Gabinete de Estratégia e Estudos

REPÚBLICA PORTUGUESA
ECONOMIA

GEE Paper

182

Julho de 2024

Distance to Export: A Machine Learning Approach with Portuguese Firms

Paulo Barbosa | João Cortes | João Amador

Gabinete de Estratégia e Estudos do Ministério da Economia
Office for Strategy and Studies of the Ministry of Economy
Avenida da República, n.º 79 - 1069-218 Lisboa, Portugal
www.gee.gov.pt
ISSN (online): 1647-6212

REPÚBLICA PORTUGUESA
ECONOMIA

Gabinete de Estratégia e Estudos

Gabinete de Estratégia e Estudos

REPÚBLICA PORTUGUESA
ECONOMIA

Distance to Export: A Machine Learning Approach with Portuguese Firms

Paulo Barbosa [1], João Cortes [2] e João Amador [3]

Abstract

This paper estimates how distant a firm is from becoming a successful exporter. The empirical exercise uses very rich data for Portuguese firms and assumes that there are non-trivial determinants to distinguish between exporters and 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 shed light on the critical factors driving the transition to successful export ventures. Neural Networks outperform the other techniques and remain highly accurate when we change the export definitions and the training and testing strategies. We show that the most influential variables for prediction are labour productivity, firms' goods and services imports, capital intensity and wages.

JEL Classification: F17, C53, C55, L21

Keywords: Machine learning, Forecasting exporters, Trade promotion, Micro level data, Portugal

[1] ISEG - University of Lisbon and Portugal Trade & Investment, paulofbarbosa5@gmail.com.

[2] Portugal Trade & Investment, cortes.joao97@gmail.com.

[3] Banco de Portugal and Nova SBE, jamador@bportugal.pt.

1. Introduction

Most countries have national strategies to boost exports, varying from economic diplomacy and trade agencies to direct grants and subsidies to firms. The empirical trade literature has been stating that it is beneficial that a large number of firms are engaged in international trade because such status is positively correlated with size, productivity, wages and tax payments (Wagner, 2007; Coad and Vezzani, 2019).

A firm chooses to export if it is productive enough to cover the fixed costs associated with selling abroad, i.e., firms selling overseas would already be larger, more productive and and pay higher wages even without being present in international markets. This is the canonical view in international trade literature and suggests a strong self-selection of the most productive firms into international markets (Roberts and Tybout, 1997; Melitz, 2003).

However, more recently, another strand of literature argues that market size matters to understanding firms' productivity, i.e., stronger access to overseas markets positively affects firms' innovation and productivity. This is the learning-by-exporting argument. Therefore, government efforts to increase the share of domestic firms that are present in international markets can lead to more investment in technology and innovation, thus increasing overall productivity and economic growth (De Loecker, 2007; Harrison and Rodrígues-Clare, 2010).

In this context, Liveeva and Trefler (2010) studied the consequences of cutting the U.S. tariffs on Canadian products after the Canada-U.S. Free Trade Agreement. They found that with the elimination of the tariff, both firms that started exporting and that already exported to the U.S. increased labour productivity, further engaged in product innovation, and adopted more advanced manufacturing technologies. Atking et al. (2017) presents further empirical evidence of learning-by-exporting. The authors selected a random sample of handmade rug producers in Egypt that had never exported before and were given the opportunity to sign a contract to sell abroad to a high-income country. They concluded that after the experiment, the firms produced higher quality products, increased total labour hours, and increased profits on average between 16% and 26%.

Several authors reported evidence of the government's success in increasing the share of firms in the international market (see Srhoj et al. (2023) for a survey). Cruz (2014), using matching difference-in-differences, showed 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. In the same way, Brooks & Van Biesebroeck (2017) and Munch & Schaur (2018) studied government efforts to increase the number of exporters in Belgium and Denmark, respectively. The first study showed that the firms that received the support had an 8.5 percentage point (pp) more probability of selling in an extra European Union market than those that had not received the treatment. The second work reports evidence that, for small firms, the Danish government support increased the chance of firms 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 effectiveness, 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-exporter manufacturer has of selling 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; Merton, 1974). 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 (2021), that explores the utilization of ML models to measure French firms' distance-to-export and predict the rise of healthy exporters. This paper uses a database that considers about 30\% containing of French manufacturers and only for medium and large firms. In contrast, our paper uses a very rich dataset containing the universe of Portuguese manufacturers, with information on its 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.

In terms of the methodology, our first step is to randomly split the sample into training and test data sets, with a proportion of 0.80 and 0.20, respectively. Then, we train our machine learning 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. With the NN, the model with the highest accuracy in our empirical exercise, we estimate the exporting scores for each non-exporter manufacturer. The estimated score indicates how likely the firm is to propose successfully in international markets. We report that most Portuguese manufacturers are far from becoming exporters, i.e., having low values of exporting scores.

Equipped With the NN -- the model with the highest accuracy in our empirical exercise -- we estimate the exporting scores for each non-exporter manufacturer. The estimated score indicates how likely the firm is to successfully engage in international markets. We conclude that most Portuguese manufacturers are far from becoming exporters, i.e., there is large density of firms with a low exporting score.

For robustness purposes, we run a large battery of tests on the ML models. Even after changing both the exporter definition and the training and forecasting strategies the estimated ML model maintains high accuracy power. Nevertheless, interestingly, although all the methods we use in this exercise have high accuracy, the predictions generated with the NN technique are quite different from those relying on regression and classification trees methods.

Another quite important distinctive feature of our paper are results on the power of different indicators included in the battery of predictors. The variables with high importance for predicting exporters, when controlling for the manufactures' size and sector, are labour productivity, whether it imports from the EU or from extra-EU countries, its capital intensity and the average wage. More specifically, high labour productivity is positively associated with high a probability of exporting. Similarly, the existence of imports from EU or extra-EU markets, high capital intensity and high wages, are highly correlated with selling in international markets. In addition, manufacturers near the median scores operate with 7.3 times fewer fixed assets than those at the top decile of exporting scores.Our empirical exercise can be helpful for all institutions that provide credit for the firms' internationalization, from private institutions to public trade promotion agencies, when targeting programs and resources. With better risk information, private funds can reduce credit prices. Export promotion agencies can focus on helping firms with higher scores to sell abroad and allocate public funds to finance manufacturers with a higher probability of becoming healthy exporters.

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 to trade promotion agencies, in order to better target programs of incentive. With higher quality information on risk, private funds can reduce credit prices. 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.

The present paper is organized as follows. The next section briefly overviews the relevant literature on machine learning in international trade. Section 3 describes the dataset and presents some descriptive statistics on the data. In Section 4, we elucidate our identification strategy and the ML techniques. Then, in Section 5, we present the results; in Section 6, we report the robustness checks; and in Section 7, we introduce time in the ML models to investigate how sensible the model is when predicting for firms with many years of experience exporting and without export experience. In Section 8, we interpret the most relevant variables; in Section 9, we provide a basic interpretation of the estimated scores; and finally, Section 10 concludes.

2. Literature review

Our paper is at the interception of two research strands, i) the export promotion literature, and ii) the utilization of machine learning techniques to address economic questions. We briefly try to survey the literature that lies in the neighbourhood 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 mostly focus on small to medium-sized firms. Authors highlight a continuous increase in sample sizes, statistical sophistication, and number of control variables. The literature started to take account further for external environmental factors as critical determinants of export performance, such as domestic market characteristics and market orientation. The increase in sample size and statistical sophistication allows for more robust empirical and theoretical frameworks, better capable of testing inherent principles and theories of international trade.

Many of the concepts related to deep learning and ML date back to the early 40's 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, Y. (2022) highlights that the ML models more widely used in international trade literature are the grey prediction model, the parallel vector machine model, and deep learning methods such as artificial neural networks, convolutional neural networks, and recurrent neural networks. These models are identified as having 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 a selected number of provisions related to technical barriers to trade, anti-dumping, trade facilitation, subsidies, and competition policy are more effective at promoting trade than other provisions appearing in preferential trade agreements. In a similar vein, Dai, C. (2023) used a Back Propagation Neural Network (BPNN) for forecasting foreign trade export volume, reaching a model 15-30% more accurate than traditional methods. Pan et al. (2018) apply supervised machine learning to predict companies' business success 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 an exporter and measure its current distance to export. Micocci, F. and Rungi, A. (2021) exploits ML techniques to evaluate whether and how close firms are to becoming successful exporters using financial information on French firms for the period 2010-2018. The authors strongly highlight the need to know about the power of different indicators included in the battery of predictors and to know 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, some of the methods used in our paper provide useful insights.

Estimating how likely firms are to become exporters is also the primary purpose of our paper. To create a model capable of identifying ex-ante whether or not a firm will be selling overseas and also to determine its distance to export allows export promotion agencies to optimize their resources and policies, using information on the business environment and firms' characteristics. While this conceptual approach is standard in finance literature to derive a distance-to-default measure, Micocci, F. and Rungi, A. (2021) is the only one we have identified that has addressed the relevance of deriving a distance to export measure, following from what financial institutions make to predict credit risk, for example, in the case of traditional Altman's Z-scores (Altman, 1968)or Merton's Distance-to-Default (Merton, 1974).

3. Data

The present paper is based on a unique firm-level dataset resulting from a merge of three different sources: (i) the Statistics Portugal - INE (Instituto Nacional de Estatística) dataset on export firms; and (ii) the Simplified Corporate Information - IES (Informação Empresarial Simplificada).

The INE dataset contains micro-level information on Portuguese exports. The firms declare export information to customs authorities, who provide this information to the statistical authorities. INE does a sample check of the accuracy of this information for firms that export as of 250,000.00 euros. Therefore, the accuracy of this dataset becomes even higher for medium and large enterprises.

The IES data set contains the mandatory information annually declared by firms for tax administration 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 essential variables such as number of employees, cost of inputs and turnover.

Table 1: Firms by industry

NACE rev. 2	code	Manufacturers non-exporters	exporters	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 products of wood and cork	16	3118	736	3854	6.5
Paper and paper products	17	488	227	715	1.2
Printing and reproduction 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
Fabricated metal prod., except machinery and equipment	25	8546	2017	10563	17.7
Computer, electronic and optical products	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, trailers and semi-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

Note: The source of the firms' main sector is the IES dataset, classified according to NACE Rev. 2. The table shows percentage values (%) corresponding to the weight of each sector and is ordered by NACE Rev. 2 code.

The dataset covers all manufacturers that may or may not have exported between 2010 and 2021. Table 1 presents the manufacturers' industries in our dataset. Note that the industry with the most significant number of 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%).

4. Identification strategy

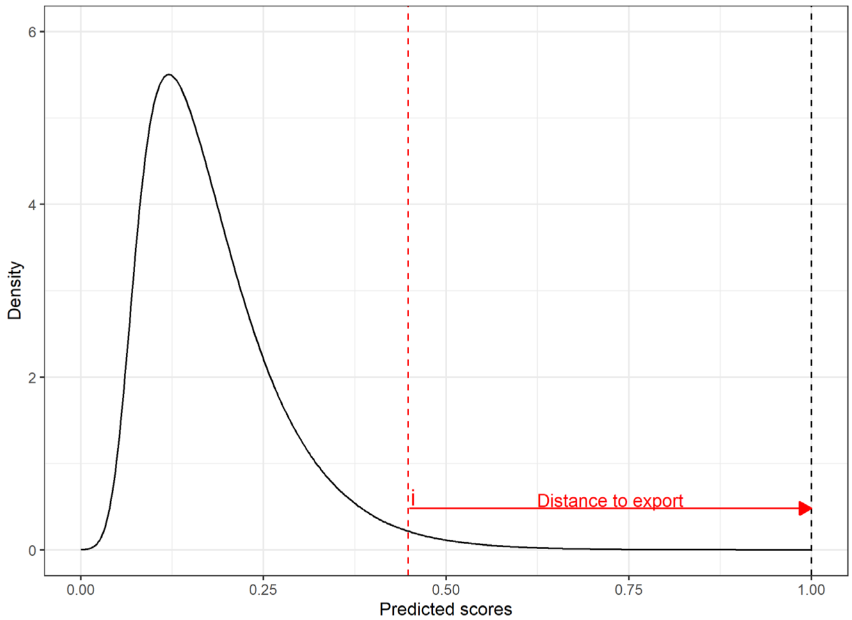
Inspired by the credit risk literature that calculates firms' distance to default (Altman, 1968; Merton, 1974; Uddin, 2021) as well as by the works of Micocci and Rungi's (2023) predicting exporters with ML, we calculate Portuguese manufacturers' distance to become successful exporters. For this purpose, we trained different ML models on our sample of firms that may have exported or not from 2010 to 2021. We randomly split the sample of manufacturers in the standard 80-20 proportion; the first part we use for training the models, and the remaining 20% for testing.

In our empirical exercise, financial information and firms' characteristics may contain non-trivial information on exporting abilities. 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 unseen data, i.e., our out-of-sample test sample. To ensure the generalizability of our model and prevent overfitting, we cease model optimization upon analyzing the results from the testing set.

Therefore, after testing and comparing different ML methods, we applied the best-performing model to calculate the export scores of Portuguese manufacturers. This score will indicate how far a non-exporting manufacturer can become a successful exporter. Figure 1 represents what we expect from the non-exporters score distribution since most non-exporters should have a low probability of exporting, and only a small share of firms should have a high probability of successfully accessing international markets. As indicated in Figure 1 through the red vertical arrow, our intuition is to calculate the distance of the ith firm from becoming an exporter.

Two concerns arise from the estimated exporting scores: i) how we classify a firm as an exporter, and ii) what threshold we should take from the estimated exporting scores in the test sample to perform out-of-sample predictions. Micocci and Rungi (2023) consider exporting firms to be those selling abroad at any value. Geishecker et al. (2019) highlight the possibility of including passive exporters with this not-so-strict definition. The authors define passive exporters as firms that systematically engage in a temporary trade. In our empirical exercise, we avoid classifying passive exporters as exporters as we aim to predict firms that sell abroad consistently, i.e., healthy exporters. In addition, our sample warrants attention because a significant portion consists of micro and small manufacturers, who are more likely to participate passively in the export market (Békés and Muraközy, 2012). Therefore, we define exporters as firms selling at least 10% of their total sales abroad.

Figure 1: Visual illustration of the export score of Portuguese non-export manufacturers.



Note: The plot shows the distribution of an illustrative simulated export score with a similar shape to what we expect for the actual score distribution of the Portuguese non-exporting manufacturers (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 firm i, indicating its fictional probability of success in international markets. On the other hand, the black vertical dashed line represents our benchmark to export, i.e., the distance the ith firm is to become an exporter. Our intuition is to calculate how far the ith firm is to becoming an exporter, i.e., how different the predicted score is from our benchmark (1)

To address the second concern, we classify the firms in the testing sample as exporters when they exceed a threshold of 0.5. This 0.5 threshold is commonly used in trade literature to classify firms' export status in the testing sample (Micocci and Rungi, 2023; Baier et al., 2014). Therefore, we estimate the exporting score from the manufacturers in the test sample, and then we apply this rule to access the models' accuracies.

4.1 Methods

We train and perform predictions over the test sample with the six ML methods, and then we choose the best ML model for predictions of the exporter. A generic representation of the ML predictions is:

$$E(Y_i \mid \boldsymbol{X}_i) = P(Y_i = 1 \mid \boldsymbol{X}_i), \qquad (1)$$

where Y_i is the binary outcome, assuming value 1 if the *ith* manufacturer is an exporter and 0 otherwise. $\boldsymbol{X}_i$ is a k-dimensional matrix that includes the k predictors we use in the models. The output of Equation 1 is what we call *exporting scores*. Moreover, the estimated model does not capture the time dimension; instead, it considers a firm's export status with the same time predictors. In this baseline model, a firm can be considered an exporter in one year and a non-exporter in another. In sections 6 and 7, we look at time-heterogeneous exporting patterns.

Following the conventional data selection 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. We keep the remaining 20% of the firms as an out-of-sample for predicting the export status, i.e., to predict whether the firms are exporters.

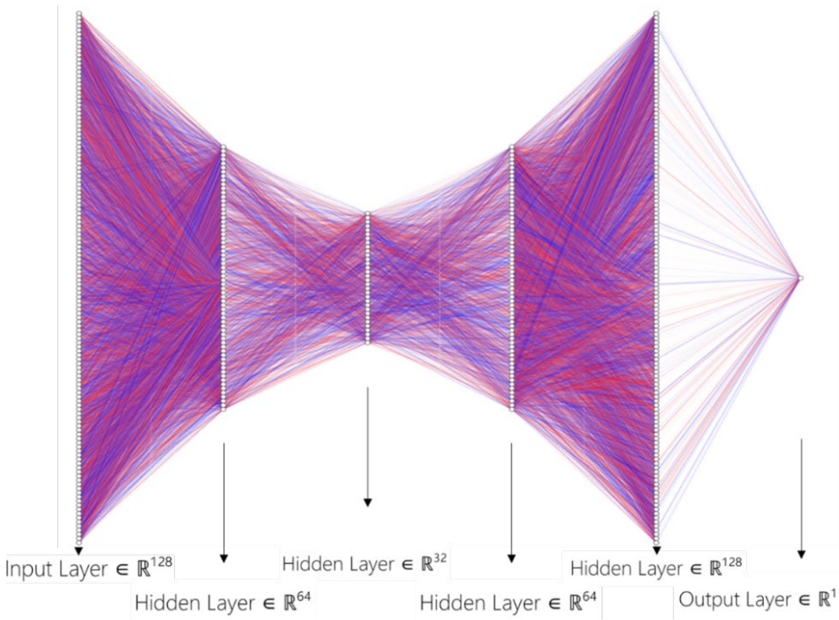
For our empirical exercise, we perform predictions with six models: Logit, Logit Lasso, Random Forest, BART, BART-MIA, and Neural Networks (NN). The BART, BART-MIA, and Random Forest (RF) are methods based on classification and regression trees. In the following sections, we show that the NN technique predicts exporters better than other ML models.

An NN simulates many interconnected processing units that resemble abstract versions of neurons. These processing units are arranged in three-layer groups: the input, the hidden layers, and the output, representing the target fields. 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. & Misra, 1996). The NN model is trained in batches of records, generating a prediction for each and adjusting weights based on its margin-of-error by iteration. This process is repeated each epoch until an optimal is reached (for a deeper discussion, see Hastie et al., 2017).

The best-performing Neural Network includes four hidden layers, each from 32 to 128 artificial neurons (nodes). These neurons include a Rectified Linear Unit (ReLU) 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 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 unless its maximum validation accuracy is reached earlier and, in that case, after eight additional epochs, the early stopping regularization prevents the model from continuing its training process. Then, the best model weights are restored for prediction.

The BART, BART-MIA, and Random Forest methods are classification tree models, where the BART-MIA creates new trees to predict when missingness in some of the features exists. The BART-MIA has high predictive power when the missingness values are non-random (Kapelner and Bleich, 2015). 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).

Figure 2: General representation of the best performing Neural Network Model.



Note: Six layers represent the NN. These layers process information received in the input layer, from left to right, up to the final 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 (ReLU) 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.

We prune the tree iteratively using a regularizer (R) if it grows excessively complex to prevent overfitting (Kapelner and Bleich, 2015; Micocci and Rungi, 2023). The result is a sum of trees as follows:

$$P(Y_i = 1 \mid \boldsymbol{X}_i) = \boldsymbol{\Phi}\Big(\xi_1(\boldsymbol{X}_i) + \cdots + \xi_q(\boldsymbol{X}_i)\Big). \quad (2)$$

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

We perform prediction with two other methods: 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 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 compare the previous methods with the LOGIT-Least Absolute Shrinkage and Selection Operator (LASSO), with the following form:

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

Where y_i the binary outcome of interest, indicating whether the manufacturers 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, to identify the most representative set of features, we use Mutual Information Classification and SelectKBest library (Pedregosa et al. 2011). The total variables generated through the method represented 90% of the dependent weight measured between the independent variables and the dependent variable. Similarly, we use the library to indicate the best subset of features for a more parsimonious model. We were able to sub select the 80% most significant features as well through this method without any additional variable selection criteria. We predicted with the more parsimonious model for robustness check and show that the decrease in accuracy is marginal. Additionally, we use this simpler model to investigate the relationship between the predictors and the output, i.e., to estimate the Shapley value discussed in Section 8.

4.2 Interpreting machine learning models

An understandable, interpretable and transparent model can provide a more informed, unbiased and ethical decision-making process. Huang et al. (2020) divide transparency into three levels: (i) Simulatability – having the model inputs, 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 behaviour 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, i.e., there is no easy way to interpret the variables's importance for predictions and how they connect the output in most ML techniques. However, one of our goals is to understand which firms' characteristics influence the probability of Portuguese manufacturers selling in international markets. Thus, to provide interpretability and identify the most relevant variables for the predictions in the chosen ML model, we use the Shapley values method, which can explain the results of any ML model.

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. Initially introduced by Lloyd Shapley, the concept of Shapley values follows the principle of averaging each marginal contribution over all possible orders in which each feature may have contributed to the predicted variable. Another way to interpret The Shapley values is as a cooperative game theory concept, corresponding to the average marginal contribution of a feature after considering all possible combinations (Alves et al., 2022). Thus, 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.

4.3 Predictors

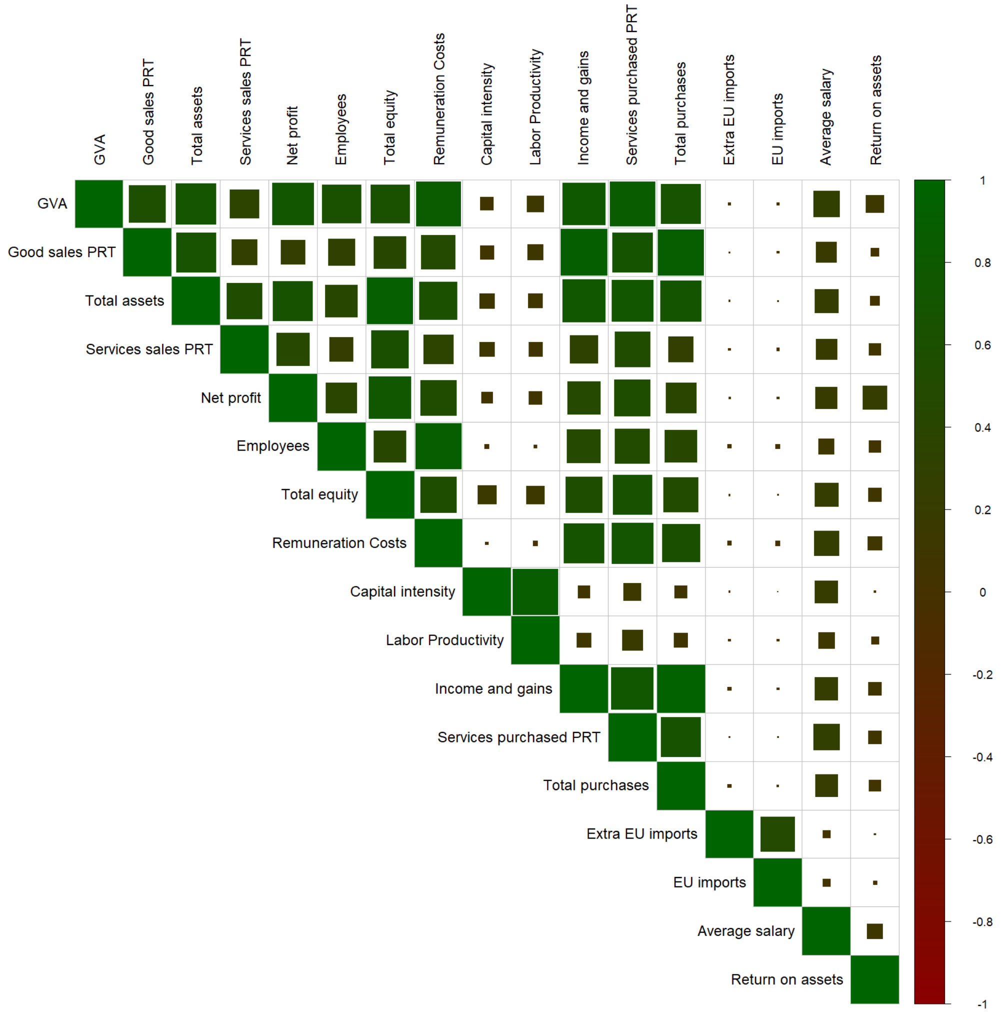
To increase the ML models' power, i.e., higher accuracy, we include all available 17 variables plus controls for sector, size, and year as predictors. The sectors are divided according to the NACE Rev. 2 code as described in Table 1. According to the European Commission[4], we create the manufacturers' size, classifying the firms into micro, small, medium, and large groups.

The correlation matrix, Figure 3, includes all numerical variables we use to predict successful exporters. Many of these variables have a high correlation, varying from 0.6 to 0.9. However, in our first empirical exercise to predict manufacturers with a high potential of selling successfully abroad, high correlation is not a concern. In other words, as we do not aim in the first empirical exercise to estimate international trade determinants, i.e., how variables are related to international trade, we use all available features for the prediction exercise despite the high correlation between some of them. Thus, we do not discriminate *ex-ante* variables in this stage for the ML models. In the robustness check, we apply the SelectKBest library (Pedregosa et al., 2011) to identify the most relevant variables in the estimated ML models, and the forecast with a subset of all variables renders a marginal decrease in the forecast accuracy. However, in our second empirical exercise to understand how the variables are

[4] More details about the definitions of micro, small, medium, and large firms can be found at the link: https://single-market-economy.ec.europa.eu/smes/sme-definition_en

associated with successful exporters, we use the most significant variables, i.e., the subset generated from the SelectKBest library.

Figure 3: Correlation Matrix



Note: The matrix reports the Pearson correlation of all numerical variables used in the ML models. Non-numerical variables, such as firm's size and sector are not reported. Positives correlations are displayed in shades of green and negatives in shades of red. In Annex A we report the variables descriptions.

5 Results

5.1 Predictors

In the present section, we report the models' out-of-sample performance. Table 2 summarises the accuracies of the model's predictions on the test sample. We use standard confusion matrix metrics to evaluate the accuracies: Specificity, Sensitivity, Balanced Accuracy, Receiver Operating Characteristic (ROC) curve, and Precision Recall (PR).

Specificity 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 gives the Balanced Accuracy. More importantly, the ROC curve, represented in Figure 5, evaluates the models' predictive performance at different classification thresholds, and it is our primary metric for evaluating the models. Finally, we also report the Precision Recall (PR), which evaluates the trade-off between precision and recall while comparing the relationship between the True Positive Rate and False Positive Rate. We display the PR for the estimated ML models in Figure B1.

Table 2. Prediction accuracies on the test sample

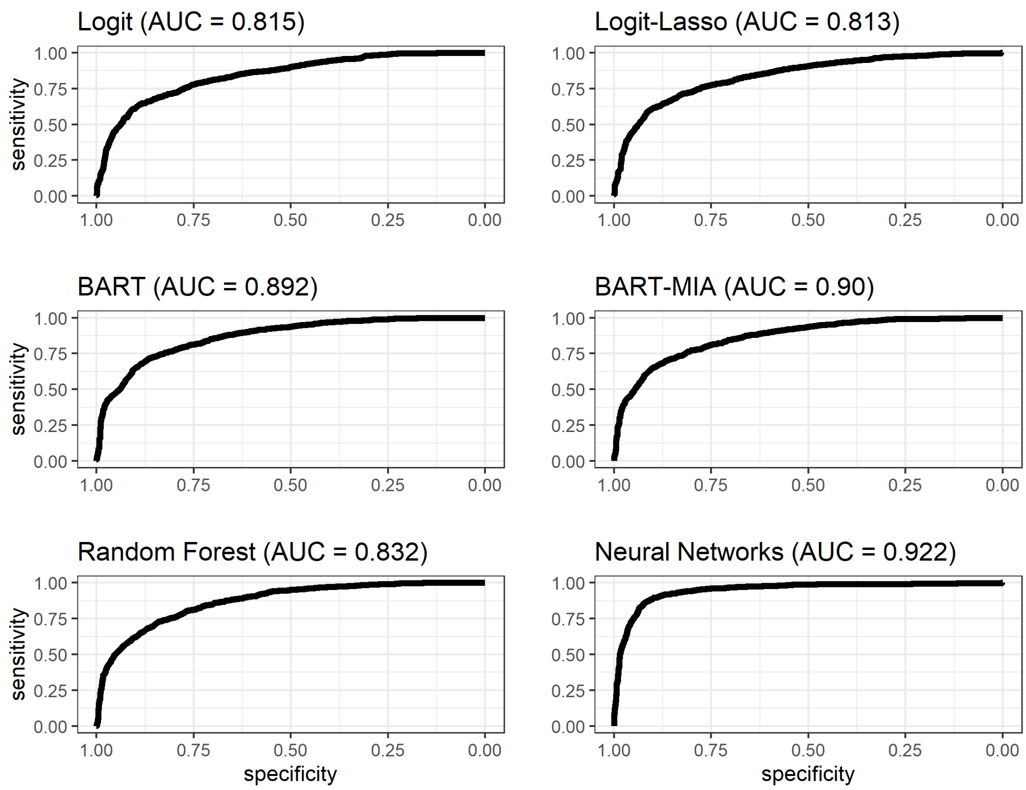
Prediction accuracies						
Method	Sensitivity	Specificity	Balanced accuracy	ROC	PR	N. obs.
Logit	0.867	0.536	0.702	0.815	0.765	40 165
Logit-Lasso	0.862	0.522	0.692	0.813	0.765	40 165
Random Forest	0.874	0.56	0.717	0.832	0.793	40 165
BART	0.902	0.619	0.7605	0.892	0.798	40 165
BART-MIA	0.897	0.615	0.756	0.9	0.798	40 660
Neural Networks	0.962	0.62	0.791	0.922	0.91	40 165

Note: The table presents the results 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 these pieces of information, i.e., the missing values, as new predictors. The remaining techniques drop missing values when calculating the score probabilities.

The standard measures reported in Table 2 show that the Neural Network (NN) model outperforms the other models with a higher ROC of 0.91. Figure 5 reports the models' ROC curves. The area under the curve (AUC) evaluates the models' performance. A higher AUC means higher accuracy, while a 0.5 AUC, i.e., a diagonal ROC line, means a random prediction.

Moreover, the NN balanced accuracy of 0.79 is the highest, with eight percentage points (pp) higher than the BART-MIA, the second-highest balanced accuracy. In terms of sensitivity, i.e., the ability of the models to classify exporters correctly in the predictions, the six machine learning methods predict exporter firms with a high accuracy varying from 0.86 to 0.902. The NN method still outperforms the other methods, whether considering the sensitivity. Regarding specificity, i.e., the ability to predict non-exporters correctly, the NN and BART accuracies are very close: 0.62 and 0.615, respectively.

Figure 4: ROC curves



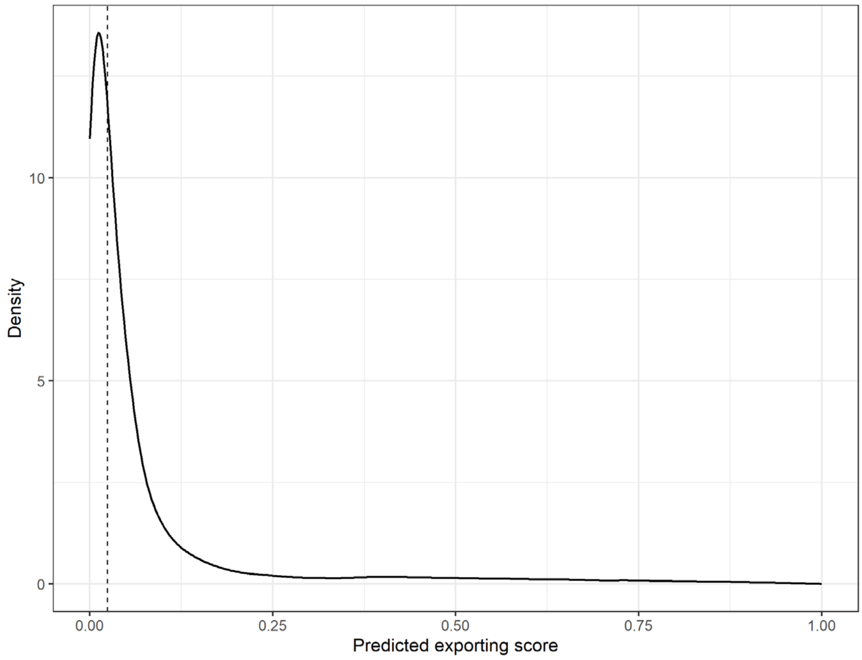
Note: The graphs report the ROC curve and the area under the curve (AUC) for the 6 machine learning techniques used to forecast exporters.

BART and BART-MIA models perform similarly. Their predictions are similar because our sample has a small percentage of missing values. 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, i.e., using BART-MIA over BART. The reason is that the sample the authors 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 nonobligatory financial reports to authorities.

5.2 Estimations

Figure 6 presents the estimated export scores of all non-exporting firms in our sample obtained from the NN. We selected all non-exporting firms without any threshold (or additional requirement) and estimated their distance to becoming successful exporters. The non-exporters score distribution is right-skewed, as most scores are on a thick left tail, i.e., have a low export score, thus far from successfully entering international markets.

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



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

The scores distribution of non-exporting manufacturers is consistent with the concept of heterogeneous firms, as shown in the trade literature. According to the international trade theory with heterogeneous firms, a company chooses to sell overseas if it is productive enough to cover its fixed cost. This is the most popular in trade literature view and suggests a strong self-selection of firms with higher productivity turning to international markets (Roberts and Tybout, 1997; Melitz, 2003).

6 Robustness check

In the results section, when comparing the models' performance for predicting exporters, we have followed the traditional rule to split firms in training and testing samples with the standard rule 80-20%, i.e., we used 80% of firms for training and the others 20% for testing the six models' predictions. We have defined exporters as firms selling at least 10% of their total sales abroad. In this section, we investigate how modifying the definitions of exporters, training and test samples, and other relevant terms affects the chosen machine learning model's predictive power.

Our primary concern regarding the robustness of our results is how sensitive our predictions are to changes in the definition of an exporter firm. Therefore, we perform additional predictions using the NN with different definitions of an exporter. We will define a firm as an exporter based on the following criteria:

1. Selling at least 5% of its total sales abroad;
2. Selling at least 15% of its total sales abroad;
3. Exporting outside the EU market;
4. Selling to at least three countries;
5. Selling to at least six countries;
6. Selling to at least ten countries.

Table C1 (in Annex C) presents the results of changing the exporter definition. Changing the exporter definition threshold from 10% to 5% and 15% of total sales abroad did not significantly impact the predictions. At 5%, the balanced accuracy is about 0.82, and the estimated ROC is 0.9. Increasing the threshold to 15% resulted in a slight improvement in precision (balanced accuracy of 0.87 and ROC of 0.95).

Similarly, changing the definition of an exporter to cases 3 to 6, i.e., considering exporter firms that sell outside the EU market, selling to at least three, six, and ten countries, the predictions' accuracies remain high for the four cases. The balanced accuracies vary from about 0.71 to 0.78. More importantly, all the estimated ROCs are over 0.8, which means that the model still predicts with good precision for these alternative definitions of an exporter.

A second concern is whether the accuracy of the predictions changes regarding firms' size. We train and perform out-of-sample predictions for micro, small, medium, and large firms with the NN method to address this issue. Results are reported in Table C1. For all firms' dimensions, prediction accuracies remain high. The estimated ROCs vary from 0.874 to 0.966, increasing with firms' dimensions, i.e., the model has higher accuracy with manufacturers of larger dimensions. The model's ability to predict exporters correctly (Sensitivity) for micro firms is 0.3, while it increases to 0.96 for large firms.

On the other hand, the model predicts non-exporters with higher accuracy for micro firms than for large firms, 0.98 versus 0.83. These differences in accuracies regarding the firms' size may be related to the smaller firms' temporary export behaviour since they are more likely to engage in short trade agreements (Békés and Muraközy, 2012). We investigate this behaviour in the next section, exploring whether temporary trade affects the predictions' accuracy.

A third concern is whether reducing the number of predictors can significantly influence our results, specifically if the NN model can have similar results with a smaller subset of predictors. This change can decrease computational costs for the estimations and result in a simpler model without relinquishing predictive power. In the same way as the original feature selection procedure elaborated in the Identification Strategy Section, the least amount of features representing the most significant dependency weight to the explainable variable were selected to attempt this exercise. Performing our predictions with only a subset of features (16 features, including controls for Year and firms' Sector) mildly affected our estimator's overall performance. We report the results in Table B2. The reduction of features does not have a high effect on predictions since the estimated ROC decreases by about 3%. However, in our empirical exercise, we do not have a reason to exclude available features, even with a high correlation between them, to estimate the exporting scores.

Another concern is whether time-variant specific relationships are affecting the general model. It can be manifested from the subset of years used for training and testing in the machine learning model, consequently affecting predictions' accuracy. In other words, our concern is whether forecasting for a new year can significantly change accuracy. Indeed, in a real-life scenario, an analyst would have data for a range of past years and want to predict his current year based on a model trained on information from past years. To answer this concern, we have performed predictions testing the NN model over different years. We present the results in Table B4. Accuracies have remained 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. Therefore, the predictions' accuracy stays mostly the same for different years, i.e., performing out-of-sample predictions over different years does not change the NN's accuracy.

Our fifth concern addresses whether the Neural Network's high accuracy remains with different training and testing samples and was not due to the train-test split used. In order to check for this effect, we cross-validated our original sample, repeating the prediction exercise four times with random training and testing samples of different sizes each time. Thus, we train the model on 90% of the firms and predict the rest, becoming the out-of-sample information. We repeat this process twice, and alternatively, we randomly split the training and testing samples in 70-30% and 50-50%. Table B5 shows that we obtained similar results in all different exercises.

Finally, to assess the sensitivity of firms' predicted export status to the machine learning technique chosen, we calculated the correlations between firms' export status based on each of the six machine learning techniques used in our empirical exercise. Results are reported in Table 4. As the Spearman's correlations rank presents, the NN approach generates predictions that are weakly correlated to predictions from the other techniques. The Neural Network varies its correlation rank with other methods from 0.58 to 0.64, while other techniques' predictions perform similarly, varying from 0.757 to 0.947.

The correlations' rank between models reveals results similar to Micocci and Rungi (2023), who also performed a similar empirical exercise to ours without considering the NN technique. The authors found a high correlation of around 0.9 between the predictions from Logit, Lasso, BART, and BART-MIA methods. As expected, these results confirm how NN predictions of exporters differ significantly from other regression models and regression-tree-based techniques, which generate less accurate predictions.

Table 4. Spearman's correlations rank of the predicted export status

	Logit	Logit-Lasso	Random Forest	BART	BART-MIA	Neural Networks
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	
Neural Networks	0.584	0.582	0.625	0.64	0.636	1

Note: The table presents the correlation rank of exporters' out-of-sample predictions to understand how connected (unconnected) each technique's prediction is to others.

7 Sensitivity to temporary trade

In a heterogeneous firm scenario, only the most productive firms sell to international markets, while others try to avoid incurring sunk costs. In this way, Békés & Muraközy (2012) demonstrate, using Hungarian data, that a substantial proportion of firms have a temporary trade behaviour. Following this idea, we do a similar exercise to Micocci and Rungi (2023) to understand if our results' precision varies in the presence of temporary exporting activity. One of our aims is to predict exporters in the presence of temporary exporters, i.e., when firms engage in a short-term export activity. This pattern can happen more frequently for micro and small firms, as well as manufacturers of capital goods. However, this behaviour could not be random, i.e., it can be associated with some of the firms' characteristics, and then our machine learning technique can incorporate this information in its predictions (Békés and Muraközy, 2012; Micocci, F. and Rungi, 2023). Thus, we perform estimates with the NN model in the presence of consistency-dependent trade patterns for five different firm group types:

1. Firms that exported in all periods of our sample (**constant exporters**);
2. Firms that never exported (**non-exporters**);
3. Firms that were non-exporters until some period and then started exporting all periods after a period t (**switching to exporters**);
4. Firms that were exporters until some period t and then became non-exporters in all periods afterward (**switching to non-exporters)**;
5. Finally, we classify those with irregular export behaviour, i.e., firms that changed between exporting and non-exporting more than once as **temporary exporters**.

Table 5 reports the predictions' accuracies for each time-relevant exporting group, estimated using the NN technique for the five groups. The results show that the original model predicts well for constant exporters and non-exporters (representing 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.

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 minor compared to predictions for the entire test sample with an ROC of 0.94, as reported in Table 3. We observe in Table 5 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 last group, the firms that switched to exporters (or to non-exporters) more than once, which we call temporary exporters, the NN model's predictions are poor. The estimated ROC is under the acceptable ROC of 0.8. Micocci and Rungi (2023) argue that firms with this characteristic, engaging in temporary trade, continue to do so systematically, which is consistent with the fewer precision estimates of the NN model.

In addition, we tested whether our results are sensitive to temporary trade using the definition proposed by Bekes and Murakozy (2012), we report the results in Table C6. 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 exporter 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.

Table 5. Prediction accuracy for time-relevant exporting profile

Firm Category	Sensitivity	Specificity	Balanced Accuracy	ROC	PR	Num. Obs.
Constant Exporters	0.855	-	-	-	-	19323
Non-Exporters	-	0.984	-	-	-	166782
Switching to Export	0.611	0.834	0.723	0.802	0.807	11829
Since t0	0.644	0.802	0.723	0.811	0.946	2588
Since t1	0.67	0.81	0.74	0.808	0.885	1767
Since t2	0.612	0.825	0.718	0.805	0.846	1550
Since t3	0.546	0.864	0.705	0.804	0.783	1063
Since t4	0.58	0.809	0.695	0.778	0.699	914
Since t5	0.528	0.877	0.703	0.793	0.66	961
Since t6	0.535	0.857	0.696	0.792	0.644	634
Since t7	0.508	0.81	0.66	0.72	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	0.33	548
Since t10	0.511	0.731	0.621	0.679	0.159	564
Switching to Non-Export	0.456	0.893	0.674	0.806	0.765	7511
Since t1	0.223	0.898	0.561	0.692	0.304	2489
Since t2	0.343	0.898	0.62	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	0.7	0.833	0.902	344
Since t8	0.708	0.83	0.769	0.868	0.944	376
Since t9	0.616	0.896	0.756	0.89	0.961	510
Since t10	0.756	0.776	0.766	0.839	0.964	409
Since t11	0.781	0.727	0.754	0.84	0.98	264
Temporary Exporters	0.437	0.845	0.641	0.736	0.637	42780
Export Experience: 1 year	0.166	0.911	0.538	0.66	0.188	10597
Export Experience: 2 years	0.283	0.85	0.566	0.671	0.338	6819
Export Experience: 3 years	0.309	0.833	0.571	0.671	0.45	5218
Export Experience: 4 years	0.388	0.789	0.589	0.68	0.555	3748
Export Experience: 5 years	0.459	0.738	0.598	0.669	0.637	3284
Export Experience: 6 years	0.506	0.692	0.599	0.678	0.712	2940
Export Experience: 7 years	0.573	0.664	0.619	0.681	0.769	2705
Export Experience: 8 years	0.624	0.639	0.632	0.698	0.835	2401
Export Experience: 9 years	0.683	0.535	0.609	0.675	0.88	2119
Export Experience: 10 years	0.758	0.479	0.619	0.684	0.929	2037
Export Experience: 11 years	0.785	0.487	0.636	0.695	0.96	912
All sample	0.54	0.74	0.614	0.614	0.441	248225

Note: The table presents the NN's prediction accuracies for firms with different export characteristics in the testing sample. The NN model is trained according to Section 5 (Results). Exporting scores are estimated for each characteristic in the table: firms having always exported, never exported, switched to exporters, non-exporters, and temporary exporters.

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 considered algorithms. However, considering the NN method, the estimated ROCs are still over the acceptable threshold of 0.8.

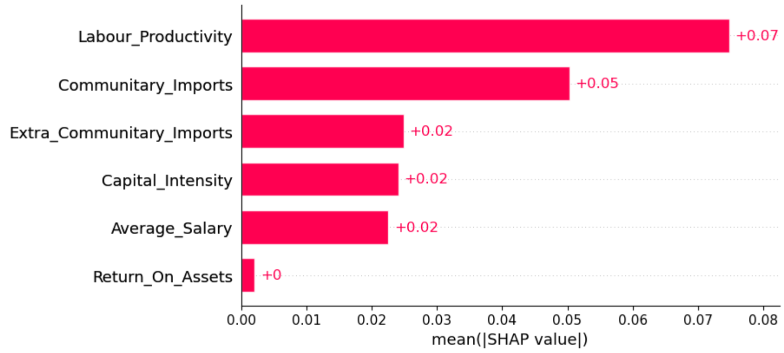
8 Interpreting predictors

In previous chapters, we have demonstrated how the Neural Network technique consistently maintained its high accuracy levels despite changes in definitions, features, sampling and other levels of robustness. However, these validation steps did not emphasize the explainability behind the model's predictions and what additional knowledge it brings to the policy discussion for each feature it includes. In this chapter, we will focus on understanding the main features impacting the model's performance and how each feature affects the model individually and at an observational level. We use the Shapley values (SHAP) discussed in Subsection 4.3 to understand the importance of the main variables of the model on the estimated exporting scores. Once again, it is essential to highlight that it is not possible to access casual effects through the SHAP. However, we can understand the importance of the variables in the predictions and how they relate to the outcome.

For the robustness check, we predict the exporting scores with a subset of variables, and the results reported in Table C3 are similar to the full set of variables. In this exercise, the estimated ROC is only 3% lower. Thus, in the present section, we focus on investigating the importance of these variables on the output. The variables in the subset are *Average Salary*, *Capital intensity*, European Union Imports (*Community imports*), *Extra-community imports*, *Labour productivity* and *Return on assets*. Moreover, we always control for the Year and firms' size and sector (2-digit code by NACE Rev. 2).

Therefore, we report in Figure 6 how each feature of the NN model affects the exporting scores, considering the absolute SHAP, i.e., without accounting if the effect of each variable is positive or negative on the predictions. *Labour productivity* is the most relevant feature for estimating exporting scores, i.e., predicting the distance needed to become a successful exporter. The second and third most important variables are the firms' imports from the European Union and the extra-European Union imports. Interestingly, from the correlation matrix (Figure 2) these variables related to the imports are not correlated with the other predictors. *Capital Intensity* and *Average Salary* are the fourth and fifth most important variables, and they have similar weights in the predictions. The ratio Return on Assets shows weights near zero on the predictions.

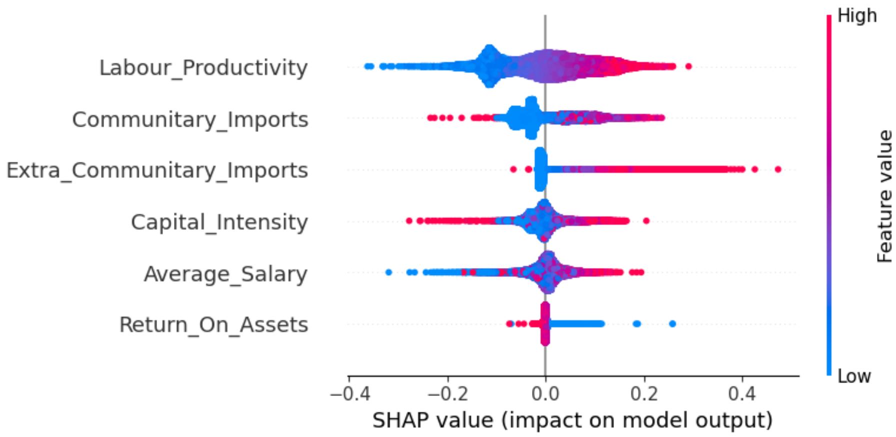
Figure 6. SHAP Bar Plot



Note: The plot summarizes the individual contribution, in absolute values, from each feature to the mean SHAP value, meaning each feature's average contribution to a firm's probability to export. The variables are in standardized values, and controls for size, sector and year are omitted.

In Figure 7, we report how each variable affected the estimated exporting scores, i.e., positively or negatively. The Y-axis denotes the variables, and the X-axis indicates the mean Shap value. Each observation is represented in dots (firm-year). The higher the observation contribution to the features that stand for a positive (negative) contribution of each value to place them inside (outside) of the category (positive or negative weight), and the colours indicate the influence with a high (red) or low (blue) value.

Figure 7. Relevance of the variables considering individual contributions



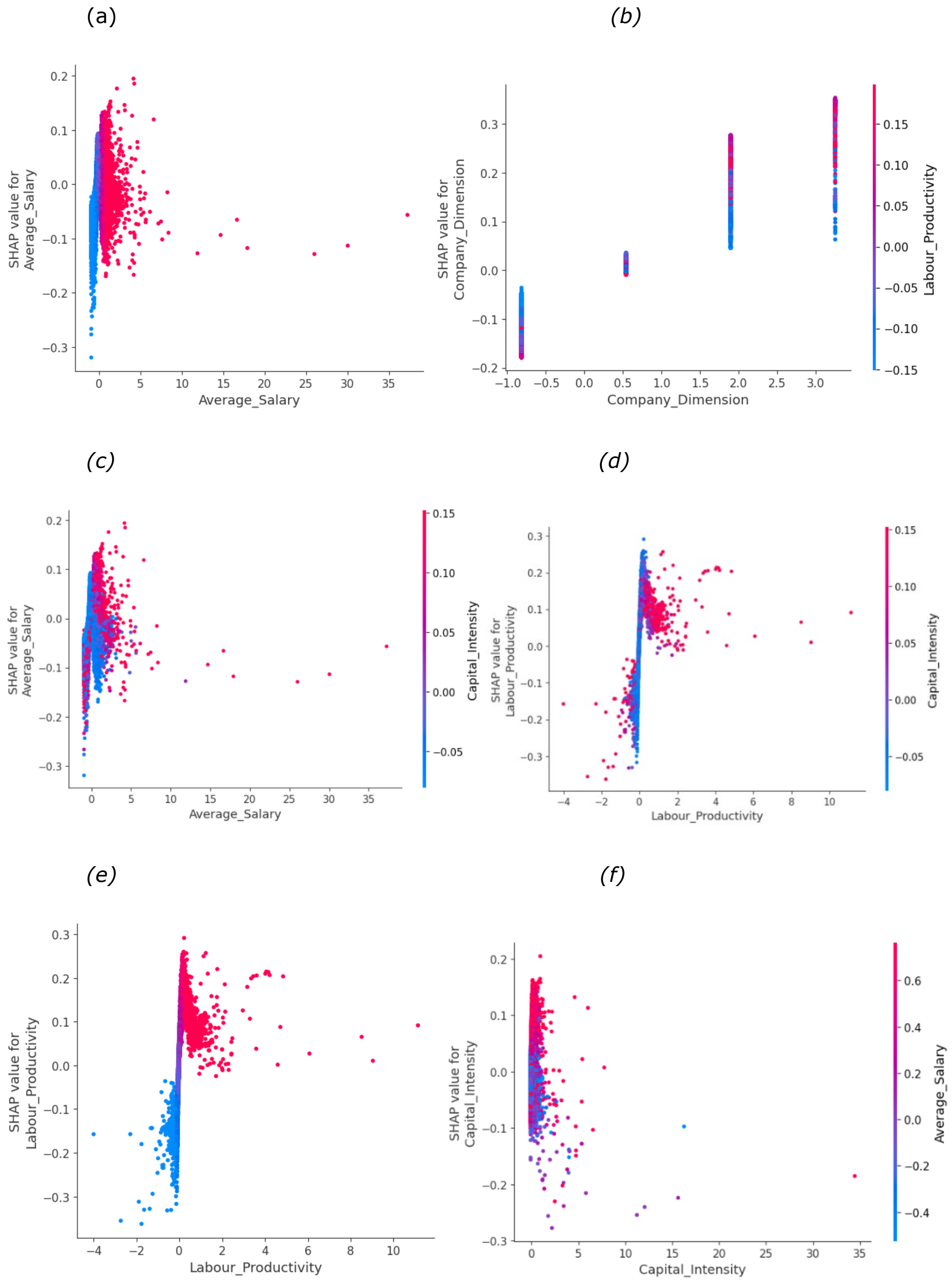
Note: Features are ordered in the Y-axis by importance from top to bottom. The X-axis denotes the SHAP value, indicating the degree of change in log odds. The chromatic spectrum of each point signifies the associated feature value, wherein red denotes high values and blue signifies lower values. Each observation is represented by a point, and the values are standardized, and controls for size, sector and year are omitted.

Labour Productivity, the variable with higher relevance on the estimated exporting scores, has a high dispersion of individual values, where the blue values represent a negative SHAP and the red ones a positive SHAP. Thus, lower labour productivity has a negative contribution, i.e., less probability of becoming an exporter. On the other hand, firms with higher labour productivity are also more likely to export. This result goes the same way as the trade international literature, where firms avoid sunk costs and only the most productive firms propose on international markets (Roberts and Tybout, 1997; Melitz, 2003).

Considering the firms' imports, firms with low values for *Community* and *Extra-community imports* appear less likely to export. Symmetrically, firms with high import values seem more likely to export. Indeed, it is expected that firms already in international markets, i.e., buying products and services overseas, are more likely to sell abroad. Indeed, firms buying intermediary products and services abroad may acquire cheaper or higher-quality products and services, which characterizes them as more productive and more likely to become successful exporters.

Firms with higher average salaries have higher SHAP, i.e., higher exporting scores. Symmetrically, the blue dots are massively on the negative SHAP side, which means that firms with lower average salaries have a lower probability of export. The SHAP of the capital intensity is ambiguous. There are high values of capital intensity on both sides, negative and positive. However, while the lower values are concentrated around zero, most higher values (red dots) are on the positive side. From these results, we can conclude that firms with higher capital intensity are more likely to sell abroad.

Figure 8. SHAP Dependences



Note: The figure shows the relationship between changes in the control variables and the SHAP, i.e., its probability of becoming an exporter. Interaction relationships are also revealed by a third variable level, distinguished from blue (lowest) to red (highest) values. In each row, colours means the same. The values are standardized.

Figure 8 presents the direct relationship between the predictors from this empirical exercise and the SHAP, i.e., how different values of the predictors are associated with the firm's export probability. For example, in Plot (a), we present the relationship between SHAP and the Average Salary, and the blue (red) colour represents low (high) Labour Productivity values. A positive relationship exists between Salary and the SHAP, i.e., high salary values are associated with a high probability of becoming an exporter, as reported in the trade literature (Melitz, 2003). Moreover, the manufacturers with higher salaries also have, on average, higher levels of labour productivity (red); symmetrically, the ones with lower salaries have lower labour productivity (blue).

Plot (b) in Figure 8, split the manufacturers into four groups based on their sizes: micro, small, medium, and large5. As expected, the manufacturers' size positively relates to the SHAP, i.e., larger firms are more likely to export. Moreover, firms with high levels of labour productivity (red colour) have higher SHAP values, especially for the groups medium and large. The Plot (c) is similar to (a). It shows the same positive relationship between the SHAP and the average Salary paid by the firms, but in (c), the colours mean the Capital intensity level. It shows that manufacturers with higher capital intensity have higher SHAP, i.e., when controlling for the Salary, firms with high Capital intensity are more likely to sell abroad.

Plots (d) and (e) present a strong positive relationship between the SHAP and Labour productivity. As reported in Figure 8, manufacturers with high Labour productivity are more likely to be present in international markets. Additionally, when controlling for Labour productivity, firms with high Capital intensity levels have, in general, higher export scores (Plot (d)). Similarly, firms with higher Salaries have higher SHAP values when controlling for Labour productivity, i.e., when controlling for Labour productivity, firms with higher salaries are more likely to be present in international markets (Plot (e)). Finally, Plot (c) presents the relationship between *Capital intensity* and SHAP. Identifying a clear relationship between *Capital intensity* and SHAP through the Plot is difficult. However, it is clear that when controlling for *Capital intensity*, firms with higher salaries have higher SHAP, i.e., are more likely to be exporting.

9 Interpreting predictions

One of our empirical exercise' aims is to predict how far the firms are to become successful exporters. Figure 3 presents the export scores distribution of the Portuguese non-export

[5] We define the firm's size according to the European Commission definition.

manufacturers according to the chosen NN model. Based on each firm's estimated probability to export, we obtain a simple distance-to-export measure for manufacturers as follows:

$$distance_i = 1 - \Pr(Y_i = 1 \mid \boldsymbol{X}_i), \tag{4}$$

where $distance_i$ is the ith's firm probability of exporting, thus bounded in (0,1). Y_i is a binary variable assuming 1 when the firm is an exporter and 0 otherwise. $\boldsymbol{X}_i$ is the matrix of covariates.

Our estimated export scores are a continuous indicator that can be useful for policymakers, export promotion agencies or credit institutions in different ways. Policymakers can identify the regions and sectors with higher potential for internationalization, i.e., where firms can propose into foreign markets more likely. In Appendix B, we report the distribution of the export scores by industry (2 digit-code by NACE Rev. 2). The industries have, in general, similar thick left distributions, i.e., the firms are located in the left-thick distribution, thus far to propose successfully on international markets. However, some sectors, like 'Food products,' have an even thicker left tail, i.e., the firms are farther from becoming exporters than the other sectors. On the other hand, some industries have a closer distance to selling in international markets, like the Furniture sector.

With the exporting scores, Export Promotion Agencies can also approach firms with high potential to become exporters, i.e., closer to our benchmark of 1, and help them sell abroad. In the same way, firms request grants from credit institutions to cover part or all costs of proposing on foreign markets. Our indicator can help intermediary institutions reduce their risk, consequently reducing firms' cost of credit.

9.1. Exporting scores and grants for exporting

Proposing on overseas markets requires a significant amount of fixed costs. In order to cover at least part of these costs, firms usually need resources from financial institutions. However, if imperfections in the financial market exist, firms may lose export opportunities, i.e., exporting at a lower level than their natural capacity (Manova, 2012). At the aggregated level, the economy is losing trade opportunities that could lead to economic growth once exporting firms pay higher wages, taxes and generate further investments (Melitz, 2003).

With this lack of information, i.e., asymmetric information in the financial market, governments have implemented several strategies for providing grants to more productive firms in order to propose on international markets (see, for example, Comi and Resmini (2020) and Srhoj and Walde (2020)).

Similarly to Micocci and Rungi (2023)[6], we estimate, from the exporting scores, the level of asset resources a firm should have as collateral to propose to international markets. By classifying non-exporting firms into different risk categories, each risk category is categorized in scoring deciles, i.e., the first risk group includes firms with scores in the range of 0-0.9, the second risk group into 0.1-0.19. The first group will thus have a higher risk, i.e., it has a higher distance to the export market (Equation 4). This higher risk means that these firms operate with a lower level of capital than other groups.4

Having each risk group generated from estimated exporting scores in Figure 5, we can estimate the following simple model:

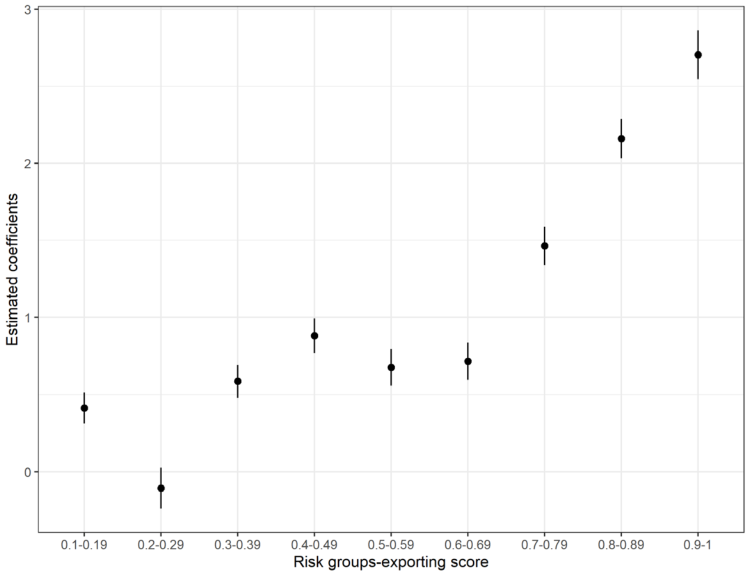
$$\log(y_{it}) = \emptyset_0 + \sum_{j=2}^{10} \emptyset_j risk_j + \alpha_1 X_{it} + \gamma_t + \delta_{ind} + v_{it}, \quad (5)$$

where y_{it} is the fixed assets of firm *i* in time t, X_{it} is control for the firms' size. γ_t is time-control, δ_{ind} is industry's control and v_{it} represents both the invariant and the time-varying error terms.

The parameters of interest of Equation 5 are $\emptyset_j$ (with j = 2,…,10), associated with the indicator variable $risk_j$ created from the estimated exporting score, displayed in Figure 5. The first risk group (with scores from 0 to 0.09) is omitted, i.e., it is the reference group and represented by the parameter $\emptyset_0$. Thus, these coefficients will represent the amount of asset resources each firm should need to operate when compared to the first risk group. As expected, the estimated coefficients are higher than zero, meaning the groups with lower risk, i.e., with lower distance to export, should, on average, be operating with a higher asset volume. Estimated coefficients are reported with a 99% confidence interval in Figure 9 and in C7.

[6] The authors call this exercise "back-to-envelope estimates" because of its simplicity. They use the logs of *total assets* and *cash available* as dependent variables. Since we do not have the *total cash*, we replicate the exercise with the *total assets* as the outcome.

Figure 9: Estimated risk coefficients



Note: The figure reports the estimated risks' coefficients from the simple linear regression in Equation 5, where the dependent variable is Total assets. We control for the firms' size, sector, and year. The errors are clustered at the firm level.

Analysing the estimated parameters of the binary variables indicating the risk groups, we can see that the first group operates with approximately $exp(\widehat{\emptyset}_0) = exp(11.32) \approx 82{,}454$ euros of fixed capital resources. The second group operates with $exp(\widehat{\emptyset}_0 + \widehat{\emptyset}_1) = exp(11.32 + 0.414) \approx 124{,}555$ euros of fixed assets. The third group is not significantly different from zero, and the 4th, 5th, 6th, and 7th groups seem similar. The 4th group operates with $exp(\widehat{\emptyset}_0 + \widehat{\emptyset}_4) = exp(11.32 + 0.587) \approx 148{,}079$, and the 7th group operates with $exp(\widehat{\emptyset}_0 + \widehat{\emptyset}_7) \approx 168{,}704$ euros of fixed assets. Therefore, the Portuguese firms with the highest risk need twice more asset resources to look similar to the medium risk level, i.e., to the 4th-7th groups.

Finally, the most significant differences start after the 8th group. The firms closer to propose on international markets successfully operate with much more level of asset resources. The 8th group operates with $exp(11.32 + 1.465) \approx 356{,}468$, and the group with lower risk, i.e., lower distance to export, operates with $exp(11.32 + 2.7) \approx 1{,}231{,}323$ euros of fixed assets, respectively. The firms in the medium level of risk, i.e., in groups from 4 to 7, operate with 7.3x less fixed capital resources compared with the lowest risk group.

11 Conclusions

In the present paper, we aim to predict how far Portuguese manufacturers are in accessing international markets successfully. For this, we perform predictions with six machine learning models and choose the neural networks method, having higher accuracy in predicting exporters. Our intuition is that there are non-trivial relations between firms' characteristics and their exporting status, and the ML models can identify these relations to predict accurately. Moreover, we aim to understand which variables are more important to the predictions and how these variables are related to the output, i.e., the probability of the Portuguese manufacturers exporting successfully.

The predictions of the NN model outperform the other ML models. In addition, we perform an exhaustive battery of robustness tests on the NN model, where we change the definition of the exporter and different training samples and strategies. Even so, the NN still maintains high predictive power across the tests. Interestingly, despite the excellent performance of all the ML models performing predictions out-of-sample, the predictions generated from the NN model are substantially different from the other ML techniques.

To understand which features are more relevant to predict exporters and investigate how they are related to the outcome, i.e., the probability of exporting, we apply the SelectKBest and SHAP libraries. Thus, when controlling for the manufacturers' size and sectors, the essential variables for prediction are *Labour productivity*, *Imports from EU* markets and *outside the EU*, *Capital intensity*, and *Average salary*. High values of Labour productivity are highly associated with a high probability of selling in international markets. Similarly, high values of imports, capital intensity, and average salary are linked to a high probability of exporting.

In addition, discuss through a simple linear model that the firms that are more distant to export, i.e., those with lower exporting scores, operate with lower fixed asset levels. On average, the group with a 10% lower probability of exporting operates with about 82,454 euros. In contrast, on average, the manufacturers in the comfort zone, i.e., the 10% with a higher probability of exporting, operate with 1.2 million euros.

Our results can be helpful for institutions that promote export grants, trade promotion programs, and trade promotion agencies in targeting manufacturers. These institutions can more effectively target manufacturers in their programs (or grants), thus reducing costs and applying resources more effectively.

One area for improvement of our empirical exercise is the impossibility of measuring causal effects, i.e., how the variables we use as controls can affect the probability of firms selling overseas successfully. Future work could investigate the impact of the Portuguese manufacturers' characteristics on the probability of them selling in international markets consistently.

References

Alves, Tiago et al. (2022) Assessing the Scoreboard of the EU Macroeconomic Imbalances Procedure:(Machine) Learning from Decisions.

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

Athey, Susan et al. (2021) Using Wasserstein generative adversarial networks for the design of monte carlo simulations. *Journal of Econometrics*.

Atkin, David et al. (2017) "Exporting and firm performance: Evidence from a randomized experiment," The Quarterly Journal of Economics, 132 (2), 551–615.

Baier, Scott et al. (2014) Economic determinants of free trade agreements revisited: Distinguishing sources of interdependence. Review of International Economics 22 (1), 31–58.

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

Bleich, Kapelner J. et al. (2014) Variable selection for bart: an application to gene regulation. The Annals of Applied Statistics 8 (3), 1750–1781.

Warner, Brad and Manavendra Misra (1996) Understanding Neural Networks as Statistical Tools, The American Statistician, 50:4, 284-293, DOI: 10.1080/00031305.1996.10473554

Breinlich, Holger et al. (2022) Machine learning in international trade research-evaluating the impact of trade agreements.

Broocks, Annette and Johannes V. Biesebroeck (2017) "The impact of export promotion on export market entry," Journal of International Economics, 107, 19–3

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, 547–581.

Chipman, Hugh A. et al (2010) Bart: Bayesian additive regression trees. The Annals of Applied Statistics 4 (1), 266–298.

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

Dai, Chenglin (2022) A method of forecasting trade export volume based on back-propagation neural network. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-022-07693-5

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

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

Hastie, T., Tibshirani, R., Friedman, J. (2017) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Springer*

Hill, Jennifer et al. (2020) Bayesian additive regression trees: A review and look forward. *Annual Review of Statistics and Its Application* 7 (1), 251–278.

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

Lileeva, Alla and Daniel Trefler (2010) "Improved access to foreign markets raises plant-level productivity. . . for some plants," The Quarterly journal of economics, 125 (3), 1051–1099.

Manova, Kalina (2012) Credit Constraints, Heterogeneous Firms, and International Trade. The 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.

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 et al. (2018) *Machine Learning Prediction of Companies' Business Success*. https://cs229.stanford.edu/proj2018/report/88.pdf

Pedregosa, Fabian et al (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 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, 545–564.

Sousa, Carlos M. P. et al (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. https://doi.org/10.1111/j.1468-2370.2008.00232.x

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.

Srhoj, Stjepan et al. (2023) "Export Boosting Policies and Firm Performance: Review of Empirical Evidence Around the World," Jahrbücher für Nationalökonomie und Statistik.

Tong, Ye (2023) Literature Review in International Trade Forecasting Based in Machine Learning Method. *SHS Web of Conferences*, *154*, 03025. https://doi.org/10.1051/shsconf/202315403025

Van Biesebroeck, Johannes V. (2016) Did export promotion help firms weather the crisis? Economic Policy 31 (88), 653–702.

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

Appendix A: Data

Table A1: List of predictors

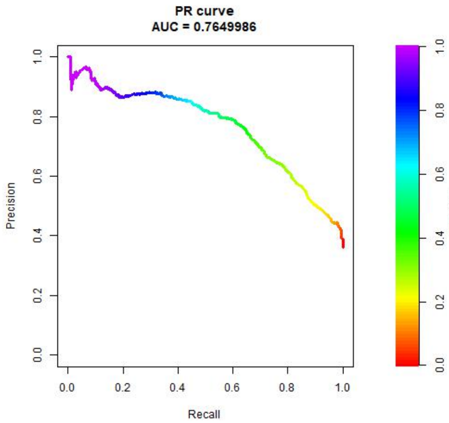
Variable	Description
Gross Value Added (GVA)	
Good Sales PRT	
Total Assets	
Service Sales PRT	
Net Profit	
Employees	
Total Equity	
Remuneration Costs	
Capital Intensity	Total Assets / Nr of Employees
Labour Productivity	GVA / Nr of Employees
Income and Gains	
Services Purchased PRT	
Total Purchases	
Extra-EU Imports	Purchases of intermediary goods of services outside EU / Total Purchases
EU Imports	Purchases of intermediary goods of services in EU markets / Total Purchases
Average Salary	Total Costs with Staff / Nr of Employees
Return on Assets	Net Profit / Total Assets

Note: Variables from our data set, IES and INE sources. We combine the data sources with the unique firm's code identifier. *The total number of products exported is calculated through the Harmonised System (HS) codes.

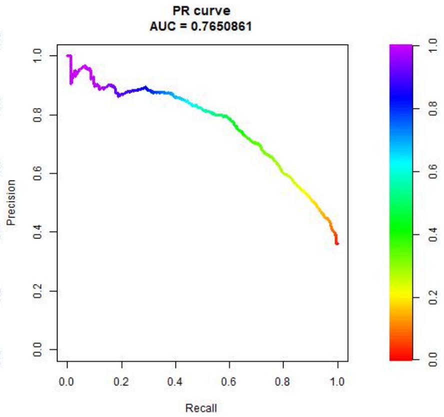
Appendix B: Figures

Figure B1: Out-of sample goodness of fit – Precision-Recall

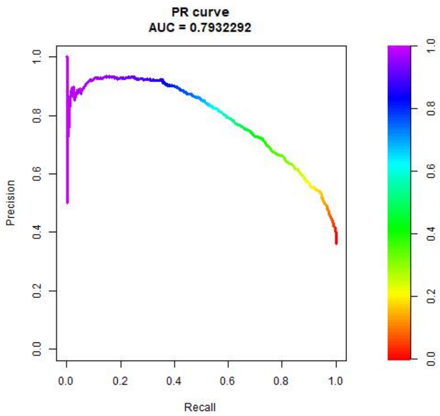
(a) Logit



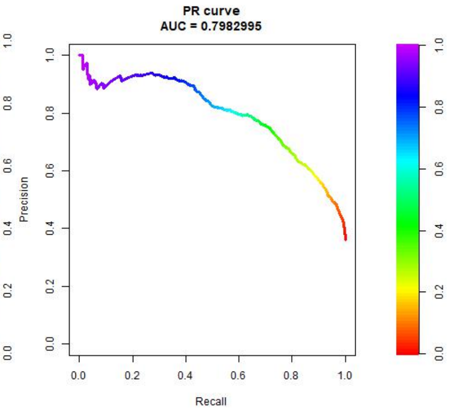
(b) Logit-Lasso



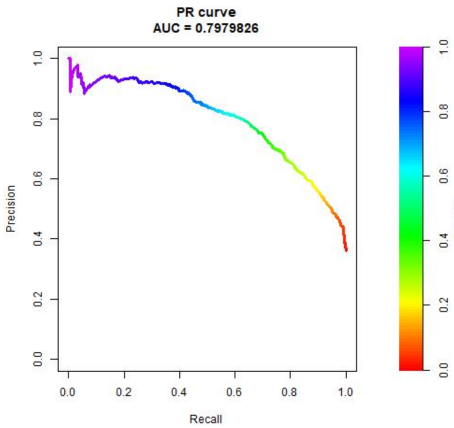
(c) Random Forest



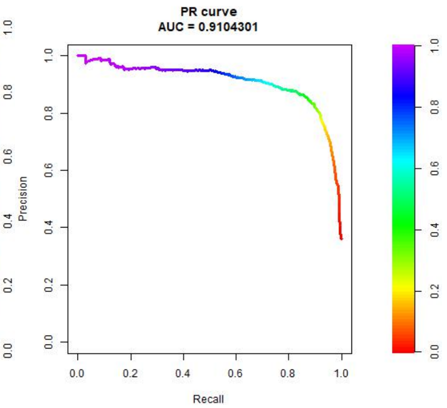
(d) BART



(e) BART-MIA

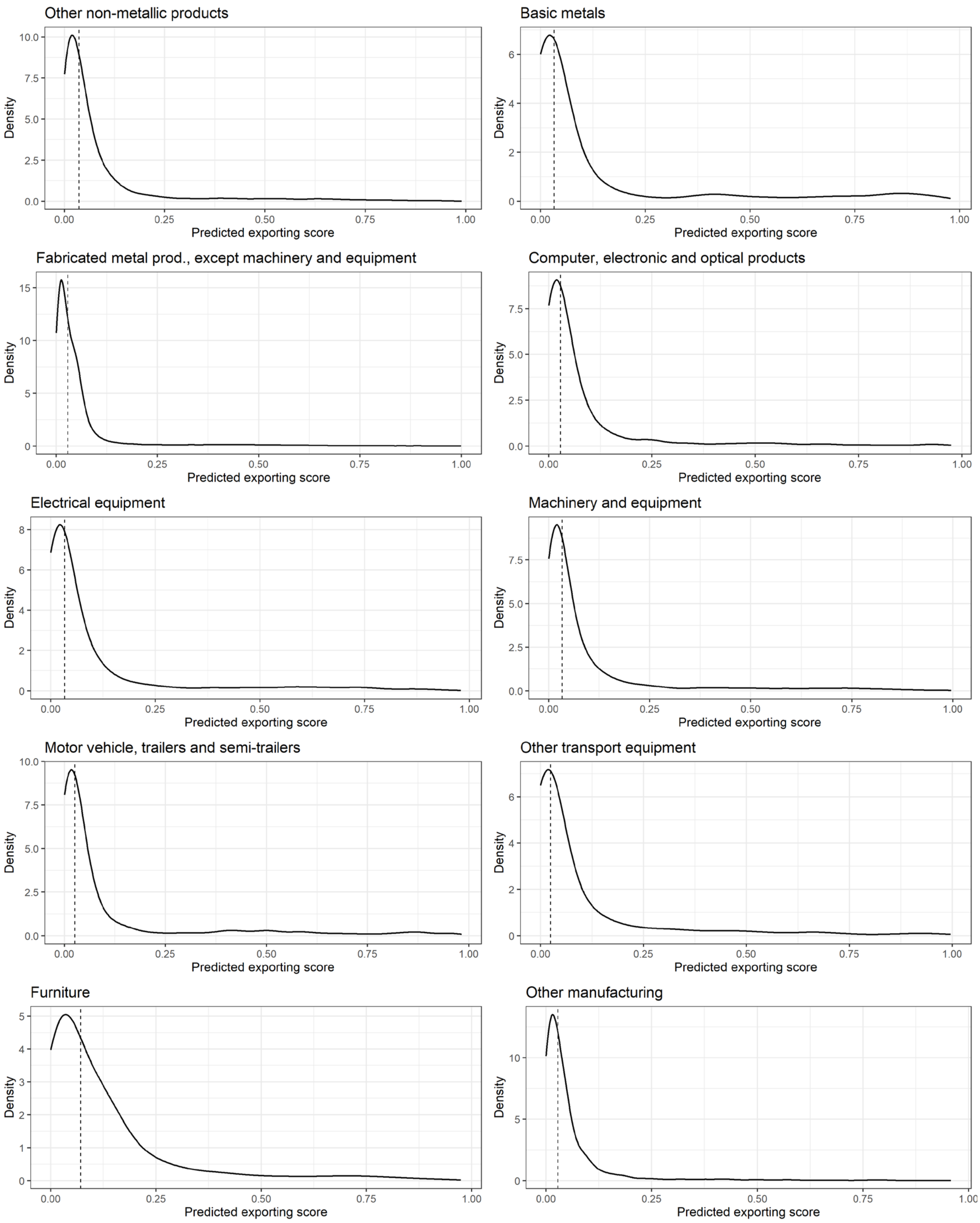


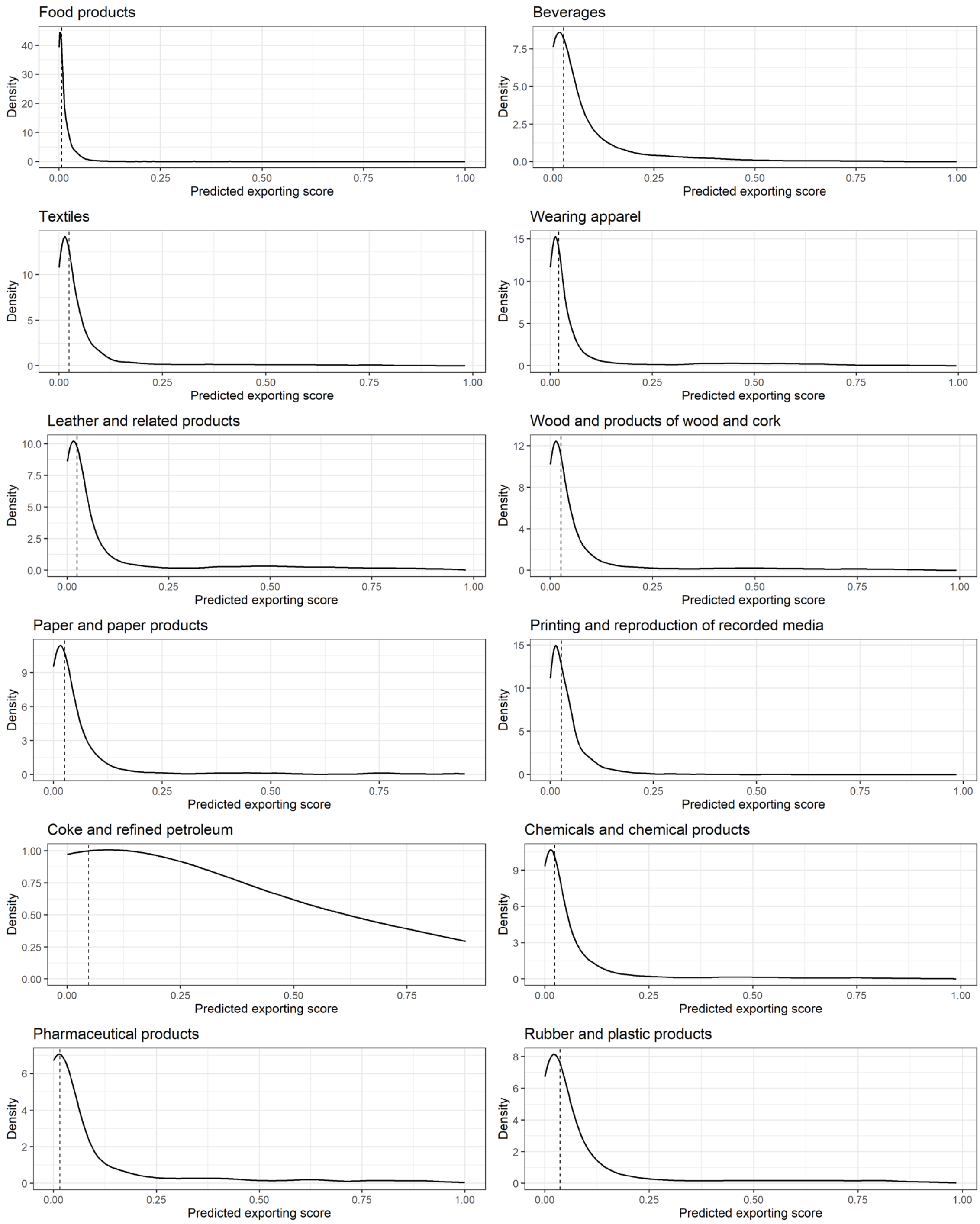
(f) Neural Networks



Nota: The graphs report the Precision-Recall curve and the area under the curve (AUC) of the 6 machine learning techniques used to forecast exporters. The colours indicate the best threshold for defining a firm as exporter.

Figura B2: Scores distributions according to firms' sector




Food products
Beverages
Textiles
Wearing apparel
Leather and related products
Wood and products of wood and cork
Paper and paper products
Printing and reproduction of recorded media
Coke and refined petroleum
Chemicals and chemical products
Pharmaceutical products
Rubber and plastic products
Density
Predicted exporting score

Appendix C: Tables

Table C1: Predictions with different definitions of exporter

Definition of an Exporter	Sensitivity	Specificity	Balanced Accuracy	ROC	PR	N. Obs
Selling at least 5% of its total sales abroad	0.895	0.741	0.818	0.901	0.924	40,165
Selling at least 15% of its total sales abroad	0.897	0.848	0.873	0.946	0.942	40,165
Exporting outside the EU market	0.496	0.942	0.719	0.865	0.730	40,165
Exporting to at least 3 markets	0.738	0.823	0.781	0.861	0.870	40,165
Exporting to at least 6 markets	0.589	0.914	0.752	0.868	0.776	40,165
Exporting to at least 10 markets	0.464	0.960	0.712	0.884	0.698	40,165

Note: Similar to the Results section, where we define an exporter as a firm exporting at least 10% of its sales, here we explore different definitions of an exporting firm. We train and test the Neural Network (NN) model out-of-sample using six different exporter definitions to assess how these definitions affect prediction accuracy.

Table C2: Prediction accuracies according to the firms' sizes

Firms' size	Sensitivity	Specificity	Banlanced Accuracy	ROC	PR	N. Obs
Micro	0.3038	0.9832	0.6435	0.8738	0.4959	164396
Small	0.6469	0.9236	0.7852	0.9102	0.7641	51087
Medium	0.9296	0.8613	0.8765	0.9129	0.9261	33517
Large	0.9643	0.8249	0.8946	0.9659	0.9822	5279

Note: We split the firms into four groups according to their sizes and performed predictions for these groups and perform predictions with the chosen NN model. We define the firms' sizes according to the European Commission.

Table C3: Forecast accuracy with a subset of variables

Sensitivity	Specificity	Balanced Accuracy	ROC	PR	N. Obs
0.654	0.958	0.806	0.910	0.814	40 165

Note: We perform predictions with the chosen NN model with a subset of variables to investigate whether prediction accuracies change

Table C4: Prediction accuracies for different years

Year	Sensitivity	Specificity	Balanced Accuracy	ROC	PR	N. 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

Note: We perform predictions separately for each year of our sample to investigate whether the accuracy of our initial prediction changes over time.

Table C5: Prediction accuracies with different samples: Cross-validation

Cross-validation	Sensitivity	Specificity	Balanced Accuracy	ROC	PR	N. Obs
Training on 90% and testing on 10%, Take 1	0.649	0.956	0.803	0.922	0.778	24,823
Training on 90% and testing on 10%, Take 2	0.660	0.953	0.807	0.923	0.783	24,823
Training on 70% and testing on 30%	0.658	0.955	0.806	0.932	0.791	74,468
Training on 50% and testing on 50%	0.665	0.949	0.807	0.920	0.778	124,113

Note: We change the sizes of the training and testing samples and, simultaneously, the firms in the training and testing samples to investigate whether the high accuracy of the NN technique remains.

Table C6: Prediction accuracies with Békés and Muraközy (2012)'s classification of exporter.

Exporter Class	Sensitivity	Specificity	Balanced Accuracy	ROC	PR	N. 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

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 remaining firms that export at least once; iii) non-exporters are firms that never export.

Table C7: Back-to-envelope estimates

Variables	Coeficients
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
Observations	202,268

Note: The table reports the linear regression coefficients according to the firms' risk categories. The outcome variable is the natural logarithm of the total of assets. We control for the firm's sector and year and the standard errors are clustered at the firm level.

Appendix D: Availability of commands

This link redirects to the github repository where all Python and R codes developed in this empirical exercise have used.

GEE Papers

1: Evolução do Comércio Externo Português de Exportação (1995-2004)
João Ferreira do Amaral

2: Nowcasting an Economic Aggregate with Disaggregate Dynamic Factors: An Application to Portuguese GDP
Antonio Morgado | Luis Nunes | Susana Salvado

3: Are the Dynamics of Knowledge-Based Industries Any Different?
Ricardo Mamede | Daniel Mota | Manuel Godinho

4: Competitiveness and convergence in Portugal
Jorge Braga de Macedo

5: Produtividade, Competitividade e Quotas de Exportação
Jorge Santos

6: Export Diversification and Technological Improvement: Recent Trends in the Portuguese Economy
Manuel Cabral

7: Election Results and Opportunistic Policies: An Integrated Approach
Toke Aidt | Francisco Veiga | Linda Veiga

8: Behavioural Determinants of Foreign Direct Investment
Ricardo Pinheiro-Alves

9: Structural Transformation and the role of Foreign Direct Investment in Portugal: a descriptive analysis for the period 1990-2005
Miguel de Freitas | Ricardo Mamede

10: Productive experience and specialization opportunities for Portugal: an empirical assessment
Miguel de Freitas | Susana Salvado | Luis Nunes | Rui Costa Neves

11: The Portuguese Active Labour Market Policy during the period 1998-2003 - A Comprehensive Conditional Difference-In-Differences Application
Alcina Nunes | Paulino Teixeira

12: Fiscal Policy in a Monetary Union: Gains from Changing Institutions
Susana Salvado

13: Coordination and Stabilization Gains of Fiscal Policy in a Monetary Union
Susana Salvado

14: The Relevance of Productive Experience in the Process of Economic Growth: an Empirical Study
Diana Vieira

15: Employment and Exchange rates: the Role of Openness and Technology
Fernando Alexandre | Pedro Bação | João Cerejeira | Miguel Portela

16: Aggregate and sector-specific exchange rate indexes for the Portuguese economy
Fernando Alexandre | Pedro Bação | João Cerejeira | Miguel Portela

17: The Macroeconomic Determinants of Cross Border Mergers and Acquisitions and Greenfield Investments
Paula Neto | Antonio Brandao | António Cerqueira

18: Does the location of manufacturing determine service sectors' location choices? Evidence from Portugal
Nuno Crespo | Maria Paula Fontoura

19: A hipótese do Investment Development Path: Uma Abordagem por Dados em Painel. Os casos de Portugal e Espanha
Miguel Fonseca | António Mendonça | José Passos

20: Outward FDI Effects on the Portuguese Trade Balance, 1996-2007
Miguel Fonseca | António Mendonça | José Passos

21: Sectoral and regional impacts of the European Carbon Market in Portugal
Margarita Robaina Alves | Miguel Rodriguez | Catarina Roseta-Palma

22: Business Demography Dynamics in Portugal: A Non-Parametric Survival Analysis
Alcina Nunes | Elsa Sarmento

23: Business Demography Dynamics in Portugal: A Semi-parametric Survival Analysis
Alcina Nunes | Elsa Sarmento

24: Digging Out the PPP Hypothesis: an Integrated Empirical Coverage
Miguel de Carvalho | Paulo Júlio

25: Regulação de Mercados por Licenciamento
Patrícia Cerqueira | Ricardo Pinheiro Alves

26: Which Portuguese Manufacturing Firms Learn by Exporting?
Armando Silva | Óscar Afonso | Ana Paula Africano

27: Building Bridges: Heterogeneous Jurisdictions, Endogenous Spillovers, and the Benefits of Decentralization
Paulo Júlio | Susana Peralta

28: Análise comparativa de sobrevivência empresarial: o caso da região Norte de Portugal
Elsa Sarmento | Alcina Nunes

29: Business creation in Portugal: Comparison between the World Bank data and Quadros de Pessoal
Elsa Sarmento | Alcina Nunes

30: The Ease of Doing Business Index as a tool for Investment location decisions
João Zambujal Oliveira | Ricardo Pinheiro Alves

31: The Politics of Growth: Can Lobbying Raise Growth and Welfare?
Paulo Júlio

32: The choice of transport technology in the presence of exports and FDI
José Pedro Ponte | Armando Garcia Pires

33: Tax Competition in an Expanding European Union
Ronald Davies | Johannes Voget

34: The usefulness of State trade missions for the internationalization of firms: an econometric analysis
Ana Paula Africano | Aurora Teixeira | André Caiado

35: The role of subsidies for exports: Evidence from Portuguese manufacturing firms
Armando Silva

36: Criação de empresas em Portugal e Espanha: análise comparativa com base nos dados do Banco Mundial
Elsa Sarmento | Alcina Nunes

37: Economic performance and international trade engagement: the case of Portuguese manufacturing firms
Armando Silva | Oscar Afonso | Ana Paula Africano

38: The importance of Intermediaries organizations in international R&D cooperation: an empirical multivariate study across Europe
Aurora Teixeira | Margarida Catarino

39: Financial constraints, exports and monetary integration - Financial constraints and exports: An analysis of Portuguese firms during the European monetary integration
Filipe Silva | Carlos Carreira

40: FDI and institutional reform in Portugal
Paulo Júlio | Ricardo Pinheiro-Alves | José Tavares

41: Evaluating the forecast quality of GDP components
Paulo Júlio | Pedro Esperança | João C. Fonseca

42: Assessing the Endogeneity of OCA conditions in EMU
Carlos Vieira | Isabel Vieira

43: Labor Adjustment Dynamics: An Application of System GMM
Pedro Esperança

44: Corporate taxes and the location of FDI in Europe using firm-level data
Tomás Silva | Sergio Lagoa

45: Public Debt Stabilization: Redistributive Delays versus Preemptive Anticipations
Paulo Júlio

46: Organizational Characteristics and Performance of Export Promotion Agencies: Portugal and Ireland compared
Inês Ferreira | Aurora Teixeira

47: Evaluating the forecast quality of GDP components: An application to G7
Paulo Júlio | Pedro Esperança

48: The influence of Doing Business' institutional variables in Foreign Direct Investment
Andreia Olival

49: Regional and Sectoral Foreign Direct Investment in Portugal since Joining the EU: A Dynamic Portrait
Irina Melo | Alexandra Lopes

50: Institutions and Firm Formation: an Empirical Analysis of Portuguese Municipalities
Simão Arouca

51: Youth Unemployment in Southern Europe
João Leão | Guida Nogueira

52: Financiamento da Economia Portuguesa: um Obstáculo ao Crescimento?
João Leão | Ana Martins | João Gonçalves

53: O Acordo de Parceria Transatlântica entre a UE e os EUA constitui uma ameaça ou uma oportunidade para a Economia Portuguesa?
João Leão | Guida Nogueira

54: Prescription Patterns of Pharmaceuticals
Ana Gonçalves

55: Economic Growth and the High Skilled: the Role of Scale Eects and of Barriers to Entry into the High Tech
Pedro Gil | Oscar Afonso | Paulo Brito

56: Finanças Públicas Portuguesas Sustentáveis no Estado Novo (1933-1974)?
Ricardo Ferraz

57: What Determines Firm-level Export Capacity? Evidence from Portuguese firms
Ana Gouveia | Ana Luisa Correia

58: The effect of developing countries' competition on regional labour markets in Portugal
Tiago Pereira

59: Fiscal Multipliers in the 21st century
Pedro Brinca | Hans Holter | Per Krusell | Laurence Malafry

60: Reallocation of Resources between Tradable and Non-Tradable Sectors in Portugal: Developing a new Identification Strategy for the Tradable Sector
Ana Fontoura Gouveia | Filipa Canas

61: Is the ECB unconventional monetary policy effective?
Inês Pereira

62: The Determinants of TFP Growth in the Portuguese Manufacturing Sector
Daniel Gonçalves | Ana Martins

63: Practical contribution for the assessment and monitoring of product market competition in the Portuguese Economy – estimation of price cost margins
Luis Folque

64: The impact of structural reforms of the judicial system: a survey
Ana Gouveia | Silvia Santos | Corinna Herber

65: The short-term impact of structural reforms on productivity growth: beyond direct effects
Ana Gouveia | Silvia Santos | Inês Gonçalves

66: Assessing the Competitiveness of the Portuguese Footwear Sector
Fábio Batista | José Matos | Miguel Matos

67: The empirics of agglomeration economies: the link with productivity
Ana Gouveia | Silvia Santos | Marli Fernandes

68: Determinants of the Portuguese GDP stagnation during the 2001-2014 period: an empirical investigation
Carlos Figueira

69: Short-run effects of product markets' deregulation: a more productive, more efficient and more resilient economy?
Ana Gouveia | Silvia Santos | Gustavo Monteiro

70: Portugal: a Paradox in Productivity
Ricardo Pinheiro Alves

71: Infrastructure Investment, Labor Productivity, and International Competitiveness: The Case of Portugal
Alfredo Pereira | Rui Pereira

72: Boom, Slump, Sudden stops, Recovery, and Policy Options. Portugal and the Euro
Olivier Blanchard | Pedro Portugal

73: Case Study: DBRS Sovereign Rating of Portugal. Analysis of Rating Methodology and Rating Decisions
Annika Luisa Hofmann | Miguel Ferreira | João Lampreia

74: For Whom the Bell Tolls: Road Safety Effects of Tolls on Uncongested SCUT Highways in Portugal
Alfredo Pereira | Rui Pereira | João Pereira dos Santos

75: Is All Infrastructure Investment Created Equal? The Case of Portugal
Alfredo Pereira | Rui Pereira

76: Why Virtuous Supply-Side Effects and Irrelevant Keynesian Effects are not Foregone Conclusions: What we Learn from an Industry-Level Analysis of Infrastructure Investments in Portugal
Alfredo Pereira | Rui Pereira

77: The Role of Gravity Models in Estimating the Economic Impact of Brexit
Graham Gudgin | Ken Coutts | Neil Gibson | Jordan Buchanan

78: Infrastructure Investment in Portugal and the Traded/Non-Traded Industry Mix
Alfredo Pereira | Rui Pereira

79: Goods and Factor Market Integration: A Quantitative Assessment of the EU Enlargement
Lorenzo Caliendo | Fernando Parro | Luca David Opromolla | Alessandro Sforza

80: Understanding productivity dynamics:a task taxonomy approach
Tiago Fonseca | Francisco Lima | Sonia C. Pereira

81: On the Effects of Infrastructure Investments on Industrial CO2 Emissions in Portugal
Alfredo Pereira | Rui Pereira

82: Assessing Competition With the Panzar-Rosse Model: An empirical analysis of European Union banking industry
Suzana Cristina Silva Andrade

83: Health Care Investments and Economic Performance in Portugal: An Industry Level Analysis
Alfredo Pereira | Rui Pereira | Pedro G. Rodrigues

84: Is deregulation of product and labour markets promoting employment and productivity? A difference-in-differences approach
Hugo Correia | Ana Fontoura Gouveia

85: Foreign acquisition and internal organization
Paulo Bastos | Natália P. Monteiro | Odd Rune Straume

86: Learning, Prices, and Firm Dynamics
Paulo Bastos | Daniel A. Dias | Olga A. Timoshenko

87: The Diffusion of Knowledge via Managers' Mobility
Giordano Mion | Luca David Opromolla | Alessandro Sforza

88: Empresas Zombie em Portugal - Os sectores não transacionáveis da Construção e dos Serviços
Gabriel Osório de Barros | Filipe Bento Caires | Dora Xarepe Pereira

89: Collective bargaining through the magnifying glass: A comparison between the Netherlands and Portugal
Alexander Hijzen | Pedro Martins | Jante Parlevliet

90: A Lower VAT Rate on Electricity in Portugal: Towards a Cleaner Environment, Better Economic Performance, and Less Inequality
Alfredo Pereira | Rui Manuel Pereira

91: Who Seeks Re-Election: Local Fiscal Restraints and Political Selection
Susana Peralta | João Pereira dos Santos

92: Assessing the Competitiveness of the Metalworking Sector
João Marinho | Pedro Carvalho

93: The efficiency of Portuguese Technology Transfer Offices and the importance of university characteristics
Aurora Teixeira | André Monteiro

94: Persistence in innovation and innovative behavior in unstable environments
Joana Costa | Anabela Botelho | Aurora Teixeira

95: The effect of entrepreneurial origin on firms' performance - The case of Portuguese academic spinoffs
Natália Barbosa | Ana Paula Faria

96: Absorptive Capacity and Firms' Generation of Innovation - Revisiting Zahra and George's Model
Dina Pereira | João Leitão

97: Innovations in digital government as business facilitators: implications for Portugal
João Martins | Linda Veiga

98: Innovation and the economic downturn: Insights from Portuguese firms
Hugo Pinto | Tiago Santos Pereira | Elvira Uyarra

99: European Funds and Firm Dynamics: Estimating Spillovers from Increased Access
João Pereira dos Santos | José Tavares

100: Corporate Leverage and Investment in Portugal
Ana Martins | José Henrique Gonçalves | João Mário Ferreira Duque

101: The effects of official and unofficial information on tax compliance
Filomena Garcia | Luca David Opromolla | Andrea Vezzulli | Rafael Marques

102: Competition effect on innovation and productivity - The Portuguese case
Anabela Santos | Michele Cincera | Paulo Neto | Maria Manuel Serrano

103: Measuring the Welfare of Intermediation in Vertical Markets
Javier D. Donna | Pedro Pereira | Tiago Pires | Andre Trindade

104: Of course Collusion Should be Prosecuted. But Maybe... Or (The case for international antitrust agreements)
Filomena Garcia | Jose Manuel Paz y Minõ | Gustavo Torrens

105: Product market competition and gender discrimination
Dudley Cooke | Ana P. Fernandes | Priscila Ferreira

106: Integration of Small Technology-Based Firms in Aeronautics
Anabela Reis | Joana Mendonça | Ligia Urbina

107: The Effects of Highway Tolls on Private Business Activity – Results from a Natural Experiment
João Pereira dos Santos | David B. Audretsch | Dirk Dohse

108: Competition and Firm Productivity: Evidence from Portugal
Pedro Carvalho

109: Do Exchange Traded Funds (ETFs) Outperform the Market? Evidence from the Portuguese Stock Index
Carlos Manuel Pinheiro | Hugo Hilário Varela

110: Assessing the Competitiveness of the Portuguese Chemical Sector
Ana Rita Marques | Cátia Silva

111: A General Equilibrium Theory of Occupational Choice under Optimistic Beliefs about Entrepreneurial Ability
Michele Dell'Era | Luca David Opromolla | Luis Santos-Pinto

112: O Mercado Segurador em Portugal: O Papel dos Gestores na Constituição de Provisões
Soraia de Sousa Bornett | Carlos Manuel Pinheiro

113: Exploring the implications of di erent loan-to-value macroprudential policy designs
Rita Basto | Sandra Gomes | Diana Lima

114: The Determinants of TFP Growth in the Portuguese Service Sector
Ana Martins | Tiago Domingues | Catarina Branco

115: Agglomeration and Industry Spillover Effects in the Aftermath of a Credit Shock
José Jorge | Joana Rocha

116: Entrepreneurial Human Capital and Firm Dynamics
Francisco Queiró

117: Global Value Chains and Vertical Specialization: The case of Portuguese Textiles and Shoes exports
Tiago Domingues

118: Firm heterogeneity and exports in Portugal: Identifying export potential
Frederico Oliveira Torres

119: Vantagens Comparativas Reveladas e suas determinantes: Uma Aplicação à Economia Portuguesa
Guida Nogueira | António Portugal Duarte

120: A Look at the main channels of Potential Impact of Brexit on the Portuguese Economy
Guida Nogueira | Paulo Inácio

121: How internationalization and competitiveness contribute to get public support to innovation? The Portuguese case
Anabela Santos, Michele Cincera, Paulo Neto | Maria Manuel Serrano

122: Grande Guerra e Guerra Colonial: Quanto Custaram aos Cofres Portugueses?
Ricardo Ferraz

123: Financing a Renewable Energy Feed-in Tariff with a Tax on Carbon Dioxide Emissions: A Dynamic Multi-Sector General Equilibrium Analysis for Portugal
Rui M. Pereira | Alfredo M. Pereira

124: Brown Sugar, how come you taste so good? The impact of a soda tax on prices and consumption
Judite Gonçalves | João Pereira dos Santos

125: ARFIMA Reference Forecasts for Worldwide CO2 Emissions and the National Dimension of the Policy Efforts to Meet IPCC Targets
José Beirute | Alfredo M. Pereira

126: Reference Forecasts for CO2 Emissions from Fossil-Fuel Combustion and Cement Production in Portugal
José M. Belbute | Alfredo M. Pereira

127: Regulated Early Closures of Coal-Fired Power Plants and Tougher Energy Taxation on Electricity Production: Synergy or Rivalry?
Alfredo Marvão Pereira | Rui Manuel Pereira

128: Picking Our Environmental Battles: Removal of Harmful Subsidies or Carbon Taxation?
Alfredo Marvão Pereira | Rui Marvão Pereira

129: Financing Future Feed-in Tariffs from Currently Installed RES-E Generating Capacity
Alfredo Marvão Pereira | Rui Marvão Pereira

130: Foreign Direct Investment, Income Inequality and Poverty in Portugal, 1973-2014: What does cointegration analysis tell us?
Aurora Teixeira | Ana Sofia Loureiro

131: On the Spillover Effects of CO2 Taxation on the Emissions of other Air Pollutants
Alfredo Marvão Pereira | Rui Marvão Pereira

132: On the Macroeconomic and Distributional Effects of the Regulated Closure of Coal-Operated Power Plants
Alfredo Marvão Pereira | Rui Manuel Pereira

133: The China Shock and Employment in Portuguese Firms
Lee Branstetter | Brian Kovak | Jacqueline Mauro | Ana Venâncio

134: Energy Taxation Reform with an Environmental Focus
Alfredo Marvão Pereira | Rui Manuel Pereira

135: ARFIMA Reference Forecasts for Worldwide CO2 Emissions and the Need for Large and Frontloaded Decarbonization Policies
José M. Belbute | Alfredo M. Pereira

136: Exporter Firms Behaviour, Evidence From Portuguese Firms Using Microdata
Luís Pedro Manso Machado

137: Collateral Value and Entrepreneurship: Evidence from a Property Tax Reform
Miguel Ferreira | João Pereira dos Santos | Ana Venâncio

138: The Financial Channels of Labor Rigidities: Evidence from Portugal
Edoardo M. Acabbi | Ettore Panetti | Alessandro Sforza

139: Can a small leak sink a great ship? A comprehensive analysis of the Portuguese household savings
Tiago Domingues | Margarida Castro Rego

140: Corporate taxes and high-quality entrepreneurship: evidence from a tax reform
Ana Venâncio | Victor Barros | Clara Raposo

141: Built Like a House of Cards? - Corporate Indebtedness and Productivity Growth in the Portuguese Construction Sector1
José Santos | Nuno Tavares | Gabriel Osório de Barros

142: Effectiveness of Simplex: The Case of Portuguese Social Security
António Alberto Nifrário de Pinho Tavares

143: Digital innovation in higher education: A questionnaire to Portuguese universities and polytechnic institutes
Paulo Nuno Vicente |Margarida Lucas | Vânia Carlos

144: Portugal in the Global Innovation Index: A panel data analysis
Marcelo P. Duarte | Fernando M. P. O. Carvalho

145: Intangible investments and productivity performance
Michele Cincera | Julie Delanote | Pierre Mohnen | Anabela Santos | Christoph Weiss

146: Digitalization in Two-sided Platform Competition
Filomena Garcia | Muxin Li

147: Collusion between two-sided platforms
Joana Pinho | Yassine Lefouili

148: Da confluência entre Big Data e Direito da Concorrência: As concentrações digitais - O caso Facebook/WhatsApp
Ana Rodrigues Bidarra

149: The Determinants of Total Factor Productivity in the Portuguese Quaternary Sector
Paulo Matos | Pedro Neves

150: Os modelos Input-Output, a estrutura setorial das economias e o impacto da crise da COVID 19
Pedro N. Ramos | João Ferreira | Luís Cruz | Eduardo Barata

151: Public Expenditure and private firm performance: using religious denominations for causal inference
Henrique Alpalhão | Marta Lopes | João Santos| José Tavares

152: Employee Training and Firm Performance: Quasi-experimental evidence from the European Social Fund
Pedro S. Martins

153: Dream Jobs
Luca David Opromolla | Giordano Mion | Gianmarco I.P. Ottaviano

154: Minimum wage and financially distressed firms: another one bites the dust
F. Alexandre | P. Bação | J. Cerejeira | H. Costa | M. Portela

155: Do short-term rentals increase housing prices? Quasi-experimental evidence from Lisbon
Susana Peralta | João Pereira dos Santos | Duarte Gonçalves

156: Economic and social policies under EMU
Ricardo Pinheiro Alves

157: International Sourcing in Portuguese Companies - Evidence from Portuguese Micro Data
Ana Martins | Guida Nogueira | Eva Pereira

158: The Impact of R&D tax incentives in Portugal
Rita Bessone Basto | Ana Martins | Guida Nogueira

159: The Determinants of Competitiveness of the Portuguese Defense Industry
Roxanne Merenda

160: How is the Minimum Wage Shaping the Wage Distribution: Bite, Spillovers, and Wage Inequality
Carlos Oliveira

161: Macroeconomy Impacts of the Covid-19 Pandemic in Some European Union Countries: a Counterfactual Analysis
António Portugal Duarte | Fátima Sol Murta

162: Digital adoption and productivity: understanding micro drivers of the aggregate effect
Natália Barbosa | Ana Paula Faria

163: Job Creation and Destruction in the Digital Age: What about Portugal?
Anabela M. Santos | Javier Barbero Jimenez | Simone Salotti | Andrea Conte

164: Is digital government facilitating entrepreneurship? A comparative statics analysis.
Joana Costa | Luís Carvalho

165: Automation trends in Portugal: implications in productivity and employment
Marta Candeias | Nuno Boavida | António Brandão Moniz

166: Digital Technologies for Urban Greening Public Policies
Maria José Sousa

167: The impact of a rise in transportation costs on firm performance and behaviour
Catarina Branco | Dirk C. Dohse | João Pereira dos Santos | José Tavares

168: Outward FDI, restructuring, performance upgrading and resilience: Firm-level evidence from Portugal
Natália Barbosa

169: Firm adaptation in COVID-19 times: The case of Portuguese exporting firms
João Capella-Ramos | Romina Guri

170: Supporting small firms through recessions and recoveries
Diana Bonfim | Cláudia Custódio | Clara Raposo

171: The Credit Channel of Public Procurement
Ricardo Duque Gabriel

172: Autonomia Estratégica Aberta na União Europeia: desafios e oportunidades na era da tecnologia digital
Gabriel Osório de Barros | Catarina Castanheira Nunes

173: R&D subsidies and Portuguese firms' performance: A longitudinal firm-level study
Inês Ferraz Teixeira | Aurora A.C. Teixeira | Luís Delfim Santos

174: Does scientific research output matter for Portugal's economic growth?
Tânia Pinto | Aurora A.C. Teixeira

175: Science and productivity in European firms: How do regional innovation modes matter?
Natália Barbosa | Ana Paula Faria

176: Employment versus Efficiency: Which Firms Should R&D Tax Credits Target?
Anna Bernard | Rahim Lila | Joana Silva

177: Forging AI Pathways: Portugal's Journey within the EU Digital Landscape
Gabriel Osório de Barros

178: Revisitar as Empresas Zombie em Portugal (2008-2021)
Ricardo Pinheiro Alves | Nuno Tavares | Gabriel Osório de Barros

179: A dependência da União Europeia no lítio e nas baterias de ião-de-lítio: análise à luz da autonomia estratégica
Beatriz Raichande

180: Artificial Intelligence in Agriculture: Revolutionizing Methods and Practices in Portugal
Maria José Sousa

181: EU-funded investment in Artificial Intelligence and regional specialization
Anabela Marques Santos | Francesco Molica | Carlos Torrecilla Salinas

182: Distance to Export: A Machine Learning Approach with Portuguese Firms
Paulo Barbosa | João Cortes | João Amador

REPÚBLICA PORTUGUESA
ECONOMIA

Gabinete de Estratégia e Estudos