Distance to Export: A Machine Learning Approach with Portuguese Firms

Paulo Barbosa
ISEG - University of Lisbon
Portugal Trade & Investment

João Amador
Banco de Portugal
Nova SBE

João Cortes
IST - University of Lisbon

10 September 2024

Abstract
This paper studies firms' distances to becoming successful exporters. The empirical exercise uses rich data on Portuguese firms and assumes that there are significant features distinguishing exporters from non-exporters. An array of machine learning models—Bayesian Additive Regression Tree (BART), Missingness Not at Random (BART-MIA), Random Forest, Logit Regression, and Neural Networks—are trained to predict firms' export probability and to shed light on the critical factors driving the transition to successful export ventures. Neural Networks outperform the other models and remain highly accurate when export definitions and training and testing strategies are changed. We show that the most influential variables for prediction are labor productivity and the share of imports from the EU in total purchases. Additionally, firms at the median distance to sell in international markets operate with about twice the assets of the group in the decile more distance from exporting. Firms in the decile closest to the export market operate with around 12 times more assets than those in the decile more distant from exporting.

JEL: F17, C53, C55, L21

Keywords: Machine learning, forecasting exporters, trade promotion, firm-level data, Portugal.

Acknowledgements: The opinions expressed in this paper are those of the authors and do not necessarily coincide with those of Banco de Portugal or the Eurosystem. The authors thank the Portugal Trade & Investment staff for their comments and insightful feedback; and to Henrique Pereira for his useful comments and suggestions. Any remaining errors and omissions are the sole responsibility of the authors. Address: Banco de Portugal, Economics and Research Department, R. Francisco Ribeiro 2, 1150-165 Lisboa, Portugal.

E-mail: paulofbarbosa5@gmail.com; jamador@bportugal.pt; cortes.joao97@gmail.com

1. Introduction

Many countries have national strategies to boost exports, ranging from economic diplomacy and trade agencies to direct grants and subsidies to firms. Firms can venture into international markets by self-financing the sunk costs or borrowing. Empirical trade literature suggests that it is beneficial for the economy when a large number of firms engages in international trade, as it is positively correlated with size, productivity, wages, and tax revenue (Wagner 2007; Coad and Vezzani 2019).

To support policies aimed to boost exports, a strand of literature argues that market size matters to understanding firms' productivity. Specifically, greater access to foreign markets positively affects firms' innovation and productivity: the *learning-by-exporting* argument. Consequently, governments' efforts to increase the share of domestic firms in international markets could lead to more investment in technology and innovation, thereby enhancing overall productivity and economic growth (De Loecker 2007; Harrison and Rodríguez-Clare 2010).

In this context, several authors provide evidence of governments' success in increasing the number of firms participating in international markets. Other authors highlight how imperfections in financial markets can make firms miss export opportunities, resulting in export levels that fall below their potential Chor and Manova (2012); Manova (2013); Srhoj and Walde (2020).

Several authors report evidence of the governments' success in increasing the share of firms operating in international markets (see Srhoj *et al.* (2023) for a survey). Cruz (2014) uses matching difference-in-differences to show that manufacturing firms supported by the Brazilian Export Promotion Agency have 2.5 times more probability of becoming successful exporters than the manufacturers that did not receive the government treatment. Besides, Broocks and Van Biesebroeck (2017) and Munch and Schaur (2018) studied government efforts to increase the number of exporters in Belgium and Denmark, respectively. The first study shows that the firms that received the support had an 8.5 percentage point (pp) higher probability of selling in an extra European Union market than those that had not received the treatment. The second paper reports evidence that, for small firms, the Danish government support increased the probability of selling abroad by 8 pp. Moreover, it reports that sales and value-added increased by 8 pp, and employment increased by 4 pp compared to firms that had never received the support.

In light of this evidence about learning-by-exporting, helping firms to sell overseas is a concern for policymakers. In this paper, we develop a machine learning (ML) model to estimate how distant manufacturing firms are to become successful exporters. In other words, our ML model estimates the probability that a non-exporting manufacturer sells in international markets. The intuition is that successful exporters have non-trivial characteristics that the algorithm can use for out-of-sample predictions. This type of approach is widespread in credit risk literature, with a range of models evaluating how distant the firm is

from bankruptcy (Altman (1968) and Merton (1974) are the precursors of this strand of research). In the same vein, we aim to understand which manufacturers' characteristics are more important in predicting those that become successful exporters and how these variables relate to the outcome, i.e., the probability of selling in international markets.

Our paper is close to Micocci and Rungi (2023), which utilizes ML models to measure French firms' distance-to-export and to predict the rise of healthy exporters. This paper uses a database that considers about 30% of French manufacturers, but only medium and large firms. In contrast, our paper uses a rich dataset containing virtually the universe of Portuguese manufacturers, with information on their financial and export status from 2010 to 2021. Therefore, we can predict the export probability of manufacturers of all sizes, while corroborating previous results in the literature.

Regarding the methodology, our first step is randomly splitting the sample into training and test data sets, with a proportion of 0.8 and 0.2, respectively. Then, we train our ML models on the manufacturers' dataset with exporters and non-exporters. We use an array of ML techniques to predict successful exporters, namely Bayesian Additive Regression Tree (BART), BART with Missingness not at Random (BART-MIA), Random Forest, Logit and Neural Networks (NN). Next, we select the best ML technique to predict exporters, i.e., the model that predicts out-of-sample with higher accuracy. It turns out that the NN outperforms the other techniques for predicting exporters.

Equipped with the NN model that shows the highest accuracy in our empirical exercise, we estimate the exporting scores for each non-exporter manufacturer. These scores provide a clear indication of the likelihood of a firm successfully engaging in international markets. Our conclusion is striking: the majority of Portuguese manufacturers are currently far from becoming exporters, as evidenced by the large density of firms with a low exporting score.

For the sake of robustness, we subject our ML models to a comprehensive battery of tests. Even after altering both the exporter definition and the training and forecasting strategies, the estimated ML model maintains its high accuracy. Notably, the predictions generated with the NN technique, while different from those relying on regression and classification tree methods, are equally accurate. This adaptability underscores the robustness of results.

Another distinctive feature of the paper are results on the power of different indicators included in the battery of predictors, as initially proposed by Shapley (1953). The variables important for predicting exporters are labour productivity, whether it imports from the EU or extra-EU countries, its capital intensity and the average wage. More specifically, high labour productivity is positively associated with a high export probability (low risk level). Similarly, the existence of imports from EU or extra-EU markets, high capital intensity, and high wages are highly correlated with sales in international markets. In addition, manufacturers near the median risk operate with around two times more assets than those with higher risk,

and the firms with the lowest risk level operate with around 12 times more asset resources than the greatest risk firms.

Beyond adding to a still very scarce strand of literature, our empirical exercise can also be helpful to private or policy institutions, notably those that grant credit for internationalization and trade promotion agencies, to better target export incentive programs. With higher-quality information on risk, private funds can also reduce the price of credit. Trade promotion agencies can focus on helping firms with higher scores to sell abroad and allocate public funds to support manufacturers with a higher probability of becoming healthy exporters.

This paper is structured as follows. Next, we briefly overview the relevant literature on ML in international trade. We then describe the dataset and present some descriptive statistics. Next, we detail our identification strategy and the ML techniques used. In the following sections we present the results, including the robustness checks. We then interpret the estimated scores and relate the results to sectors and the size of the firm's assets. Finally, we present some concluding remarks.

2. Literature review

Our paper is at the interception of two research strands: i) the export promotion literature and ii) the utilization of ML techniques to address economic questions. We briefly survey the literature that lies in the neighborhood of this interception.

Sousa *et al.* (2008) reviews 52 articles published between 1998 and 2005 on the determinants of export performance, showing that most studies have focused on manufacturing firms, only some have included the service sector, and they mainly focus on small to medium-sized firms. The authors highlight a continuous increase in sample sizes, statistical sophistication, and in the number of control variables. The literature started to consider external environmental factors as critical determinants of export performance, such as domestic markets' characteristics. However, the increase in sample size and statistical sophistication allows for more robust empirical and theoretical frameworks, which can better test inherent principles and theories of international trade.

Many concepts related to deep learning and ML date back to the early 40s when Walter Pitts and Warren McCulloch created a computer model based on the human brain called NN. However, the application of ML to economics is still in its early stages of development. Tong (2023) highlights that the ML models more widely used in international trade literature are the grey prediction model, the parallel vector machine model, deep learning methods such as artificial neural networks, convolutional neural networks, and recurrent neural networks. These models have brought significant advancements in prediction accuracy to the field.

For example, Breinlich *et al.* (2022) evaluates the effects of trade policies (TP) and agreements (TA) on trade flows, examining both preferential and non-preferential trade relationships. The authors apply the latest ML techniques and

variable selection methods to quantify TA's impact on trade flows. They conclude that several provisions related to technical barriers to trade, anti-dumping, trade facilitation, subsidies, and competition policy are more effective at promoting trade than other provisions in preferential trade agreements. Similarly, Dai (2023) used a Back Propagation Neural Network (BPNN) to forecast the volume of exports, reaching a model that is 15% to 30% more accurate than traditional prediction methods. Pan *et al.* (2018) applies supervised ML to predict firms' success, defined as an event that gives a large sum of money to the company's founders, investors and early employees (e.g., an M&A or an IPO), by comparing the performance of different ML techniques such as K-NN, Random Forest and Logit regression.

Despite the most recent advancements in deep learning and ML, to the best of our knowledge, only one other paper has explicitly used those methods to estimate whether a firm will become a successful exporter and to measure its current distance to export. Indeed, Micocci and Rungi (2023) exploits ML techniques to evaluate whether and how close firms are to becoming successful exporters using financial information on French firms from 2010 to 2018. This conceptual approach is standard in the finance literature to derive a distance-to-default measure, as in the seminal contributions by (Altman 1968) and (Merton 1974).

Estimating how likely firms are to become exporters is also the primary purpose of our paper, but we try to go beyond that goal. Micocci and Rungi (2023) highlights the need to know about the power of different indicators included in the battery of predictors and how and why firms are in a condition to export. Although we agree that predictive models do not substitute structural economic models, policy evaluation methods or impact analyses, further insights are possible within the scope of the ML techniques. Therefore, we compute the power and impact of different indicators included in the battery of predictors using Shapley values (SHAP). This cooperative game theory concept corresponds to a variable's expected average marginal contribution after all possible combinations have been considered. This method explains the local importance of the variable and how it changes with lower or higher sample values. Importantly, SHAP provides the importance of each indicator in the forecast but does not evaluate the quality of the forecast itself.

3. Data

The paper is based on a rich and unique firm-level dataset obtained from merging two sources with a common anonymous firm identifier: i) a Portuguese transaction-level exports database and ii) a balance sheet and income statement dataset for Portuguese firms.

Statistics Portugal collects the export database, including detailed information on export and import transactions of goods by firms located in Portugal, which is used to compute the official international trade statistics. The data originally comes from customs returns forms for extra-EU trade, and for intra-EU trade, it

NACE rev.2	Code	Not selling abroad	Selling abroad	Total	%
Food products	10	6677	979	7656	12.8
Beverages	11	1278	601	1879	3.1
Textiles	13	2272	716	2988	5.0
Wearing apparel	14	5612	1192	6804	11.4
Leather and related products	15	2513	807	3320	5.6
Wood and prod. of wood and cork	16	3118	736	3854	6.5
Paper and paper products	17	488	227	715	1.2
Printing and rep. of recorded media	18	2245	606	2851	4.8
Coke and refined petroleum	19	30	7	37	0.1
Chemicals and chemical products	20	955	347	1302	2.2
Pharmaceutical products	21	223	64	287	0.5
Rubber and plastic products	22	1137	613	1750	2.9
Other non-metallic products	23	2707	1080	3787	6.3
Basic metals	24	346	150	496	0.8
Metal products except machinery	25	8546	2017	10563	17.7
Computer, electronic and optical	26	400	148	548	0.9
Electrical equipment	27	613	280	893	1.5
Machinery and equipment	28	1682	706	2388	4.0
Motor vehicle and trailers	29	529	241	770	1.3
Other transport equipment	30	314	89	403	0.7
Furniture	31	2987	1068	4055	6.8
Other manufacturing	32	1909	414	2323	3.9
Total		46581	13088	59669	100

Table 1. Number of firms by NACE rev.2 manufacturing sector in the database

comes from the Intrastat reports. The dataset includes all transactions with non-EU countries. However, the intra-EU transactions are subject to the assimilation threshold, i.e., the annual trading value below which Intrastat declaration is not mandatory. The assimilation thresholds in Portugal are 350 thousand euros for EU imports and 250 thousand euros for EU exports.

The Simplified Corporate Information (Informação Empresarial Simplificada - IES) is maintained by Statistics Portugal (Instituto Nacional de Estatística).The IES data set contains the annual mandatory information that firms must report to tax administration, social security and statistical authorities. The information consists of economic, financial, and accounting balances for the respective fiscal year and covers the population of Portuguese non-financial corporations. Firms report detailed balance sheet figures and information concerning other key variables such as the number of employees, cost of inputs and turnover.

The final dataset covers all manufacturers that may or may not have exported goods between 2010 and 2021. Table 1 presents manufacturers' distribution in the database along sectors using a NACE 2-digit classification. Note that the manufacturing sector with the most firms is "Fabricated metal prod., except machinery and equipment", with 10 573 firms (17.7%), followed by "Food products" with 7 656 firms (12.8%). The sectors with the smallest number of

firms are "Coke and refined petroleum" and "Pharmaceutical products" with 7 and 64 firms (0.1% and 0.5%), respectively.

4. Empirical strategy

Inspired by the credit risk literature, which computes firms' distance to default and credit risk (Altman 1968; Merton 1974; Angelini *et al.* 2008; Uddin 2021), as well as the contribution by Micocci and Rungi (2023), which predicts exporters with ML, we proceed to compute Portuguese manufacturers' distance to becoming successful exporters. To achieve this goal, we trained different ML models on the set of Portuguese manufacturers, including exporters and non-exporters, in 2010-2021. We randomly split the observations into two samples using the standard 80-20 proportion, based on the entity level. The first sample was used to train the models, and the remaining 20% observations were used for testing. It should be noted that we never mix firms for test and training groups, and observations for a firm in different periods are taken as independent.

Our empirical exercise assumes that information about firms' characteristics contains non-trivial information on their exporting abilities. Additionally, we assume that observations are independent, meaning that there is no temporal dependence between firm i at different moments of time t (we perform a series of robustness tests to assess the impact of this assumption on results). Then, we use the trained models to generate distributions of predictions for new data (out-of-sample predictions). These distributions help us assess how close manufacturers are to being able to export. The final model is chosen based on its accuracy in predicting outcomes for data it has not seen, i.e., our test sample. Therefore, after testing and comparing different ML methods, we applied the best-performing model to calculate the export scores of Portuguese manufacturers.

Figure 1 is a stylized illustration of what we expect to be the score distribution for non-exporters in the data. Most non-exporters are expected to have a low probability of becoming successful exporters, resulting in a high-density mass in the left low score segment. As illustrated in Figure 1 by the red solid horizontal arrow, our objective is to calculate the distance of the firm i to becoming an exporter (a score equal to 1).

Two difficulties arise from estimating exporting scores with ML methods: i) how to classify a firm as a successful exporter, and ii) what threshold to take from the estimated exporting scores into the test sample to perform the out-of-sample predictions.

As for the first difficulty, the empirical exercise aims to predict firms that consistently sell abroad, i.e., labeled as successful exporters. In our benchmark exercise, we define those exporters as firms that sell at least 10% of total sales abroad. Micocci and Rungi (2023) considers exporting firms to be those that sell abroad at any positive value. However, Geishecker *et al.* (2019) highlights that this definition may not be strict enough, as passive exporters may be included.

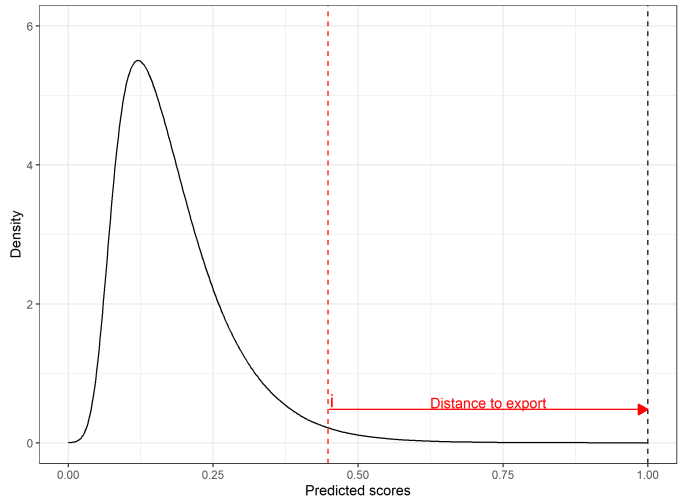


Figure 1: Visual illustration of the exporting scores of non-exporters

Note: The Figure shows the distribution of an illustrative simulated export score. This simulated random process is generated from a chi-squared distribution with three degrees of freedom bounded in an interval [0,1]. The red vertical dashed line represents the predicted score of a firm i, indicating its probability of success in international markets. The black vertical dashed line represents the benchmark to export, thus the horizontal solid red arrow represents distance firm i is to become a successful exporter.

In fact, these authors define passive exporters as firms that engage in exports only temporarily. As pointed out by (Békés and Muraközy 2012), micro and small manufacturers are more likely to participate passively in the export market, which is an important feature to consider given the prevalence of these firms in our database. As for the second difficulty, to assess the accuracy of the different models, we classify the firms in the testing sample as successful exporters when they exceed a threshold of 0.5. This value for the threshold is also used by Micocci and Rungi (2023) and Baier *et al.* (2014).

Addressing the above-mentioned difficulties also consists of repeating the exercises, taking different definitions of successful exporters, and using different proportions for the training and testing samples. Such robustness tests are presented in detail below, in Subsection 5.3.

4.1. Machine learning models

We train and compute predictions on the test sample with the six ML methods and choose the one that performs best to predict successful exporters. A generic representation of the ML predictions is as follows:

$$E(Y_i|\mathbf{X}_i) = P(Y_i = 1|\mathbf{X}_i), \tag{1}$$

where Y_i is the binary outcome, assuming the value 1 if the firm i is a successful exporter and 0 otherwise. $\mathbf{X}_i$ is a matrix that includes the k predictors used in the models. The output of Equation 1 is the *exporting scores*. The estimated model does not capture the time dimension, instead it considers a firm's export status with contemporaneous predictors. Therefore, in this baseline model, a firm can be considered an exporter in one year and a non-exporter in another. In Subsection 5.3, we look at dynamic export participation to assess the robustness of the results.

Following the conventional data splitting principle (Athey *et al.* 2021), 80% of the universe of Portuguese manufacturers are randomly selected as in-sample information for training the six ML models, and the remaining 20% are kept out-of-sample for predicting the export status. We perform predictions with six ML models: Logit regression, Logit lasso, Random forest, BART, BART-MIA and NN.

The NN model simulates many interconnected processing units resembling abstract neurons. These processing units are arranged in three-layer groups: input, hidden layers, and output. These units are connected with varying weights based on how the information is fed into the algorithm from the input to the output layer (Warner and Misra 1996). The NN model is trained in batches of observations, generating a prediction for each and adjusting weights based on its margin-of-error by iteration. This process is repeated until an optimal is reached (for a deeper discussion, see Hastie *et al.* (2009) and Angelini *et al.* (2008)). Figure 2 presents a stylized illustration of the NN model.

The best-performing NN includes four hidden layers, each ranging from 32 to 128 artificial neurons (nodes) that include a rectified linear unit activation function connecting each layer to the next with a relatively weighted impact defined during training. The model is compiled with a binary cross-entropy loss and an optimizer based on stochastic gradient descent's first and second-order moments. The model is trained in batches of 10 observations per iteration, with a validation split of 20% from the original training sample. The model is prepared to train up to 100 epochs (i.e., complete passes of the training data set through the algorithm) unless its maximum validation accuracy is reached earlier and, in that case, after eight additional epochs. Then, the best model weights are restored for prediction.

The BART, BART-MIA, and random forest methods are classification tree models. A classification tree is generally constructed using if-then statements that divide the training data based on predictor values, which allows for capturing non-linear relationships between predictors and outcomes. The algorithm for building a classification tree follows a top-down approach, recursively dividing the primary sample into distinct sub-samples (nodes and leaves). The tree is pruned iteratively, using a regularizer if it grows excessively complex to prevent overfitting (Kapelner and Bleich 2015). The result is a sum of trees as follows:

$$P(Y_i = 1|\mathbf{X}_i) = \mathbf{\Phi}(\xi_1(\mathbf{X}_i) + \ldots + \xi_q(\mathbf{X}_i)). \tag{2}$$

Where $\mathbf{\Phi}$ is the cumulative density function of the standard normal distribution and the q binary trees are represented by ξ.

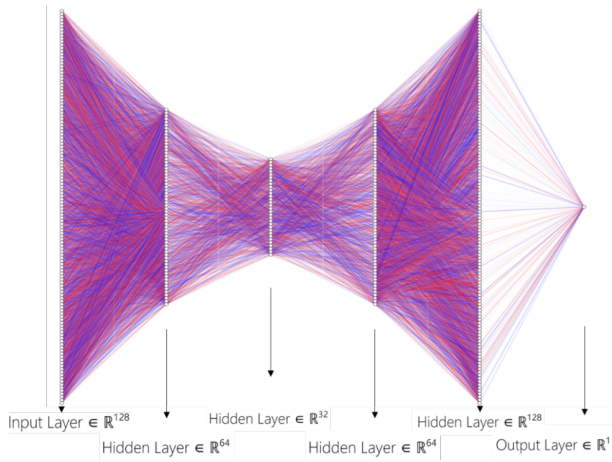


Figure 2: Stylized representation of a neural network

Note: A representation of a six-layer NN. Intermediate layers process information received in the input layer, from left to right, up to the output layer, where a firm's export status is attributed. The first layer (128 nodes, each represented by a circle) is connected to the second layer (64 nodes) with a rectified linear unit activation function determining the weights for each node. This process occurs consecutively for all layers, henceforth, up to the final layer, which has only one node.

Notably, the general representation of the relationship between each feature and the predicted variable is unknown, and each model attempts to estimate it. These ML models allow us to capture a potential non-linear relationship between the variables and the output.

Finally, the other methods used are logistic regression (LOGIT) and logistic Lasso regression. The LOGIT is a classical econometric method for binary response variables, with an ex-ante assumption about the relationship between predictors and the outcome. The LOGIT - Least Absolute Shrinkage and Selection Operator (LASSO), has the following form:

$$arm\min_{\beta} \frac{1}{2N}\sum_{1}^{N}(y_i(X_i^{'}\beta) - log(1+e^{(X_i^{'}\beta)}))^2 \quad subject\ to \quad \sum_{p=1}^{j}|\beta|_p \leq k. \quad (3)$$

Where y_i is the binary outcome of interest, indicating if the firm is an exporter. $X_i^{'}$ is the transposed matrix of predictors subject to the indicated condition, that limits the complexity of the estimated model, avoiding overfitting.

In addition, in order to identify the most representative set of features, including its functions, we use mutual information methods such as *Mutual Information Classification* and *SelectKBest* library (Pedregosa *et al.* 2011). The variables generated through the method represented 90% of the dependent weight. We also use these libraries to indicate the best subset of features for a more parsimonious model.

4.2. Interpretation of machine learning models

An understandable, interpretable, and transparent model can provide a more informed, unbiased, and ethical decision-making process. Barredo Arrieta *et al.* (2020) divides transparency into three levels: i) simulatability – having the model input, its calculations can be made at a human level in a reasonable time; ii) decomposability – each part of the model has an intuitive explanation, and iii) algorithmic transparency – there are theoretical guarantees about the convergence or behavior of the algorithm. Many widely used ML and deep learning models lack simulatability and decomposability since the parameters in the hidden layer lack an intuitive explanation.

Interpreting each feature's importance in an ML model is computationally intensive and laborious. Initially introduced by Shapley (1953), the Shapley values are a cooperative game theory concept, corresponding to the average marginal contribution of a feature after considering all possible combinations. More recently, Lundberg and Lee (2017) classify Shapley additive explanations (SHAP) as an extension of a framework for global and local model-agnostic explanations using feature attribution. Therefore, through this method, it is possible to understand the local importance of the feature and how this importance changes with different values of the feature itself. This method has been used to unveil economic interpretation from ML models (Alves *et al.* 2022).

4.3. Features and accuracy of predictions

In order to increase the accuracy of the ML models, we include all 17 available variables as predictors, plus sector, size and year dummies. Sectors are defined along a NACE Rev.2 classification (Carré 2008), as previously described in Table 1. Firms are classified as micro, small, medium, and large in accordance with the definition proposed by the European Commission.[1]

Figure 3 presents the correlation matrix for the 17 variables that predict successful exporters. These variables are highly correlated, with the coefficient ranging from 0.6 to 0.9. This excessive number of variables is not a concern because our first empirical exercise does not aim to discuss the determinants of export activity. Thus, we do not discriminate variables ex-ante. Nevertheless, in the robustness check, we select a subset of ratios: average salary, capital intensity, intra-EU imports, extra-EU imports, labour productivity, and return on assets, plus information for size, sector and year, and perform a forecast with these variables as

1. According to the Recommendation of the European Commission 2003/361/EC, the category of micro-sized comprises firms that employ fewer than 10 workers and an annual turnover or total annual balance sheet not in excess of 2 million euros. In turn, small firms employ fewer than 50 people and have an annual turnover or total annual balance sheet not in excess of 10 million euros. The medium-sized firms employ fewer than 250 people and have an annual turnover not in excess of 50 million euros or a total annual balance sheet not in excess of 43 million euros. Thus, the large firms are those that do not belong to any of the aforementioned categories.

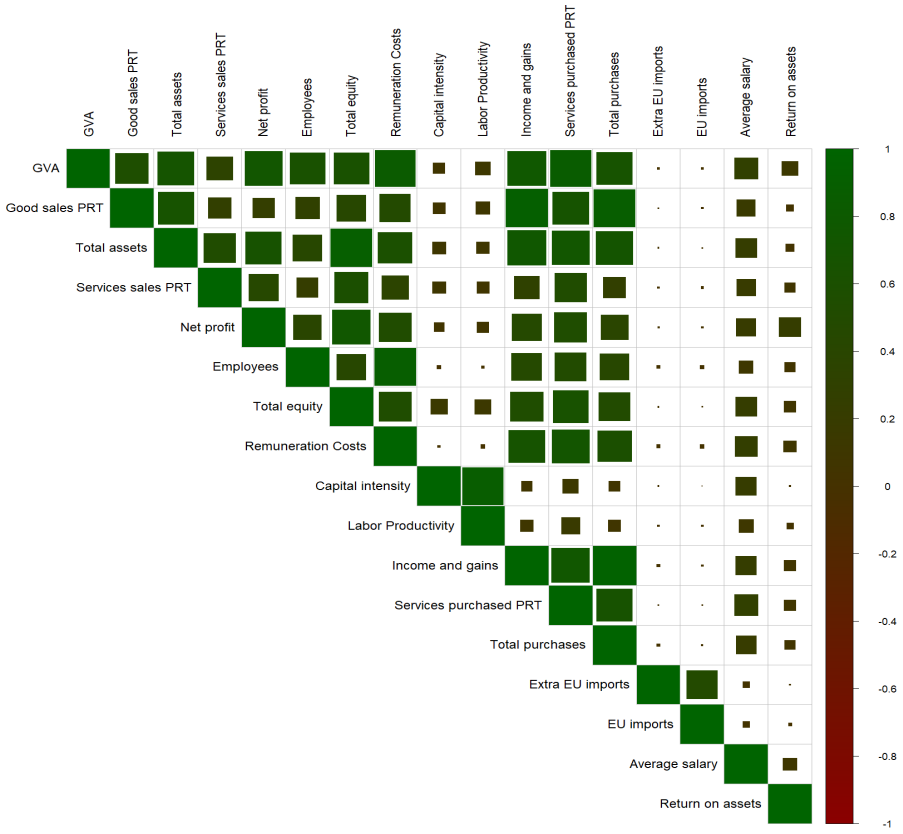


Figure 3: Correlation matrix

Note: The matrix reports the Pearson correlation of the 17 numerical variables used in the ML models. Positive correlations are displayed in shades of green and negative in shades of red. The EU and extra-EU imports are expressed as a ratio on total purchases of the firm.

predictors. We show that this lower number of variables causes a marginal decrease in the forecast accuracy. We also use these variables in our second empirical exercise to understand how they are associated with successful exporters, thus understanding their local importance.

5. Results

5.1. Accuracy of the models

In this section we report the performance of the different models on the test sample. Table 2 summarises the accuracy of their predictions with standard confusion matrix metrics. The five indicators are: specificity, sensitivity, balanced accuracy, receiver operating characteristic (ROC), and precision recall (PR).

Method	Sensitivity	Specificity	Balanced accuracy	ROC	PR	Nb. obs.
Logit	0.867	0.536	0.702	0.815	0.765	40 165
Logit-Lasso	0.862	0.522	0.692	0.813	0.765	40 165
Random Forest	0.874	0.560	0.717	0.832	0.793	40 165
BART	0.902	0.619	0.761	0.892	0.798	40 165
BART-MIA	0.897	0.615	0.756	0.900	0.798	40 660
Neural Network	0.962	0.620	0.791	0.922	0.910	40 165

Table 2. Prediction accuracies on the test sample

Note: The table presents the accuracy of the six machine-learning models for the test sample. We train the models with 80% of firms from the data set and perform out-of-sample predictions over the remaining 20% of the firms. The BART-MIA does not drop missing values in the features to generate predictions; instead, the model uses this information, i.e., the missing values, as new predictor. The remaining techniques drop missing values when calculating the score probabilities.

The Specificity indicator measures the ability to find true negatives, i.e., how accurate the model is in predicting non-exporting firms. Sensitivity represents the proportion of true positives correctly assigned by the model, i.e., the ability to predict exporters. The arithmetic mean of Specificity and Sensitivity is the balanced accuracy. The ROC curve, represented in Figure 4, evaluates the models' predictive performance at different classification thresholds, and it is our primary metric to evaluate the models. Finally, we also report the PR, which evaluates the trade-off between precision and recall while comparing the relationship between the true and false positive rates.

The standard measures reported in Table 2 show that the Neural Network (NN) outperforms the other models with a ROC of 0.92. The panels of Figure 4 plot the models' ROC curves. The area under the curve (AUC) evaluates the models' performance. A higher AUC means higher accuracy, with a 0.5 AUC, i.e., a diagonal ROC line, meaning a random prediction.

The NN balanced accuracy of 0.79 is the highest, followed at a distance of 3 and 2 percentage points by the BART and BART-MIA, respectively. As for sensitivity, i.e., the ability of the models to classify exporters correctly in the predictions, the six ML methods predict exporter firms with an accuracy ranging from 0.86 (Logit-Lasso) to 0.96 (NN). Regarding specificity, i.e., the ability to predict non-exporters

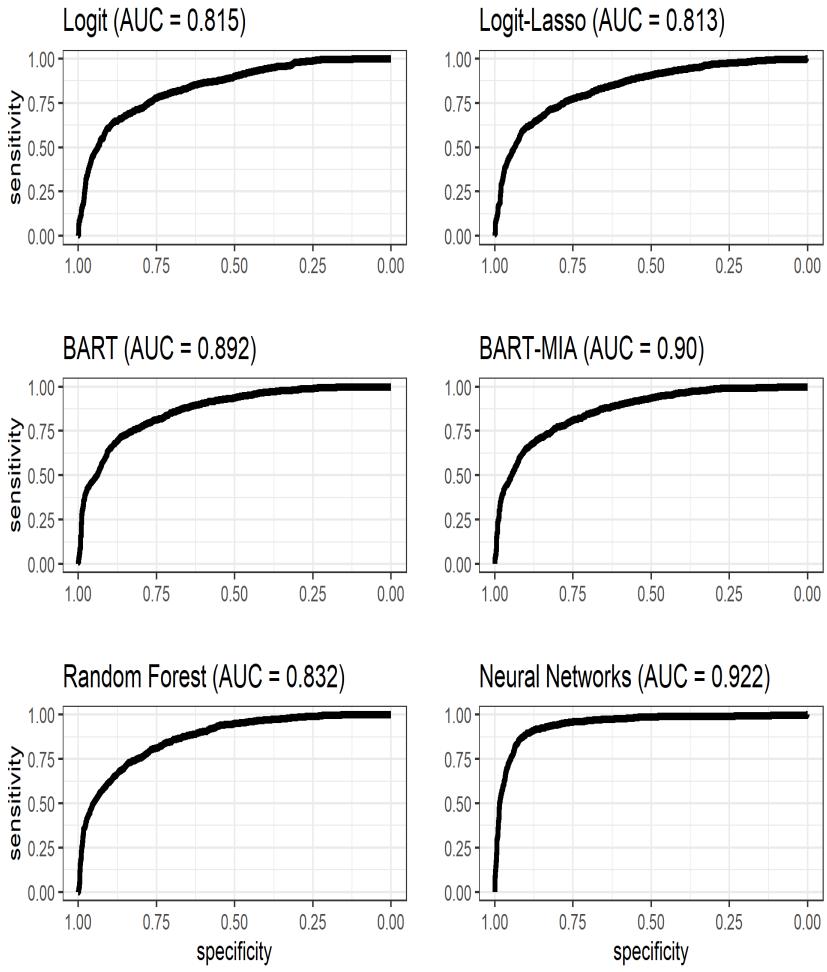


Figure 4: Receiver Operating Characteristic (ROC) curves

Note: The graphs report the ROC curve and the area under the curve (AUC) for the six ML techniques used.

correctly, the accuracies of NN and BART are very close, with values of 0.620 and 0.615, respectively.

According to Micocci and Rungi (2023), BART-MIA is better than other ML methods in the presence of many non-random missing values. In this case, BART-MIA includes the missingness of each predictor of the observation as an additional feature, i.e., another branch in the regression tree. The authors report a 14.4% higher accuracy when including the missingness of the features as a new predictor, that is when using BART-MIA over BART. The reason is that the sample used to predict exporters only includes 23% of observations with complete data (no missing

values). In addition, this missingness is not random, as the authors report a lack of information on micro and small firms due to the non-mandatory financial report to authorities. In our case, the BART and BART-MIA models perform similarly because our sample has a small percentage of missing values.

5.2. Estimations

Figure 5 presents the estimated export scores of all non-exporting firms in our sample obtained from the NN – the model with greater accuracy – as discussed in the previous section. We selected all non-exporting firms without any threshold (or additional requirement) and estimated their distance to becoming successful exporters. Similarly to (Micocci and Rungi 2023), the non-exporters score distribution is right-skewed with a heavy left tail. Thus, many firms are far from successfully entering international markets.

The economic interpretation of the shape of the distribution of export scores justifies some considerations. The fact that almost all non-exporters are very distant from becoming exporters primarily means that these firms do not share the characteristics associated with those that actually entered successfully. According to the literature, their productivity levels are not high enough to cover the fixed costs of exporting (Roberts and Tybout 1997; Melitz 2003; Melitz and Redding 2014). Additionally, the results indicate that firms with the right conditions to enter foreign markets do not delay that decision, i.e., it is unlikely that firms stay in the domestic market for a preparation period before exporting. Another explanation for the high density of firms with low export scores is that firms may belong to economic groups where export activity is concentrated in a specific firm, even if other branches actively contribute to production.

5.3. Robustness

In this section, we investigate how modifying the definitions of exporters, training and test samples, and other relevant terms affects the chosen ML model predictive power. We take the different robustness dimensions sequentially.

The primary concern regarding the robustness of our results is how sensitive the predictions are to changes in the definition of an exporter firm. We perform additional predictions using the NN with the following criteria to define an exporter:

1. Exports at least 5% of its turnover
2. Exports at least 15% of its turnover
3. Exports outside the EU market
4. Exports to at least 3 countries
5. Exports to at least 6 countries
6. Exports to at least 10 countries

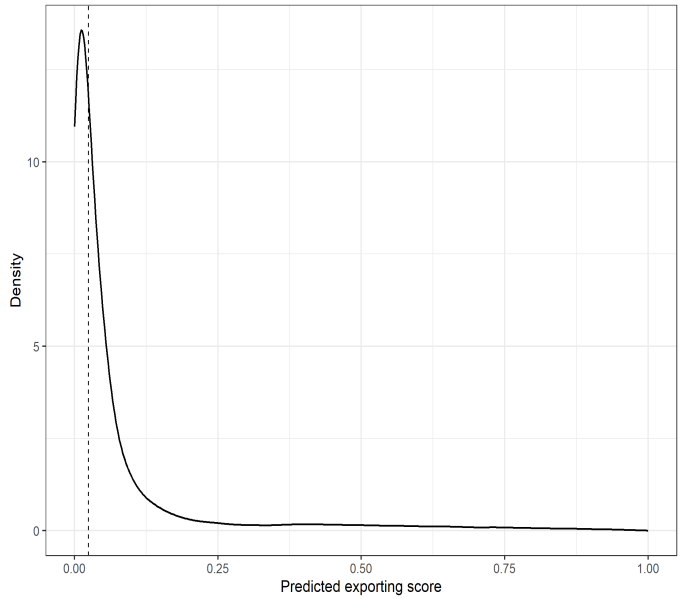


Figure 5: Distribution of the estimated export scores of the non-exporting firms with the NN model

Note: The vertical dashed line is the median of the predicted scores.

Table A.1 (in Appendix A) presents the results under these alternative definitions. Changing the export on turnover threshold from 10% to 5% or 15% does not significantly impact the predictions. With a 5% threshold, the balanced accuracy is 0.82, and the estimated ROC is 0.9, and increasing the threshold to 15% results in a slight improvement in precision (balanced accuracy of 0.87 and ROC of 0.95).

Similarly, changing the definition of an exporter to that of criteria 3 to 6, i.e., considering as exporters firms that sell outside the EU market or that sell to at least 3, 6, and 10 countries, the accuracy of the model remain high (Table A.1 in Appendix A). The balanced accuracies vary from 0.71 to 0.78, and, more importantly, all estimated ROCs are over 0.8.

In addition, we tested whether our results are sensitive to temporary trade using the definition proposed by Békés and Muraközy (2012) and we report the results in Table A.2. The authors consider exporters to be those that export permanently. To assign a firm as an exporter, they take the threshold of 4 consecutive years, i.e., they consider firms exporting at least four consecutive years. Temporary exporters are the remaining firms that export at least once. Our results show that the NN method is still suitable when predicting with this definition of exporter. Sensitivity and Specificity are over 0.7, and, more importantly, the estimated ROC is 0.83.

A second question is whether the accuracy of the NN changes for firms of different sizes. We train and perform out-of-sample predictions for micro, small, medium, and large firms with our NN model. Results are reported in Table A.3 in Appendix A. For the four dimension classes, prediction accuracies are high. The estimated ROCs vary from 0.874 to 0.966 and increase with firms' size, i.e., the model has higher accuracy on larger manufacturers. The model's ability to predict correctly exporters (Sensitivity) is 0.3 for micro firms, but it increases to 0.96 for large firms. In addition, the model predicts non-exporters (Specificity) with higher accuracy for micro firms than large firms, with values of 0.98 versus 0.83, respectively. This difference in accuracy along firms' size may be related to smaller ones being more likely to engage in short export spells (Békés and Muraközy 2012). Next, we investigate this behavior in detail by exploring whether the length of trade spells affects the accuracy of the predictions.

The third question is whether reducing the number of predictors significantly affects our results and, more specifically, whether the NN model can produce similar results with fewer predictors. If this were the case, computational costs would be reduced without relinquishing predictive power. To test this hypothesis, we select six features – salary, capital intensity, intra-EU imports, extra-EU imports, labour productivity, and return on assets, maintaining the dummy predictors indicating the year, firm's size and sector – and rerun the NN model. This exercise with partial information only mildly affects our estimator's overall performance. The estimated ROC decreases by about 3 percentage points, as reported in Table A.4 in Appendix A.

Our fourth question is whether time-variant-specific relationships affect the general model. To answer this concern, we have performed predictions testing the NN model over subsets of years. Results are presented in Table A.5 in Annex A. Accuracies remain similar to the one achieved on the original testing sample. Balanced accuracies are close to 0.8, and ROCs are always above 0.9 over the timeline.

Békés and Muraközy (2012) document that many Hungarian firms export sparsely in time. This is a stylized fact in a world of heterogeneous firms, and it is also observed in Portugal. This feature occurs more frequently in micro and small firms and in manufacturers of capital goods whose production process tends to be longer. If the time pattern of exports is not random, i.e., if it can be associated with some of the firms' characteristics, the ML techniques can incorporate such information in their predictions (Békés and Muraközy 2012; Micocci and Rungi 2023). Therefore, we perform an exercise similar to that suggested by Micocci and Rungi (2023), where the NN model is used to predict five different types of firms:

1. Firms that exported in all periods of our sample (constant exporter)
2. Firms that never exported (non-exporter)
3. Firms that were non-exporters until some period t and then started exporting in all periods afterwards (switch to exporter)

4. Firms that were exporters until some period t and then became non-exporters in all periods afterwards (switch to non-exporter)
5. Firms with irregular export behaviour, i.e., firms that switch between exporting and non-exporting more than once (temporary exporter)

Table A.6 in Appendix A reports the prediction accuracies for each export-permanence group (adding further detail on the length of export spells), all estimated with the NN model. Results show that the original model predicts well for constant exporters and non-exporters (they represent almost 75% of our sample). When predicting exporters in the sample with constant exporters only, the Sensitivity is 0.855, whereas in the sample with non-exporters, the Specificity is 0.984.

These results are similar to Micocci and Rungi (2023), who also performed a comparable empirical exercise. Our predictions suggest that firms with a gap in export activity, i.e., discontinuity to exporting, bring less precision to the algorithms.

When performing out-of-sample predictions where firms switched to become exporters, the estimated ROC is about 0.8. In comparison, for a group of firms that switched to non-exporters, the estimated ROC is about 0.81. These results are poor compared to predictions for the entire test sample with a ROC of 0.92, as reported in Table 2. We observe in Table A.6 that the accuracy of the NN algorithm increases proportionally with the exporting experience of the firms in the test sample. In other words, the accuracy is higher (smaller) in a sample where firms switched to exporters earlier (later). Similarly, the accuracy is higher (smaller) in a sample where the firms turned to non-exporters later (earlier).

Looking at the group of firms that switched to exporters (or to non-exporters) more than once, which we label temporary exporters, the NN model's predictions are poor. The estimated ROC is under the acceptable ROC of 0.8. Micocci and Rungi (2023) argues that firms that engage sparsely in trade continue to do so systematically, consistent with the lower precision estimates of the NN model.

Another question is whether the NN remains highly accurate with different training and testing samples. In order to assess the robustness of results to this type of choice, we cross-validated our original sample, repeating the prediction exercise four times with random training and testing samples of different sizes. We trained the model on 90% of the firms and predicted the remaining 10%, which became the testing sample, and repeated this process twice (labeled as take one and take 2). Additionally, we randomly split the training and testing samples with proportions 70%-30% and 50%-50%. Table A.7 in Appendix A shows that the results are similar in the four alternative exercises.

Finally, we assessed the sensitivity of our results to the choice of model. To do this, we calculated the correlations between firms' export status based on each of the six ML models used in our empirical exercise. As shown in Table 3, the NN predictions exhibit weak correlations with those from the other models, ranging from 0.58 with logit regression to 0.64 with BART. In contrast, the results obtained for the other models show higher correlations, ranging from 0.76 between

logit regressions and random forests to 0.95 between BART and BART-MIA. This comprehensive comparison provides a signal of the different NN predictions versus other models.

	Logit	Logit-Lasso	Random forest	BART	BART-MIA	NN
Logit	1					
Logit-Lasso	0.928	1				
Random forest	0.766	0.757	1			
BART	0.828	0.837	0.854	1		
BART-MIA	0.824	0.828	0.85	0.947	1	
NN	0.584	0.582	0.625	0.640	0.636	1

Table 3. Spearman's correlations rank of the predicted export status between models

6. Main features, sectors and firm's assets

In this section, we identify the variables that are the main features of the scoring exercise and explore two additional dimensions. Firstly, we analyze the results along manufacturing sub-sectors. Secondly, we compute the level of assets associated with firms in different deciles of the scoring distribution in connection with export support activities by public agencies.

6.1. Main features

In previous subsections, we discussed how the NN model consistently maintains high accuracy levels despite changes in definitions, features and sampling. However, these validation steps do not emphasize the explainability behind the model's predictions and what additional insight it brings to the policy discussion. In this section, we focus on identifying the main features impacting the model's performance and how each affects the model individually. In this way, we use the Shapley values (SHAP), discussed in Subsection 4.2. However, it is essential to reiterate that it is not possible to assess causal effects through the SHAP.

The robustness check carried out above has shown that a subset of variables is enough to accurately score exporters, as presented in Table A.4. In fact, in the exercise with six features, the estimated ROC is only 3 percentage points lower than the entire set of 17 features (in both cases, including time, the 2-digit sector of NACE Rev.2, and the size of the firm as predictors). Therefore, in this section, we focus on investigating the importance on prediction of these six features: wage, capital intensity, imports from the EU (as a percentage of firms' purchases), extra-EU imports (as a percentage of firms' purchases), labour productivity, and return on assets.

Figure 6 illustrates how each feature of the NN model influences the exporting scores, considering the absolute SHAP. Notably, labour productivity emerges as the

most crucial for estimating export scores, indicating its pivotal role in predicting the likelihood of becoming a successful exporter. The second most important variable is firms' imports from the EU as a percentage of total firm's purchases, which signals export activities. The next three variables-imports from extra-EU countries as a percentage of firms' total purchases, capital intensity, and wages-all carry similar weights in the predictions. The return on assets ratio, on the other hand, shows negligible influence on the predictions.

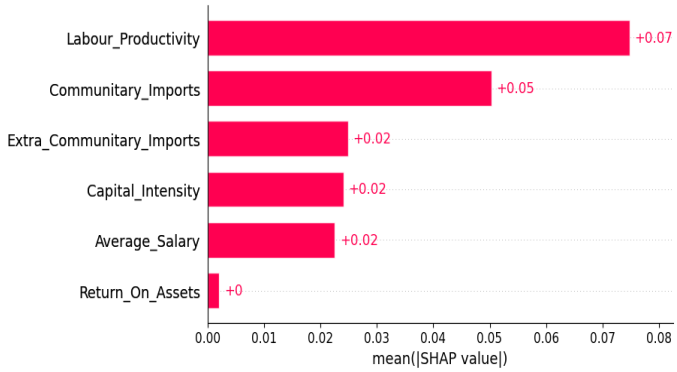


Figure 6: Absolute SHAP bar plot

Note: The plot summarizes the absolute individual contribution of each feature to the mean SHAP value, i.e., each feature's average contribution to a firm's probability to export. The variables are in standardized values.

Figure 7 provides a detailed visualization of how different variables influence the estimated exporting scores, either positively or negatively. Each dot represents a firm-year observation. If a feature's contribution to the prediction is positive, the dot is placed inside the category; if negative, it is outside. Each dot's color indicates the influence's magnitude, with a gradient scale ranging from high (red) to low (blue). This chart helps identify which features have the most significant influence on the model's predictions, offering clear insights into the role of each variable in the exporting scores.

Labour Productivity emerges as the pivotal variable in determining the likelihood of a firm's export. Lower labour productivity negatively influences the firm's export potential, while higher labour productivity enhances the likelihood of export. This finding aligns with the empirical international trade literature mentioned earlier and is visually represented in Figure 7 by placing blue and red dots in the negative and positive SHAP segments, respectively.

Firms with low values for EU and extra-EU imports appear less likely to export. Symmetrically, firms with high import values seem more likely to export, especially those importing from extra-EU countries. Buying products and services overseas increases a firm's performance and involves knowledge similar to the one required to sell abroad. The SHAP of the capital intensity is ambiguous. While lower values

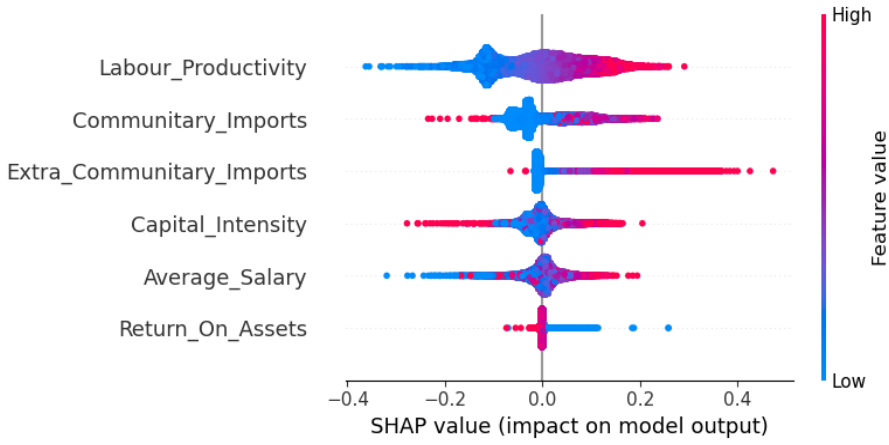


Figure 7: Relevance of the variables considering individual contributions

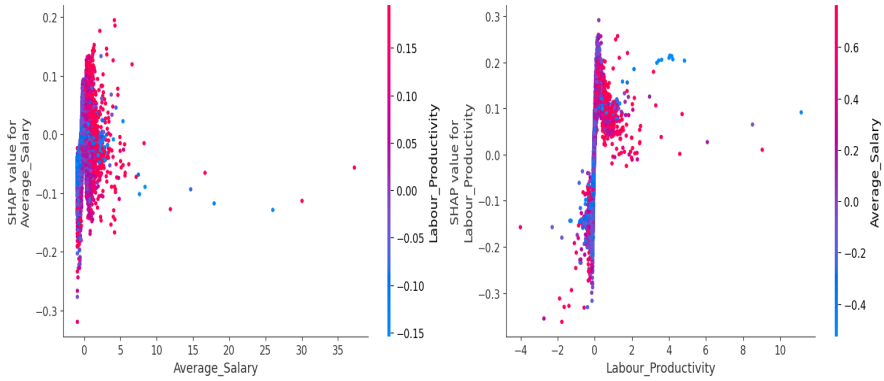
Note: Features are ordered in the Y-axis by importance from top to bottom. The X-axis denotes the SHAP value, indicating the degree of change in log odds. The chromatic spectrum of each point signifies the associated feature value, wherein red denotes high values and blue signifies lower values of the variables. A point represents each observation (firm-year). Values are standardized.

are concentrated around zero, most higher values (red dots) are on the positive side. In addition, firms with higher wage levels have higher exporting scores as the blue dots are massively on the negative SHAP segment. This is not surprising as wages are positively correlated with labour productivity. This relation is visible in the context of the SHAP.

Figure 8 presents the relationship between wages, labour productivity and capital intensity along SHAP values, i.e., how different variables are associated with the firm's export probability while signaling the levels of the other variable. In panel (a), we present the relationship between SHAP and wages, and the blue (red) colour represents low (high) levels of labour productivity. As expected, a positive relationship exists between wage levels and SHAP, i.e., high wages are associated with a high probability of becoming an exporter. Moreover, firms with higher wages also have, on average, higher levels of labour productivity (going from blue to red). Panel (b) presents the mirror of this relationship with higher productivity and higher wage firms (in red) posting higher SHAP.

6.2. Sectors

We claimed above that the estimated export scores are a valuable indicator for policymakers, export promotion agencies, and credit institutions. One use is identifying the sectors with higher potential for internationalization. Figure C.1 shows the distribution of the exporting scores by sector along a 2-digit NACE Rev.2 classification. Almost all industries show a large density in the left tail of the



(a) Average salary x Labour Productivity (b) Labour productivity x Average salary

Figure 8: SHAP wage and labour productivity interactions

Note: The figure shows the interactions between the wages, labour productivity and capital intensity in the SHAP. The values are standardized.

distribution, meaning their non-exporters are far from performing successfully in international markets. One extreme case is "food products," with the distribution almost collapsing near zero, while "furniture" has a higher density of firms closer to becoming successful exporters. The shape of the distribution for "coke and refined petroleum" differs from that of other sectors due to the relatively larger size and smaller number of firms. A larger share of non-exporting firms in this sector are close to becoming successful exporters.

6.3. Firm's assets

As discussed above, entering overseas markets implies significant fixed costs. Therefore, firms typically need resources from financial institutions to cover at least part of these costs. However, financial market imperfections may conduce firms to miss out on export opportunities, resulting in export levels below their potential (Chor and Manova 2012; Manova 2013). Due to this asymmetry of information between firms and financiers that may appear, public authorities try to mitigate market failure, enabling better equipped firms to compete in international markets (Comi and Resmini 2020; Srhoj and Walde 2020). In particular, export promotion agencies can approach firms with high potential to become exporters and help them sell abroad either with grants or credit guarantees. Our indicator of distance to exports can help these agencies to better target their support, i.e. to increase the effectiveness of public support.

Similarly to Micocci and Rungi (2023), we utilize exporting scores to estimate the difference in total assets Portuguese firms need to enter international markets successfully. This exercise involves comparing the assets of firms closer to exporting

with those farthest from exporting. By classifying non-exporting firms into different risk categories, each of them based on scoring deciles, we apply a fixed-effect estimator to estimate the following model:

$$\log(y_{it}) = \varphi_0 + \sum_{j=2}^{10} \varphi_j risk_j + \alpha_1 X_{it} + \gamma_t + \delta_{ind} + v_{it}, \tag{4}$$

where y_{it} is the total assets of firm *i* in moment *t*, X_{it} is a factor indicating the firm's size group, γ_t is a time fixed effect, δ_{ind} is a sector fixed effect and v_{it} represents both the invariant and the time-varying error terms. This approach allows us to control for time-invariant unobserved heterogeneity across firms and capture industry-specific and temporal trends in the analysis.

The parameters of interest in Equation 4 are φ_j (with $j = 2, \ldots, 10$), associated with each scoring decile dummy $risk_j$ created from the estimated exporting score of the firms. The first risk group (with scores from 0 to 0.09) is omitted, serving as the reference group. Thus, these coefficients represent the amount of assets each firm operates with compared to the first risk group. As expected, the estimated coefficients are greater than zero, indicating that the groups with lower risk (closer to being successful exporters) operate with higher assets. Estimated coefficients are reported with a 99% confidence interval in Figure 9 and Table B.1.

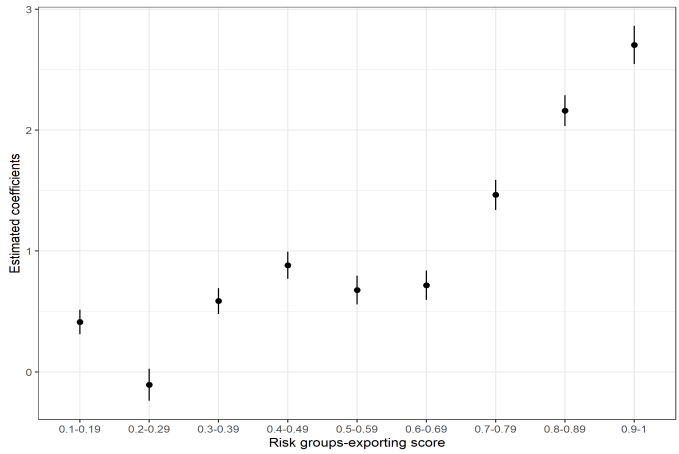


Figure 9: Estimated risk coefficients

Note: The figure presents the fixed-effects coefficients estimated from Equation 4. The model uses the log of total assets as the dependent variable. We control for firm size, industry sector, and year fixed-effects. Standard errors are clustered at the firm level to address potential correlations within firms over time, and 99% confidence intervals are reported.

Analyzing the estimated parameters of the dummy variables that indicate the risk groups, the second group operates with 51.3% $[(\exp(\widehat{\varphi}_2) - 1) * 100\%]$ more assets than the reference group. The third risk group is not significantly different from zero at a 5% level. Figure 9 shows that from the 4th to the 7th groups, the

asset amounts at a 99% confidence level are similar. For instance, group 4 has 80% more assets than group 1, and group 7 has 105% more assets. Finally, the most significant differences are observed in the last three risk groups (8^{th}-10^{th}). These groups operate with substantially higher levels of assets. The 8^{th} group operates with 333% more resources than the first one, while the group with the lowest risk (10^{th}) operates with 1,388% more assets than the reference group. Therefore, firms in the medium risk level (4^{th}-7^{th} groups) operate with over ten times fewer assets than the lowest risk group (10^{th} decile).

These results are strong and, although they should be compared with those for other countries, they reinforce the long standing view that the small size of Portuguese firms is a major obstacle to competitiveness, notably in external markets. Nevertheless, firm size is itself an endogenous variable, turning it hard to identify the key policies to address this issue.

7. Concluding remarks

The paper aims to predict how far Portuguese non-exporting manufacturing firms are from successfully selling abroad. To this end, we make predictions and assess the accuracy of predicting exporters of six machine learning models, then choose the neural network method. Our hypothesis is that there is a nontrivial relationship between firms' characteristics and their exporting status, which the ML models can identify to predict accurately. Moreover, we try to understand which variables are more important to the predictions and how they relate to the probability of Portuguese manufacturers becoming successful exporters.

We perform an extensive battery of robustness tests on the neural network model, and it maintains high predictive power across all of them. Interestingly, the predictions generated from the NN model are substantially different from those of the other techniques despite all having excellent performance in out-of-sample predictions.

We compute the SHAP values to identify the features that are most relevant to predicting exporters and investigate how they are related to the probability of exporting. The essential variables for prediction are labour productivity and, at a distance, imports from EU markets as a share of the firm's total purchases. As expected, high levels of labor productivity are associated with a high probability of selling in international markets.

Our findings are helpful for agencies that design internationalization policies and promote exports and financial institutions. In this vein, we assess the relationship between firms' assets, interpreted as potential collateral, and exports. By understanding the asset resource requirements of different potential exporters, these institutions can more effectively target their programs and grants, thus reducing costs and increasing effectiveness. In this way, we classify the firms into ten risk groups, and as expected, the result shows that firms that are more distant from exporting, those with higher risks, operate with lower levels of assets. Firms

in groups 4^{th}-7^{th} have similar asset levels, ranging from 80% to 105% more than group 1. Portuguese firms with the highest risk need twice as many asset resources as medium-risk firms (groups 4^{th}-7^{th}).

Moreover, the most significant differences are in groups 8^{th}-10^{th}, where the 8^{th} and 10^{th} groups have 333% and 1,388% more assets than the first group, respectively. Therefore, firms in the median risk level operate with over ten times fewer assets than the lowest risk group (10^{th}). This links with the long-standing view that small size is a barrier to the external competitiveness of Portuguese firms.

This strand of research provides some important insights, and results are usable regularly by public agencies to effectively enact export promotion policies. In addition, it is another interesting illustration of the utilization of machine learning techniques to economic research. However, it is essential to acknowledge its limitations. The main one is the inability to unveil causal effects. The layers of relationships in neural networks are inscrutable and operate as black boxes, even if they are powerful. Future research should add new variables to the analysis and assess which internationalization preparation strategies work best for firms with different characteristics, products and destination markets.

References

Altman, Edward I (1968). "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." *The journal of finance*, 23(4), 589–609.

Alves, Tiago, João Amador, and Francisco Gonçalves (2022). "Assessing the Scoreboard of the EU Macroeconomic Imbalances Procedure:(Machine) Learning from Decisions."

Angelini, Eliana, Giacomo di Tollo, and Andrea Roli (2008). "A neural network approach for credit risk evaluation." *The Quarterly Review of Economics and Finance*, 48(4), 733–755.

Athey, Susan, Guido W Imbens, Jonas Metzger, and Evan Munro (2021). "Using wasserstein generative adversarial networks for the design of monte carlo simulations." *Journal of Econometrics*, p. 105076.

Baier, Scott L, Jeffrey H Bergstrand, and Ronald Mariutto (2014). "Economic determinants of free trade agreements revisited: Distinguishing sources of interdependence." *Review of International Economics*, 22(1), 31–58.

Barredo Arrieta, Alejandro, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera (2020). "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." *Information Fusion*, 58, 82–115.

Békés, Gábor and Balázs Muraközy (2012). "Temporary trade and heterogeneous firms." *Journal of International Economics*, 87(2), 232–246.

Breinlich, Holger, Valentina Corradi, Nadia Rocha, Michele Ruta, JMC Santos Silva, and Thomas Zylkin (2022). "Machine learning in international trade research-evaluating the impact of trade agreements."

Broocks, Annette and Johannes Van Biesebroeck (2017). "The impact of export promotion on export market entry." *Journal of International Economics*, 107, 19–33.

Carré, H (2008). "NACE Rev. 2, Statistical classification of economic activities in the European Community." *General and regional statistics*.

Chor, Davin and Kalina Manova (2012). "Off the cliff and back? Credit conditions and international trade during the global financial crisis." *Journal of international economics*, 87(1), 117–133.

Coad, Alex and Antonio Vezzani (2019). "Three cheers for industry: Is manufacturing linked to R&D, exports, and productivity growth?" *Structural Change and Economic Dynamics*, 50, 14–25.

Comi, Simona and Laura Resmini (2020). "Are export promotion programs effective in promoting the internalization of SMEs?" *Economia Politica*, 37(2), 547–581.

Cruz, Marcio (2014). "Do export promotion agencies promote new exporters?" *World Bank Policy Research Working Paper*, (7004).

Dai, Chenglin (2023). "A method of forecasting trade export volume based on back-propagation neural network." *Neural Computing and Applications*, 35(12), 8775–8784.

De Loecker, Jan (2007). "Do exports generate higher productivity? Evidence from Slovenia." *Journal of international economics*, 73(1), 69–98.

Geishecker, Ingo, Philipp JH Schröder, and Allan So rensen (2019). "One-off export events." *Canadian Journal of Economics/Revue canadienne d'économique*, 52(1), 93–131.

Harrison, Ann and Andrés Rodríguez-Clare (2010). "Trade, foreign investment, and industrial policy for developing countries." *Handbook of development economics*, 5, 4039–4214.

Hastie, Trevor, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer.

Kapelner, Adam and Justin Bleich (2015). "Prediction with missing data via Bayesian additive regression trees." *Canadian Journal of Statistics*, 43(2), 224–239.

Lundberg, Scott M and Su-In Lee (2017). "A unified approach to interpreting model predictions." *Advances in neural information processing systems*, 30.

Manova, Kalina (2013). "Credit constraints, heterogeneous firms, and international trade." *Review of Economic Studies*, 80(2), 711–744.

Melitz, Marc J (2003). "The impact of trade on intra-industry reallocations and aggregate industry productivity." *econometrica*, 71(6), 1695–1725.

Melitz, Marc J and Stephen J Redding (2014). "Heterogeneous firms and trade." *Handbook of international economics*, 4, 1–54.

Merton, Robert C (1974). "On the pricing of corporate debt: The risk structure of interest rates." *The Journal of finance*, 29(2), 449–470.

Micocci, Francesca and Armando Rungi (2023). "Predicting Exporters with Machine Learning." *World Trade Review*, 22(5), 584–607.

Munch, Jakob and Georg Schaur (2018). "The effect of export promotion on firm-level performance." *American Economic Journal: Economic Policy*, 10(1), 357–387.

Pan, Chenchen, Yuan Gao, and Yuzi Luo (2018). "Machine Learning Prediction of Companies' Business Success." *CS229: Machine Learning, Fall*.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, *et al.* (2011). "Scikit-learn: Machine learning in Python." *the Journal of machine Learning research*, 12, 2825–2830.

Roberts, Mark J and James R Tybout (1997). "The decision to export in Colombia: An empirical model of entry with sunk costs." *The american economic review*, pp. 545–564.

Shapley, Lloyd (1953). "A value for n-person games."

Sousa, Carlos MP, Francisco J Martínez-López, and Filipe Coelho (2008). "The determinants of export performance: A review of the research in the literature between 1998 and 2005." *International journal of management reviews*, 10(4), 343–374.

Srhoj, Stjepan, Vanja Vitezić, and Joachim Wagner (2023). "Export boosting policies and firm performance: Review of empirical evidence around the world." *Jahrbücher für Nationalökonomie und Statistik*, 243(1), 45–92.

Srhoj, Stjepan and Janette Walde (2020). "Getting ready for EU Single Market: The effect of export-oriented grant schemes on firm performance." *Structural Change and Economic Dynamics*, 52, 279–293.

Tong, Ye (2023). "Literature Review in International Trade Forecasting Based in Machine Learning Method." In *SHS Web of Conferences*, vol. 154, p. 03025, EDP Sciences.

Uddin, Mohammad Shamsu (2021). "Machine learning in credit risk modeling: empirical application of neural network approaches." *The Fourth Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success*, pp. 417–435.

Wagner, Joachim (2007). "Exports and productivity: A survey of the evidence from firm-level data." *World economy*, 30(1), 60–82.

Warner, Brad and Manavendra Misra (1996). "Understanding neural networks as statistical tools." *The american statistician*, 50(4), 284–293.

Appendix A: Robustness results

Definition of exporter	Sensitivity	Specificity	Balanced accuracy	ROC	PR	Nb. Obs
Exports at least 5% of its turnover	0.895	0.741	0.818	0.901	0.924	40,165
Exports at least 15% of its turnover	0.897	0.848	0.873	0.946	0.942	40,165
Exports outside the EU market	0.496	0.942	0.719	0.865	0.730	40,165
Exports to at least 3 countries	0.738	0.823	0.781	0.861	0.870	40,165
Exports to at least 6 countries	0.589	0.914	0.752	0.868	0.776	40,165
Exports to at least 10 countries	0.464	0.960	0.712	0.884	0.698	40,165

Table A.1. Predictions with different definitions of exporter

Note: In the benchmark case we define an exporter as a firm selling abroad at least 10% of its turnover. Here we train and test the Neural Network (NN) model out-of-sample using six different exporter definitions to assess how they definitions affect prediction accuracy.

Cross-validation	Sensitivity	Specificity	Balanced accuracy	ROC	PR	Nb. Obs
Permanent Exporters	0.751	0.736	0.743	0.831	0.956	42,005
Temporary Exporters	0.329	0.867	0.598	0.698	0.459	39,438
Non-Exporters	-	0.984	-	-	0.5	166,782

Table A.2. Prediction accuracies with Békés and Muraközy (2012) classification of exporter

Note: The table reports prediction accuracies for the NN model with firm's classification according to Békés and Muraközy (2012): i) permanent exporters are firms that export at least four consecutive years; ii) temporary exporters are the remaining firms that export at least once; iii) non-exporters are firms that never export.

Firms size class	Sensitivity	Specificity	Balanced accuracy	ROC	PR	Nb. Obs
Micro	0.3038	0.9832	0.6435	0.8738	0.4959	164,396
Small	0.6469	0.9236	0.7852	0.9102	0.7641	51,087
Medium	0.9296	0.8613	0.8765	0.9129	0.9261	33,517
Large	0.9643	0.8249	0.8946	0.9659	0.9822	5,279

Table A.3. Predictions according to firms' size class

Note: We split the firms into four groups according to their size class and performed predictions for these groups with the chosen NN model. We define the firms' size classes in accordance with the definition by the European Commission.

Subset of variables	Sensitivity	Specificity	Balanced accuracy	ROC	PR	Nb. Obs
–	0.654	0.958	0.806	0.910	0.814	40,165

Table A.4. Predictions with a subset of variables

Note: We perform predictions with the chosen NN model with a subset of variables: Salary, capital intensity, intra-EU imports, extra-EU imports, labour productivity, and return on assets.

Year	Sensitivity	Specificity	Balanced accuracy	ROC	PR	Nb. Obs
2010	0.684	0.937	0.811	0.918	0.792	15,154
2011	0.561	0.955	0.758	0.901	0.779	15,257
2012	0.652	0.942	0.797	0.913	0.799	15,541
2013	0.643	0.945	0.794	0.913	0.807	15,838
2014	0.558	0.955	0.757	0.879	0.778	15,864
2015	0.725	0.912	0.818	0.905	0.805	16,711
2016	0.657	0.942	0.799	0.918	0.818	16,961
2017	0.674	0.943	0.808	0.924	0.823	17,566
2018	0.664	0.969	0.816	0.941	0.799	29,952
2019	0.681	0.973	0.827	0.951	0.808	34,740
2020	0.696	0.970	0.833	0.940	0.785	33,779
2021	0.626	0.981	0.803	0.940	0.795	20,862

Table A.5. Predictions for different blocks of years

Note: We perform separate predictions for each year using all information up to the final year of the sample, aiming to assess whether the accuracy is stable over time.

Firms category	Sensitivity	Specificity	Balanced accuracy	ROC	PR	Nb. Obs
Constant exporter	0.855	-	-	-	-	19,323
Non-exporter	-	0.984	-	-	-	166,782
Switch to exporter	0.611	0.834	0.723	0.802	0.807	11,829
Since t0	0.644	0.802	0.723	0.811	0.946	2,588
Since t1	0.670	0.810	0.740	0.808	0.885	1,767
Since t2	0.612	0.825	0.718	0.805	0.846	1,550
Since t3	0.546	0.864	0.705	0.804	0.783	1,063
Since t4	0.580	0.809	0.695	0.778	0.699	914
Since t5	0.528	0.877	0.703	0.793	0.660	961
Since t6	0.535	0.857	0.696	0.792	0.644	634
Since t7	0.508	0.810	0.660	0.720	0.421	576
Since t8	0.452	0.862	0.657	0.709	0.382	664
Since t9	0.294	0.915	0.604	0.736	0.330	548
Since t10	0.511	0.731	0.621	0.679	0.159	564
Switch to non-exporter	0.456	0.893	0.674	0.806	0.765	7,511
Since t1	0.223	0.898	0.561	0.692	0.304	2,489
Since t2	0.343	0.898	0.620	0.745	0.579	973
Since t3	0.477	0.844	0.661	0.788	0.719	531
Since t4	0.514	0.874	0.694	0.801	0.813	522
Since t5	0.449	0.855	0.652	0.813	0.829	533
Since t6	0.541	0.887	0.714	0.855	0.893	560
Since t7	0.551	0.848	0.700	0.833	0.902	344
Since t8	0.708	0.830	0.769	0.868	0.944	376
Since t9	0.616	0.896	0.756	0.890	0.961	510
Since t10	0.756	0.776	0.766	0.839	0.964	409
Since t11	0.781	0.727	0.754	0.840	0.980	264
Temporary exporter	0.437	0.845	0.641	0.736	0.637	42,780
Export experience: 1 year	0.166	0.911	0.538	0.660	0.188	10,597
Export experience: 2 years	0.283	0.850	0.566	0.671	0.338	6,819
Export experience: 3 years	0.309	0.833	0.571	0.671	0.45	5,218
Export experience: 4 years	0.388	0.789	0.589	0.680	0.555	3,748
Export experience: 5 years	0.459	0.738	0.598	0.669	0.637	3,284
Export experience: 6 years	0.506	0.692	0.599	0.678	0.712	2,940
Export experience: 7 years	0.573	0.664	0.619	0.681	0.769	2,705
Export experience: 8 years	0.624	0.639	0.632	0.698	0.835	2,401
Export experience: 9 years	0.683	0.535	0.609	0.675	0.880	2,119
Export experience: 10 years	0.758	0.479	0.619	0.684	0.929	2,037
Export experience: 11 years	0.785	0.487	0.636	0.695	0.960	912

Table A.6. Prediction accuracy for time-relevant exporting profile

Note: The table presents the NN prediction accuracies for firms with different export characteristics in the testing sample. Exporting scores are estimated for each characteristic in table's lines: constant exporters, non-exporters, switch to exporter, switch to non-exporter and temporary exporter.

Cross-validation	Sensitivity	Specificity	Balanced accuracy	ROC	PR	Nb. Obs
Train on 90% & test on 10%, take 1	0.649	0.956	0.803	0.922	0.778	24,823
Train on 90% & test on 10%, take 2	0.660	0.953	0.807	0.923	0.783	24,823
Train on 70% & test on 30%	0.658	0.955	0.806	0.932	0.791	74,468
Train on 50% & test on 50%	0.665	0.949	0.807	0.920	0.778	124,113

Table A.7. Prediction accuracies with different samples: Cross-validation

Note: We change the sizes of the training and testing samples and, simultaneously, the firms in the training and testing samples to asess if the high accuracy of the NN model remains.

Appendix B: Coefficients estimated for risk groups

Variables	Coefficients
Intercept	11.3188***
	(0.0566)
risk 2	0.4137***
	(0.0392)
risk 3	-0.1028*
	(0.0556)
risk 4	0.5867***
	(0.0415)
risk 5	0.8814***
	(0.0436)
risk 6	0.6775***
	(0.0458)
risk 7	0.7171***
	(0.0471)
risk 8	1.4652***
	(0.0483)
risk 9	2.16***
	(0.0495)
risk 10	2.7048***
	(0.0516)
Adj R2	0.41
Nb. Obs	202,268

Table B.1. Coefficients of dummies for risk categories defined from the deciles of scores

Note: The table reports estimated coefficient for the dummies associated to firms' risk categories, defined from the deciles of exporting scores. The dependent variable is the natural logarithm of total assets. We consider sector and year fixed-effects, and the standard errors are clustered at the firm level. ***, **, and *, indicate the 1%, 5%, and 10% significance levels, respectively.

Appendix C: Sectors

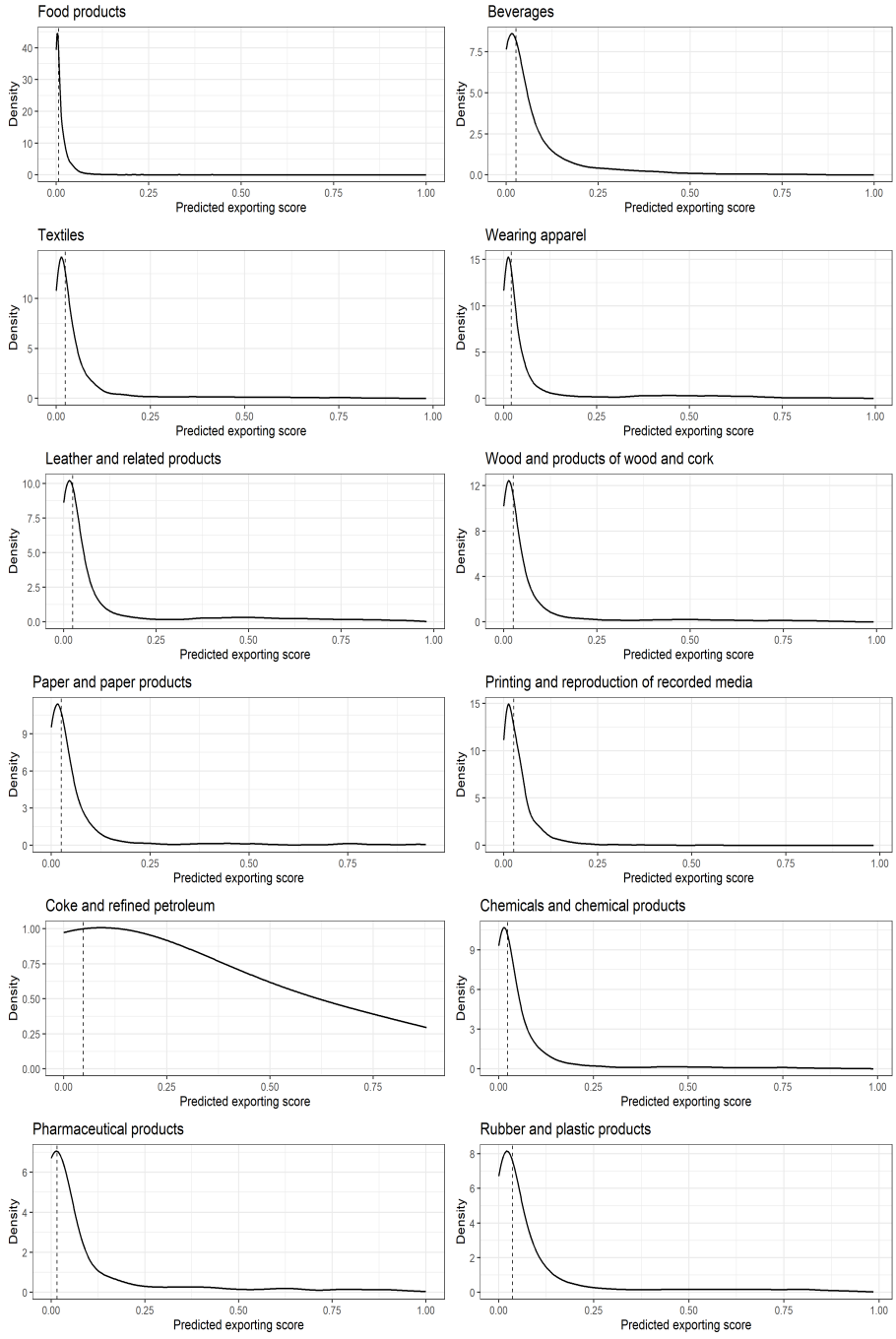


Figure C.1: Scores distributions according to firms' sector

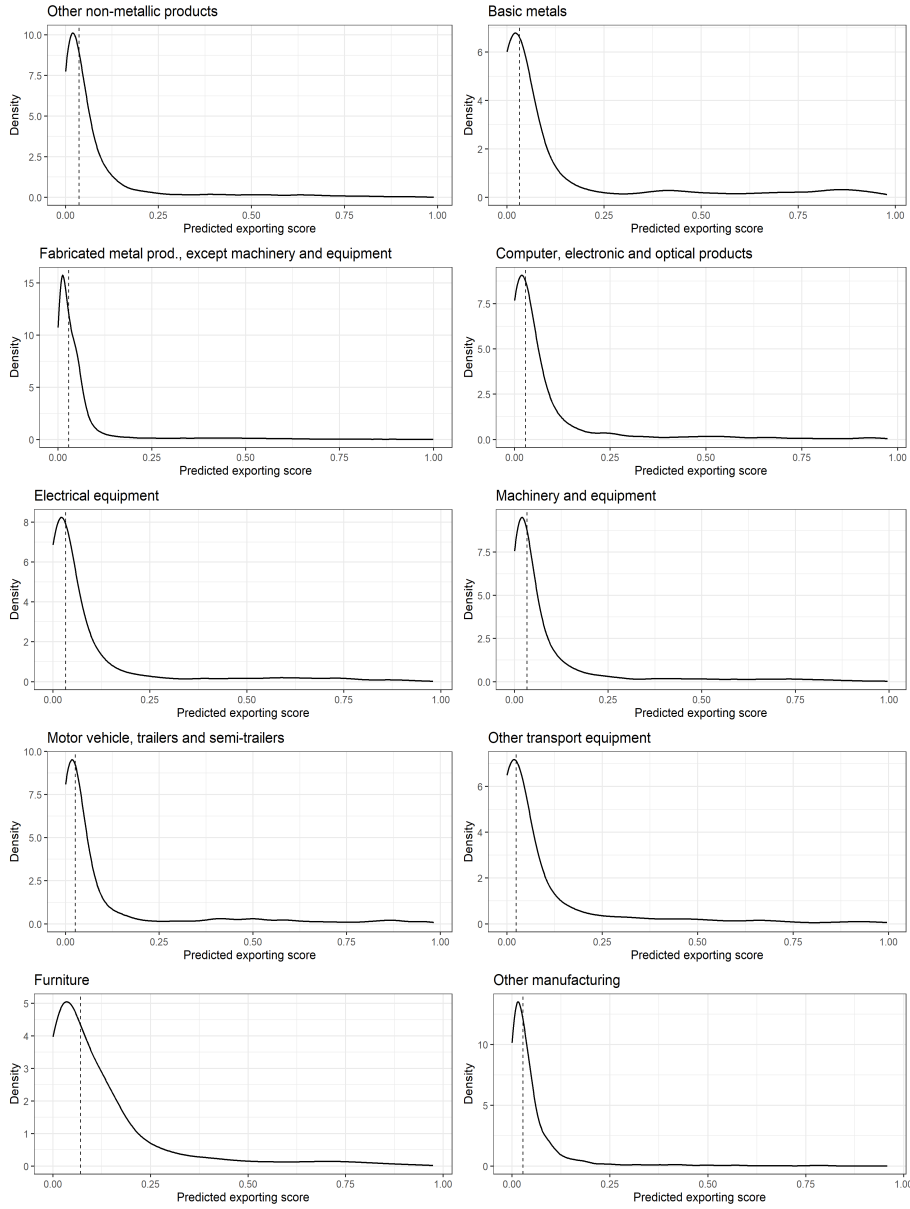


Figure C.1: Scores distributions according to firms' sector

Note: The plots present the distribution of the estimated export scores of all non-exporting firms in our sample by sector, generated by the model with higher accuracy, NN. The vertical dashed line is the median of the predicted scores.