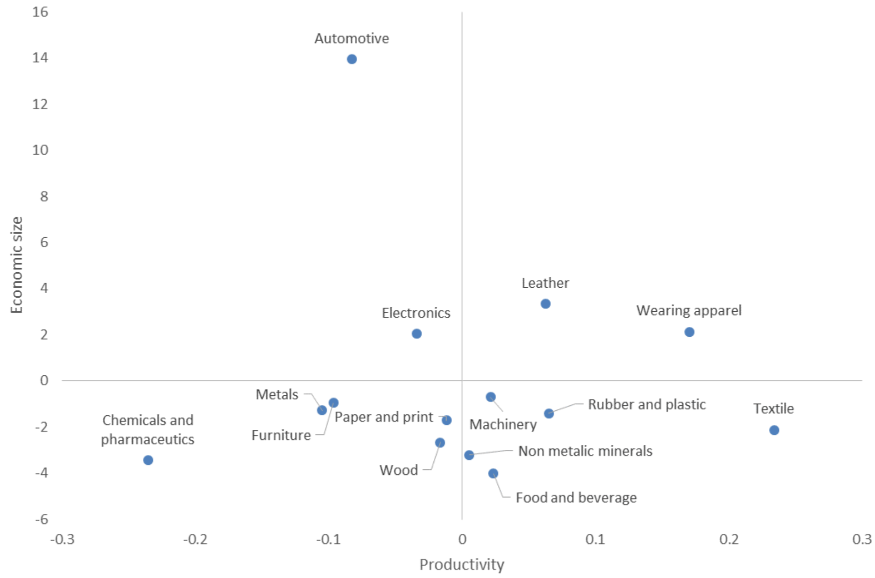
3.4. Results: across the export threshold

The higher capability of the Z-model in individuating exporters implies that in manufacturing sector, for a firm to become an exporter, a given degree of complementarity might exist between productivity and economic size. In particular, in some cases, a sort of compensation could emerge between these two factors in order to cope with sunk costs related to enter international markets.

While this occurs in all industries, a relevant degree of sectoral heterogeneity exists according to the relative role of productivity and economic size in determining the export threshold for the given industry. In this respect, Figure 3 shows the position of each industry according to the relative relevance of productivity and economic size in determining the export threshold.

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

Figure 3. Relative importance of productivity and economic size in determining the "export thresholds", by industries (*Effect of productivity – economic size – for industry i minus effect of productivity – economic size – for whole manufacturing*)

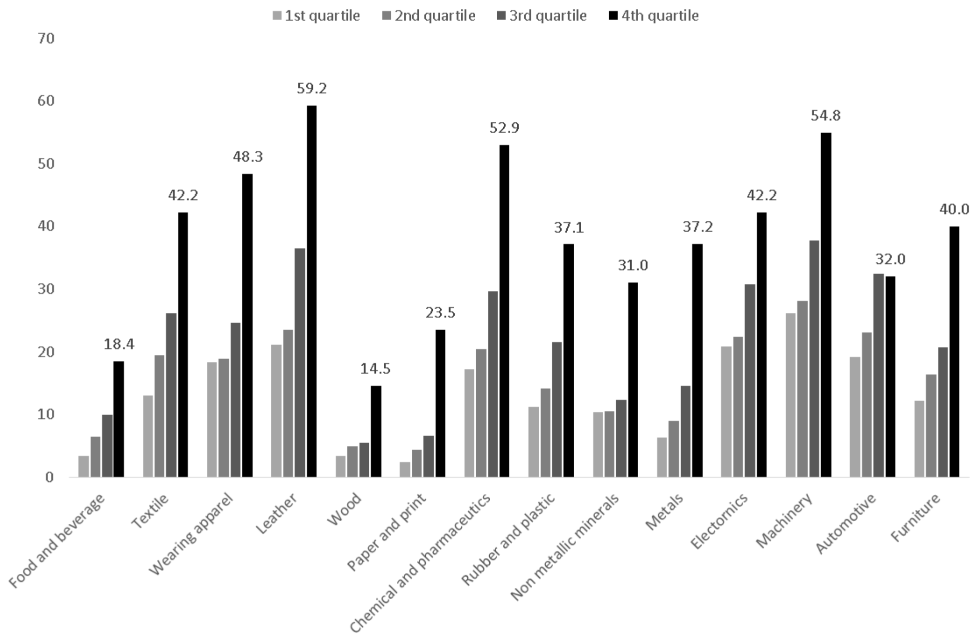


Source: Authors' calculations on Istat data

The capability to export for firms in industries laying in the first quadrant (i.e. leather and wearing apparel) depends on higher-than-average levels of both productivity and economic size. On the opposite, lower-than average levels of both variables characterises industries laying in the third quadrant (e.g. metals, chemical and pharmaceutics). This implies that the conditions for these firms to export appear structurally more favourable with respect to the ones prevailing in the other industries. In the second and the fourth quadrants, instead, the export threshold results more sensitive with respect to one or the other of the two variables. In particular, internationalisation appears to be productivity-driven for industries laying in the fourth quadrant (e.g. Machinery, Textile). In other terms, given the technology prevailing in the Italian manufacturing system, for firms operating in these industries to reach the export threshold an increase in productivity would result more effective than one in economic size. A symmetric picture characterizes industries laying in the second quadrant (i.e. Electronics and especially Automotive). These are more economic size-driven activities, meaning that to reach the threshold a growth in size-related variables (workforce, turnover, capital intensity) may turn out to be more effective than an increase in labour productivity.

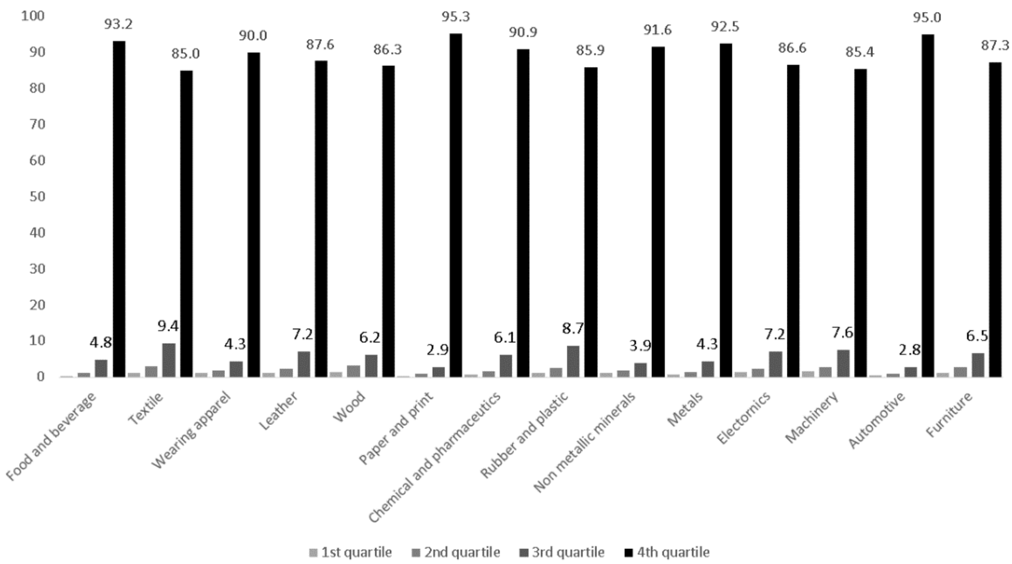
A remarkable heterogeneity also characterises the distribution of firms across the export thresholds within every industry. Indeed, productive units above the export threshold show relevant differences in terms of size, productivity, export (even higher than that characterizing below-threshold units). Furthermore, as reported in Figure 4, in all industries (with the exception of Automotive) for half of firms (2nd quartile) the export-to-turnover ratio is below 25%, confirming the stylized fact that the Italian business system is characterized by a relatively low intensive margin (Berthou *et al.*, 2015; ISTAT, 2017). Moreover, the firms most distant from the export threshold (4th quartile) are notably the most export-oriented (with the exception of Automotive), with the largest increase in the intensive margin occurring between the 3rd and 4th quartiles. The very low values – even in correspondence of the 4th quartile – observable in Wood, Food and beverage, and Paper and print basically reflects the typically low propensity of these industries to sell abroad.

Figure 4. Export propensity of firms above the export threshold, by manufacturing industry (*Average of export-to-turnover ratio by quartiles of distance from the export threshold, percentage values*)



Source: Authors' calculation on Istat data.

Figure 5. Share of total export accounted for by firms above the export threshold, by manufacturing industry and quartiles of distance from the export threshold (*Percentage values*)



Source: Authors' calculation on Istat data

The role of firms laying farther above the threshold (4th quartile) is significant also in terms of their weight on total manufacturing export (Figure 5): in all industries the share of export of these top exporters ranges from 85% (Textiles, Machinery) to 95% (Paper and print, Automotive). On the one hand, this result confirms that exporting firms tend to be very different from each other in terms of size, productivity and the value of their exports, with a minority of them which is generally responsible for the bulk of total trade (Meyer and Ottaviano, 2007) or for a substantial contribution to the business cycle co-movements among countries (the "granularity approach" originated by Gabaix, 2011 and further developed, among others, by Acemoglu *et al.*, 2012; and di Giovanni *et al.*, 2018). On the other hand, it has to be reminded that the top quartile of firms above the export threshold includes over 17,000 units. This reflects the fact Italy actually has a high number of exporting firms (over 195,000 in 2016; more than 88,000 in manufacturing), even though they account for less than 5% of the total number of enterprises (22.8% in manufacturing). The distribution of exports (especially in manufacturing industries) is less right-skewed than that observed in other advanced countries: according to the official trade statistics, in 2016 the top 100 Italian exporters accounted for 27% of Italy total export, against 57.4% in Germany and 50.6% in France (the top 10 exporters shares were 10.8%, 25.3% and 44.5% respectively). Furthermore, while in main EU countries the export is mostly due to large firms (i.e. units with no less than 250 workers), in Italy the relatively high contribution of medium-sized firms (50-249 workers) stands out (Bugamelli *et al.*, 2018).

4. Definition of the "Technology line"

The export threshold points out the minimum combination of productivity and economic size that manufacturing firms needs to acquire in order to become exporters. Furthermore, the methodology underlying the definition of this threshold also provides – on the basis of the composite indicator Z – an instrument to measure every firm's distance from that minimum combination, therefore allowing for a map of how, in each industry, business units distribute across the export threshold.

It is now possible to dig deeper into the analysis of the positioning of firms by explicitly considering the technological conditions prevailing in each industry, in order to assess each firm distance from the export threshold also in the light of its own position within the technological pattern of the industry. To do so, we identify the firm-level combination of productivity and economic size, which corresponds to the median level of technical efficiency.[9] We define this combination as "technology line".

The literature has clearly showed that firms export orientation is closely linked to their technical efficiency (see, among others, Verschelde et al., 2016), namely the firms choices about the joint use of productive factors. In this paper, following a well-established approach (Aigner et al., 1977; Meeusen and Van den Broeck, 1977), firms efficiency is measured on the basis of a model of stochastic production frontier which estimates the level of value added a firm is able to generate given its factors endowment. More in detail, a logarithmic transformation of the Cobb-Douglas production function is estimated (the use of different functional specifications, such as the translogarithmic one, led to similar results) having the added value as a dependent variable and the input of labour and the value of depreciation (as a proxy of the input of capital) as explanatory variables. Error decomposition has been estimated following the model by Battese *et al.* (1998).[10] Successively, a logit model of the probability for a firm to have an efficiency level higher than the median value of the industry is estimated, using the same covariates and controls as in Equation [5]:

$$\text{Prob (Tech. Efficiency} > \text{Median} \mid \mathrm{S}, \pi, \mathrm{G}, \mathrm{I}) = \Lambda(\alpha_1 \mathrm{E} + \alpha_2 \pi + \alpha_3 \mathrm{G} + \alpha_4 \mathrm{I}) \qquad [7]$$

Applying the coefficients of covariates E and π to the values of economic size and productivity of the firm identifying the export threshold (E^e and π^e, respectively), we obtain a new composite indicator Z^t,

[9] Throughout this paper the terms "efficiency", "technical efficiency", "productive efficiency" are considered as synonyms.
[10] See also Kumbhakar and Lovell (2000).

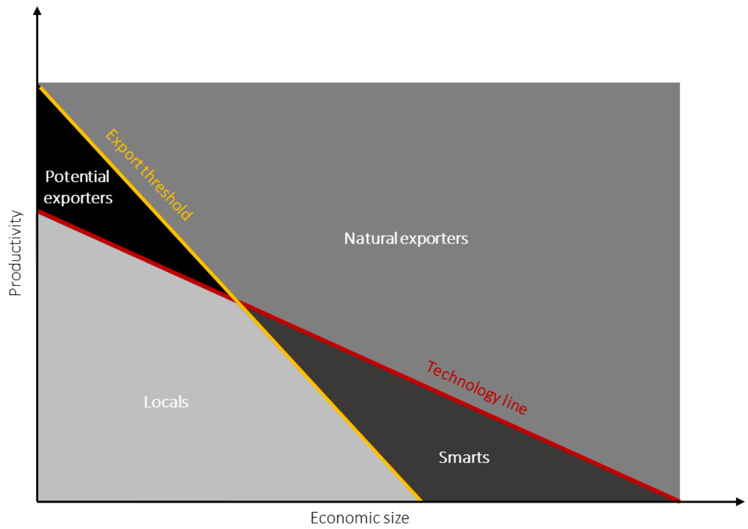
indicating the combination of economic size and productivity which is representative of the technology prevailing in the given industry. Formally, then, the "technology line" is:

$$Z^t = \widehat{\alpha}_1 E^e + \widehat{\alpha}_2 \pi^e \qquad [8]$$

5. Mapping the business system: a new taxonomy of firms according to their export orientation

The interaction between the technology line and the export threshold allows to better qualify the comparison between exporting and non-exporting firms, to shed new light on the relative role of economic size and productivity in accompanying firms internationalization, and to provide new insights (and tools) for policies aimed at increasing the extensive margins of export.[11] In fact, within the space defined by the combination of productivity and economic size axes that interaction ideally defines four areas as depicted in Figure 6.

Figure 6. The taxonomy of firms export orientation



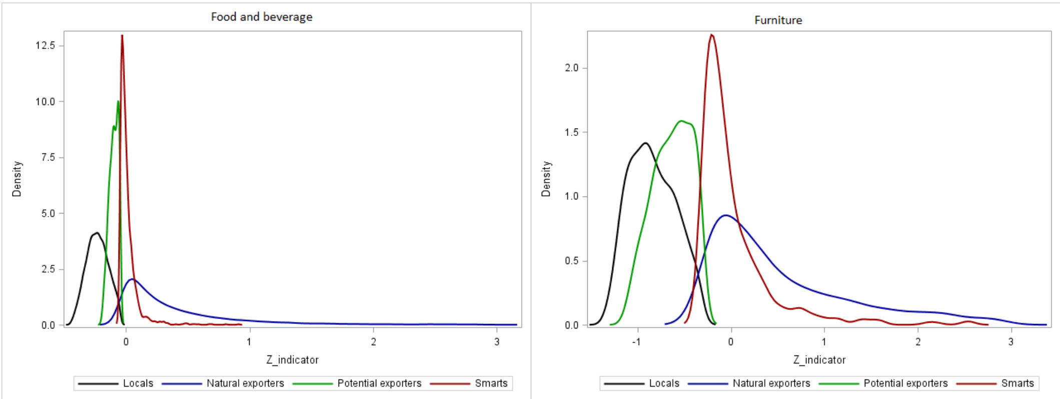
Depending on how the firms distribute across these areas, it is possible to derive the following four-class taxonomy of firms export orientation.

- "Natural exporters": firms with values of Z (combination of productivity and economic size) higher both than Z^e (the export threshold) and Z^t (the technology line). These units are productive and/or "large" enough to produce efficiently and export.
- "Smarts": firms with values of Z higher than Z^e but lower than Z^t. These units are classified as exporters notwithstanding their combination of productivity and size corresponds to an efficiency lower than the median value of their own sector.
- "Potential exporters": firms with values of Z lower than Z^e but higher than Z^t. These units have levels of productivity and size consistent with an over-the-median efficiency, but insufficient to export.
- "Locals": firms with values of Z lower than both Z^e and Z^t. These units do not reach the minimum combinations of productivity and economic size required to be relatively efficient or to export.

[11] In Appendix B we reports kernel density graphs of the distribution Z indicator for exports and technology and the value of Z^t and Z^e for each industry.

The close relationship between the Z indicator and firms' export orientation is still valid for this taxonomy. As Figure 7 shows, in all industries the values of the Z indicator tend to increase as we move from Locals to Natural exporters, also confirming the limited overlap between curves related to firms above the export threshold (i.e. Smarts and Natural exporters) and below it (i.e. Locals and Potential exporters). Moreover, this framework further specifies the previous analysis by clarifying what targets, in terms of size-productivity combinations, firms and policy-makers need to point at to enhance firms participation in international markets.

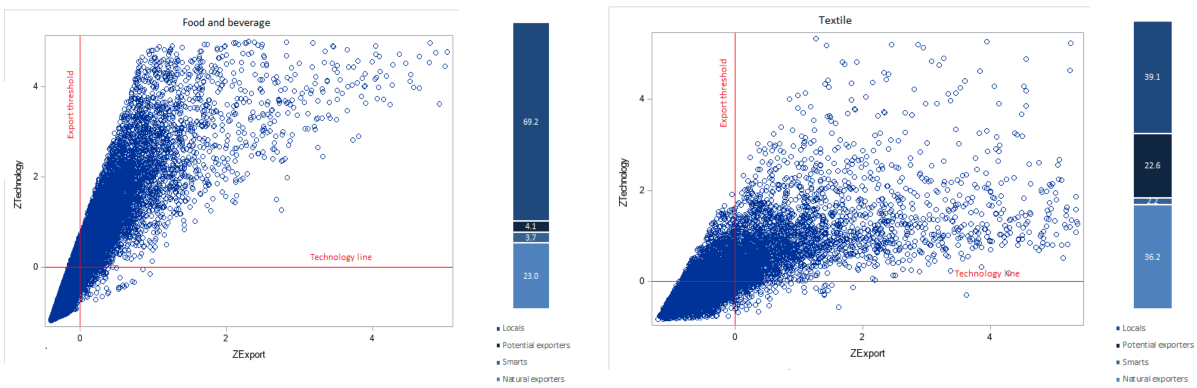
Figure 7. Z indicator and interaction between export threshold and technology line, by classes of firms[12]



Source: Authors' calculation on Istat data

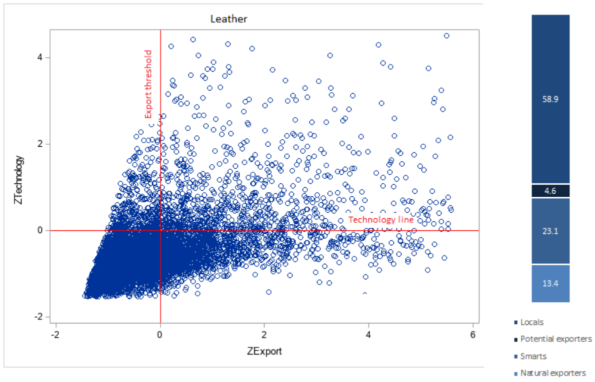
The distribution of firms in the four classes are plotted in Figure 8 according to their respective values of the Z indicators ($Z_{h,i}$) referring to export and technology. The noticeable heterogeneity among the exporters (both potential and natural ones) clearly emerges. Moreover, in all industries, the class of Locals tends to outnumber the others with the exceptions of Machinery and Chemical and pharmaceutics, that is 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).

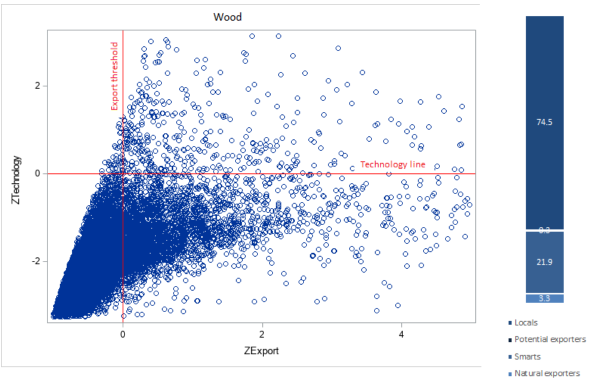
Figure 8. Mapping business system with respect to the interaction between export threshold and technology line

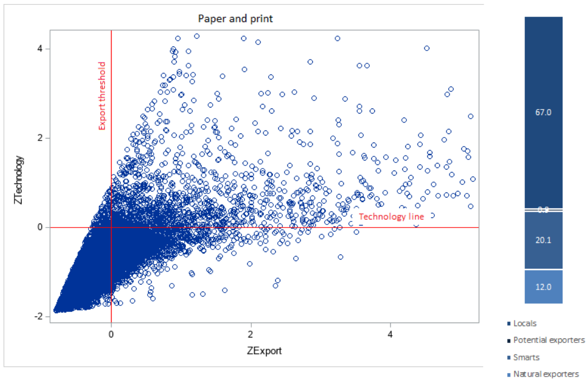


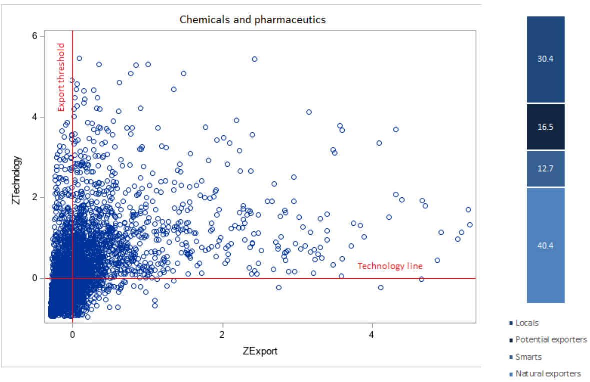
[12] We included in the text only Food and beverage and Furniture. Figures for all industries in Appendix C.

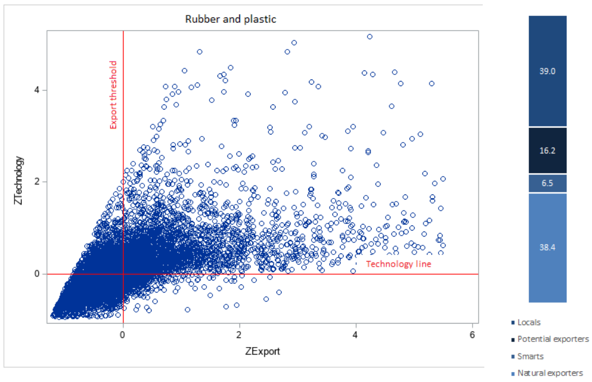

Wearing apparel
ZTechnology
ZExport
Export threshold
Technology line
38.3
24.7
9.7
27.2
Locals
Potential exporters
Smarts
Natural exporters

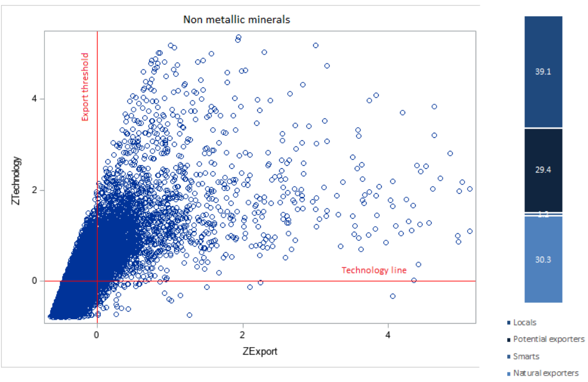

Leather
ZTechnology
ZExport
Export threshold
Technology line
58.9
4.6
23.1
13.4
Locals
Potential exporters
Smarts
Natural exporters

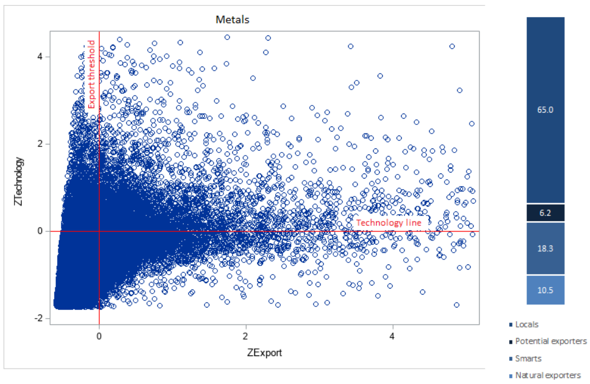

Wood
ZTechnology
ZExport
Export threshold
Technology line
74.5
21.9
3.3
Locals
Potential exporters
Smarts
Natural exporters

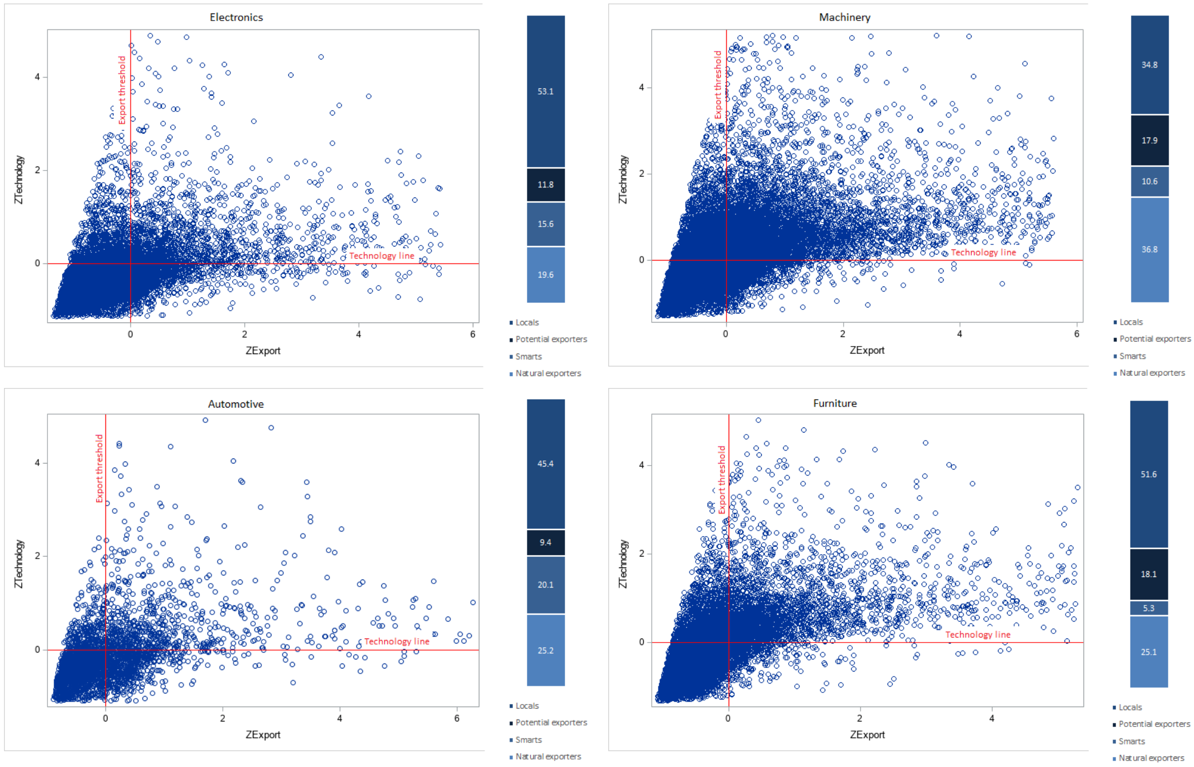

Paper and print
ZTechnology
ZExport
Export threshold
Technology line
67.0
20.1
12.0
Locals
Potential exporters
Smarts
Natural exporters


Chemicals and pharmaceutics
ZTechnology
ZExport
Export threshold
Technology line
30.4
16.5
12.7
40.4
Locals
Potential exporters
Smarts
Natural exporters


Rubber and plastic
ZTechnology
ZExport
Export threshold
Technology line
39.0
16.2
6.5
38.4
Locals
Potential exporters
Smarts
Natural exporters


Non metallic minerals
ZTechnology
ZExport
Export threshold
Technology line
39.1
29.4
30.3
Locals
Potential exporters
Smarts
Natural exporters


Metals
ZTechnology
ZExport
Export threshold
Technology line
65.0
6.2
18.3
10.5
Locals
Potential exporters
Smarts
Natural exporters



Source: Authors' calculation on Istat data.

Table 4 reports the some descriptive related to the different classes by industry. In all industries, Locals and Natural exporters are the classes with the highest weight in terms of both value added and turnover, this suggesting a strong polarization, in terms of both number of observation and economic relevance. Taking into account the average value of size and productivity, in almost all industries the average size and productivity tend to increase moving from Locals to Natural exporters, passing by Smarts and Potential exporters. In this context, it is interesting to note that the distance between the values of Potential and Natural exporters tends to be explained by size rather than productivity. Finally, the last indicator, which approximate the degree of internationalization of the groups in which eventually firms are involved. In this case, the trend of the average value of the indicator monotonically increase moving from Locals to Natural exporters, this involving that Potential exporters are more involved in MNEs strategies with respect to Smarts.

From analytical and policy-making points of view, the most important groups of firms are those of Smarts and Potential exporters. The formers, which manage to export despite size-productivity combinations lower than the technology line, are numerous especially in Leather, Wood, Automotive, Paper and print and Metals, that is in activities among which there are some important industries of the Italian specialization models. Only partially is this class of firms linked to the belonging to multinational groups: in every industries their incidence among Smarts firms is below 15%. The characteristics of these firms rather seem to identify the "bright stars" of the Italian internationalized firms.

The Potential exporters are even more linked to the typical activities of Italian manufacturing, as they are relatively numerous (with shares ranging from 18 to over 29%) in Furniture, Textile, Wearing apparel, Machinery and Non metallic minerals. This class is the most relevant for policy purposes, because represent the target that measures aiming at increasing the number of exporting firms (i.e. to stimulate domestic units to cross the export threshold) should actually focus on. These firms, in fact, have levels of productivity and economic size corresponding to a relatively high technical efficiency, but for some reason they do not export. Independently from the possible influence of factors different from productivity and size (also exogenous to the firms), such as lack of demand, finance constraints, regulation barriers and, as we stressed, peculiar intra-group strategies, our analysis indicates that the efficiency of these units is sufficient for them to thrive on domestic market, but not to successfully sell abroad. They actually appear to need an increase of productivity, of size, or both, depending, among others, on the technological characteristics of their respective industries, and on how much, in each industry, the ability to reach the export threshold depends on each one of the two factors.

Table 4. Characteristics of firms by typology and industry

Industry	Taxonomy	Value added (Shares of total industry)	Turnover (Shares of total industry)	Employment (Mean, Workers)	Labor productivity (Mean, thousand euro)	Internationalization of group indicator
Food and beverage	Locals	39.5	37.4	7.8	26.2	0.033
	Smarts	6.2	5.8	7.7	27.9	0.074
	Potential Exporters	21.1	22.6	17.2	39.9	0.153
	Natural exporters	33.2	34.2	16.0	41.5	0.427
Textile	Locals	35.3	33.3	9.1	33.9	0.036
	Smarts	8.3	8.4	10.7	38.0	0.057
	Potential Exporters	19.5	19.6	18.9	41.3	0.286
	Natural exporters	36.9	38.7	19.2	42.1	0.480
Wearing apparel	Locals	43.8	43.3	10.1	28.0	0.018
	Smarts	7.0	6.6	8.9	29.9	0.049
	Potential Exporters	19.5	18.6	19.5	34.5	0.150
	Natural exporters	29.8	31.6	20.3	34.9	0.352
Leather	Locals	48.0	47.3	12.0	32.8	0.034
	Smarts	8.4	9.8	12.9	33.8	0.092
	Potential Exporters	16.9	16.4	20.0	40.2	0.257
	Natural exporters	26.7	26.5	20.8	41.0	0.599
Wood	Locals	41.3	39.9	4.3	26.8	0.039
	Smarts	8.0	7.7	5.4	27.9	0.089
	Potential Exporters	18.0	18.2	7.5	30.3	0.235
	Natural exporters	32.8	34.1	8.6	32.5	0.330
Paper and print	Locals	38.5	37.3	8.7	33.4	0.059
	Smarts	7.2	7.4	8.8	35.4	0.133
	Potential Exporters	17.3	17.8	15.7	41.3	0.268
	Natural exporters	37.0	37.6	16.9	41.7	0.671
Chemicals and pharmaceutics	Locals	31.2	29.9	32.8	68.8	0.146
	Smarts	11.7	9.3	39.2	84.5	0.430
	Potential Exporters	17.4	16.1	48.6	80.0	0.525
	Natural exporters	39.7	44.7	62.1	85.6	1.369
Rubber and plastic	Locals	32.3	32.0	15.9	48.3	0.104
	Smarts	7.7	7.8	17.7	50.1	0.173
	Potential Exporters	21.4	20.9	29.5	55.3	0.320
	Natural exporters	38.6	39.3	29.8	56.0	0.752
Non metalic minerals	Locals	35.2	35.8	9.0	31.5	0.047
	Smarts	6.2	6.4	9.4	33.9	0.073
	Potential Exporters	25.0	24.1	18.6	41.7	0.416
	Natural exporters	33.5	33.6	16.6	40.0	0.544
Metals	Locals	41.1	39.2	9.9	39.6	0.061
	Smarts	7.9	7.5	10.6	44.7	0.131
	Potential Exporters	17.1	15.7	16.4	46.8	0.380
	Natural exporters	33.9	37.6	18.3	47.5	0.655
Electronics	Locals	35.1	34.5	18.1	48.7	0.131
	Smarts	5.9	5.6	15.1	53.5	0.310
	Potential Exporters	15.6	15.2	28.7	52.9	0.517
	Natural exporters	43.4	44.8	43.4	52.2	1.077
Machinery	Locals	38.0	38.7	19.4	54.7	0.115
	Smarts	7.2	7.2	17.1	57.4	0.231
	Potential Exporters	18.0	17.0	32.3	63.3	0.353
	Natural exporters	36.8	37.2	33.3	62.7	0.847
Automotive	Locals	30.7	44.0	58.2	47.9	0.126
	Smarts	6.3	6.6	33.2	47.3	0.388
	Potential Exporters	38.4	27.7	185.8	54.2	0.553
	Natural exporters	24.6	21.7	84.2	53.0	1.136
Furniture	Locals	39.3	38.4	7.8	30.5	0.047
	Smarts	10.3	10.4	8.9	33.2	0.067
	Potential Exporters	18.7	18.3	14.0	35.1	0.245
	Natural exporters	31.7	33.0	14.9	35.1	0.415

Source: Authors' calculation on Istat data

In this respect, further insights come from Figure 3 above, which helps identify which variable is to be stimulated in order to most effectively induce Potential exporters to actually export. In particular, according to the approach here developed, a possible measure addressing Potential exporters of Textile, Machinery and Non metallic minerals should primarily foster a growth in productivity. In turn, given the technology prevailing in the Italian manufacturing, to increase the extensive margin in the Furniture industry a more comprehensive action may be needed, as the effectiveness of both productivity and size is relatively low in promoting the cross of the export threshold. On the opposite, for Potential exporters of Wearing apparel industry a recovery of economic size and/or labour productivity could be very effective to become actual exporters.

6. Conclusions

In this paper, we use, and further develop, a ROC-based approach that we introduced in a previous work to measure the role of firms' productivity and economic size in determining their ability to access international markets. More in detail, we provide a methodology that allows to cluster business units according to their export orientation and capability, so that it becomes possible to distinguish, beside the "natural exporters" and the somehow "natural domestic" units, what firms are able to export despite their relatively (within the sector) low efficiency, and, even more importantly, what domestic firms have more potential to enter international markets.

To do so, we firstly estimate, for every industry, the "export threshold", defined as the firm-level minimum combination of (levels of) productivity and economic size (the latter defined on the basis of a firm's turnover, workers, capital, and age) corresponding to the transition from the non-exporter to exporter status. Like in our previous paper, this provides a map of how far every firm lays from the export threshold, based on the two factors (productivity and economic size) that the literature on trade revealed as the main determinants of export. Successively, we introduce the "technology line", that is the estimated combination of productivity and economic size which corresponds to a median level of technical efficiency within the industry. In this case, the methodology allows for a map of the firms distribution according to their efficiency levels permitted by their size-productivity combination. The interaction between the export threshold and the technology line offers new promising insights both on the characteristics of Italian firms' participation in international markets, and on the needed and possible policy measures to encourage an increase in extensive margin.

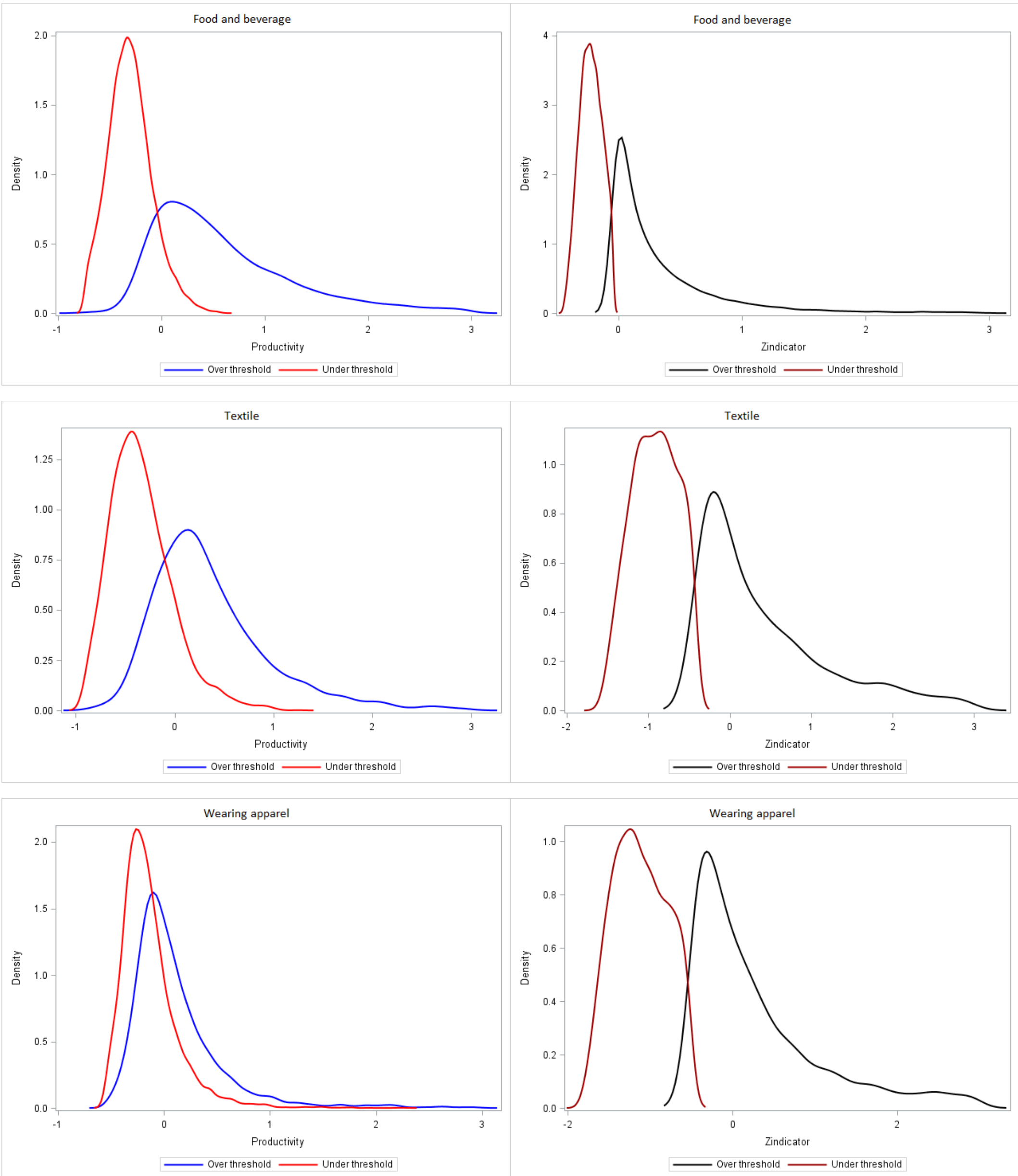
In particular, the methodology here proposed is especially useful with reference to two aspects: *a)* the detection of what units are to be taken as relevant targets for policies aimed at fostering firms internationalization; *b)* the indication of which lever – productivity or economic size – is better to stimulate in order to maximize the effectiveness of those policies.

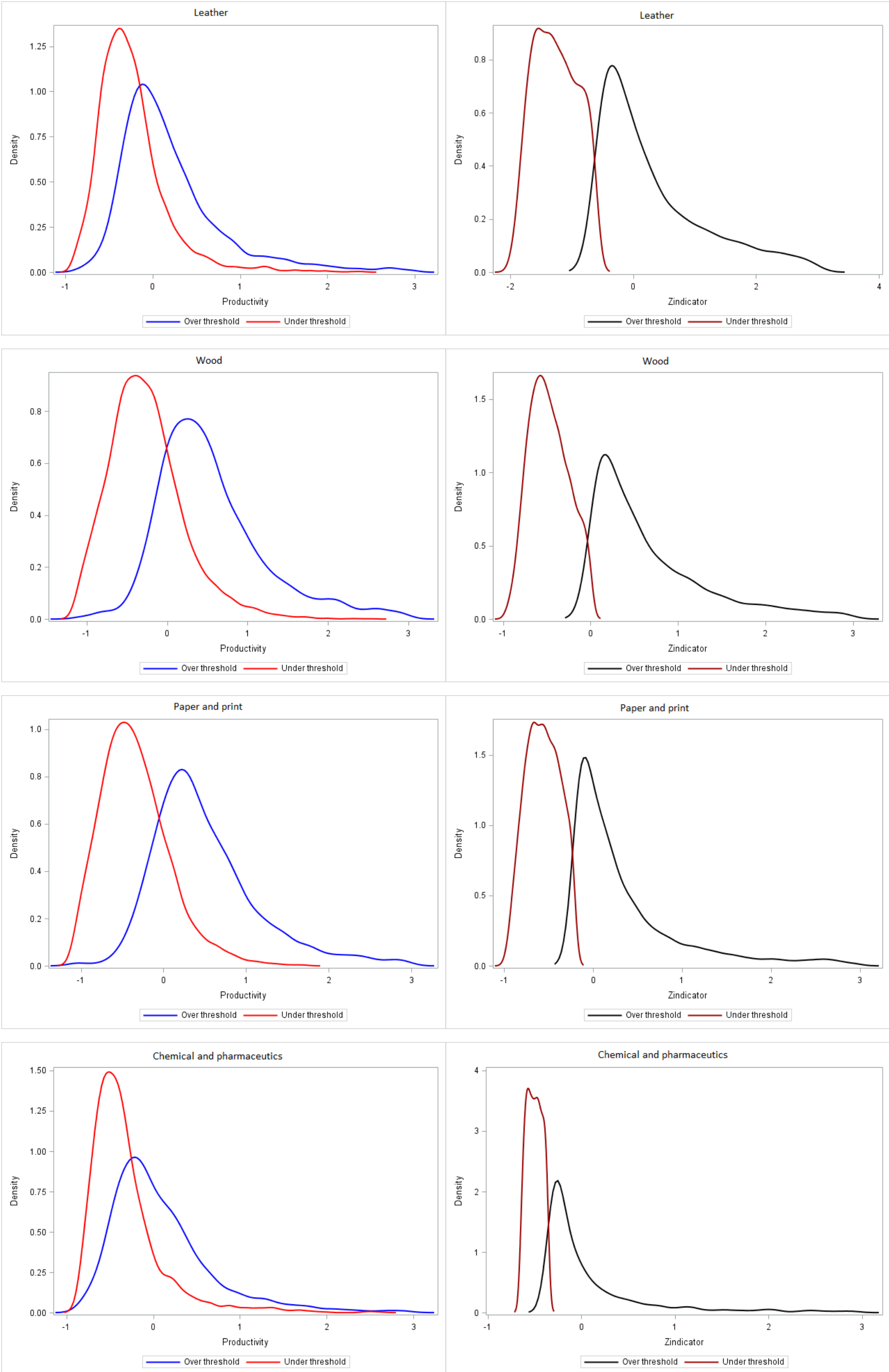
As for the first aspect, the interaction between the two "frontiers" allows to identify, among the domestic enterprises, the Potential exporters, namely those units whose levels of economic size and productivity are consistent with a relatively high efficiency within the industry (i.e. the firms lay over the technology line) but appear insufficient to export (i.e. the firms lay under the export threshold). This is the very set of business units on which policies aimed at fostering firms internationalization should focus.

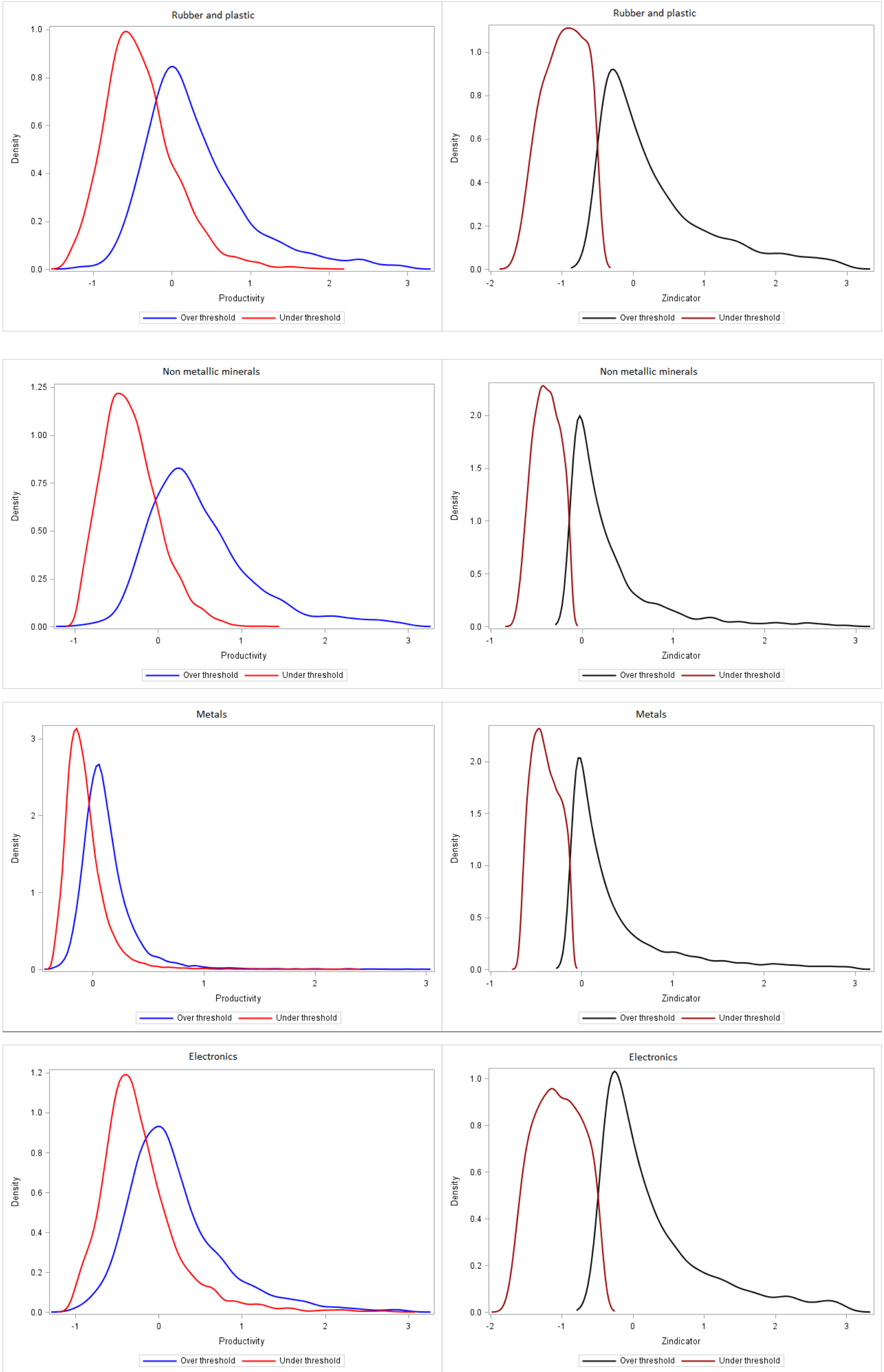
With regard to what type of intervention to implement, this will be different depending on the characteristics of the industry considered, particularly on whether the export threshold results more responsive to an increase in productivity or in size. In the first case (that involves some important industries of the Italian specialization model), a recovery in labour productivity should be encouraged; in the second one the policies should promote firms growth.

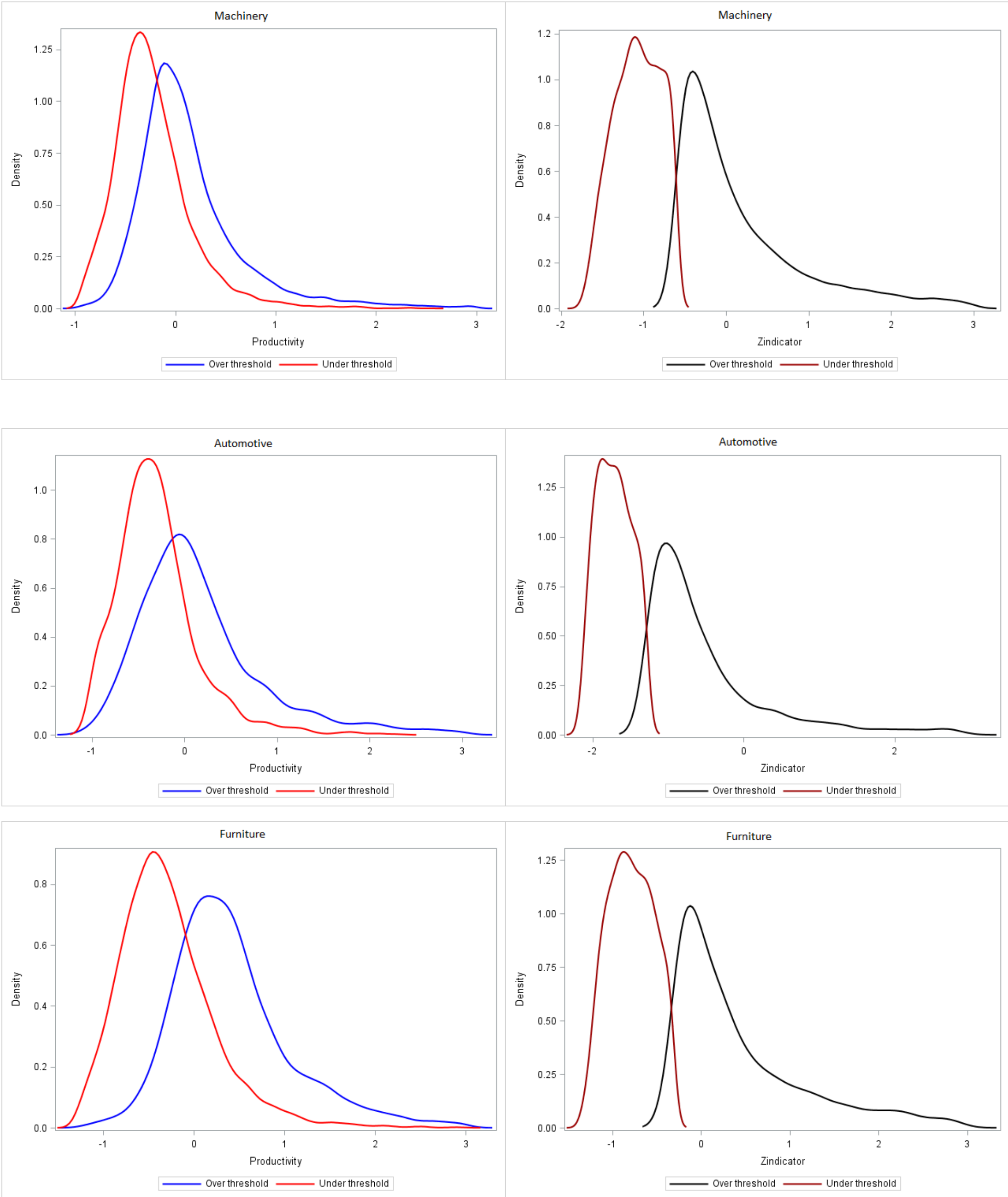
Finally, our approach may be further extended in several ways. Among the most promising ones, our indicator Z can be augmented taking into consideration "exogenous" barriers to the internationalization of firms that could be reduced by some type of policy intervention. This is the case, for instance, of different kinds of trade costs both beyond-national borders (such as transport costs, tariff and non-tariff regulatory measures, market access restrictions, trade finance availability), and in crossing borders (such as documentation and customs compliance requirements, lengthy administrative procedures and other delays, transport infrastructure and logistics). Taking into account these aspects would make it possible to calculate how much a reduction in these barriers would increase the number of "new" exporting firms by lowering the export threshold.

Appendix A. Kernel density graphs of labour productivity and Z indicator for firms over and under the export threshold

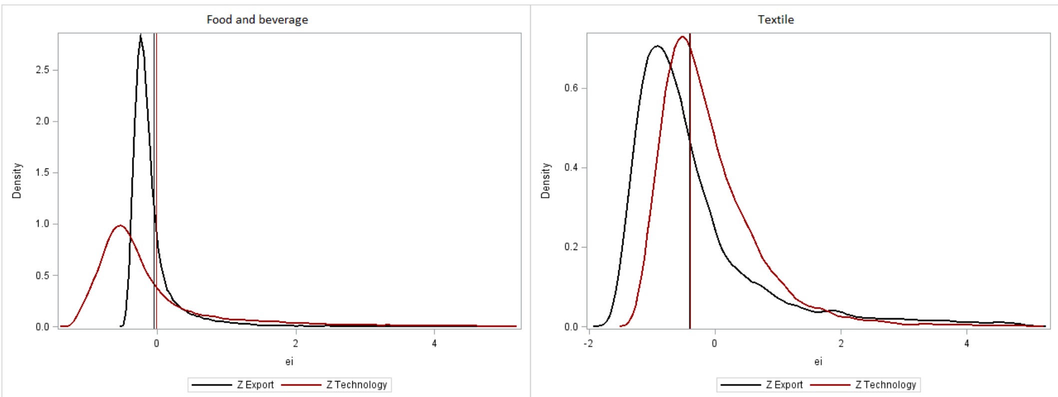


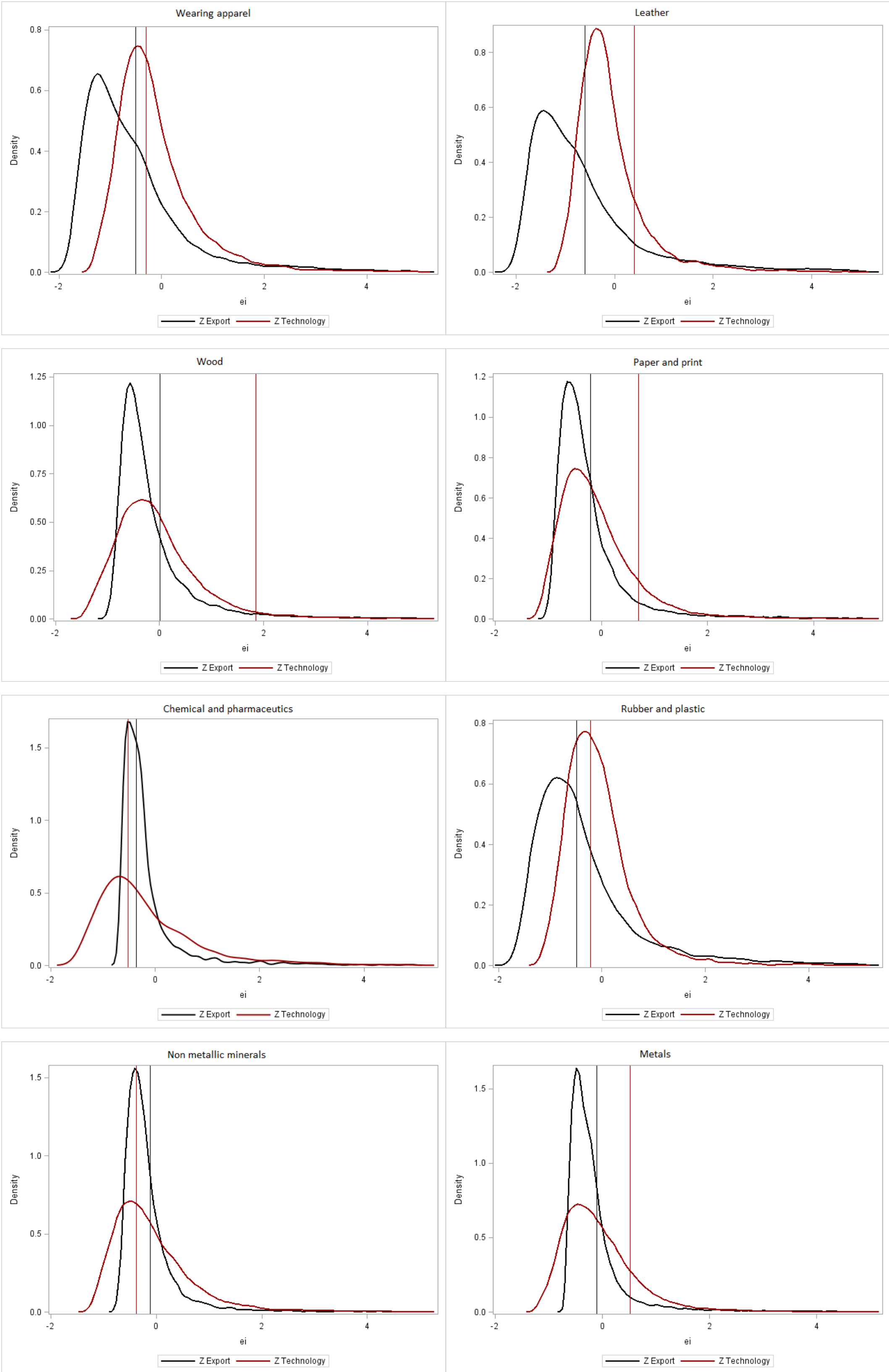

Leather
Density
Productivity
Over threshold
Under threshold
Leather
Density
Zindicator
Over threshold
Under threshold
Wood
Density
Productivity
Over threshold
Under threshold
Wood
Density
Zindicator
Over threshold
Under threshold
Paper and print
Density
Productivity
Over threshold
Under threshold
Paper and print
Density
Zindicator
Over threshold
Under threshold
Chemical and pharmaceutics
Density
Productivity
Over threshold
Under threshold
Chemical and pharmaceutics
Density
Zindicator
Over threshold
Under threshold

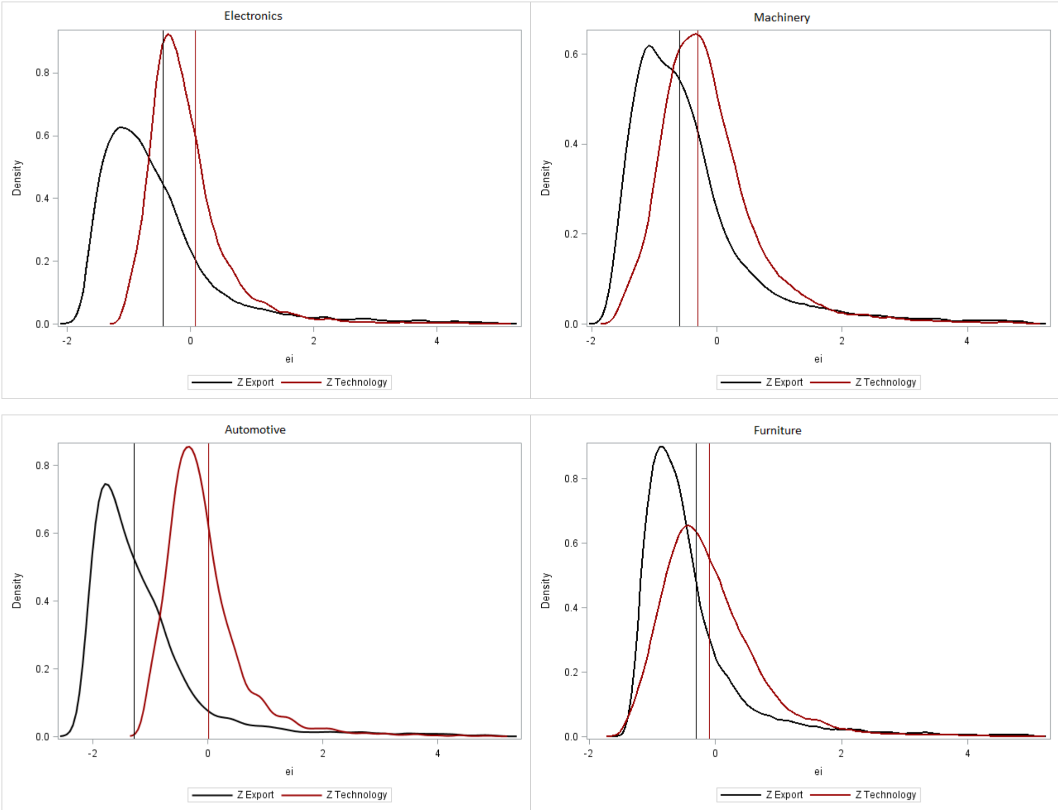

Rubber and plastic
Density
Productivity
Over threshold
Under threshold
Rubber and plastic
Density
Zindicator
Over threshold
Under threshold
Non metallic minerals
Density
Productivity
Over threshold
Under threshold
Non metallic minerals
Density
Zindicator
Over threshold
Under threshold
Metals
Density
Productivity
Over threshold
Under threshold
Metals
Density
Zindicator
Over threshold
Under threshold
Electronics
Density
Productivity
Over threshold
Under threshold
Electronics
Density
Zindicator
Over threshold
Under threshold



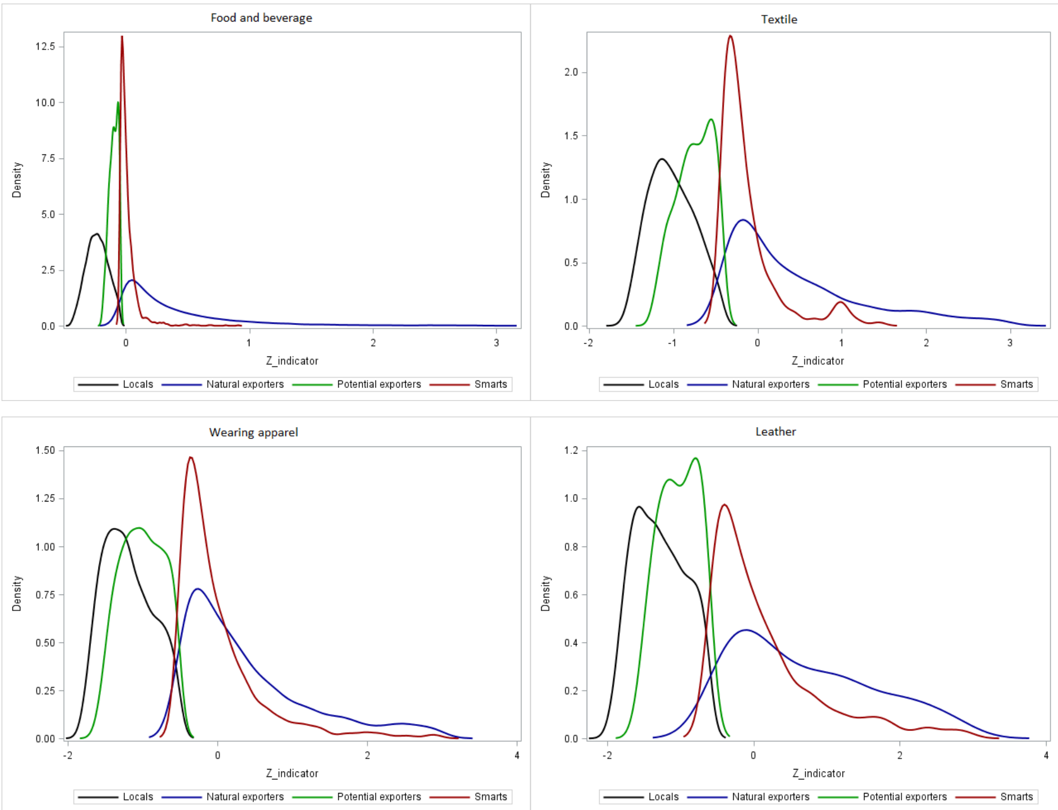
Appendix B. Kernel density graphs of the distribution Z indicator for exports and technology and the value of Z^t and Z^e for each industry

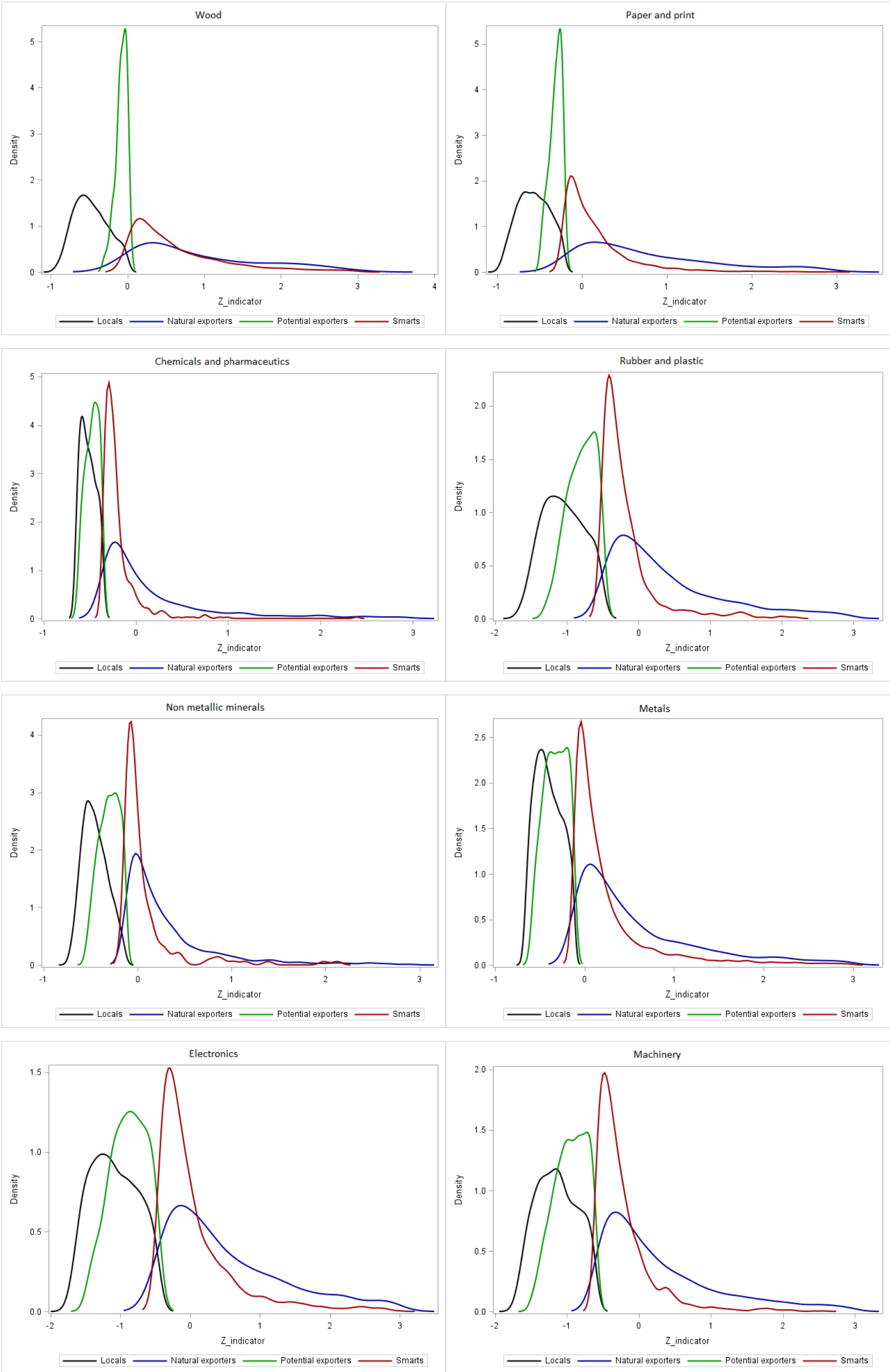


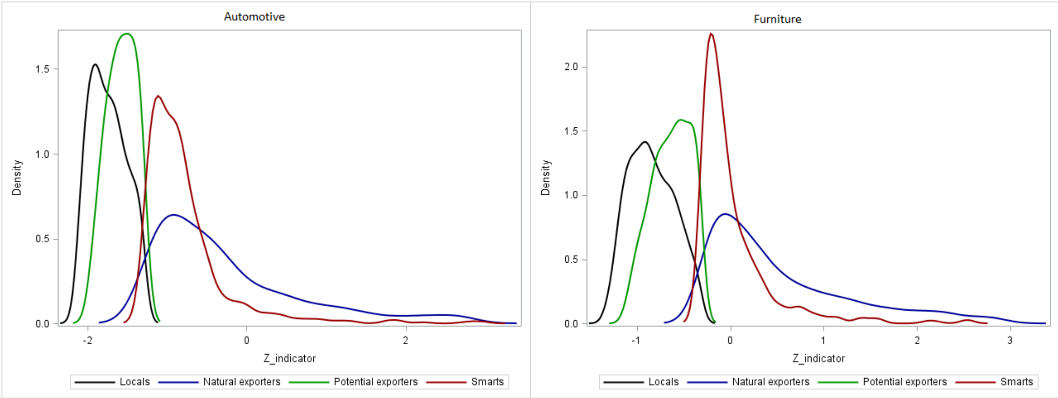

Wearing apparel
Leather
Wood
Paper and print
Chemical and pharmaceutics
Rubber and plastic
Non metallic minerals
Metals
Density
ei
Z Export
Z Technology



Appendix C. Kernel density graphs of the distribution of Z indicator and interaction between export threshold and technology line, by classes of firms




Wood
Paper and print
Chemicals and pharmaceutics
Rubber and plastic
Non metallic minerals
Metals
Electronics
Machinery
Density
Z_indicator
Locals
Natural exporters
Potential exporters
Smarts



References

Acemoglu D., V.M. Carvalho, A. Ozdaglar and A. Tahbaz-Salehi (2012), The network origins of aggregate fluctuations, *Econometrica*, 80(5): 1977-2016.

Aigner D.J., Lovell C.A.K., Schmidt P. (1977), "Formulation and Estimation of Stochastic Frontier Production Functions", *Journal of Econometrics*, vol. 6, pp. 21-37.

Battese, G.E., T.J. Coelli and D.S.P. Rao (1988), *An Introduction to Efficiency and Productivity Analysis*, Kluwer Academic Publishing, Boston.

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

Berthou, A., E. Dhyne, M. Bugamelli, A.-M. Cazacu, C.V. Demian, P. Harasztosi, T. Lalinsky, J. Meriküll, F. Oropallo, A.C. Soares (2015). "Assessing European firms' exports and productivity distributions: The CompNet trade module", *ECB Working Paper*, No. 1788, May.

Bugamelli, M., S. Fabiani, S. Federico, A. Felettigh, C. Giordano, A. Linarello (2018), Back on Track? A Macro–Micro Narrative of Italian Exports, *Italian Economic Journal*, March 2018, 4(1): 1-31.

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

Costa, S., F. Sallusti, C. Vicarelli, 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.

di Giovanni, J., A.A. Levchenko and I. Méjean (2018), The micro origins of international business-cycle comovement. American Economic Review, 108(1): 82-108.

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

Gabaix X. (2011), The granular origins of aggregate fluctuations. Econometrica, 79: 733-772.

Geishecker, I., P.J.H. Schröder and A. Sørensen (2017), Explaining the size differences of exporter premia: theory and evidence, *Review of World Economics*, *153* (2), 327-351.

ISTAT – Italian National Institute of Statistics (2017). *Report on competitiveness of business sectors* (in Italian), March.

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

Kumbhakar, S.C. and C.A.K. Lovell (2000), *Stochastic frontier analysis*. Cambridge University Press, Cambridge.

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

Mayer, T. and G.I.P. Ottaviano (2007), *The Happy Few: New Facts on the Internationalisation of European Firms*. Bruegel CEPR EFIM Report. Bruegel Blueprint Series.

Meeusen, W. and J. Van den Broeck (1977), “Efficiency Estimation From the Cobb-Douglas Production Functions with Composed Errors”, *International Economic Review*, vol. 18, pp. 435-444.

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.

Verschelde, M., M. Dumont, G. Rayp and B. Merlevede (2016), Semiparametric stochastic metafrontier efficiency of European manufacturing firms, *Journal of Productivity Analysis*, 45 (1), 53-69.

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.