IMT School for Advanced Studies, Lucca
Lucca, Italy

Policy evaluation and machine learning in international economics

PhD Program in Systems Science

Track in Economics, Networks and Business Analytics

XXXV Cycle

By

Francesca Micocci

2024

The dissertation of Francesca Micocci is approved.

PhD Program Coordinator: Ennio Bilancini, IMT School for Advanced Studies Lucca

Advisor: Prof. Armando Rungi, IMT School for Advanced Studies Lucca

The dissertation of Francesca Micocci has been reviewed by:

¡Reviewer 1¿, ¡Reviewer 1 institution¿

¡Reviewer 2¿, ¡Reviewer 2 institution¿

IMT School for Advanced Studies Lucca
2024

Contents

List of Figures

List of Tables

Acknowledgements

Vita

September 17, 1994	Genzano di Roma (Rome), Italy
2016	Bachelor degree in Economics Final mark: 109/110 Università degli Studi di Trento (TN), Italy
2018	Msc in Economics Final mark: 110/110 cum laude Università di Pisa and Scuola Superiore degli Studi Sant'Anna di Pisa (PI), Italy

Publications

Publications

1. F. Micocci, A. Rungi (2023) “Predicting Exporters with Machine Learning.” *World Trade Review* 22.5: 584-607

Working papers

1. L. Fontagné, F. Micocci, A. Rungi (2024) “The heterogeneous impact of the EU-Canada agreement with causal machine learning” *preprint* available at *arXiv:2407.07652*
2. G. Cerulli, F. Micocci, A. Rungi (2024) “A dose-response function for learning-by-exporting.”

Others

1. F. Micocci, A. Rungi (2023) ”On the utility of predicting the next exporters with machine learning.” *VoxEu column - frontiers of Economic Research, international Trade*. August 2023
2. M. Ghodsi, F. Micocci (2024) ”The impact of foreign direct investment on innovation in the EU.” *Wiiw Monthly Report* No. 6, June 2024
3. C. Castelli, R. Davis, F. Flòrez-Mendoza, M. Ghodsi, F. Micocci (2024) ”Mapping innovation in climate mitigation technologies across Europe: a regional perspective.” *Wiiw Monthly Report* No. 6, June 2024

Presentations

1. F. Micocci, "Predicting exporters with machine learning," at Sardinia Empirical Trade Conference (SETC) held by the Forum for Research in Empirical International Trade (FREIT) in *University of Cagliari*, Cagliari, Italy, 2023.
2. F. Micocci, "Predicting exporters with machine learning," at EEA-ESEM 2022 in *Bocconi University*, Milan, Italy, 2022.
3. F. Micocci "The heterogeneous effects of CETA on French export", at the Research Group on the analysis of economic policies in *CNR*, Rome, Italy, 2022
4. F. Micocci, "Predicting exporters with machine learning," at 14th FIW-Research Conference 'International Economics' in *WU Vienna*, Wien, Austria, 2022.
5. F. Micocci, "Predicting exporters with machine learning," at International Trade and Interdependence in global production organised by the Italian Trade Study Group (ITSG) in *University of Florence*, Florence, Italy, 2022.
6. F. Micocci, "Predicting exporters with machine learning," at ETSG 2021 22nd Annual Conference in *University of Ghent*, Ghent, Belgium, 2021

Abstract

This thesis explores innovative empirical models in international economics, leveraging machine learning techniques and a dose-response method to address issues of multidimensionality, heterogeneity, and nonlinearity, while exploiting detailed firm- and product-level microdata.

Firstly, we investigate the capacity of machine learning techniques to forecast the firm's exporting status. Analyzing comprehensive financial accounts and firm-and industry-specific data from French manufacturing firms (2010-2018), we demonstrate that machine-learning methodologies can accurately forecast a firm's exporting status with up to 90% accuracy. Unlike traditional econometrics, our method handles multidimensional data and exploits it to model non-linear relationships among endogenous predictors, thus proving a valuable tool for targeted trade promotion programs.

Next, we assess the heterogeneous impacts of the EU-Canada Comprehensive Economic and Trade Agreement (CETA) on French trade using a causal machine learning approach. Employing a non-parametric matrix completion algorithm rooted in potential outcome models, we predict multidimensional counterfactuals at the firm, product, and destination levels, capturing complex interactions without assuming functional forms. Using predicted potential outcomes allows us to uncover significant heterogeneity in the trade agreement's effects, which conventional average effects models might overlook. Furthermore, our methodology is suitable to evaluate spillover effects. Within our framework, these manifest as classical Vinerian diversion effects, wherein trade to Canada

partially substitutes for trade outside Canada, especially for products with a higher elasticity of substitution.

Lastly, we examine the learning-by-exporting phenomenon by isolating the effect of export intensity on firm productivity from the endogenous selection into exporting status. Using a dose-response model that treats export intensity as a continuous treatment affecting firm productivity, we move beyond traditional binary treatment models to provide insights into how this relationship evolves across the full spectrum of export intensity values. Our findings indicate that productivity gains from exporting are non-linear, with firms needing to achieve a 60% export intensity threshold to fully capitalize on knowledge spillovers and effectively compete in international markets.

Overall, this research expands the frontier of empirical research in international economics, revealing insights into the complex dynamics of trade through innovative methodologies.

Introduction

Trade is inherently complex, with various stakeholders responding differently to technological changes, consumer preferences, regulations, and geopolitical shifts. This multi-dimensionality and heterogeneity complicate the identification of causal relationships and the disentanglement of the impact of shocks from other contextual dynamics. The advent of detailed firm and product microdata has revolutionized international economics by offering granular insights into trade patterns and firm behavior previously unattainable with traditional aggregate data. However, this disaggregated data introduces new challenges, complicating empirical analyses and necessitating advanced empirical methods. While invaluable, indeed, granular data exacerbate estimation complexity by revealing intricate relationships that, if not properly addressed, could lead to biased results.

For instance, a firm's decision to enter the export market results from a complex interplay of firm-specific characteristics, industry conditions, and external influences. Larger firms are often better able to manage the costs associated with entering foreign markets, while innovative firms have a greater capacity to create products that appeal to global consumers. High-tech industries typically produce goods with international appeal and encourage export engagement, whereas firms in competitive sectors seek new markets to diversify domestic dependencies. Each factor holds significance on average, yet their interactions vary across specific industries and competitive environments.

Once involved in international trade, firms face decisions on export volume and target destinations. Factors such as demand strength, growth potential, and technological advantages incentivize higher export intensities, while market saturation, competitors, and consumer preferences influence profitable export destination choices. Additionally, tariffs, non-tariff barriers, and trade agreements significantly impact the costs and benefits of entering foreign markets. The interactions among these factors affecting export intensity and target destinations exhibit significant nonlinearity, meaning that changes in one variable can yield disproportionate and sometimes unpredictable effects on others. For example, a tariff reduction might boost exports of advanced products, but this could be mitigated by market saturation or competition. Similarly, increased export intensity from strong demand growth might be offset by higher production costs or logistical challenges.

Understanding these complex dynamics is essential for formulating effective trade policy. However, this necessitates sophisticated empirical models that can accommodate the nonlinear and multifaceted nature of international trade dynamics.

In this context, machine-learning algorithms offer robust tools for identifying patterns and extracting meaningful insights from noisy data (Athey & Imbens, 2019). These techniques handle heterogeneity and nonlinear relationships (Athey, 2018; Mullainathan & Spiess, 2017), offering the potential to uncover hidden structures and manage the multidimensionality of firm and product attributes. Moreover, recent developments in causal machine learning have provided helpful identification strategies in program evaluation, accounting for heterogeneous causal effects and endogeneity (Athey & Imbens, 2017; Chernozhukov et al., 2018; Wager & Athey, 2018).

Machine learning applications in economic research have grown substantially over the last decade. For example, Deryugina et al. (2019) used Cox-Lasso machine learning to estimate the causal impact of air pollution on medical costs in the context of predicted versus observed counterfactuals. Handel and Kolstad (2017) explored the heterogeneity of treatment effects on health behaviors induced by access to wearable technologies using a recursive partitioning model developed by Athey and

Imbens (2016). The strengths of machine learning in dimensionality reduction and data extraction have also enhanced classification tasks. Notable examples include terrorism risk assessment (Limodio, 2022), textual analysis of political speeches (Gentzkow et al., 2019), measuring CEO performance (Bandiera et al., 2020), understanding health behavior (Chandra et al., 2024), and economic specialization (Bartelme et al., 2024). Moreover, machine learning methods have proven instrumental in forecasting, identifying financial and banking crises (Alessi & Detken, 2018; Bluwstein et al., 2023; Joy et al., 2017) and using new data sources for prediction, including scanner data for demand forecasting (Bajari et al., 2015) and human resources data for employee performance (Chalfin et al., 2016). Policy targeting has also consistently improved through data-driven decision-making models such as personalized pricing strategies (Dubé & Misra, 2023) and credit institutions' lending decisions (Dobbie et al., 2021).

Despite these advancements, machine learning applications in international economics remain limited, likely because of the complexity and specificity of the trade data. However, some promising examples have recently emerged. Breinlich et al. (2022) and Kim and Steinbach (2023), for example, apply machine learning techniques (lasso and several extensions) to identify PTA provisions that are most important for increasing trade flows. Gordeev and Steinbach (2024) employ machine learning to identify the most critical determinants of the countries' inclusion in PTA provisions, such as competition for export markets, geographic proximity, and governance quality. Focusing on forecasting, Jaax et al. (2024) developed a model to nowcast aggregate services imports and exports using monthly services trade data. Similarly, Gnecco et al. (2023) used a matrix completion algorithm to predict the revealed comparative advantages (RCAs) of countries in different product categories.

This thesis aims to contribute to this emerging literature by providing three case studies that apply novel methodologies to address distinct challenges in international trade studies arising from multi-dimensionality, nonlinearity and heterogeneity.

In Chapter 1[1], we employ machine learning methods to predict a firm's extensive margin of trade while trying to identify the drivers of export potential. Motivated by a substantial body of literature linking firm heterogeneity with their trading status (Bernard & Jensen, 1999; Bernard et al., 2012; Hottman et al., 2016; Lin, 2015; Melitz, 2003; Melitz & Ottaviano, 2008; Melitz & Redding, 2014), we argue that exporters' financial profiles differ significantly from those of non-exporters because of the unique cost structures required to sustain export fixed costs and navigate foreign market regulations and consumer preferences (Aw et al., 2023). Therefore, we commence by assembling an expansive and inclusive array of economic and financial predictors capturing diverse firm and industry attributes. Leveraging advanced machine learning techniques, we achieve robust prediction accuracy, surpassing 90% in discerning between firms engaged in export activities and those that are not.

The inherent endogeneity of our predictors, posing challenges in conventional econometric frameworks, enhances the explanatory power of our models by offering insights into the degree to which a firm mirrors a successful exporter. Notably, among the algorithms tested, tree-based models demonstrated the highest precision, emphasizing the intricate interplay of non-linear interactions among firm characteristics. Defining a successful exporter reveals to be a challenging task, particularly considering the variable relevance of firm attributes contingent upon the dynamics of respective industries and geographical contexts.

The final outcome of our predictive exercise manifests in an export score ranging from zero to 100, serving as a metric akin to credit scoring (Altman, 1968; Altman et al., 2000; Merton, 1974), thereby offering insights into a firm's internationalization strategies and creditworthiness. Following rigorous validation against diverse definitions of exporters and various training methodologies, we perform a detailed examination of the predictive power of individual predictors, highlighting how they offer valuable information on trade potential at different levels of aggregation.

[1]This chapter is based on the paper: F. Micocci, A.Rungi "Predicting Exporters with Machine Learning." World Trade Review 22.5 (2023): 584-607.

Chapter 2[2] introduces a novel causal machine learning approach to estimate the impact of a free trade agreement (FTA). Trade agreements wield significant influence over international trade patterns, yet their impact estimation is complicated by self-selection and heterogeneity. Firms that choose to export under a trade agreement may differ systematically from those that do not, leading to biased estimates. Additionally, products included in trade agreement provisions may already have larger markets for trading partners before the treaty.

Conventional empirical methods such as difference-in-differences, regression discontinuity designs, and structural models have traditionally addressed these challenges (Baier & Bergstrand, 2007; Head & Mayer, 2014), but frequently yield unstable and fragile estimates (Baier et al., 2019). Our methodological innovation proposes a causal machine-learning approach to investigate the impact of the EU–Canada Comprehensive Economic and Trade Agreement (CETA) on French trade, using monthly customs data on the universe of French exports. Specifically, we adapt a matrix completion algorithm tailored for causal panel data (Athey, Bayati, et al., 2021) and grounded in potential outcome models, to estimate causal effects after predicting unobserved counterfactual outcomes. Notably, using non-parametric methods allows us to predict potential outcomes effectively amidst non-linearities without imposing stringent assumptions on functional forms or the data-generating process.

By treating French customs data as an observed outcomes matrix partitioned between treated and untreated observations, both pre- and post-CETA, we strategically exclude entries corresponding to treated units post-treatment. Leveraging information encapsulated within remaining observed entries, we derive counterfactual predictions in the absence of treatment, thereby obtaining estimates of multidimensional treatment effects across products, firms, and export destinations.

Our analysis at the product level reveals a positive impact of CETA on French exports, evidenced by an average increase of 1.28% in prod-

[2]This chapter is based on L. Fontagn'e, F. Micocci, A. Rungi "The heterogeneous impact of the EU-Canada agreement with causal machine learning", available at *arXiv:2407.07652*

uct flows to Canada following the agreement's implementation. Concurrently, approximately 13.1% of newly introduced French products accessed the Canadian market for the first time, with 11.9% exiting due to the new trade provisions.

At the firm level, multiproduct firms exhibited increased exports of their already most exported products to Canada after the CETA. This result aligns with the theoretical framework proposed by Mayer et al. (2021) and Eckel and Neary (2010), which suggests that multiproduct exporters reallocate their product mix in response to demand shocks in the export markets.

Crucially, our matrix-completion methodology unveils heterogeneous treatment effects associated with trade agreements, enabling complex evaluations across entire distributions of estimated treatment effects. In our analysis, we observed both positive and negative impacts. Moreover, we identify positive associations between treatment effects on individual products and a metric of revealed comparative advantage for French exporters relative to global peers, while product churn outcomes correlate positively with elasticity of substitution. Notably, such heterogeneity would be masked in more traditional estimations, such as DID, which are frequently employed by international offices to evaluate the impact of FTAs.

Furthermore, our approach detects the classical Vinerian diversion effect (Viner, 1950) whereby intra-PTA trade partially substitutes for trade with non-PTA members, underscoring the policy spillovers inherent in trade agreements.

In Chapter 3, we investigate the impact of export intensity on firm performance by building on the learning-by-exporting (LBE) hypothesis, positing that firms enhance their operational efficiency through exposure to international markets. The theoretical foundation of the LBE includes knowledge spillovers (Eaton & Kortum, 2002; Grossman & Helpman, 1991), competitive pressure (Atkeson & Burstein, 2010; Clerides et al., 1998), and resource reallocation exploiting economies of scale (Helpman et al., 2004). However, empirical evidence on the LBE effect is mixed, with some studies finding significant productivity gains from exporting (De Loecker, 2007), and others reporting minimal or even negative effects (Bernard & Jensen, 1999; Greenaway & Kneller, 2008; Wagner, 2007).

Our study offers a distinct perspective, aiming to disentangle the impact of export intensity on firm performance from the confounding effects of self-selection biases inherent in exporting behaviors driven by firm heterogeneity. Adopting a potential outcome framework, we estimate a dose-response function for permanent exporters, considering export intensity as a continuous treatment that impacts firm productivity. This approach enables us to go beyond single average effect estimation by mapping the effect as a function across varying levels of treatment intensity, thereby revealing the underlying pattern of the causal relationship across the entire spectrum of export intensity.

Our findings substantiate the hypothesis that firms accrue productivity benefits from exporting only upon attaining critical mass in export volumes. At lower export intensities, firms necessitate to develop absorptive capacities and logistical efficiencies to derive productivity gains from foreign markets. As export intensity escalates, firms streamline production processes to sustain competitiveness, ultimately improving production processes. Consequently, substantial productivity gains associated with exporting, fueled by LBE mechanisms, manifest only beyond a minimum threshold of export intensity.

Empirical analysis of French firm-level data spanning 2010-2018 corroborates this hypothesis. Following Cerulli (2015), we estimate a dose-response function that maps export intensity to a firm's productivity, finding a nonlinear relationship. Exporting firms do not immediately experience benefits from increased export intensity, but significant rewards are observed when the export-sales ratio surpasses 60%. Beyond this threshold, productivity notably escalates as exporting becomes a primary revenue driver. Conversely, for values below 5%, exporting exhibits minimal impact on production processes, likely reflecting passive exporting behaviors.

Additionally, we identify a "low-productivity trap" within the range of 5-35% export intensity. In this interval, exports negatively affect productivity, as firms allocate resources towards exporting infrastructure without seeing corresponding returns. We further show that the 35% threshold distinguishes groups of exporters who then maintain similar levels of foreign activities in subsequent years. Firms below this threshold struggle to exceed it, maintaining moderate export intensity levels over time, while those surpassing the threshold consistently sustain heightened export intensities in subsequent periods.

Furthermore, sector-specific and technological trajectory analyses following Pavitt's Taxonomy, reveal that the impact of export intensity on firm performance varies significantly. This highlights the importance of considering industry-specific factors when assessing the benefits of export intensity.

In conclusion, our studies demonstrate the potential of machine learning techniques and dose-response methods in addressing the complexities of international trade analysis. By leveraging comprehensive firm-level and product-level data, we uncover intricate interactions and causal relationships that traditional methods might overlook. Our findings contribute to the literature by offering novel insights into predicting export potential, estimating the effects of trade agreements, and understanding the non-linear impact of export intensity on firm performance.

Chapter 1

Predicting Exporters with Machine Learning

This chapter is based on the paper: Francesca Micocci, Armando Rungi "Predicting Exporters with Machine Learning." World Trade Review 22.5 (2023): 584-607 available at https://doi.org/10.1017/S1474745623000265.

1.1 Introduction

Building trade capacity is a purpose of many international and national agencies. The World Trade Organization provides special support programs for developing countries to better integrate into the multilateral trading system. On the other hand, many developing and developed economies prefer to establish their facilitative agencies to provide firms with information, technical advice, marketing services, and policy advocacy about access to foreign markets.

The general idea is that there are opportunities for gains from trade, yet not all firms have the same ability to sell their goods and services abroad. Exporting activity entails beach-head costs when handling different regulatory environments, meeting different consumer tastes, and establishing marketing and logistics channels. Only some more productive firms may be able to self-select into exporting status. In contrast, other companies may not have the necessary skills or resources to propose in foreign markets[1]. Hence, the necessity to resort to trade promotion programs to fill the gap and help firms build trade capacity to take advantage of open markets. Eventually, openness to trade is a determinant of economic growth insofar as it allows exploiting differential comparative advantages and economies of scale. Companies can benefit while tapping into foreign technology and raising aggregate productivity in the home countries [2].

Against the previous background, our simple intuition is to adopt machine learning techniques to evaluate how far a company is from reaching an export status based on the assumption that firms' accounts convey non-trivial information on firm-level trade capacity. In other words, we propose to train an algorithm on in-sample financial statements to predict out-of-sample firms' ability to start exporting. Our intuition follows 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). Unlike credit risk literature, our problem is not to check if a company is proximate to bankruptcy. On the contrary, our challenge is to measure how far a company is from being healthy enough to start and propose on foreign markets.

We begin by training different machine learning techniques on a sample of 57,016 manufacturing firms in France, which may have exported or not in 2010-2018. Following statistical standards, we randomly partition the initial sample in an 80-20 proportion to separate it into a training and a testing set. Therefore, we train different models armed with a battery of 52 predictors that we believe may contain non-trivial information on exporting abilities. Then we use the trained models to obtain distributions of out-of-sample predictions that can be useful to assess a company's distance from exporting capability. In simple terms, the exporting score

[1] For a review of the arguments according to which only the most efficient firms can self-select into an export status and the consequences on the sources of gains from trade, see among others Bernard and Jensen, 1999; Bernard et al., 2012; Hottman et al., 2016; Melitz and Redding, 2014

[2] Seminal works identify macroeconomic linkages between trade openness, technological progress, and economic growth. See Grossman and Helpman, 1990, Rivera-Batiz and Romer, 1991, Romer, 1994, Barro and Sala-i-Martin, 1997.

summarizes how much a non-exporter looks like an exporter.

Crucially, we find that our procedure correctly separates exporters from non-exporters with an accuracy of up to 90%. The latter is a figure we obtain from a horse race among different algorithms. We find that a Bayesian Additive Regression Tree with Missingness not at Random (BART-MIA) (Kapelner & Bleich, 2015) is the procedure that provides the most robust predictions. The BART-MIA is a regression tree with a Bayesian component for regularization through a prior specification that allows flexibility in fitting various regression models while avoiding strong parametric assumptions (Hill et al., 2020). What makes BART-MIA especially useful for our case is the possibility of exploiting additional predictive power from non-random missing values on predictors. The latter is a feature that is especially useful in catching business dynamics when coverage of financial accounts is likely to be correlated with other dimensions, e.g., firms' size or productivity, which, in turn, can correlate with firms' export status. In our case, we assess that considering non-random missing values helps us increase prediction accuracy by about 14.4%. Eventually, we ensure that prediction accuracies are robust to different definitions of exporters and to the presence of discontinuous exporting activity (Békés & Muraközy, 2012; Geishecker et al., 2019). The last check is especially relevant in the case of smaller exporters, or when exporters specialize in manufacturing capital goods, whose relationships with customers entail several breaks in the time series.

Our framework is also robust to different cross-validation strategies since we obtain similar performance by randomly picking training and testing subsets in different ways, albeit from a unique sample. Finally, we test that reducing the set of predictors brings lower levels of accuracy after we perform a Least Absolute Shrinkage and Selection Operator (LASSO) for dimensionality reduction (Ahrens et al., 2020; Belloni, Chernozhukov, et al., 2013; Belloni et al., 2014, 2016).

After assessing which tool is better at predicting exporters, we delve into the prediction power of single predictors, i.e., how much they contribute to getting good predictions. The practical utility of this exercise is to show that there may be, indeed, some dimensions of the firms' economic activity that correlate relatively more with their trade potential. Thus, following Chipman et al. (2010), we implement a procedure to derive *Variable Inclusion Proportions* (VIPs), which can be interpreted as posterior probabilities (Bleich et al., 2014). Crucially, we discuss how VIPs have a relevant internal validity since they catch predictive power within the given testing vs training sets. Yet, we may not attribute them any external validity because predictors can change their power in different contexts. Indeed, we discuss how such changes in different contexts and sub-populations could actually be informative of the changing resilience of firms and from where it comes. For example, in the French case we study, the difference we observe in the model's selection of influential predictors between Île-de-France and the rest of France suggests there are geographic-specific firms' dynamics. The same predictors may or may not play a major role in the probability of exporting, depending on the specific technological characteristics of the production environment.

The final sections discuss how we see exporting scores applied in practice. We suggest looking at baseline predictions to derive a probabilistic exporting score to a firm, i.e., a score summarising how similar a non-exporter is to benchmark exporters on a scale from 0 to 1. We argue that exporting scores could be helpful for trade promotion or trade finance programs. After aggregation, we show how they can represent an additional tool to describe the trade competitiveness of regions or industries.

Finally, to briefly illustrate the practical utility of exporting scores, we classify firms into risk categories and provide simple back-of-the-envelope estimates of how much cash resources and capital expenses they would need to reach export status. We find that increasing cash and capital is required to reduce the distance from export status. For example, in the case of medium-risk firms, i.e., firms that have just below 50% probability of exporting, we show a need for up to 44% more cash resources and up to 246% more capital expenses to reach full export status.

The remainder of the paper is organized as follows. We relate to previous literature in Section 1.2. We introduce data and sample coverage in Section 1.3, whereas Section 1.4 discusses the empirical strategy. Results are commented on in Section 1.5, while robustness checks are discussed in Section 1.6. A specific Section 1.7 tests for the sensitivity of predictions to the phenomenon of temporary trade, while a practical use of exporting scores is presented in Section 1.10. Section 1.11 concludes.

1.2 Related literature

Most countries worldwide implement trade promotion programs that envisage the expenditure of substantial amounts of public funds. Thus, it is hardly surprising that there have been concerns about the efficacy and effectiveness of those support programs. Interestingly, Volpe Martincus and Carballo, 2008 show how export promotion actions are usefully associated with increased exports by already trading firms and traded products, i.e., the intensive margin. In terms of extensive margins, i.e., the increase of firms and products crossing national borders, Volpe Martincus et al., 2010 show that an influential role is often played by the establishment of diplomatic representations, especially in the case of producers of homogeneous goods. In general, activating new trading relationships may require various services bundled into more complex export promotion programs (Volpe Martincus & Carballo, 2010a). Eventually, a majority of studies investigate how effective a policy is on the *ex-post* companies' exporting performances while controlling for cherry-picking Volpe Martincus and Carballo (2010b). In general, Van Biesebroeck et al., 2016 demonstrate how trade promotion programs have been a vital tool to overcome economic crises, such as recovery after the global recession in 2008-2009.

In this context, our contribution focuses explicitly on the possibility of increasing the trade extensive margin proposing a measure of the ability of non-exporters to start exporting. From this perspective, what we propose is a pure prediction exercise based on the intuition that exporters are statistically different from non-exporters. In this sense, we rely on a two-decades-long strand of research that has established a connection between firms' heterogeneity and trading status (Bernard & Jensen, 1999; Bernard et al., 2012; Hottman et al., 2016; Lin, 2015; Melitz, 2003; Melitz & Ottaviano, 2008; Melitz & Redding, 2014). Our intuition is that a prediction of export status is possible only because we know that exporters have different cost structures than non-exporters. After all, they have to sustain the fixed costs to gain access to foreign markets, where regulations and consumer tastes can differ much from home (Aw et al., 2023), and where shipping is costly. Thus, we demonstrate that starting from a comprehensive battery of economic and financial predictors allows indeed separating exporters from non-exporters with a relatively high prediction accuracy, up to 90%.

Please note that ours is not a classic policy evaluation exercise nor a structural model to understand the determinants of export status. We do not want to assess whether any specific policy design works to support would-be exporters. Ours is a simple scoring exercise in the fashion of what one can find in previous literature about credit scoring. There is a long tradition to try and spot firms in financial distress based on the disclosure of financial accounts. See seminal attempts with Z-scores by Altman, 1968; Altman et al., 2000, and Distance-to-Default by Merton, 1974, where some specific threshold is set as a rule of thumb to say whether a firm is financially sound and worthy of credit. Nowadays, most financial institutions adopt predictive models to evaluate credit risk, including machine learning (Uddin, 2021). A statistical learning exercise to spot financially distressed firms, i.e., so-called zombie firms, is reported in Bargagli-Stoffi et al., 2020. See also the exercises on firm-level correlations to spot investment-to-cash-flow sensitivities and assess time-varying financial constraints (Almeida et al., 2004; H. (Chen & Chen, 2012; Fazzari et al., 1988).

The additional difficulty in our exercise is that we want to score success, i.e., the ability of a firm to outreach across national borders. In contrast, credit risk analyses take as reference previous firms' failures, i.e., their distance-to-default. Yet, we argue, the intuition is the same: to get as benchmark firms that realized an outcome, in our case, an export status, and thus measure how far we are from that outcome. Eventually, we can also relate to literature on trade finance. We know very well that routine access to trade credit is needed to outlive foreign markets, and well-functioning financial markets are crucial to export performance (Lin, 2017; Manova, 2012). Eventually, external finance helps firms gain and keep access to foreign markets despite the high beach-head costs, especially for smaller producers who have a reduced ability to provide collateral to financial institutions (Chor & Manova, 2012). In this context, we believe exporting scores are potentially valuable to better target financial institutions' credit policies in a familiar way, e.g., by considering credit risk classes. To better grasp our previous intuitions, we propose a simple back-of-the-envelope exercise that estimates, *ceteris-paribus*, how much cash resources and capital expenses firms need to switch across low, medium and high-risk classes.

Moreover, from a macroeconomic viewpoint, one can use firms' scoring as yet another indicator of the competitiveness of an economy (or lack thereof). Inspired by so-called growth diagnostics, international and national statistics offices have developed frameworks for assessing the potential of countries, regions, and industries to compete in international markets. See, for example, works on measuring trade competitiveness (Gaulier et al., 2013; Reis et al., 2010). In the case of French manufacturing, we show how potential exporters are unevenly distributed across industries and regions. We believe there is no reason why an indicator like ours about the potential of extensive margins should not find room in a standard trade diagnostic kit.

Finally, we want to remark on how ours is one of the first attempts to exploit statistical learning techniques in international economics. As far as we know, only a few notable efforts are in progress (see M. Gopinath et al., 2020 and Breinlich et al., 2021). Yet, we believe that statistical learning exercises have great potential and should find their way in a field like international economics, where one often needs to extract valuable information from big and complex datasets, which can be dealt with by a combination of both predictive tasks and standard causal inference exercises (Athey, 2018; Mullainathan & Spiess, 2017).

1.3 Data

We source firm-level information from ORBIS[3] compiled by the Bureau Van Dijk. Notably, France is a much-explored case study of firm-level trade data, allowing us to confront previous literature. See among others Crozet et al., 2012 and Fontagné et al. (2018). Our main outcome of interest is the export status of a firm that we derive from information on export revenues [4]. *Prima facie*, we will consider a firm as an exporter if it reports positive export revenues. Then, in Sections 1.6 and 1.7, we will challenge our baseline definition to comply with the phenomenon of temporary trade (Békés & Muraközy, 2012) when it is optimal for firms to export every once in a while. As for firm-level predictors of exporting status, we employ a battery of 52 indicators elaborated on original financial accounts that we use to train our models. Further details on our choice are discussed in Section 1.4.2, while we include the list of predictors with a complete description in the Data Appendix.

To grasp the coverage of our sample, we draw Figure A2.1 and Table A2.1, reported in the Appendix. Figure A2.1 shows how relevant exporters are in every NUTS-2 region in France, as from our sample. Table A2.1 compares sample industry coverage with the one provided by Eurostat census in 2018. We do find that we have fair coverage by 2-digit industries since the correlation by industry shares is about 0.90. Yet, according to Eurostat business demographics, our sample covers 32.6% of firms' population which represents about 75% of total operating revenues in France. As largely expected, we cannot retrieve the financial accounts of smaller firms because they are not required to comply with accounting regulations in the same way as medium and larger ones. See also a comparison by class categories with Eurostat in Appendix Table

[3]The ORBIS database has become a standard source for global firm-level financial accounts. For a previous usage of this database, among others, see G. Gopinath et al., 2017, Cravino and Levchenko, 2016, Del Prete and Rungi, 2017, and Del Prete and Rungi, 2018. It complements financial accounts with other information from different sources on ownership, corporate governance, and intellectual property rights, which we also use for predictions in the following analyses.

[4]Interestingly enough, French firms must report revenues from exports separately, as from the subsequently amended *Règlement n. 99-03 du Comité de la réglementation comptable.*

A2.2. In the following paragraphs, we will show how our baseline analysis can handle non-random missing values in financial information.

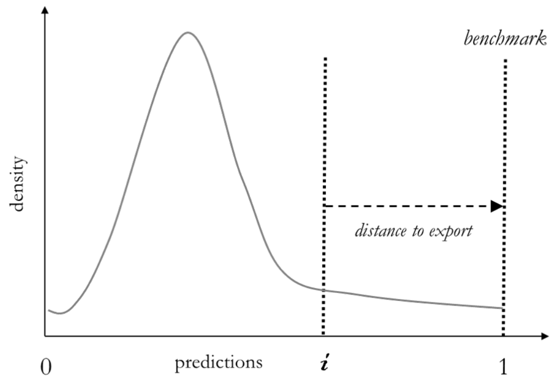
1.4 The empirical strategy

Our main intuition is that we can predict out-of-sample exporting capability based on the in-sample experience of both exporters and non-exporters. The first step is to find the best algorithm that is able to separate exporters and non-exporters after conditioning on financial information. Our prior is that exporters and non-exporters are statistically different, as acknowledged by previous literature reported in Section 1.2. Thus, once we assess the method that assures the best predictive accuracy with the minimum numbers of false positives and false negatives (see Section 1.5.1), we can test out-of-sample and use the distribution of predictions to assign each firm an exporting score that is bounded, by construction, in an interval from 0 to 1. The higher the score, the better the chances a firm is able to make it on foreign markets.

In Figure 1, we report a visual fictional representation of our intuition. Assuming that we did a good job in training and that prediction accuracy is acceptable, we can reasonably test on new firms and locate actual exporters at the end of the right tail of the distribution of exporting predictions. Thus, any ith non-exporting firm located on the left of predicted exporters will come with a positive distance, which will convey non-trivial information on how viable that firm is to start exporting. In other words, we take as a reference point the export status at 1 and, thus, we check how far a company is from that reference point.

Eventually, in Section 1.8 we provide a framework for the interpretability of predictors by catching the influence of each of them in getting the exporting scores. That is, we are able to sum up how important one predictor is with respect to the entire set in any out-of-sample exercise we may run. Obviously, given the predictive nature of our analyses, we won't be able to attach any causal interpretation to our exercise. For our purpose, we will make use of *Variable Inclusion Proportions*, i.e., the proportion of times a predictor is selected as a splitting rule for the construction of the random trees. The construction and interpretation of VIP are discussed in section 1.8. Notably, selected predictors are contingent on the trained sample, i.e., their role won't have any external validity. Yet, we argue identifying the drivers of the model performance helps further comment on the nature of exporting scores.

Figure 1: Visual intuition of an exporting score.



Note: We represent a fictional distribution of predictions of exporting status by definition bounded in an interval $[0, 1]$. Along the distribution, we could spot an i-th non-exporting firm. We reasonably assume that actual exporters locate at the end of the right tail. By definition, non-exporters are less likely to start exporting at an increasing distance from predicted exporters.

1.4.1 Methods

We train and compare different statistical learning techniques to get our best predictions. Thus, we make use of the generic predictive model for firms' export status in the form:

$$f(\mathbf{X}_i) = Pr(Y_i = 1 \mid \mathbf{X}_i = x) \tag{1.1}$$

where Y_i is the binary outcome that assumes value 1 if the ith firm is exporting and zero otherwise. $\mathbf{X}_i$ is a matrix that includes a full battery of firm-level predictors, which we discuss in detail in the following Section 1.4.2. Please note that, at this stage, we do not consider the time dimension, i.e., we train the predictive model considering the export status of a firm in relation to present predictors. In this baseline model, it is entirely possible that a firm is considered an exporter in one year and a non-exporter in another year. See Section 1.7, where we consider heterogeneous exporting patterns.

The functional form that links predictors to outcomes is *ex-ante* unknown and looked for by the generic supervised machine learning technique. We provide an overview of our different methods in Section 1.4.1. The advantage of any of them is to extract information from many predictors while catching non-linearities that may be present in the association with export status. Briefly, the generic predictive model has to pick the best in-sample loss-minimizing function in the form:

$$\arg\min \sum_{i=1}^{N} L(f(x_i), y_i) \quad over \quad f(\cdot) \in F \qquad s.\,t. \qquad R\big(f(\cdot)\big) \leq c \quad (1.2)$$

where F is a function class from where to pick the specific function $f(\cdot)$. Importantly, $R\big(f(\cdot)\big)$ is the generic regularizer that summarizes the complexity of $f(\cdot)$. The latter is a tool that allows us to solve the common trade-off between an as high as possible in-sample fit and an as high as possible flexibility of the prediction model, able to take on board new out-of-sample information. It is the solution to the so-called bias-variance trade-off. The set of regularizers, R's, will change following the standards proposed by each method that we compare in the following paragraphs. Eventually, any method shall minimize the constrained loss function represented in eq. 1.2, while searching for the function that can be better used to process new out-of-sample information.

As a common strategy across our different models, we will pick at random 80% of our French firms to be considered as in-sample information. We will then use it to train the generic statistical learning algorithm. We will keep the remaining 20% as out-of-sample information to predict export status. Hence, we will be able to assess the accuracy of our predictions within the limit of our data sources. As it is standard in similar exercises, we perform a cross-validation check described in Section 1.6, to verify that a specific segment of the sample does not affect prediction accuracy.

In the following paragraphs, we show how a specific variant of the Bayesian Additive Regression Tree (BART) performs better than others because it is able to consider the presence of non-random missing values as further predictors for the outcome. The variant we use is the BART with Missingness In Attributes (BART-MIA). For more details, see also Kapelner and Bleich, 2015. For a previous application to firms' dynamics, see Bargagli-Stoffi et al., 2020.

In general, any classification tree $\mathcal{T}$ is built on *if-then* statements that split the training data according to the observed values of predictors, allowing for non-linear relationships between the predictors and the outcomes. Thus, the generic algorithm for the construction of a classification tree, $\mathcal{T}$, is based on a top-down approach that recursively splits the main sample into non-overlapping sub-samples (i.e. the nodes and the leaves). Therefore, the tree is pruned iteratively with the generic regularizer R to improve its predictive ability while avoiding overfitting, in case trees develop along too many layers [5].

As in the baseline version (Chipman et al., 2010), BART-MIA is a sum-of-trees ensemble with an estimation approach relying on a fully Bayesian probability model. The algorithm elaborates the ensemble by imposing a set of Bayesian priors that regularize the fit by keeping the individual trees' effects small in an adaptive way. The result is a sum of trees, each of which explains a small and different portion of the predic-

[5] It is beyond the scope of this paper to get into further details of single techniques. We refer to Hastie et al., 2017 for a deeper introduction to statistical learning.

tive function. The BART-MIA variant we adopt can be expressed as:

$$\mathbb{P}(Y = 1|\mathbf{X}) = \Phi\left(\mathcal{T}_1^{\mathcal{M}}(\mathbf{X}) + ... + \mathcal{T}_q^{\mathcal{M}}(\mathbf{X})\right), \tag{1.3}$$

where Φ denotes the cumulative density function of the standard normal distribution and the q distinct binary trees are denoted by $\mathcal{T}$, each being a single tree coming with an entire structure made of nodes and leaves. The sum-of-trees model serves as an estimate of the conditional probit at $\mathbf{X}$, which can be easily transformed into a conditional probability estimate of $Y = 1$.[6]. The Bayesian component of the BART includes three priors that have demonstrated to use the data at disposal efficiently:

1. the prior on the probability that a node will split at depth k is $\beta(1+k)^{-\eta}$, where $\beta \in (0,1), \eta \in [0,\infty)$, and the hyper-parameters are chosen to be $\eta = 2$ and $\beta = 0.95$;

2. the prior on the probability distribution in the leaves is a normal distribution with zero mean: $\mathcal{N}(0, \sigma_q^2)$, where $\sigma_q = 3/d\sqrt{q}$ and $d = 2$;

3. the prior on the error variance is $\sigma^2 = 1$.

Thus, the regularization parameter $R(\cdot)$ in the general formulation of ML algorithm 1.2 corresponds to the priors themselves. Finally, the BART-MIA algorithm employs a Metropolis-within-Gibbs sampler (Geman & Geman, 1984; Hastings, 1970) to generate draws from the posterior distribution of $\mathbb{P}(\mathcal{T}_1^{\mathcal{M}}, ..., \mathcal{T}_m^{\mathcal{M}}, 1|\Phi(Y))$. [7] Let us denote with K the size of the sample of the draws $\{p_1^*, \ldots, p_K^*\}$ from the posterior distribution.

[6]Note that each classification probability $P(Y = 1|\mathbf{X})$ is obtained as a function of a sum of regression trees. At the same time, standard classifier approaches use a majority or an average vote based on an ensemble of classification trees. See, for example, Breiman (2001).

[7]This passage involves introducing small perturbations to the tree structure: growing a terminal node by adding two child nodes, pruning two child nodes (rendering their parent node terminal), or changing a split rule.

Then, the prediction $p(x) = P(Y = 1|\mathbf{X})$ at a particular x, is

$$p^*(x) = \sum_{k=1}^{K} p_k^*(x)$$

In addition to the Bayesian component, the BART-MIA variant augments the original algorithm by exploiting information on missing values and splitting on *missingness* features that are used as additional predictors in each binary-tree component.

Eventually, the BART-MIA is chosen in the following paragraphs as the baseline method after a comparison with four other alternatives. At first, we compare with a simple logistic regression (LOGIT). The latter is a classical econometric technique for binary outcomes with a specific *ex-ante* assumption on the functional form linking predictors with the outcome. Then, we perform three other methods based on regression trees, namely a Classification and Regression Tree (CART) (Breiman et al., 1984), a Random Forest (RF) (Breiman, 2001), and the original unaugmented BART. CART is the most basic regression tree, while RF is an ensemble method that aggregates different regression trees to get a stronger predictive power, as the BART does, but without a Bayesian framework. Finally, we compare previous regression trees' models with the Least Absolute Shrinkage and Selection Operator (LASSO) in the form:

$$\underset{\beta \in \mathbb{R}^p}{\arg\min} \ \ \frac{1}{2N} \sum_{i=1}^{N} \left(y_i(x_i^T \beta) - log(1 + e^{(x_i^T \beta)}) \right)^2 \quad \text{subject to } \|\beta\|_1 \leq k. \quad (1.4)$$

where y_i is a binary variable equal to one if a firm i is an exporter and zero otherwise. Any x_i is a predictor chosen in $\mathbb{R}^p$, whereas $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$ and $k > 0$. The constraint $\|\beta\|_1 \leq k$ limits the complexity of the model to avoid overfitting, and k is chosen, following Ahrens et al. (2020), as the value that maximises the Extended Bayesian Information Criteria (J. Chen & Chen, 2008). To account for the potential presence of heteroskedastic, non-Gaussian and cluster-dependent errors, we adopt the rigorous penalization introduced by Belloni et al., 2016.

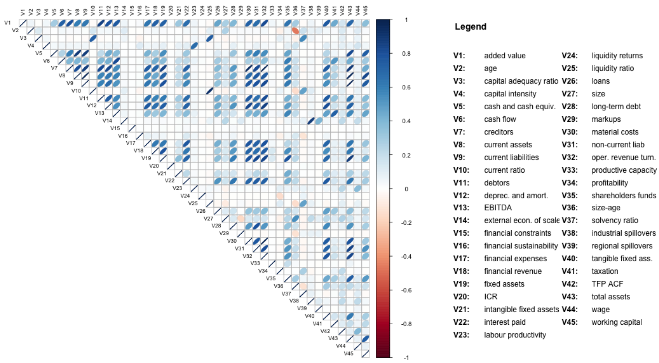
1.4.2 Predictors

To increase models' predictability, we include a full battery of 52 predictors that we derive from firms' balance sheets and profit and loss accounts. A detailed description is reported in the Data Appendix. Broadly speaking, we choose to include:

1. original financial accounts without any elaboration;
2. financial ratios and other proxy indicators (e.g., productivity, economies of scale, spillovers) that we expect to be correlated with exporting activity;
3. firms' locations, ownership status, and industry affiliations, which can help in spotting categories of firms at a competitive advantage or disadvantage.

Usefully, in Figure 2, we show a correlation matrix including all numeric predictors. Please note how some of them are indeed much cross-correlated with values well above 0.6. Yet, high correlations are not that relevant to our case since, in a context of pure prediction like ours, we do not (want to) estimate coefficients. At this stage, we also do not need a prior on which financial information conveys the highest predictive power. Hence, we choose not to discriminate among predictors *ex ante*, although we do have information provided by previous literature that some variables more than others are associated with exporting activity (productivity, firm size, financial constraints, etc.). See also a specific robustness check in Section 1.6, where we show what happens when we reduce our set of predictors. In another words, we are well aware that our long list of predictors entails a great deal of endogeneity among variables that are otherwise studied in different structural relationships. As we are not interested in obtaining estimates for determinants of trade, such endogeneity is not relevant for our purpose. What we need to do is to minimize the prediction errors given albeit marginally useful observable information. In Section 1.9, we further discuss the limits and benefits of a pure predictive exercise when it comes to the interpretability of predictors.

Figure 2: Correlation matrix of predictors



Note: We report a correlation matrix of the predictors we use. Non-numeric predictors are excluded here but included in the following analyses: NUTS-2 locations, NACE Rev.2 industries, a categorical variable for consolidated accounts, patents' dummy, inward FDI, outward FDI, and corporate control. Positive correlations are reported as upward-sloping ellipses, while negative correlations are reported as downward-sloping ellipses. The color intensity and the ellipse width indicate the strength of the correlation.

1.5 Results

1.5.1 Models' horse race

In Table 1, we compare measures of standard prediction accuracy across the methods we test. For details on how these metrics are constructed, please see Appendix C. Briefly, what we can see is that Sensitivity focuses on the ability to predict exporters, i.e., the amount of *true positives*, while Specificity focuses on the ability to predict non-exporters, i.e., the amount of *true negatives*. Balanced Accuracy is an arithmetic mean between Sensitivity and Specificity. Importantly, the ROC curve (receiver operating characteristic curve) evaluates the predictive performance at different classification thresholds, as reported in Figure A2.2, and it is our baseline measure of performance across different models. Finally, Precision-Recall is of help to us in assessing the trade-off between returning accurate results (high precision) vis á vis returning a majority of positive results (high recall).

Table 1: Prediction accuracies

	Specificity	Sensitivity	Balanced Accuracy	ROC	PR	N. obs.
LOGIT	0.6642	0.7776	0.7210	0.7940	0.8053	86,754
LOGIT-LASSO	0.6606	0.7722	0.7164	0.7847	0.7891	86,754
CART	0.5700	0.7896	0.6796	-	-	86,754
Random Forest	0.6078	0.8276	0.7178	0.7947	0.8010	86,754
BART	0.6272	0.8048	0.7158	0.7911	0.7998	86,754
BART-MIA	0.9064	0.6496	0.7782	0.9054	0.7375	382,606

Note: We report standard measures of prediction accuracies (by column) for different methods we train (by row). For details on how prediction accuracies are constructed, see Appendix C. Any observation is a firm-year present in the sample. All methods but BART-MIA do not train or test on observations when at least one predictor is missing. Hence, a larger number of observations in testing BART-MIA.

From Table 1, we immediately notice that BART-MIA outperforms other methods with an ROC equal to 0.9054, a value that is considerably higher than in the case of other methods. In fact, BART-MIA is in general more able than others to predict both exporters and non-exporters, with a Balanced Accuracy of 0.77.

Yet, when we look at Specificity *vis á vis* Sensitivity values, we realize it predicts relatively better non-exporters rather than exporters. The reason is that the boost in overall prediction accuracy by BART-MIA is largely due to an efficient use of the non-random missing values on smaller firms reporting incomplete financial accounts. See also the specific robustness checks performed in 1.6. As largely expected, smaller firms with partial information are also the ones that are more likely to be classified as non-exporters, because: i) larger size is more likely associated with an export status, and ii) smaller firms do not have to report financial information as complete as it is required to bigger companies.

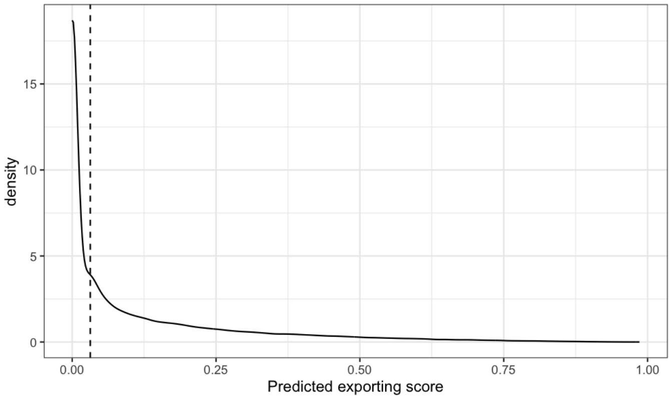
Since BART-MIA is able to include the *missingness* of any single feature as an additional predictor (i.e., as yet another *branch* of the regression tree), we understand why it outperforms other methods, which instead simply drop from computation companies that have any missing values in predictors.

Finally, a simple comparison between the accuracy of BART and the one of BART-MIA allows us to quantify what is the gain in considering the predictive power of missing values. Overall, we observe a 14.4% increase in ROC, which we take as our baseline measure of prediction accuracy. We will further discuss the trade-off between Specificity and Sensitivity once we challenge our results in Section 1.7. Suffice it to say here that, in general, predicting true exporters is made difficult by the presence of temporary trade, i.e., when firms export in some years and not in others, thus breaking the time series.

1.5.2 Predictions

In Figure 3, we report the entire distribution of predicted scores for non-exporters that we obtain from our baseline BART-MIA. Without any selection threshold, these are the values that one could consider for evaluating how far a company is from export status. What is relevant to observe here is that the distribution is much skewed, hence the majority of non-exporters in France is located on a thick left tail, thus far from being able to propose on foreign markets. Briefly, the distribution of scores that we obtain here is consistent with the idea of firm heterogeneity that we take from trade literature, as introduced in Section 1.2. In other words, only a relatively small number of non-exporters is proximate to reaching the right tail's goal. The observation that firms are heterogeneous also in exporting scores is relevant for taking informed policy decisions that we discuss in Section 1.10.

Figure 3: Distributions of exporting scores of non-exporters after BART-MIA



Note: We report the distribution of the score after implementing BART-MIA on the entire sample and selecting all non-exporting firms. The vertical line identifies the median non-exporting firm.

1.6 Robustness checks

So far, we adopted a relatively standard $80-20$ random partition of the firms in the sample at our disposal when training our model (Athey, Imbens, et al., 2021). Therefore, our first concern here is to cross-validate our choice by repeating the prediction exercise other four times with a similar random partition. We want to check that our high prediction accuracy is not due to a fortunate selection of the training-and-testing partition. Any time, we train on a random 80% of the dataset that we consider as in-sample information, then we test the accuracy of our predictions on the rest 20%, which we take as out-of-sample information. We show in Table A2.3 how we obtain similar performance scores across all exercises, and we pick BART-MIA once again as the most predictive algorithm. We conclude that previous results had not been driven by a specific selection of training *vis á vis* testing data.

Our second concern is that prediction accuracies are robust to different definitions of exporters. So far, we defined an exporter as any firm with positive exporting revenues. Here, we will define an exporter as a firm whose export share over total revenues is higher than a specific minimum threshold, to make our results robust to the presence of so-called *passive exporters* (Geishecker et al., 2019), i.e., domestic firms that engage in one-off exporting events.

Appendix Table A2.8 shows prediction accuracies after we run simulations by excluding from the category of exporters those firms that report export shares lower than the first, second, and fifth percentile, respectively. Prediction accuracies are similar in magnitude to those of our benchmark definition. Latter evidence suggests that baseline predictions are not affected by the presence of a few less proactive firms.

A third concern we have is to verify the robustness to changes in predictors. Our problem here is whether we could obtain similar prediction accuracy with a minor effort, once neglecting variables that contribute with a relatively little predictive power. For this purpose, we perform a Logit-LASSO exercise before running again the models described in 1.4.1. As in standard applications (Belloni et al., 2017), the Logit-LASSO selects a subset of best predictors (in our case, 23 out of 52) to contribute relatively more to predict export status. Once again, BART-MIA outperforms other statistical learning techniques. However, when we perform BART-MIA including only such a subset of predictors, we obtain lower accuracy than baseline results, as reported in Appendix Table A2.5. Yet, we gather there is no reason to exclude available predictors despite the high cross-correlations we observed in Figure 2.

A fourth concern we have is to check whether the time of training and testing matters for predictions. So far, we considered firms and their export status throughout the entire period at our disposal, between 2010 and 2018. In Appendix Table A2.6, we train and test our predictive model separating each year. It is evident how predictions do not change dramatically over the timeline.

A fifth concern is that performance measures are robust to different probability thresholds for predicting the exporting status. In baseline analyses, we adopt a quite standard cut-off value set at 0.5 to separate exporters and non-exporters in prediction. Yet, we know that exporting is a relatively rarer event than non-exporting, and our prediction accuracies can suffer from a bias. The choice of the threshold is, indeed, crucial for the computation of most prediction accuracies because the values in Table 1 are threshold-specific. For a similar case in trade literature, see Baier et al., 2014. Here we want to check that a different threshold does not alter the ranking of methodologies obtained by comparing prediction accuracies in Table 1. Therefore, in appendix Table A2.4, we show how performance measures vary when we choose, for each model, the optimal cut-off value obtained following Liu, 2012, who aims at maximizing the product of sensitivity and specificity. When an optimal threshold is set, the evidence of BART-MIA superiority is even more striking as it outperforms the others by all measures of prediction accuracy except for PR. We will discuss in section 1.7 how the latter is negatively affected by the presence of discontinuous exporters. Note, however, that both PR and ROC are not affected by the change in cut-off values because they are independent of thresholds by construction. The latter is also the reason why we consider them as baseline measures of performance.

A final concern is that baseline predictions improve mechanically only because the sample size is bigger in BART-MIA than in other exercises. In fact. we want to investigate whether improvements actually come from missing values. For our purpose, we perform two different exercises: i) we add *ex ante* a predictor to our original set that catches the relative *missingness* of financial information at the firm-level; ii) we impute missing values on single predictors based on median values available as from other companies' financial accounts. From a combined reading of both exercises, we better understand the role of *missingness*.

Results for the latter exercise are reported in Appendix Table A2.9.

Interestingly, prediction accuracies do increase overall for all methods after predictors' imputation, although classification trees (BART[8], Random Forest), perform relatively better along the different segments of the distribution (ROCs are 0.907 and 0.905, respectively). Eventually, when we check for the relative importance of a predictor on *missingness*, we find that it is always selected as the best predictor no matter what procedure we choose. We conclude that missing values do have a prediction power, yet our baseline BART-MIA better catches their role without introducing unnecessary data manipulation.

Eventually, we consider useful also reporting Spearman's rank correlations in Table 2, to test whether rankings in predictions are sensitive to the choice of predictive models in Table 1. Please note how, by construction, the Spearman's rank correlations can be performed only on the subset of the data where every technique obtains predictions.

As a matter of fact, we get relatively high rank-correlations across predictive models with a minimum of 0.87 and a maximum of 0.96. In general, models do not dramatically alter the relative positions of firms on the distribution of predictions. Interestingly, please note that rank-correlation between the simpler BART and the BART-MIA is about 0.92. Although the latter is just a variant of the first with *missingness* of values as an additional feature, the rankings in predictions are different. The latter is a significant result that allows us to further qualify the difference between the simpler BART and its variant. The bottom line is that information from firms with missing values in predictors allows BART-MIA to identify different thresholds on predictors' distributions, which in turn change the relative positions of firms on the distribution of predictions.

[8]At this stage, computing BART-MIA or BART is equivalent, since we filled in missing value with imputations. The BART-MIA won't find any missingness, and won't include missing values among predictors, thus reversing to a more traditional BART procedure

Table 2: Spearman's rank correlations of predicted probabilities from different models

	LOGIT	LOGIT-LASSO	Random Forest	BART	BART-MIA
LOGIT	1	0.9657	0.8773	0.8841	0.9012
LOGIT-LASSO		1	0.8925	0.9030	0.9118
Random Forest			1	0.9112	0.9167
BART				1	0.9179
BART-MIA					1

Note: We report a Spearman's rank correlation among out-of-sample predictions to show how rankings in export status are sensitive to changes in predictive models. All models, including BART-MIA, are thus trained and tested on the same observations.

1.7 Sensitivity to temporary trade

We investigate in this section the sensitivity of our results to the presence of discontinuous exporting activity, i.e., when firms engage in trade relationship that are temporary (Békés & Muraközy, 2012). Indeed, the biggest challenge we face when predicting exporters is that firms can export in some years and then lay idle for a while before re-proposing (or not) on foreign markets. This is especially true for smaller firms or for firms that are specialized in manufacturing capital goods. Thus, our prior is that discontinuity is not at random; it could be correlated with some firms' attributes, and our previous predictions could be therefore sensitive to the relevance of temporary trade within our sample.

For our purpose, we perform separate checks by classifying firms into five categories:

1. firms that always export, which we call *constant exporters*;
2. firms that never export, which we call *non-exporters*;
3. firms that start exporting at some period t and always export afterwards, which we call *switching exporters*;
4. firms that export in all periods until t and never export afterwards, which we call *switching non-exporters*[9];

[9] Please note how we may have had more switching non-exporters if we were able to

5. *discontinuous exporters*, which export with an irregular pattern with more than one gap along the timeline.

Prediction accuracies are eventually reported in Table 3, after testing out-of-sample our baseline BART-MIA algorithm. As expected, we observe that our predictive model performs quite well in separating constant exporters from non-exporters, since Sensitivity and Specificity are about 0.86 and 0.95, respectively[10]. On the other hand, predictions become relatively less accurate when we look at out-of-sample information on firms that show gaps along the timeline. In general, we still have acceptable accuracies as ROCs reach up to 0.86 and 0.81, respectively, in the case of *switching exporters* and *switching non exporters*. In line with our priors, the quality of predictions is proportional to the number of years that the firms actually exported. Predictions are more accurate when firms started (stopped) exporting sooner (later) in our data.

Finally, we focus on the category what we define discontinuous exporters, when firms have more than one break in the time series, entering and exiting the export status. In this case, at the bottom of Table 3, we find that prediction accuracy reached a relatively lower albeit acceptable threshold (ROC: 0.80). The accuracy is lower than the one obtained in predicting constant exporters and non-exporters. Interestingly, we do register that our procedure is less and less able to predict the export status in the case of firms that have less experience of foreign markets. This is however consistent with the idea that firms engaging in temporary trade may continue to do so systematically, hence their lower predictability on a year-by-year basis.

Eventually, a final sensitivity check to temporary trade is performed by introducing a more liberal definition of exporters proposed by Békés

zoom out on a longer timeline. We cannot exclude that firms that do not export in our sample did so in previous unobserved periods. The latter is an element of imperfection that we cannot expunge from our prediction exercise.

[10]Please note that we cannot estimate other measures of prediction accuracy when we focus exclusively on either positive or negative outcomes. See Appendix C for a definition of different measures of prediction accuracies.

and Muraközy (2012), according to whom only firms with at least four years of consecutive exporting can be actually considered as *permanent exporters* vis á vis *temporary exporters*. As largely expected, we find in Appendix Table A2.7 that prediction accuracies for *permanent exporters* are relatively higher (AUC: 0.849; PR: 0.934) than in the case of temporary exporters. In particular, the model fails at predicting the export status of temporary exporters, i.e., it reports a relatively lower true positives' rate, as shown by the low scores on sensitivity, PR and ROC.

From our viewpoint, it makes sense that exporters with irregular exporting patterns represent intermediate cases somewhere between firms that always export and firms that never export. Therefore, classification algorithms struggle to separate intermediate cases on a binary outcome. Based on financial accounts, such firms can be seen neither as fit for exporting as constant exporters, nor as unfit as non-exporters. Yet, it is more likely that such intermediate cases are of less interest in policy applications because trade promoters or financial institutions need instead to understand whether a firm that never exported at all needs some support or not.

1.8 Interpretability of predictors

In line with our empirical strategy, we focused so far on prediction accuracy while neglecting the role of single predictors. We discussed in Section 1.4 how our choice is driven by the necessity to maximize prediction accuracy; therefore we have been using an as complete as possible list of predictors, even though we are aware that we carried on with us a compound of endogenous variables that are highly cross-correlated, as commented after Figure 2.

What we want to do now is to show how predictors do have different influence on the outcome, and we can still discuss their influence on predictions without implicating any causality. On the contrary, the inter-

Table 3: Prediction accuracies and temporary trade

Firm category	Sensitivity	Specificity	Balanced Accuracy	ROC	PR	Num. Obs.
Constant Exporters	0.856	-	-	-	-	21,834
Non-exporters	-	0.951	-	-	-	158,625
Switching to export	0.629	0.849	0.739	0.864	0.764	15,084
Since t_0	*0.749*	*0.682*	*0.716*	*0.794*	*0.954*	*1,980*
Since t_1	*0.729*	*0.694*	*0.712*	*0.808*	*0.914*	*1,296*
Since t_2	*0.711*	*0.751*	*0.731*	*0.838*	*0.888*	*1,179*
Since t_3	*0.618*	*0.806*	*0.712*	*0.832*	*0.821*	*1,215*
Since t_4	*0.582*	*0.796*	*0.689*	*0.812*	*0.73*	*1,323*
Since t_5	*0.585*	*0.819*	*0.702*	*0.823*	*0.638*	*1,683*
Since t_6	*0.463*	*0.835*	*0.649*	*0.804*	*0.45*	*2,187*
Since t_7	*0.262*	*0.903*	*0.583*	*0.792*	*0.251*	*4,221*
Switching to non-export	0.599	0.802	0.7	0.819	0.786	27,891
Until t_0	*0.269*	*0.81*	*0.539*	*0.643*	*0.152*	*3,915*
Until t_1	*0.376*	*0.745*	*0.561*	*0.65*	*0.291*	*2,511*
Until t_2	*0.419*	*0.725*	*0.572*	*0.689*	*0.443*	*2,124*
Until t_3	*0.479*	*0.737*	*0.608*	*0.733*	*0.599*	*2,412*
Until t_4	*0.508*	*0.815*	*0.662*	*0.816*	*0.757*	*2,844*
Until t_5	*0.563*	*0.925*	*0.744*	*0.929*	*0.924*	*5,409*
Until t_6	*0.664*	*0.843*	*0.754*	*0.877*	*0.931*	*3,996*
Until t_7	*0.742*	*0.813*	*0.778*	*0.874*	*0.97*	*4,680*
Discontinuous exporters	0.547	0.807	0.677	0.796	0.686	85,023
export experience: 1 year	*0.216*	*0.873*	*0.544*	*0.686*	*0.171*	*19,152*
export experience: 2 years	*0.313*	*0.823*	*0.568*	*0.702*	*0.334*	*12,816*
export experience: 3 years	*0.387*	*0.796*	*0.592*	*0.718*	*0.483*	*10,962*
export experience: 4 years	*0.478*	*0.736*	*0.607*	*0.719*	*0.595*	*8,910*
export experience: 5 years	*0.519*	*0.74*	*0.63*	*0.753*	*0.72*	*9,297*
export experience: 6 years	*0.593*	*0.721*	*0.657*	*0.755*	*0.808*	*8,460*
export experience: 7 years	*0.662*	*0.7*	*0.681*	*0.774*	*0.886*	*7,758*
export experience: 8 years	*0.757*	*0.658*	*0.708*	*0.781*	*0.951*	*7,668*
All sample	0.6491	0.9080	0.7785	0.9048	0.7383	308,457

Note: We report prediction accuracies after BART-MIA for firms with different exporting patterns. For switching-exporters and switching-non-exporters we identify the year when they are observed changing status, i.e., the year when the firm passes from never exporting to always exporting, and vice versa. For discontinuous exporters we distinguish by number of exporting years over the sample timeline.

nal validity of our 'influential predictors' is to us more important than an external validity. They are relevant because we can interpret them in relationship with the specific prediction exercise we want to comment. If we consider a different sample, those 'influential predictors' will be almost certainly different.

Our baseline method for the interpretability of a BART-MIA exercise is called Variable Inclusion Proportions (VIP)[11]. The Variable Inclusion Proportion for any given predictor represents the proportion of times that variable is chosen as a splitting rule out of all splitting rules among the posterior draws of the sum-of-trees model (Kapelner & Bleich, 2013). It is computed as follows: (1) Across all q trees in the ensemble (1.3), we examine the set of predictor variables used for each splitting rule in each tree; (2) For each sum-of-tree model we compute the proportion of times that a split using x_p as a splitting variable appears among all splitting variables $\mathbf{X}$ in the model; (3) with K being the number of the sum-of-tree models f_k^*, drawn from the posterior distribution $\mathbb{P}(\mathcal{T}_1^{\mathcal{M}}, ..., \mathcal{T}_m^{\mathcal{M}}, 1|\Phi(Y))$, and z_{pk} being the proportion of all splitting rules that use the p^{th} component of $\mathbf{X}$ in model f_k^*, the Variable Inclusion Proportion is computed as

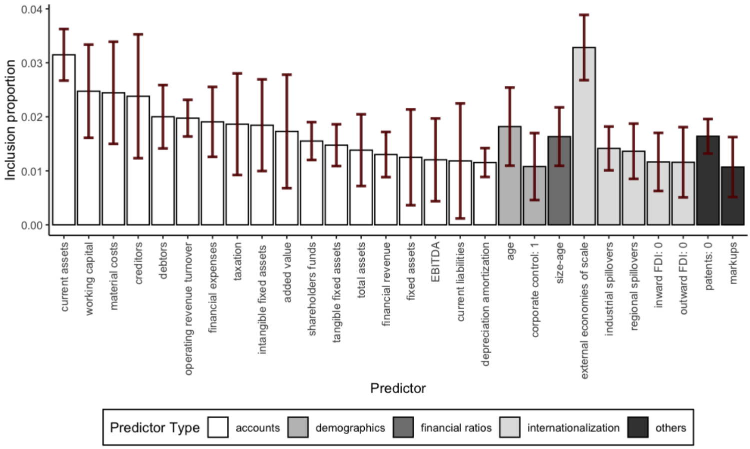
$$v_p = \frac{1}{K} \sum_{k=1}^{K} z_{pk} \tag{1.5}$$

Thus, we report in Figure 4 a visualization of the VIPs accompanied by a standard deviation that is computed after running five different random tests. Please note how averaging across multiple trials allows us to improve the stability of estimates, as suggested by Kapelner and Bleich, 2013. For the sake of visualization, we report in Figure 4 only those predictors that register a VIP equal or higher than 1%.

When we look at Figure 4, we document that the best predictor in our baseline exercise is the proxy we use for the existence of external economies of scale, which indicates the presence of other firms in the

[11]For a different choice of methods to catch the relative importance of predictors, see also Joseph (2020) and the case of neural networks.

Figure 4: Variable inclusion proportions after BART-MIA



Note: We report Variable Inclusion Proportions (VIPs), i.e., the proportion of times each predictor is chosen for a splitting rule in BART-MIA. Of all the predictors in baseline, we visualize only those with a VIP higher than 1%. The bars represent standard deviations of inclusion proportions obtained by replicating five different times the BART-MIA on the same random training set.

same industry and in the same region, as suggested by Bernard et al., 1995. Once again, we want to stress that since we are in a pure prediction framework, we cannot say whether external economies of scale, measured in this way, are an actual determinant of export status. We cannot exclude reversal causality. On the one hand, it is indeed possible that local spillovers help neighbouring firms to start exporting after, for example, sharing infrastructures or intangible knowledge about foreign markets: Dhyne et al. (2023) found such a dynamic using buyer-seller linkages in the Belgian production network. On the other hand, it is possible that firms in industries at a comparative advantage locate in geographical proximity before becoming exporters. In any case, it is beyond the scope of our analysis to unravel the endogeneity of this specific relationship or any other we know we have among predictors and the outcome. Suffice it to say that the industrial concentration of exporting firms in a region of France is a good albeit not unique predictor of export status for the representative firm located in that area.

Notably, we observe in Figure 4 how original accounts altogether provide an important contribution to predict export status. Yet, no single predictor contributes more than 4% in any of the tests we performed. Besides financial accounts, business demography has predictive power: firm age has an inclusion proportion higher than 2%. It also makes perfect sense that the activities of multinational enterprises play a role in export status. Being either a foreign subsidiary (inward FDI) or owning a subsidiary abroad (outward FDI) affects the probability of exporting. As expected, the ability to innovate and register patents is also related to the likelihood of becoming an exporter.

Eventually, we want to bring the attention on the absence of Total Factor Productivity (TFP) in Figure 4, which we however included following the methodology by Ackerberg et al., 2015. Although TFP is a much-studied determinant of export status, we do not find it to be among the most relevant predictors in a machine learning exercise. Our educated guess is that the role of TFP is already captured by the sample variation in raw financial accounts that are also needed to compute it as a residual from a firm-level production functions (turnover, costs of materials, employees, etc.).

1.9 Internal vs. external validity

In this Section, we discuss the reproducibility of our predictive exercise in different contexts, i.e., the external validity of our results.

A first concern we want to address is the possibility to replicate our study in the case of other countries, e.g. in the case of countries with different economic development. In this contribution, we investigated the case of France mainly because French firm-level data had been used extensively in related literature. Yet, we argue that our predictive setup can be applied to any country, regardless of its economic development, provided that financial accounts have predictive power on a firm's export status. We already discussed in Section 2 how we rely on extensive literature that supports the evidence that exporters are significantly different from non-exporters when we look at financial accounts (Bernard & Jensen, 1999; Bernard et al., 2012; Hottman et al., 2016; Lin, 2015; Melitz, 2003; Melitz & Ottaviano, 2008; Melitz & Redding, 2014). Therefore, in the case of developing countries, we do expect exporters and non-exporters to be at least as statistically different in financial accounts as in the case of a developed country. In the case of developing countries, we actually expect domestic allocative inefficiencies to be higher and exporters to be relatively larger and more productive than non-exporters than in developed ones, very concentrated at the top of the distribution (Alfaro et al., 2009; Tybout, 2000). In this case, we expect our algorithm, if anything, to perform at least as good in a developing country as in the case of France.

A second concern relates the external validity of our results on the prediction power of single financial accounts in Section 1.8. Can we assume that they will have a similar predictive power in other contexts? We argue they will not. VIPs constitute a posterior probability that the variable x_k has a (linear or nonlinear) association with the response variable (Bleich et al., 2014). Variables selected through VIPs would be almost certainly different if we considered different countries or regions. Yet, we argue that the relevance of VIPs resides in their internal validity, given the peculiarity of each predictive exercise. For example, one could compare across different countries or regions how the relative importance of predictors changes and use that information to take solid policy decisions. To make our point, we replicate our exercise after separating Île-de-France from the rest of the country. We show VIPs for both subsets in Appendix Figure A2.3.

We observe that not only the set of influential predictors differs, but also that the relative importance of predictors changes from one exercise to the other. This hints at the presence of locally different dynamics. For example, the predictor *(log of) number of employees* is selected in the sample excluding Île-de-France, but not in Île-de-France, where there is possibly more homogeneity in terms of firm size. In contrast, the predictor *patent* is influential in Île-de-France, but not elsewhere, possibly indicating that in the first there is a comparative advantage in more innovative activities that have the potential to reach foreign markets. *Prima facie*, the latter evidence is consistent with our prior knowledge about the landscape of the French economy.

A third concern we want to address is the validity of our methodology in presence of structural breaks or external shocks, e.g., in the case of policy changes. In this regard, please note that ours is a cross-sectional classification exercise: we use information on both exporters and non-exporters to understand how non-exporters are statistically different from exporters. We may pool data over longer periods to only increase the training set's size. However, it is unnecessary for our scope, and we include a few robustness checks in Section 1.7, when we change the pooling strategy. Eventually, in our case, the levels of prediction accuracy depend only on the ability of predictors to capture the statistical difference between exporters and non-exporters within the same period in different contexts. Structural breaks or policy shocks are of no concern to us as far as we do not use variation from the past to predict the future. Our only concern is that our list of predictors includes the different dimensions that can contribute to the gap between exporters and non-exporters in different policy environments. A discussion of the rationale for single predictors is included in Section 1.4.2.

1.10 How to use exporting scores

We provide now examples of possible applications of exporting scores as either indicators for trade credit or a tool for assessing the trade potential of regions and industries. Based on the prior knowledge that exporters and non-exporters are statistically different across financial attributes, we use in-sample information to predict out-of-sample capability to export. Thus, it is immediate to build a continuous indicator that provides an exporting score based on our baseline predictions to indicate the potential of companies to successfully propose on foreign markets, i.e., their distance from export status. We visualize our intuition in Figure 1.

Briefly, we can get a basic and simple export (probabilistic) score for any out-of-sample non-exporting ith firm in the form:

$$distance_i = 1 - Pr(Y_i = 1 \mid \mathbf{X}_i = x) \tag{1.6}$$

which is by definition bounded in a range $(0, 1)$, and made conditional on the set of predictors, $\mathbf{X}_i$, as from previous exercises.

To illustrate our idea of the relationship with creditability, we perform back-of-the-envelope estimates here to predict how much capital and cash resources may be needed by a company to become fit for export. We classify firms in different *risk categories*, i.e., categories based on a partition of the distribution of exporting scores obtained in Figure 3. For simplicity, let us consider all firms included in a decile of predictions as belonging to the same *risk category*. Obviously, the higher the distance from export status, $1 - Pr(Y_i)$, the higher the risk for trade credit. We obtain symmetric segments of length equal to 0.1, i.e., about ten percentage points of lower risk in each category when approaching export status. Therefore, we can run the following simple specification:

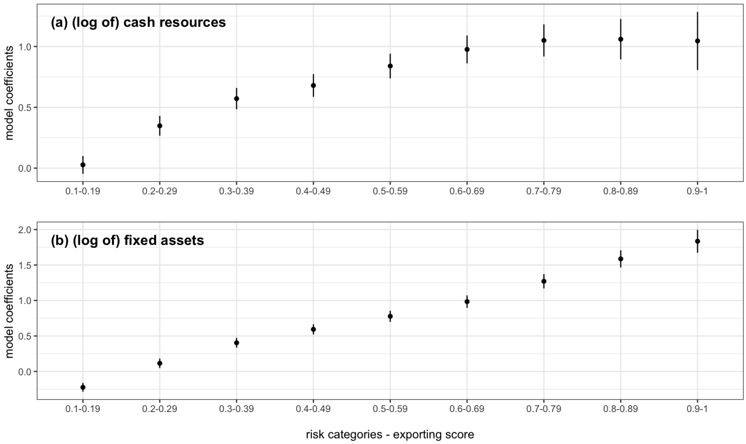
$$\log Y_{it} = \beta_0 + \sum_{risk=1}^{10} \theta_{risk} + \beta_1 x_{it} + +\phi_t + \delta_s + \eta_r + \epsilon \tag{1.7}$$

where Y_{it} is either cash resources or fixed assets for firm i at time t, and x_{it} is its firm-level size. We will always control for time (ϕ_t), four-digit NACE sector (δ_t), and two-digit NUTS region (η_r) fixed effects. We cluster standard errors at the firm level. Crucially, our coefficients of interest are the ones on θ_{risk}, as these are the risk classes we built on exporting scores. We report them in decreasing order of risk in Figure 5 together with 99% confidence intervals. Once we omit the first segment $[0, 0.09]$, the estimated intercepts of eq. 1.7 will indicate (logs of) cash resources and fixed assets needed by a representative firm that is more distant from export status. To obtain what is on average needed by a firm in a *risk category*, we predict (log) premia with respect to the baseline omitted first segment. For example, the representative firm with exporting scores lower than 0.1 operates with $exp(\hat{\beta_0}) = exp(11.6338) \approx 112,850$ euro of cash resources and $exp(\hat{\beta_0}) = exp(13.4027) \approx 661,790$ euro of fixed assets. Firms in the fifth category, when exporting scores are in a range $[0.4, 0.5)$, will need $exp(\hat{\beta}_0 + \hat{\theta_5}) = (11.6338 + 0.6797) \approx 222,690$ euro of cash resources and $exp(\hat{\beta}_0 + \hat{\theta_5}) = exp(13.4027 + 0.5933) \approx 1,197,800$ euro of fixed assets. To put it differently, we can say that a firm that is in a medium-risk category needs about 97% more cash resources and about 81% more fixed assets if compared with a firm with the lowest exporting scores.

On the other hand, if we look at firms in a comfort zone with exporting scores in a range $[0.9, 1]$, we see that they operate with $exp(\hat{\beta}_0 + \hat{\theta_{10}}) = exp(11.6338{+}1.0459) \approx 321,160$ euro of cash and $exp(\hat{\beta}_0 + \hat{\theta_{10}}) = exp(13.4027 + 1.8348) \approx 4,145,360$ euro of fixed assets. Please note that the higher the probability that a firm starts exporting, the higher the cash resources and capital expenses it needs. In the latter case, if we compare with average exporting scores in the fifth risk class, we find that medium-risk firms need 44% more cash resources and up to 246% more capital expenses to look like firms that have been classified under the lowest risk category.

We observe that there is an increasing need for financial resources to climb risk categories and reduce the distance from export status. Based on predictions made on the experience of both exporters and non-exporters, a financial institution could evaluate whether it's worth the effort of investing in internationalization and, in case, how many resources a firm needs to reach its target. Finally, we spend a few words to show how

Figure 5: Premia on relevant firm dimensions across exporting scores



Note: Fixed effects on segments of exporting scores after linear regressions where the outcomes are (log of) cash resources and (log of) fixed assets, respectively. We always control for firm size, NUTS 2-digit regions, NACE 2-digit industries, and time fixed effects. Errors are clustered at the firm level.

exporting scores can help assess the potential for expanding the set of exporters in a region or an industry, i.e., the potential for a trade extensive margin. Openness to international trade is a determinant of economic growth. Consumers can gain from trade thanks to differential comparative advantages and economies of scale. Both developed and developing economies have benefited from integration into the global economy through export growth and diversification. Thus, export performance has been long used as yet another proxy for measuring countries' competitiveness by a consolidated tradition in economic literature and by international organizations (Gaulier et al., 2013; Leamer & Stern, 1970; Richardson, 1971a, 1971b).

To make our point, we follow a dartboard approach as in Ellison and Glaeser, 1997 and propose location quotients in Appendix Figure A2.4. See Appendix D for further details on computations. Regions with location quotients greater than one are the ones where potential exporters are more concentrated than what one would expect. Eventually, we do find a geographic pattern since non-exporters with the highest potential are mainly present in North-Eastern regions. In contrast, Southern regions and overseas territories lag behind in trade potential.

Eventually, more sophisticated analyses on the distribution of exporting scores in industries and regions can be performed to evaluate trade potential. For example, one could exploit the variation in time to understand how much competitive in trade a region or an industry is becoming. Also, one could compare across countries to check whether there is a different potential for trade beyond actual export performance. We believe any of them could be a useful tool in the kit of the analyst that aims at assessing the trade competitiveness of an economy.

1.11 Conclusions

This paper exploits statistical learning techniques to predict firms' export ability. After showing how financial accounts convey non-trivial information to separate exporters from non-exporters, we propose predictions as a tool that can be useful for targeting trade promotion programs, trade credit, and assessing firms' trade potential. The central intuition is that exporters and non-exporters are statistically different in their financial structures since they have to sustain the sunk costs of gaining access to foreign markets, where regulations and consumer tastes differ. Thus, we train and test various algorithms on a dataset of French firm-level data from 2010-2018. Eventually, we find that the Bayesian Additive Regression Tree with Missingness In Attributes (BART-MIA) outperforms other models due to efficient use of the non-random missing information on smaller firms reporting incomplete financial accounts.

Notably, prediction accuracy is rather high, up to 90%, and robust to both changes in the definition of exporters and different training strategies. Interestingly enough, our framework allows handling cases of discontinuous exporters, as they are intermediate cases between permanent exporters and non-exporters. Eventually, we discuss how predictions can be used as scores to catch firms' internationalization strategies and creditability. For example, imitating what a financial institution would professionally do, we order firms along *risk categories*. Thus, we show back-of-the-envelope estimates of how much cash resources and capital a firm would need to climb risk classes and become fit for foreign markets.

To conclude, we argue that exporting scores obtained as predictions from firm-level financial accounts can be yet another useful tool in the analyst kit to evaluate trade potential at different levels of aggregations. As we show in the case of France, for which we provide summary statistics where a high heterogeneity of trade potential is detected across regions.

Chapter 2

The heterogeneous impact of the EU-Canada agreement with causal machine learning

This chapter is based on the paper: Lionel Fontagné, Francesca Micocci and Armando Rungi "The heterogeneous impact of the EU-Canada agreement with causal machine learning" Papers 2407.07652, arXiv.org, revised Jul 2024. Preprint available at https://doi.org/10.48550/arXiv.2407.07652.

2.1 Introduction

Ex-post estimates of the impact of Free Trade Agreements (FTAs) have been shown to be both unstable and fragile (Baier et al., 2019). This can primarily be attributed to the challenges of effectively addressing issues of endogenous selection in trade agreements and the design of sensible counterfactuals. Due to the phasing-in of tariff reductions, staggered treatment adoption, where groups of products are treated over different periods, is an issue often raised when evaluating trade agreements (Nagengast & Yotov, 2024). And even if the design is not staggered, "forbidden comparisons" can be problematic if the treatment is not binary (De Chaisemartin & d'Haultfoeuille, 2023). These empirical challenges are all the more aggravated by the presence of heterogeneous firms in trade, which can sell multiple products and operate in multiple destinations.

In this contribution, we propose a causal machine-learning approach to uncover the impact of an FTA at the product and firm level. We apply this method to investigate the impact of the CETA (EU-Canada Comprehensive Economic and Trade Agreement) on French trade, using monthly customs data on the universe of French exports. Therefore, our empirical strategy evaluates multidimensional counterfactuals at the product, firm and destination levels. Following our proposed strategy, multidimensional counterfactuals are made possible by adapting a matrix completion algorithm for causal panel data originally suggested by Athey, Bayati, et al. (2021).

Notably, machine learning methods are increasingly used in economics for causal inference (Athey & Imbens, 2019; Mullainathan & Spiess, 2017). The simple intuition is that non-parametric methods are better at predicting potential outcomes in the presence of non-linearities by adopting less stringent assumptions on functional forms and the data-generating process. More specifically, we consider the French customs data as a matrix of observed outcomes to be partitioned between: i) treated vs. untreated observations, depending on whether the units of observation had seen a reduction of tariffs or a change in the quotas thanks to the CETA; ii) observations before and after the signature of the CETA.

Crucially, we can follow the application of the CETA agreement with monthly trade data from 2015M01 to 2018M12. As the signature occurred in September 2017, we split the timeline around that threshold. Then, we perform our exercise first at the product level, considering as treated the manufacturing products that have been included in CETA, and then at the firm level, this time considering multiproduct firms that have been concerned by the CETA because at least one of their products is enlisted by the treaty. In the product-level case, the matrix has cells identified by 5,118 products at the HS 6-digit level, 18 alternative destinations, including Canada, and time. In the second case, the matrix has each cell identified by 3,791 multiproduct firms, 18 destinations, up to three of their most exported products, and time. Preliminary evidence suggests an endogenous product selection in the treaty, given that the products covered by the new CETA provisions already had, on average, a larger market for French producers before the treaty was signed. The products on which the parties negotiated were already exported by a greater number of firms in France, more frequently, with a lower average transaction value and a lower average value dispersion. We argue that such an endogenous selection needs to be monitored as it may be relevant for different trade policy environments. In our case, we implement a placebo test and confirm that matrix completion is capable of handling endogenous selection.

Eventually, once the matrix of observed trade outcomes is designed, we can drop the observations of the treated units after the agreement entered into force and, thus, predict their trade values as if the CETA had not been signed. Crucially, predictions are obtained by exploiting all the information left in the matrix, including two years before the treaty. On the other hand, we can control the prediction accuracy of the method by looking at the elements of the matrix that were not treated. Following standard approaches in machine learning methods, we train our model on five random folds of the part of the matrix that includes untreated units, and then we check out-of-sample how far our predictions are from actual realizations of the outcomes.

For our purposes, CETA is a compelling case of an FTA whose negotiation has been intricate, lengthy and contrasted. It took ten years since the first discussions[1] to have the agreement provisionally entered into force in 2017. According to its provisional enforcement, most of the trade provisions in the agreement have already been applied, although it is still awaiting final ratification by all EU members[2]. During the negotiations, France emerged as one of the main proponents of establishing a closer trading relationship with Canada. A shared colonial past, a common language,[3] and similar consumer preferences give Canada more than an incentive to trade with France. Ratification by the French Assembly was voted on in July 2019, and the agreement was examined and eventually rejected by the Senate in March 2024.

Yet, an asymmetry was evident from the beginning for all parties involved in the negotiation. The treaty would have *prima facie* been more relevant for Canada than for European countries. However, the EU's interest was to foster unprecedented economic cooperation with new partners in the face of the rise of emerging markets like China (Hübner et al., 2017) and to have a testing ground for *deep trade agreements* covering areas beyond tariffs. Notably, an asymmetry in the size of parties involved in the Treaty makes the local competition among European exporters potentially larger compared to the relatively smaller positive demand shock induced by the trade liberalization.[4] Therefore, by looking from the perspective of a single exporting country, France, we would expect a non-negligible impact, possibly magnified by the competition of French exporters with other European producers. We proceed with our investigation in three steps. At first, we evaluate the overall impact of CETA at the product level. Crucially, at this stage, we find that CETA

[1] It dates back to a Canada-EU bilateral summit in Berlin in 2007.

[2] Even if the European Commission is solely in the competence of the trade policy of the European Union, in July 2016 it was decided that CETA qualified as a *mixed agreement* because it touches upon other policy domains different from trade, and thus it needed to be ratified also through national procedures.

[3] English and French have been established as joint official languages since 1969.

[4] Please note that Canada's GDP is similar in size to Italy's. France is Canada's ninth-largest trading partner and the fourth-largest among EU members. At the same time, Canada stands as only the thirtieth-largest partner, amounting to a share of only 0.8% total exports.

positively impacted French exports at both the intensive and extensive margins. On the one hand, product-level flows to Canada increased on average by 1.28%. On the other hand, we find that there has been a relevant product churning due to the treaty beyond regular entry-exit dynamics, as about 13.1% of new French products reached Canada for the first time, and 11.9% of them abandoned the market thanks to the new provisions. Importantly, our matrix completion approach allows us to expose the relevant heterogeneity of the impact of a trade treaty. We argue that it is an advantage with respect to other more synthetic empirical strategies. In fact, we can evaluate the full distribution of treatment effects that emerge from the matrices, i.e., on each product or firm that is concerned by the CETA. In doing so, we observe that we have both cases of positive and negative impacts on observed units and that, for example, the treatment effects on single products are positively associated with a measure of revealed comparative advantage for French exporters vs. the rest of the world. That is, the increase in the export flow has been higher for those products for which French producers had a competitive edge before the treaty signature. Similarly, when we consider the heterogeneity at the extensive margin, we find that product churning is positively associated with the elasticity of substitution. In other words, as largely expected, the French products that either enter or exit the Canadian market as a direct consequence of the new treaty are also the ones that have an elasticity of substitution that is higher if compared with products that just continue to be exported. In the second stage of the analysis, we investigate the firm-level dimension with a special focus on the strategies of multi-product firms. Trade theory tells us that the latter can adjust their portfolios after the signature of a trade treaty. After we rank products within firm-level portfolios, we find that multi-product firms, on average, sell relatively more of the already first-sold products to Canada after the CETA. We believe this result is consistent with the theoretical framework proposed by Mayer et al. (2021) and Eckel and Neary (2010), according to which multiproduct exporters tend to reallocate their product mix as a response to the demand shock in the export markets. In fact, trade liberalization generates relatively higher competition for French exporters, who find it convenient to invest relatively more and focus on the products on which they have a higher competitive advantage. Finally, we follow best practices in the trade literature dealing with general equilibrium effects of a change in bilateral trade costs between parties to a trade agreement (Anderson & Yotov, 2016; Head & Mayer, 2014). Cancellation of tariffs between the parties increases relative trade costs between the parties and third countries, leading to indirect trade effects. This is indeed consistent with a classical Vinerian diversion effect (Viner, 1950), whereby trade between parties to a PTA partially substitutes for trade between third parties that do not participate in the PTA. Following this logic, reducing trade costs with Canada is equivalent to a relative increase in the costs of exporting to other destinations. In our context, trade diversion takes the form of indirect policy spillovers: we detect a significant and negative association between the effects on the export of products from France to Canada enlisted by the CETA and the changes in the exports of the same products from France to alternative destinations. The correlation is all the more significant for products with a relatively higher substitution elasticity. The remainder of the paper is structured as follows. We begin with a short review of the relevant literature in Section 2.2. Section 2.3 presents the data and offers preliminary evidence. In Section 2.4, we outline the empirical strategy. Results are displayed in Section 2.5, while robustness and sensitivity checks are presented in Section 2.6. Section 2.7 concludes.

2.2 Related Literature

Ex-post evaluation of free trade agreements is challenging (Baier et al., 2019) because they often entail an endogenous selection of partners or products (Baier & Bergstrand, 2004, 2009), on the one hand, and a self-selection of heterogeneous exporters (Melitz, 2003), on the other hand. Hence, Goldberg and Pavcnik (2016) consider this endogeneity a major hurdle to the causal identification of the economic impact of FTAS.

This endogeneity of PTAs has been addressed by using various approaches, including gravity equations with additional controls (e.g. bilateral fixed-effects) for unobserved characteristics (Abrams, 1980; Aitken, 1973; Bergstrand, 1985; R. C. Feenstra et al., 2001; Soloaga & Wintersb, 2001), instrumental variable (IV) or control-function techniques with cross-sectional data (Baier & Bergstrand, 2002, 2009; Magee, 2003), panel data models with a rich set of fixed effects (Baier & Bergstrand, 2007; Head & Ries, 1998; Westerlund & Wilhelmsson, 2011; Yang & Martinez-Zarzoso, 2014), or matching techniques (Baier & Bergstrand, 2009; Egger & Tarlea, 2021).[5] In this paper, we explore the scope for using a potential outcome model to assess the causal impact of preferential trade agreements[6]. In particular, we draw from the most recent advances in causal machine learning, whose aim is to estimate average causal effects after predicting the missing potential outcomes with non-parametric methods (Abadie et al., 2010, 2015; Arkhangelsky et al., 2019; Chernozhukov et al., 2021). Specifically, we leverage the literature on matrix completion that originally exploited observed information to predict unobserved information when matrices are sparse (E. Candes & Recht, 2012; E. J. Candes & Plan, 2010; Mazumder et al., 2010). For our purpose, we adapt the algorithm initially proposed by Athey, Bayati, et al. (2021), whose intuition is that a matrix approach can also be used for causal inference while allowing for time dependence, unregularized units and time-fixed effects. All properties that, according to Athey, Bayati, et al. (2021), help boost the quality of potential outcomes' predictions.

[5]See the reviews by Limão (2016) and Larch and Yotov (2023) of the empirical exercises estimating the impact of trade agreements.

[6]The framework for causal inference that uses 'potential outcomes' to define causal effects at the unit level in the context of randomized experiments and quasi-experiments is dubbed Rubin Causal Model (Rubin, 2005). The introduction of this framework in economics helped comply with the so-called *credibility revolution* cited by Angrist and Pischke (2010).

On top of empirical challenges, we know from trade theory that opposing mechanisms may hinder an accurate estimate of the impact of tariff reduction. On the one hand, a tariff reduction implies greater market access because the demand increases in the liberalized market. On the other hand, tariff reductions under trade agreements may have procompetitive effects. When Marshall's second law of demand does not apply, monopolistic exporters may reduce their markups in response to reduced tariffs (Mrázová & Neary, 2017) or preferential market access (Crowley & Han, 2022). This induces, in turn, selection effects. Market size and trade openness affect the intensity of competition in a market, which reinforces the selection of exporters to that market (Melitz & Ottaviano, 2008). Against this background, we design our empirical strategy encompassing multidimensional counterfactuals, both at the product and firm level, which enable us to discuss competing mechanisms.

Crucially, our empirical design acknowledges the role of heterogeneous firms in trade agreements, especially in a world where multi-product firms dominate trade flows (Bas & Bombarda, 2013; Eckel & Neary, 2010; R. Feenstra & Ma, 2007; Iacovone & Javorcik, 2010). In this, we refer to Mayer et al. (2014) and Bernard et al. (2010, 2011), who incorporate multi-product firms into models of heterogeneous firms while building upon the pioneering work by Melitz (2003). They show that tougher competition in a liberalized market leads firms to skew their export sales towards their better-performing products. On a similar line of research, Dhingra (2013) and Qiu and Zhou (2013) predict that falling trade costs make the most productive firms expand their product scope, and the least productive firms contract theirs. According to J. Baldwin and Gu (2009), the net effect could be ambiguous because tariff cuts can both increase exporters' plant size by extending the production-run length of the exported portion of the product line and reduce the exporters' plant size by reducing the total number of products. A final layer of complexity that we consider in this contribution arises when considering the adjustment mechanisms of firms to multiple destinations. Two mechanisms concur with third-country effects, i.e., on destination markets that are not part of the signed PTA. On the one hand, reducing trade costs between the EU and Canada increases the relative cost of exporting to countries that are not parties to the agreement. General equilibrium effects of a change in the matrix of bilateral trade costs are conducive to indirect trade effects (Anderson & Yotov, 2016; Head & Mayer, 2014). Trade between parties to a PTA partially substitutes for trade between parties and third countries, which should appear at the aggregate level (Viner, 1950). On the other hand, at the firm level, the determinants of exporters' geographical expansion reveal patterns of entry, sales distribution across markets, and export participation (Eaton & Fieler, 2019; Eaton et al., 2004, 2011, 2012). Notably, Arkolakis and Muendler (2013) and Arkolakis et al. (2021) found that the scope of exporters is unrelated to the size of destination markets, but it is related to geographic distance. As a result, after trade liberalization, we expect to observe a larger effect on the intensive rather than the extensive margin of trade depending on the geographical distance of the trading partner.

2.3 Data and preliminary evidence

2.3.1 Customs data and trade regime changes

Our primary data source is the French Customs (*Direction Générale des Douanes et Droits Indirects*)[7], where we have records of trade values at the product, firm, and month levels. Products are originally classified by the 8-digit Combined Nomenclature (CN8), and firms are identified by their *SIREN* number, i.e., the 9-digit identifier assigned to every registered business in France by the National Institute of Statistics and Economic Studies. Moreover, we rely on the WTO tariff databases to retrieve information on those products at the HS 6-digit level whose tariffs or tariff

[7]The database was accessed through the CASD, French Secure Data Access Center (project DYNAMEX).

quotas have been modified by the EU-Canada Comprehensive Economic and Trade Agreement (CETA)[8].

Original customs data are first aggregated from monthly to yearly levels in September-August segments, following the timeline of the trade treaty, which became operational in September 2017. In addition, we align the product classification from the 8-digit Combined Nomenclature (CN) to the 6-digit Harmonized System (HS) classification to match the original information on products whose tariff or tariff quota has been changed by CETA. Since the HS classification was revised in 2017, we converted the codes of entries back to HS 2012.

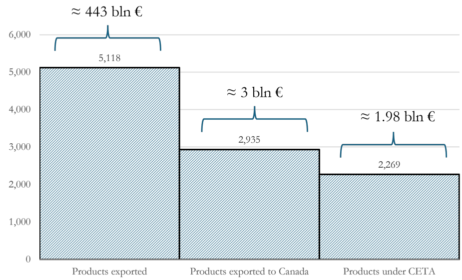
So far, we have identified the perimeter of the product-level analyses we perform in Section 2.5.1. Our investigation encompasses all products that France exports to Canada regardless of the firms' characteristics. In the second part of the empirical strategy, we will focus on the impact that CETA has on multiproduct firms; therefore, we need to eliminate from our sample perimeter[9]: i) firms that do not export to Canada, ii) firms that export only one product to Canada.

In Figures 6 and 7, we provide waterfall charts to visualize the relevance of products and firms included in our study. On the one hand, when we separate products liberalized after CETA, we observe that they make up 77% of the total product lines exported from France to Canada. On the other hand, the list of products that have seen a change in the tariff or non-tariff regime thanks to CETA coincides for about 57% with the list of product lines that French exporters already trade with the rest of the world.

[8] Appendix Table B1.1 briefly summarizes the extent of tariff changes for French exporters in Canada due to CETA

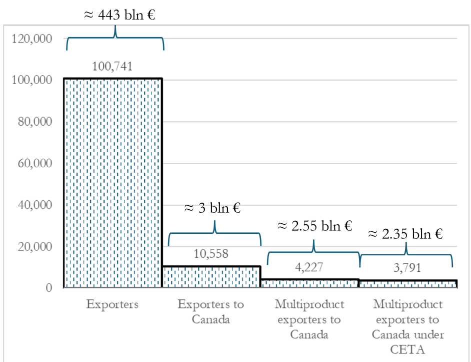
[9] In the original data, we find firms that are active in service industries and occasionally export goods. We eliminate these cases from our firm-level sample perimeter because they conceal a delivery of materials needed to proceed with the service supply (e.g., building materials for construction firms, laboratory equipment for an R&D company, etc.).

Figure 6: Products' coverage in 2016



Note: The figure shows sample coverage of products in 2016. The y-axis indicates the number of products, whereas the text boxes on top of the bars indicate the total trade value in 2016. On the left is the number of products exported from France to any destination. In the centre is the number of products exported to Canada. On the richt is the number of products that are both exported to Canada and fall under the provisions of the Canada-EU Trade Agreement.

Figure 7: Firms' coverage in 2016



Note: The figure shows sample coverage of exporters in 2016, while text boxes on top of the bars indicate the total trade value in 2016. On the left is the number of French firms that exported to any destination. Then, we report the number of exporters to Canada and, among the latter, the number of multiproduct firms because they export at least two products to Canada. On the right is the number of multiproduct exporters to Canada, with at least one product enlisted by the Canada-EU Trade Agreement, for which we indicate the value of their total exports to Canada, encompassing both products with and without a trade regime change.

From our perspective, either stylized fact is worth further investigation. In the first case, we expect an endogenous selection of products in the treaty negotiation, and we test it in the following paragraphs. In the second case, we expect general equilibrium effects inducing indirect trade effects on alternative destinations as it will be *relatively* more costly to export the same products to alternative destinations: see Section 2.5.3.

As for firms, we first need to drop those that have never exported to Canada because they are not directly concerned about the signature of the CETA. Then, following a basic definition of multiproduct firms, we will consider only those that export at least two products to Canada. In this case, as from Figure 7, we can see that only about 10.5% of French exporters reach Canada as an export destination. Among them, about 40% are multiproduct firms and can sell a portfolio of at least two products in Canada. Finally, among the latter, 79.8% have seen a tariff or non-tariff change in at least one of their products exported to Canada after CETA.

In the second part of the paper, the subset of multiproduct firms is of special interest to us not only because they are relevant in terms of aggregate trade flows (2.55 billion euros vs 3 billion euros of total exports to Canada) but also because they are a segment that potentially shows adjustments in product scope, which would be otherwise hidden if we do not consider the firm-level dimension. In Appendix Figure B1.1, we show French exporters' distribution of product portfolios to Canada.

2.3.2 Preliminary evidence

In the following paragraphs, we investigate whether products and firms that have seen a change in the trade regime significantly differ from those that have not. The obvious intuition is that negotiators could have picked production segments that could show higher gains from trade. Alternatively, it is possible that bigger firms had the power to impose their own agenda on negotiators. In Table 4, we investigate the issue with two sequences of t-tests on the difference in means of indicators that could possibly capture the peculiar differences between products included and not included in the CETA. First, we test our indicators considering bilateral exports from France to Canada. Then, we consider the same partition of products under the CETA, this time looking at the features of products and producers at the global level after aggregating over destinations.

Table 4: Characteristics of trade flows before CETA - 2015M01-2016M12

	products in the CETA	products not in the CETA	difference in means
Exports to Canada			
Avg. trade value	30231.8	54023.6	-35700.5***
Avg. dispersion	65579.8	122571.7	-78671.4***
Avg. number of transactions	2571.4	599.9	1971.4***
Avg. number of firms	212.1	100.2	111.8***
Avg. firm's exports	509,037.5	207,466.9	301,507.6***
All exports			
Avg. trade value	35265.7	60645.2	-25379.5***
Avg. dispersion	162147.3	301385.0	-188687.6***
Avg. number of transactions	42852.1	23216.9	19635.2***
Avg. number of firms	1290.5	1278.3	12.18***
Avg. firm's exports	8,150,142	1,412,479	6,737,762***

Note: The table reports t-tests computed on average indicators of the export matrix in 2015-2016 considering products that will see a change with CETA in 2017 (column 2) vs. products whose trade regime will not change (column 3). Column 4 reports differences in the means considering unequal variances. *** stands for $p \leq 0.001$, hence the average means are significantly different. In the first half of the table, we consider only export flows to Canada, i.e., the destination involved in the treaty. In the second bottom half of the table, we enlarge the matrix to consider export flows to all export destinations, although they are not parties in the CETA.

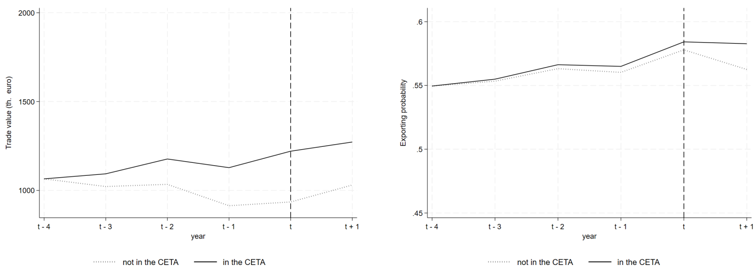
The first three indicators we test in Table 4 refer to features of the product-level monthly flows observed in the period 2015M01 to 2016M12, while the other two indicators refer to the firm-level dimension. Starting from the top of the table, we observe that the average trade value of products included in the CETA had a lower magnitude, a lower dispersion around the sample means, and its transactions were more frequent in the two years preceding the treaty's signature. If we look at exporters, the product was usually traded by more firms, which had, on average, a relatively higher exposure to Canada as an export destination.

If we look at the bottom of the table, we see that the same differences observed in the bilateral relationship between France and Canada are confirmed by aggregate flows between France and the rest of the world. Briefly, products included in the CETA are usually traded by firms whose export size is, on average, bigger, while single monthly flows are smaller, more frequent, and with lower volatility around the mean value in the two years before the CETA.

Eventually, preliminary evidence shown in Figure 4 motivates the choice of an empirical strategy that is capable of handling an endogenous selection of product lines in a trade treaty, thus making policy evaluation unbiased by the political economy of the bigger firms or by the tendency of negotiators to cherry-picking products that already have a higher potential.

Figure 8: Time trends at the product level, intensive and extensive margins

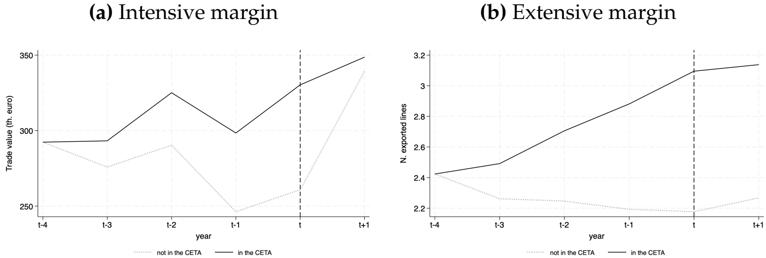
(a) Intensive margin **(b)** Extensive margin



Note: We report in panel (a) linear trends for trade values of product lines exported to Canada, separating those that are included in the CETA and those that are not. In panel (b), we report linear trends for the probability that a new product line is exported to Canada, separating those that are included in the CETA and those that are not. The graphs are generated using the predictions of a difference-in-difference model augmented with interactions of time with an indicator of treatment when products are enlisted by the CETA.

Our preferred empirical strategy should also be capable of handling

Figure 9: Time trends at the firm-level on the intensive and extensive margins



Note: We report in panel (a) linear trends for trade values of firms that exported to Canada, separating those that have a product enlisted by the CETA and those that have not. In panel (b), we report linear trends for the number of lines a firm exports to Canada, separating those that have a product included in the CETA and those that have not. The graphs are generated using the predictions of a difference-in-difference model augmented with interactions of time with an indicator of treatment when firms have a product enlisted by the CETA.

the presence of heterogeneous time trends. It is, in fact, possible that products and firms concerned by the CETA were already on paths to growth before the treaty was signed. The presence of un-parallel time trends could possibly confound the actual impact of the trade treaty. In Figures 8 and 9, we display linear trends after the estimation of simple difference-in-difference models[10] of the intensive and extensive margins for both products and firms, separating when they are concerned by the CETA and when they are not.

After a graphical inspection, we can observe that intensive margins at the product and firm levels (panels (a) in Figures 8 and 9) were already on

[10] We estimate simple difference-in-difference models augmented with terms that capture the differences in slopes across the products/firms that are concerned by the CETA and those that are not. See Appendix B for more details. Results of the difference-in-difference models are reported in Appendix Table B2.1. Please note how diff-in-diff results suggest that the CETA had only an effect on the firm-level extensive margin, whereas no significant impact is registered on the intensive margins at the product and firm levels. While serving as a valuable reference point, a simple diff-in-diff methodology cannot be valid if the assumption of parallel trends is violated, as from Figures 8 and 9, and when the treatment is not orthogonal to relevant characteristics of the treated units, as from Table 4.

diverging paths. In the case of products, those not included in the CETA were already on a downward trend. In the case of firms, those that do not have a product enlisted by the CETA had been on a decreasing trend in the years before the treaty and then increased significantly thereafter. In the case of extensive margins, product flows do not show significant differences, while firm-level pre-trends were significantly diverging.

2.4 Empirical strategy

2.4.1 Treated products and treated firms

In the following paragraphs, we develop an empirical strategy to evaluate the impact of CETA. For the sake of generalization, we will define a generic u-th unit of observation at time t, such that the exposure to CETA, i.e., our treatment, can be defined as W_{ut}. Yet, for our purpose, we need to introduce two different definitions of policy treatment: one at the product level and one at the firm level.

At the product level, we will consider the treated population, $\mathcal{T}$, consisting of all the products that experienced a tariff or a quota change after CETA. Let p denote the product, d represent the destination, and t indicate time. Notice that d can indicate either Canada, as it is the only destination in which treated products are exported with a tariff or quota change, or it can indicate alternative destinations different from Canada. Please note that we consider a product as treated regardless of the destinations in which it is exported. The latter setup will turn out to be useful when we evaluate general equilibrium effects later in the paper.

Since CETA entered into force in September 2017, we aggregate monthly flows by year τ in the period September-August[11]. In this case, the treat-

[11]In the following, $\tau - 2$ refers to the period from September 2015 to August 2016, $\tau - 1$ refers to the period from September 2016 to August 2017, and τ refers to the period from September 2017 to August 2018. Our dataset provides information up to December 2018, which means we can only observe one period (τ) ahead of CETA. Nonetheless, our approach is also suitable for analyzing a staggered adoption scheme across multiple post-

ment indicator is defined as follows:

$$W_{pdt} = \begin{cases} 1 & p \in \mathcal{T}, t > \tau \\ 0 & otherwise \end{cases} \tag{2.1}$$

When we switch to the firm level, our population consists of multi-product firms that export to Canada at least two distinct products[12]. Among them, the set of treated firms Θ is defined as:

$$\Theta = \{i: \ \Psi_{itCA} \cap \mathcal{T} \neq \emptyset, t = [\tau - 2, \tau], \}$$

where Ψ_{itCA} represents the set of products p exported to Canada by firm i in year t, and $|\Psi_{itCA}| \geq 2$. Briefly, we consider as treated any firm that, before or after the entry into force of CETA, exported at least two products to Canada, with at least one of them enlisted by the CETA. Conversely, we will consider non-treated firms that exported at least two products to Canada before CETA but do not have in their portfolio any products included in the CETA.

Once we have defined the set of *treated firms*, Θ, we can establish the treatment at the firm-*per*-product level. Let i denote the firm, p indicate the product, and t represent the year. The treatment indicator at the firm level is defined as:

$$W_{ipt} = \begin{cases} 1 & \forall \, |\Psi_{itCA}| \geq 2, \ i \in \Theta, \ t = \tau \\ 0 & otherwise \end{cases} \tag{2.2}$$

.

Therefore, in the following paragraphs, when we deem it not necessary to specify it, our generic indicator of treatment W_{ut} for the u-th unit will suffice. When presenting results, we will indicate which of the eqs.

treatment periods

[12]See Section 2.3 for a description of the firm-level sample selection strategy

2.1 or 2.2 defines the treatment.

2.4.2 Matrix completion

At this point, we are ready to illustrate the details of our causal machine-learning application on trade policy evaluation. Originally, matrix completion methods were used to recover lost information in highly sparse matrices. In the context of statistical and computer science exercises, the task has been to fill in the missing entries of a matrix that was only partially observed (E. Candes & Recht, 2012; E. J. Candes & Plan, 2010; Mazumder et al., 2010). The novel intuition by Athey, Bayati, et al. (2021) is that one could instead frame a matrix completion algorithm in the context of potential outcome models with predictions of missing multidimensional counterfactuals. We adapt the framework by Athey, Bayati, et al. (2021) to our case of trade policy evaluation, when we have N units of observations (products or firms), T time periods, and there exists a pair of potential outcomes, $Y_{ut}(0)$ and $Y_{ut}(1)$, with unit u exposed in period t to the entry into force of the CETA. The generic treatment has been defined in the previous section as a matrix with entries $W_{ut} \in \{0, 1\}$, and the realized outcomes are thus equal to $Y_{ut} = Y_{ut}(W_{ut})$.

In our case, the fundamental problem of causal inference is that a set $\mathcal{M} < NT$ of potential outcomes is not observed. Specifically, we do not observe the outcomes of the treated units as if the treatment did not occur. In our context, we will never observe the potential exports of products or firms concerned by the CETA as if the latter was not signed. Briefly, we need valid counterfactuals for the set $\mathcal{M}$, and the solution is to predict them using the information available in the trade matrix from entries $\mathcal{O} \equiv NT - \mathcal{M}$, which are observed. Once we obtain valid counterfactuals, we can compute the relative treatment effect on the treated (TET) expressed in monetary values as:

$$\forall \{u, t\} \in \mathcal{M} : TET_{ut} = Y_{ut}(1) - \hat{Y}_{ut}(1) \tag{2.3}$$

.

Then, we can manipulate the latter expression to find the best solution, in levels or in percentage points, depending on whether we want to comment on the intensive or extensive margin, as we explain in the following paragraphs.

Effects on the intensive margin

We can evaluate the impact of the new trade regime on the intensive margin after looking at the moments of the entire distribution produced by the entries we obtain from the matrix of counterfactuals. In this case, we prefer to express the treatment effect on treated from eq. 2.3 as a ratio, to comment in relative terms and on percentage points, in the form:

$$\forall \{u,t\} \in \mathcal{M} : TET^*_{ut} = \frac{Y_{ut}(1) - \hat{Y}_{ut}(1)}{Y_{ut}(1)} \times 100 \tag{2.4}$$

Finally, we can compute the weighted average treatment effect on the treated (WATET), also expressed in relative terms, in the form:

$$WATET = \sum_{\{u,t\} \in \mathcal{M}} s_{ut} TET^*_{ut} \tag{2.5}$$

,

where s_{ut} indicates the salience of the export flows. For the sake of simplicity, we can use for salience the share of the trade flows of unit u at time $t-1$, i.e., before the signature of the CETA, on the total export flows for each entry $\{u,t\} \in \mathcal{M}$.

Effects on the extensive margin

In the evaluation of the extensive margin of trade, the potential outcomes are binary, $Y_{ut}(1) = \{0,1\}$, i.e., they are equal to one if the product is exported and zero otherwise. Our matrix completion application reduces to a classification problem, and we obtain predictions in a binary form,

$\hat{Y}_{ut}(1) = \{0,1\}$, such that treatment effects can have three alternative values, $TET_{ut} \in \{-1, 0, 1\}$. A value -1 means that our counterfactual predicts that a trade flow existed in that entry of the trade matrix, but it actually did not. We will define the latter as the negative extensive margin. A value of 1 implies that our counterfactual indicates that the product should not have been traded, but it actually was. We will call the latter the positive extensive margin. On the other hand, every time that we find a $TET_{ut} = 0$, it means that our counterfactuals and the observed outcomes corresponded. Please note that, against the previous background, products can still enter or exit the foreign market following regular product churning, regardless of a change in the trade regime. The latter cases would all be flagged with a zero in the set of treatment effects.

The estimator

Let us start by representing the entire trade matrix from the original data. In the product-level analysis, we will have a matrix with entries defined by the trade value of each 6-digit product-*per*-destination (i.e., the u-th observation) and time in a cell. In the firm-level analysis, we report each matrix cell's trade by firm-*per*-product (i.e., the u-th observation) and time. Next, we empty the set $\mathcal{M}$ of matrix entries where we have exports with tariff and tariff-quota changes after the CETA signature, i.e., $Y_{ut}(1)$ when $\geq$ 2017, and we ask the algorithm to reconstruct the full matrix while feeding it information from the set $\mathcal{O}$, including:

1. treated and untreated observations before the treatment, when CETA did not exist (i.e., $Y_{ut}(1)$ and $Y_{ut}(0)$ when $t < 2017$)

2. untreated observations after the treatment (i.e., $Y_{ut}(0)$ when ≥ 2017)

Further details on the product-level and firm-level trade matrices are described in Sections 2.5.1 and 2.5.2, respectively. In our context, the value of a matrix completion approach lies in its ability to leverage non-parametrically all available information without making stringent assumptions on joint distributions and functional forms. By predicting each unobserved potential outcome, we obtain multidimensional counterfactuals for each cell in a matrix that pertains to treated units, therefore taking on board all the heterogeneity that can possibly derive from a trade policy treatment.

We obtain predictions from a decomposition of the $N \times T$ matrix $\mathbf{Y}$, such that:

$$\mathbf{Y} = \tilde{\mathbf{Y}} + \tilde{\gamma} + \tilde{\delta} + \varepsilon \tag{2.6}$$

where we can collect $\hat{\mathbf{Y}} = \tilde{\mathbf{Y}} + \tilde{\gamma} + \tilde{\delta}$, as these are the components we want to estimate. Among them, $\tilde{\mathbf{Y}}$ is a low-rank matrix with respect to the original $N \times T$. Then, we have $\tilde{\gamma}$, which is the $N \times 1$ vector of row-fixed effects, and $\tilde{\delta}$, which is the $1 \times T$ vector of time fixed effects[13]. In our context, the $N \times 1$ vector of row-fixed effects can represent either product-destination or firm-level fixed effects, respectively. Eventually, we leave ε as an $N \times T$ matrix of random noise values.

Our $\tilde{\mathbf{Y}}$ is the result of a singular value decomposition (SVD), such that $\tilde{\mathbf{Y}} = \mathbf{S\Sigma R}^{\top}$, where $\mathbf{S}$ and $\mathbf{R}$ are unitary matrices, and Σ is a rectangular diagonal matrix with singular value entries $\sigma_u(Y)$. The latter entries are substituted by $\max(\sigma_i(\tilde{\mathbf{Y}}) - \lambda - Y, 0)$ after regularization. In fact, we introduce regularization on the $\tilde{\mathbf{Y}}$ component, $\lambda_Y||\tilde{\mathbf{Y}}||$, to avoid overfitting. In our context, overfitting would imply that the model corresponded too closely to the training matrix, and its power would be poor in predicting counterfactuals. Indeed, overfitting problems more likely arise in cases like ours where we have a high $N \times T$ dimensionality. Finally, the estimator can be written as the result of an optimization problem in the general form:

[13]Note that the row and column-fixed effect can be subsumed in matrix $\tilde{\mathbf{Y}}$. However, Athey, Bayati, et al. (2021) already pointed out that separating fixed effects without regularization greatly improves prediction quality. In our case, we confirm that prediction power deteriorates when we do not separate fixed effects.

$$\min_{\tilde{Y},\gamma,\delta}\left[\sum_{(u,t)\in\mathcal{O}}\frac{1}{|\mathcal{O}|}\left(Y_{ut}-\tilde{Y}_{ut}-\gamma_i-\delta_j\right)^2+\lambda_Y||\tilde{\mathbf{Y}}||_*\right] \tag{2.7}$$

where $\mathcal{O}$ includes any pair (i,t) in the set of observed export outcomes, and $||\tilde{\mathbf{Y}}||_*$ is the nuclear norm of the matrix $\tilde{\Sigma}$ resulting from shrinking the scaling matrix with the singular value decomposition (SVD) by λ_Y. We select the optimal value of λ_Y after cross-validation[14] on K different random subsets $\mathcal{O}_k \subset \mathcal{O}$ of the original matrix, having a fraction of observed data equal to the one in the original sample. Finally, once we have predicted matrix $\hat{\mathbf{Y}}$, we obtain the counterfactuals we need to estimate treatment effects as in eq. 2.3.

2.5 Results

In this section, we discuss the findings of our application to both a product-level and a multiproduct firm-level investigation. For each case, we introduce separate exercises for the intensive and extensive trade margins. In each case, we start by describing the specific design of the matrix structure that we draw before running the estimator. Then, we report the prediction accuracies always needed to validate the model. Finally, we comment on the results with the help of a few post-estimation statistics.

2.5.1 Product-level analysis

The unit of observation is the product p at the 6-digit level of the HS classification exported at time τ to different destinations d. A product is *treated* if its tariff or quotas have changed after CETA since September 2017[15]. Therefore, in this section, we are interested in evaluating treat-

[14] As we choose a nuclear norm for regularization, the estimator can be computed using fast convex optimization programs like the one proposed by Mazumder et al., 2010

[15] Please note how, since eq. 2.1, we consider the treatment to be product-specific and not destination-specific. The reason is that we will also investigate policy spillovers in

ment effects on the treated in percentage points, which we now write as TET^*_{pdt} because the general u-th unit of observation is now represented by a product p, at destination d, and time t.

For our purpose, besides Canada, we aggregate and rank major destinations of French exports to avoid matrix sparsity[16]. We compute two separate destination rankings, and then we consolidate them. At first, we rank importing countries based on the average total trade value they received from France in 2010-2016. In a second exercise, we rank destinations after counting the number of products received from France in the same period. Finally, we include in our selection those countries that are in the top ten in either ranking. The remaining destinations are mainly aggregated by continent (e.g., the rest of Europe, the rest of Asia, etc.). In Appendix Table B1.3, we record the relative trade importance of each destination in our final ranking.

As for products, we ensure we can properly separate the intensive and the extensive margin. In the first case, we only consider the subset of products that were exported to Canada in either of the two years before the treatment and were still exported after the CETA[17].

In Figure 10, we visualize our matrix structure. In the case of the intensive margin, the P rows of the matrix correspond to the HS 6-digit products exported by France. The TD columns of the matrix, instead, correspond to the set of D possible export destinations in T different times. Then, each matrix element Y_{pdt} is the total export value for product p at destination d and time t.

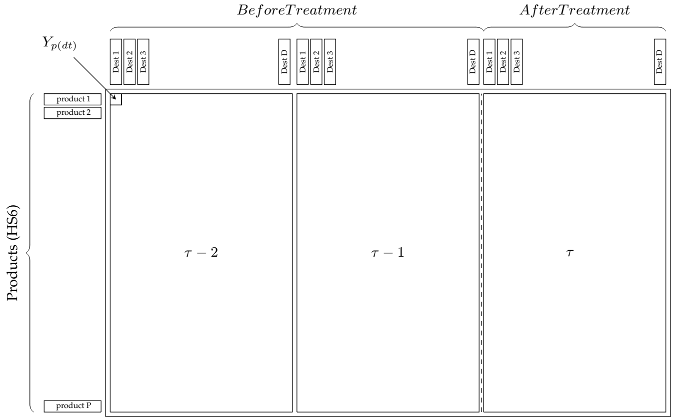
In the case of the extensive margin, our focus is the effect on the export probability of treated products. In this case, we will consider all

destinations that are not directly affected by the CETA, as it will become evident in Section 2.5.3.

[16]As in Fontagné et al. (2018), we also observe a high sparsity because the selection of products at each destination is stringent. In the original data, the vector of products exported to each destination contains, on average, at least 80% of zeros. A highly sparse matrix with an inflation of zeros complicates calculations while saturating computer memory.

[17]For a visual representation of the trade patterns included in the intensive margin, see Appendix Table B1.2.

Figure 10: Matrix Structure for the product-level analysis



possible products $\mathcal{P}$ exported by France anywhere, and each matrix element is a binary variable, $Y_{pdt} = \{0, 1\}$, which takes the value 1 if product $p \in \mathcal{P}$ is exported at destination d in time t, and 0 otherwise.

Table 5: Prediction accuracy at the product level

model	min RMSE	$\overline{Y}$	SI	NRMSE
Intensive Margin	7.12126	7,060.71	0.000001	0.00027
Extensive Margin	0.25861			0.25861

Note: The table reports standard measures of prediction accuracy. $\overline{Y}$ is the average trade of a line p in a year for any destination d, and it is used to compute the normalised version of the RMSE and the Scatter Index (SI). The value of $\overline{Y}$ indicates the average predicted counterfactual in monetary values. On the extensive margin, no normalization is required, as the predicted outcomes are already in a range $0, 1$.

We estimate the model by solving the minimization problem described in the generic eq. 2.7, and we obtain two matrices of predicted outcomes: one for the intensive margin and one for the extensive margin. Then, crucially, Table 5 reports some measures of the prediction accuracy. Briefly, a certain level of prediction accuracy guarantees that our empirical model returns valid counterfactuals. If the predicted values are close enough to the observed values, then we expect a minimum bias when we evaluate the impact of the policy. As in a standard machine learning framework, the algorithm is first trained on different in-sample subsets and then evaluated on out-of-sample segments. In our specific case, the evaluation is made with a minimum average Root Mean Squared Error (RMSE) obtained after five random folds[18].

Notably, we record a high prediction quality in both cases of the intensive and extensive margins, as indicated by the small values of the Normalized Root Mean Squared Error (NRMSE) and the Scatter Index (SI). For the intensive margin, the average difference between the predicted and observed values is 7.12 in the case of the intensive margin and 0.26 in the case of the extensive margin.

Products' intensive margin

Let's start by looking at the heterogeneity of the treatment effects on the intensive margin for products exported to Canada in Figure 11. We can find either products that experienced a reduction in trade following the implementation of CETA or products that consistently benefited from the new trade regime. Visually, we can realize that positive treatment effects slightly prevail. In Table 6, column (1), we report the average weighted treatment effects on the treated products, following eq. 2.5, which is our synthetic number to evaluate how product-level trade responded to the new trade regime. We find a positive and significant value of 1.28% on export flows[19]. Interestingly, other moments of the distribution help us

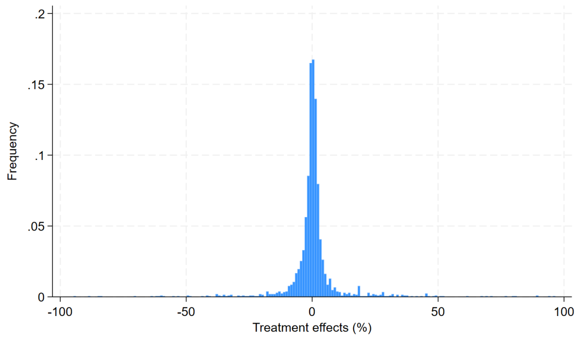
[18]Following the original procedure by Athey, Bayati, et al. (2021), five random folds are used as cross-validation to derive the optimal λ_Y^* of eq. 2.7. For each λ_Y, we train our model in-sample on each k-th random training subset, $\mathcal{O}_k \subset \mathcal{O}$, and we compute $\hat{\mathbf{Y}}(\lambda_{Y(k)}, \mathcal{O}_k)$. We then calculate the RMSE for each out-of-sample k^{th} testing set. We pick the λ_Y corresponding to the minimum RMSE, which guarantees better prediction accuracy. Thus, Table 5 reports the minimum average RMSE corresponding to the optimal λ_Y^*.

[19]The statistical significance is derived from the computation of a weighted standard deviation computed as $\sqrt{\frac{\sum_{i=1}^{N} w_{pdt}\left(TET_{pdt}^{*} - WATET_{pdt}\right)^2}{(\mathcal{M}-1)\backslash\mathcal{M}\sum_{i=1}^{N} w_{pdt}}}$, where we take into account the distribution of weights, $\mathcal{M}$ is the number of the treatment effects on the treated products

in evaluating the impact of the CETA. The simple average (ATET), the median, and the skewness all point to an overall positive yet asymmetric impact on product-level export flows.

Yet, the great degree of heterogeneity of the treatment effects is worth special attention, as it is a piece of evidence that has been neglected in trade policy literature. We argue that exposing heterogeneity is one important advantage of implementing matrix completion for trade policy evaluation, whereas the otherwise typical empirical test would have summarized the policy's effectiveness with a unique synthetic coefficient. For example, if we implemented a simple diff-in-diff strategy, as in Appendix B, we would obtain a unique statistically non-significant coefficient, on which we would have concluded that the treaty did not have any impact. In reality, positive and negative effects could cancel out, and the unique coefficient can conceal relevant heterogeneity.

Figure 11: Distribution of the relative Treatment Effects on the Treated (TET) - intensive margin



Note: The figure reports a histogram for the distribution of relative treatment effects, TET^{*}_{pdt}, following eq. 2.5, which have been computed for each HS 6-digit product exported to Canada that has seen a change in the trade regime after CETA, and then they are weighted for the relevance each product had in the year before the treaty signature.

---

that we computed, and $WATET_{pdt}$ is the weighted average we get from 2.5.

Table 6: Weighted Treatment Effects on the Treated (WATET) products to Canada - intensive margin

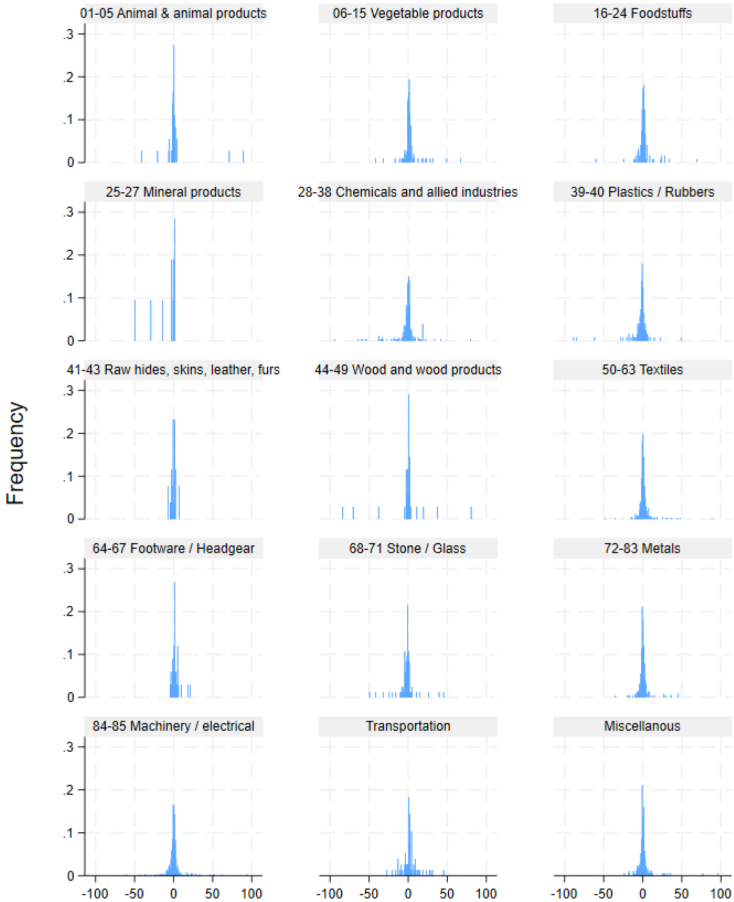
Model	WATET (1)	weighted st. dev. (2)	N. products (3)
Intensive margin	1.278***	0.524	2,165

Note: The table reports the Weighted Average Treatment Effects on the Treated (WATET) products, obtained from TET^{*}_{pdt}, considering the relevance each product had in the year before the treaty signature. The weighted standard deviations are computed as $\sqrt{\frac{\sum_{i=1}^{N} s_{pdt}\left(TET^{*}_{pdt}-WATET\right)^{2}}{(\mathcal{L}-1)\backslash\mathcal{L}\sum_{i=1}^{N} s_{pdt}}}$, where $\mathcal{L}$ is the number of counterfactuals in the trade matrix for Canada. *** stand for $p < 0.001$.

The heterogeneity is still pronounced when we group single products by main classes, as in Table 7 and Figure 12. Apparently, Animal & Animal Products, Foodstuffs, Plastic/Rubbers, Stone/Glass, and Machinery/Electrical are the classes that register a positive impact, while no negative impact is found on any other class. The impact is positive and higher in Foodstuffs with a weighted average treatment effect (WATET) of 1.9%, and it is lower in the case of Animal & Animal Products with a WATET of 0.3 %. Notably, almost all distributions are positively skewed with an asymmetry in favour of the positive quadrant, with the exceptions of Mineral Products (HS 25-27) and Wood & Wood Products (HS 44-49), whose WATETs are anyway non-significantly different from zero.

Nonetheless, when we evaluate the entire distribution of each product class, we always observe a fringe of products for which the signature of CETA has brought a negative impact. Even if such negative effects do not dominate the distributions, where the impact is either positive or statistically non-significant, they are still relevant and require a discussion. As a matter of fact, unweighted standard deviations are high, and they indicate huge variations around the average treatment effect. Therefore, we introduce in Section 2.5.1 a few descriptive statistics that help and qualify the positive and negative variation around the albeit positive average effect.

Figure 12: Distribution of the relative Treatment Effects (TE) on the intensive margin by main product classes



Note: The figure reports histograms for the distribution by main product classes of relative treatment effects, TE_{pdt^*}, following eq. 2.5, which have been computed for each HS 6-digit product exported to Canada that has seen a change in the trade regime after CETA, and then they are weighted for the relevance each product had in the year before the treaty signature.

Table 7: Weighted Average Treatment Effects on the Treated (WATET) products to Canada - intensive margin of main product classes

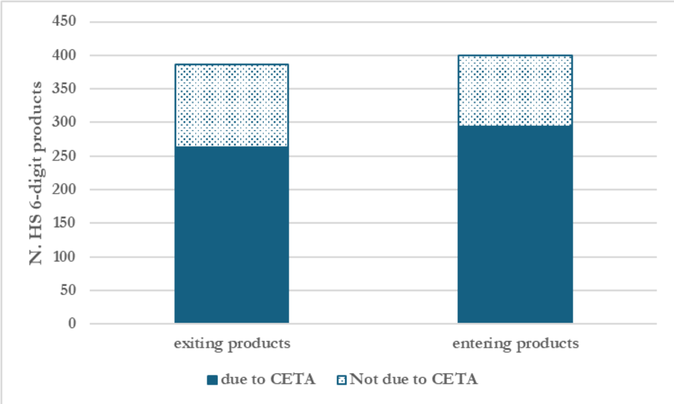
Product class	Class name	WATET	weighted st. dev.	N. products
01-05	Animal & Animal Products	0.307***	0.096	43
06-15	Vegetable Products	0.990	0.989	109
16-24	Foodstuffs	1.911***	0.095	130
25-27	Mineral Products	1.000	1.070	11
28-38	Chemicals & Allied Industries	1.161	1.346	268
39-40	Plastics / Rubbers	0.454***	0.208	131
41-43	Raw Hides, Skins, Leather & Furs	0.678*	0.370	27
44-49	Wood & Wood products	0.849	0.741	36
50-63	Textiles	1.372	1.874	442
64-67	Footwear / Headgear	1.337	1.835	36
68-71	Stone / Glass	0.436***	0.189	88
72-83	Metals	1.410	1.987	230
84-85	Machinery / Electrical	0.927***	0.356	417
86-89	Transportation	1.301	1.638	83
90-97	Miscellaneous	0.908	0.824	186

Note: The table reports the Weighted Average Treatment Effects on the Treated (WATET) exports by main product classes to Canada. Treatment effects in percentage points, TET^*_{pdt}, are weighted for the relevance each product had in the year before the treaty signature to obtain the unique $WATET$. The weighted standard deviations are computed as $\sqrt{\frac{\sum_{i=1}^{N} s_{pdt}\left(TET^*_{pdt}-WATET\right)^2}{(\mathcal{L}-1)\backslash\mathcal{L}\sum_{i=1}^{N} s_{pdt}}}$, where $\mathcal{L}$ is the total number of the treatment effects on the treated units for the reference population of each row. *** stand for $p < 0.001$.

Extensive margin

Figure 13 provides a snapshot of the impact on the extensive margin, while corresponding numbers are reported in Table 8. The impact is evaluated by considering the additional entry-exit dynamics due to CETA on top of the regular entry-exit that we would have seen in any case in the absence of any treatment. In Figure 13, we start by separating the exiting products on the left and the entering products on the right. The light-coloured areas indicate, in both cases, the share of entry-exit that we do not attribute to the CETA because it is regularly predicted by the matrix of potential outcomes we obtain after our algorithm. The dark-coloured area represents instead the cases of treatment effects (TET) that are different from zeros, as from eq. 2.3. If we compare with the number of incum-

Figure 13: Positive and negative extensive margin



Note: The figure reports the numbers of exiting (on the left) and entering products (on the right) that we observe after the signature of the CETA. The light-coloured areas indicate products that would have entered or exited in any case without the CETA, i.e., they are predicted as such in the matrix of potential outcomes. The dark-coloured area includes products that enter or exit Canada as a result of the CETA signature, i.e., they are obtained as non-zero treatment effects after the matrix of potential outcomes.

bent products[20] in 2017; the bar on the left indicates a positive extensive margin of about 14.5%. That is, in 2017, we had an additional 14.5% of products exported from France to Canada for the first time, thanks to CETA. On the other hand, we register a negative extensive margin equal to 13.1% if we compare it with incumbent products. That is, in 2017, we had an additional 13.1% of products that were not exported anymore due to CETA.

[20]We consider as incumbent the 2,031 products exported in Canada after the signature of the treaty, and that were also exported at least two years before the signature of the CETA. If we consider the demography predicted by the algorithm in the absence of the CETA, we would have about 5.2% of regular entries and 6% of regular exits. These numbers are close to what we find in entry/exit in previous years, before CETA.

Table 8: Positive and extensive margins - with and without CETA

	with CETA	without CETA	Total
Negative extensive margin	263	123	386
Positive extensive margin	294	106	400

Note: The table reports the numbers of exiting (first row) and entering products (second row) that we observe after the signature of the CETA. In the first column, we report the numbers of products that have entered or exited due to the CETA, i.e., they are obtained as non-zero treatment effects after the matrix of potential outcomes. In the second column, we report the numbers of products that have entered or exited not due to the CETA, i.e., they are predicted as such in the matrix of potential outcomes.

Table 9: Extensive margin by main product classes

HS class	Product class	Exiting	Entering	Net entry
01-05	Animal & Animal Products	19	24	5
06-15	Vegetable Products	41	22	-19
16-24	Foodstuffs	6	11	5
25-27	Mineral Products	12	8	-4
28-38	Chemicals & Allied Industries	29	71	42
39-40	Plastics / Rubbers	3	1	-2
41-43	Raw Hides, Skins, Leather & Furs	1	4	3
44-49	Wood & Wood products	12	21	9
50-63	Textiles	60	31	-29
64-67	Footwear / Headgear	0	0	0
68-71	Stone / Glass	5	16	11
72-83	Metals	35	34	-1
84-85	Machinery / Electrical	31	37	6
86-89	Transportation	5	5	0
90-97	Miscellaneous	4	9	5
	Total	263	294	31

Note: The table reports the numbers of exiting (first column) and entering products (second column) by main HS product class. The focus is on the extensive margins we observe as they are due to the CETA, i.e., they are obtained as non-zero treatment effects after the matrix of potential outcomes. The third column represents the difference between the entry and the exit.

In Table 9, we further separate negative and positive extensive margins by main product classes. Here, we explicitly focus only on the entry-exit dynamics we attribute to CETA. Notably, the product class that has by far benefited the most from the treaty is the Chemicals & Allied Industries (HS 28-38), with an entry of 71 more products, followed by Machinery/Electrical (HS 84-85) with 37 new products, and Textiles (HS 50-63) with 31 new products. If we look at the negative extensive margin, we find that the group with the highest number of exits is Textiles (HS 50-63) with 60 products, followed by Vegetable Products (HS 06-15) with 41, and Metals (HS 72-83) with 35. Notably, Textiles (HS 50-63) is the class for which the net extensive margin has been most negative, with a loss of 29 products, whereas Chemicals & Allied Industries is the one with the highest gain from the net entry, with a total of 42 products.

Post-estimation analysis

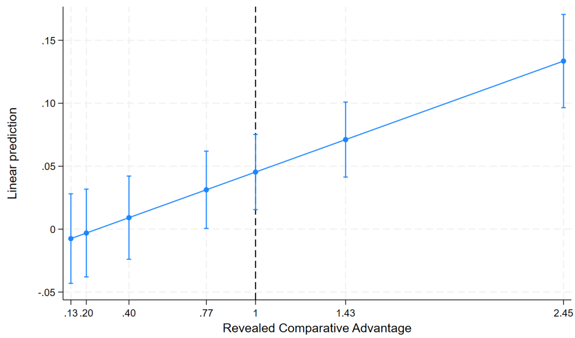
In this section, we explore a few additional descriptive statistics that help qualify the relevant heterogeneity we detected in the previous paragraphs. We investigate the intensive and the extensive margins in Canada in relationship with a few dimensions that we deem important to describe the heterogeneity we observe.

Let us start with the results of the intensive margin. Most interestingly, we record a positive correlation between the treatment effects expressed as percentage points, TET^*_{pdt}, and a measure of revealed comparative advantage (RCA) computed in the year before treatment considering the universe of French customs data[21]. Eventually, in Figure 14, we visualize the statistical association with a 95% confidence interval. We observe that the correlation is positive and statistically significant after the threshold value when RCA is equal to one.

Briefly, Figure 14 shows that the higher the previous comparative advantage of the product in Canada, the higher the positive impact of the

[21] The standard measure of revealed comparative advantage (RCA) that we compute is in the form: $RCA_{pt} = \frac{\frac{X_{CA,pt}}{X_{CA,t}}}{\frac{X_{W,pt}}{X_{W,t}}}$, where $X_{CA,pt}$ is the export flow of the single p HS 6-digit product from France to Canada at time t, $X_{CA,t}$ is the total export to Canada at time t, $X_{W,pt}$ is the export of the same p product from France to the world at time t, and finally $X_{W,t}$ is the total export from France at time t.

Figure 14: Treatment Effects on the Treated (TET %) and comparative advantage - intensive margin



Note: The figure reports a plot of the predicted margins after a linear regression between the set of treatment effects on the treated in percentage points TET^{*}_{pdt} when the destination is Canada and a standard measure of Revealed Comparative Advantage computed in the year before the CETA. The reference line, when RCA is equal to one, indicates that products below it were at a comparative disadvantage and products above it were at a comparative advantage. Bars indicate a 95% confidence interval.

CETA. When tariffs are reduced or quotas are extended, the response in percentage points is higher for those products that were already selling well on the Canadian market. In a nutshell, a good portion of product-level heterogeneity in the effects of CETA is finally explained by initial comparative advantage positions. The latter is an interesting result that we can record because we can rely on an array of counterfactuals thanks to matrix completion.

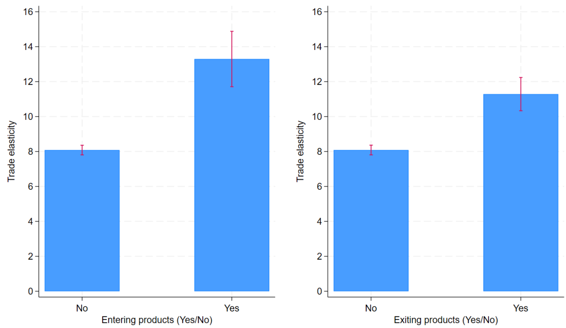
Please note, however, that when RCA is lower than one, the association is not statistically significant. In cases of products that were at a comparative disadvantage, when a product was not selling well in Canada, it is not clear what impact we should expect after the treaty signature. At this point, we can proceed with investigating the estimates we obtained for the extensive margin in Canada. Figure 15 reports the results of two binary regressions. In both cases, we visualize the result of a linear regression model whose dependent variable is the trade elasticity of the single HS 6-digit product sourced from Fontagné et al., 2022. On the left panel, a binary variable (Yes/No) declares whether the product entered the Canadian market due to the CETA or was already exported. On the right panel, a binary variable (Yes/No) declares whether the product exited the Canadian market due to the CETA or survived after the treaty. What we see is that entering and exiting products have, in general, a higher trade elasticity if compared with incumbent products. We believe it makes sense that products whose response to changes in trade costs is relatively higher are also the ones that react the most to a tariff reduction or a quota extension, eventually contributing to the extensive margin. In the case of the negative extensive margin, a fringe of exporters who face a relatively higher trade elasticity observe the changes in the relative costs and decide to reduce export values up to the point of exiting the Canadian market. Similarly, in the case of the positive extensive margin, a fringe of producers who face a relatively higher trade elasticity were not able to export in Canada and decided to enter the market when they observed an albeit small change in tariffs or quotas.

2.5.2 Firm-Level Analysis

Our choice is to investigate the peculiar category of multi-product firms. The latter is an interesting category of firms that is certainly relevant, as we have seen in Figure 7 that multiproduct firms are responsible for about 85% of export flows from France to Canada. From another perspective, multiproduct firms are also an interesting case to follow after trade liberalization events because we want to test whether they adjust their portfolios of products as predicted by trade theory.

From the original data, we select only those firms exporting more than one product to Canada within our time frame. Then, we generate a ranking for each firm by ordering products based on their trading values, from the most to the least traded by the single firm in the year before the treaty. We will report results only on firms that trade at least two or three

Figure 15: Extensive margin and trade elasticity

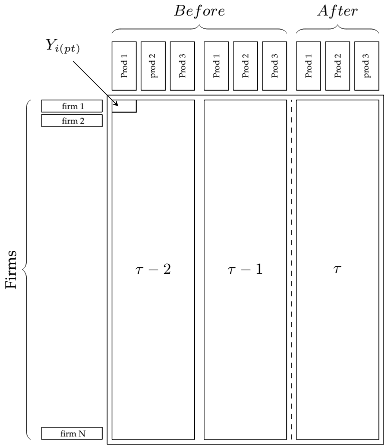


Note: The figure reports a plot of the predicted margins after two linear probability models (LPMs), whose dependent variable is the trade elasticity of the single HS 6-digit product that is exposed to the CETA. In the left panel, the comparison is between incumbent and the exiting products. In the right panel, the comparison is between the incumbent and the entering products. Trade elasticities are sourced from Fontagné et al., 2022. Bars indicate a 95% confidence interval.

product lines to reduce the noise caused by yearly volatility in bigger portfolios of products. Notably, the first most traded products at the firm level account, on average, already for 70% of that firm's exports.

In Figure 16, we report the design of a firm-level matrix to study the intensive margin by multiproduct firms. Please remember that, consistently with eq. 2.2, we consider as treated any (multiproduct) firm with at least one product line whose tariff or quota has been affected by the signature of the CETA. In Figure 16, rows correspond to the N multi-product French exporters. Among them, Θ is the population of treated firms, and $(N - \Theta)$ is the set of untreated firms. Each column represents a different combination of time t and product p. The product is identified at the HS 6-digit level, and we include only the three most traded lines for each firm before τ, i.e., the year of treatment, among those exported in each of the three years in the panel. The matrix element $Y_{i,(pt)}$ measures the observed outcome of firm i for the product p at time t.

Figure 16: Matrix Structure for the firm-level analysis



Similarly to what we did at the product level, we reconstruct the matrix of observed outcomes and predict the counterfactuals following the estimator in eq. 2.7. Table 10 presents summary statistics of the prediction quality of our firm-level exercise. The percentage of expected error for the parameter of interest (i.e. the Scatter Index) is 29%. Prediction power indicates that the algorithm successfully replicates the dynamics of the original matrices of outcomes in the observed entries[22]. At this point, we can validly use predicted values of unobserved potential outcomes as counterfactuals for what would have happened if CETA was not signed.

[22] As in a classic machine-learning predictive framework, the algorithm is first trained on different in-sample subsets and then tested out of the sample. See also footnote 19 for further details.

Table 10: Prediction quality - Firm-level analysis

Model	n. obs.	$\overline{Y}$	min Av(RMSE)	SI	NRMSE
Intensive	3,177	203,345.61	59,069.2	29.04	42.93

Note: The table collects quality indicators for the predictions of observed values in the multiproduct firm-level exercise. The following columns indicate the average predicted value, the root mean squared error (RMSE), the scatter index, and the normalized RMSE.

Multiproduct firms and product scope

Results on the impact of CETA on multiproduct firms are reported in Table 11, while Figure 17 reports a visualization of the distributions of treatment effects for the first, second and third exported products, respectively. Please note that, in these paragraphs, we are considering the multiproduct firms exposed to CETA and that exported at least three products in Canada vs. a control group of untreated firms, as described in eq. 2.4.1. Therefore, our quantities of interest are the treatment effects on the treated, TET^{*}_{ipt}, expressed in percentage points with reference to products ordered, $p = \{1, 2, 3\}$, after considering their trade values in the firm's portfolio before CETA.

If we look at the first part of Table 11, we find that the weighted average treatment effect on the treated (WATET) first products is 0.87%, although weakly significant. At the same time, the WATET on the second product is not significantly different from zero, while the WATET on the third product indicates a tiny yet significant increase of 0.012%. Briefly, the CETA has, on average, a positive impact on at least two products out of three in the portfolio of multiproduct firms exposed to CETA. Yet, the impact is bigger for products already performing better in the Canadian market. Visually, our results are confirmed by the three graphs we included in Figure 17 where, however, we can observe relevant heterogeneity in the positive and negative quadrants.

Table 11: Weighted Average Treatment Effects on the Treated (WATET) products ranked by the multiproduct firms

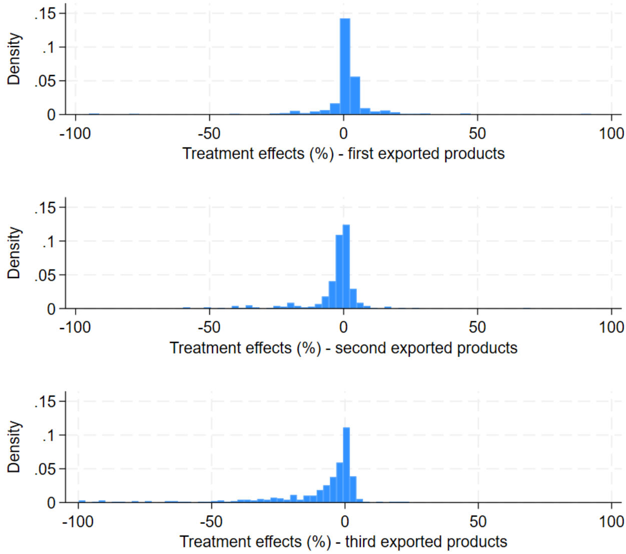
Type of firm/product	WATET	weighted st. dev	N. obs
All firms			
First exported product	0.886*	0.481	418
Second exported product	0.001	0.001	418
Third exported product	0.012***	0.001	418
Manufacturing firms			
First exported product	0.729***	0.296	298
Second exported product	-0.025***	0.001	298
Third exported product	0.001	0.001	298
Trade intermediaries			
First exported product	0.157***	0.003	120
Second exported product	0.027***	0.001	120
Third exported product	0.011***	0.001	120

Note: The table reports the Weighted Average Treatment Effects on the Treated (WATET) exports for the first, second and third products in the multiproduct firms' portfolio. The $WATET$'s are computed considering products' trade shares in the year before the CETA. The weighted standard deviations are computed as $\sqrt{\frac{\sum_{i=1}^{N} s_{ipt}\left(TET^{*}_{ipt}-WATET\right)^{2}}{(\mathcal{L}-1)\backslash\mathcal{L}\sum_{i=1}^{N} s_{ipt}}}$, where $\mathcal{L}$ is the total number of the treatment effects on the treated units for the reference population of each row. *, **, *** stand, respectively, for $p < 0.05$, $p < 0.01$, $p < 0.001$.

Importantly, the second and third parts of Table 11 differentiate firms separating manufacturing firms from those firms that professionally act as intermediaries on behalf of other firms[23]. Our separation is based on

[23]Originally, our data also included firms in primary markets, like agricultural products and other commodities, in the NACE rev. 2 sectors 01-09. However, none of these firms are

Figure 17: Distribution of treatment effects (%) by product ranked in multi-product firms



Note: The Table shows the distribution of the treatment effects on the treated in percentage points, TET^{*}_{ipt}, for the first, second and third exported products in the multi-product firms' portfolio.

the NACE rev. 2 core activities of the firms, on which we assume that wholesalers and retailers (NACE 45, 46 and 47) work as trade intermediaries in our data. It is interesting to see that, in the case of manufacturing firms, the first exported products sell about 0.73% more, whereas the second exported products sell an almost negligible 0.03% less after the CETA. When we look at trade intermediaries, we confirm that the impact on exported products is, on average, higher, but we still find positive albeit minor effects on second and third products.

Finally, we believe previous results are in line with a mechanism of

multiproduct if we follow the definition we introduced, and they are excluded from this part of the analysis.

portfolio adjustment predicted by trade theory, as in Mayer et al., 2014 and Eckel and Neary, 2010. According to trade theory, liberalization events also entail more competition in an export market. More firms can access the Canadian market, and competitive pressure induces exporters to concentrate their efforts on their best-performing products, thus focusing on their core competencies. Our findings are also confirmed by a quick check on aggregate flows. According to our data, after trade liberalization between Canada and France with CETA, the first products by French exporters concentrated about 77% of the total firms' exports, which is an increase with respect to a share of 70% registered just before the treaty signature.

2.5.3 General equilibrium trade impacts

Our approach allows us to consider destinations different from Canada and, hence, to test whether CETA has brought about any trade diversion effects. In fact, the product-level matrix we designed in Figure 10 included fifteen alternative destinations, of which ten top partners of France and the rest are continental aggregates[24], while we always have considered the treatment to be product-specific, to have the possibility to evaluate what happens in the destinations alternative to Canada. As a consequence, our matrix completion algorithm returns us counterfactuals on sixteen destinations, including Canada, and we can check the treatment effects on the treated, TET^{*}_{pdt}, for each HS 6-digit product p exposed to the CETA, which is exported to a destination d different from Canada in time t.

The mechanism is that any trade liberalization event, including CETA, changes the distribution of relative costs incurred by exporters. A tariff decrease in Canada increases the relative cost of exporting to other destinations. This is especially true when we are in the presence of bigger

[24]The complete list is reported in Table B1.3. Alternative trade destinations have been picked considering a combination of two ranks: export values and numbers of exported products.

exporters, which have the possibility to adjust their portfolio of destinations once they internalize the new distribution of relative costs across the globe. Eventually, this is the classical Vinerian diversion effect Viner (1950), whereby trade between parties to a PTA partially substitutes for trade between partied and third countries.

We test this mechanism in Table 12, where the dependent variable is the treatment effects on the treated expressed in monetary values, TET_{pdt}, where d is different from Canada. This time, we consider treatment effects in monetary values because we want to check whether there is a correlation in the magnitudes with the treatment effects on the treated in Canada, $TET_{CA,pt}$. Our coefficients of interest are, indeed, on the first row. When we control for the initial value of the trade flows in the alternative destination (column 2), we find a negative association equal to 1.042 between the export change in Canada and the export changes of the same products in the alternative destinations. This association is robust to the inclusion of a double clustering of errors by country and by product classes (column 3). Notably, when we separate between products by their trade elasticity sourced from Fontagné et al., 2022, we discover that the association is mainly driven by the most elastic products (column 5), i.e., the ones whose elasticity value is above the median computed on the entire distribution. Briefly, export flows of products listed by the CETA adjust in alternative destinations as a consequence of the expected general equilibrium effects. We believe the latter is a powerful result that confirms the existence of mechanisms of reallocation on a global scale, as in the case of trade diversion, to take into account the changing distribution of relative trade costs after a liberalization event.

2.6 Robustness and sensitivity checks

Our first concern is that products could have been endogenously selected by the parties during the treaty negotiations, and we may pick a positive

Table 12: CETA and alternative destinations - general equilibrium trade effects

Dependent variable TET_{pdt}	(1)	(2)	(3)	(4)	(5)
$TET_{CA,pt}$	-0.552	-1.042**	-1.042***	-0.101	-1.745***
	(0.449)	(0.437)	(0.364)	(0.154)	(0.639)
$Value_{pdt-1}$		0.019***	0.019***	0.006***	0.019***
		(0.004)	(0.002)	(0.002)	(0.001)
Constant	52,182.54***	-56,199.12**	-56,199.12***	-10,115.56*	-51,568.57***
	(11,497.63)	(24,598.21)	(17,027.99)	(4,845.29)	(16,505)
N. obs.	31,758	31,758	31,758	15,445	16313
R squared	0.0012	0.8123	0.8123	0.1890	0.8294
Clusters by country	No	Yes	Yes	Yes	Yes
Clusters by product class	No	No	Yes	Yes	Yes
Elasticity of subst.	All	All	All	below median	above median

Note: The Table shows results after a linear regression model whose dependent variable includes the treatment effects on the treated in monetary values, TET_{pdt}, where destination d is different from Canada. The main regressor of interest is the vector of treatment effects on the treated in monetary values, TET_{pdt}, where destination d is instead Canada. The unique control variable is the value of the product p export flow in destination d different from Canada in the period before the CETA, $t-1$. Errors are double-clustered by country and product class. Trade elasticity is sourced from Fontagné et al., 2022 **, *** stand, respectively, for $p < 0.05$, $p < 0.01$, $p < 0.001$.

impact just because selected products already showed a higher trade potential. Clues of an endogenous selection into the treaty were offered in Table 4. Products in the CETA were already exported by a greater number of French firms, more frequently, with a lower average transaction value and a lower average value dispersion. To address this concern, we conduct a placebo test by replicating the matrix completion analysis using the same definition of treated products as in the baseline, but for the period September 2012-August 2015. In Appendix Table B1.4, we report no significant effect, and we argue that this is supporting evidence for our empirical approach, which is capable of handling cherry-picking selections into the treaty.

A second concern is that specific matrix configurations can drive different results. The concern is specifically relevant to the validity of our findings on trade diversion when we search for possible general equilibrium effects. In this case, we test different configurations for how destinations alternative to Canada are included in the baseline matrix. In

Appendix Table B1.5, we show results when:

1. we consider the popularity of alternative destinations classified by the number of French exporters that serve them;
2. we adopt a measure of import structure similarity to Canada, computed considering the sums of the absolute values of the distances between the share of each product p in destination d and the corresponding share of imports in Canada;
3. we select destinations based on the size of their import market.

Interestingly, the baseline estimates of the WATET for the products' intensive margin consistently fall in an interval $[0.94, 1.22]$, which is only slightly lower than our baseline estimates at 1.28%. Importantly, Appendix Table B1.6 confirms also the robustness of general equilibrium effects when we select destinations based on either the number of French exporters or the size of the import market. When we consider similar import structures to Canada, the coefficient of interest is not statistically significant anymore, and we argue that it makes sense because the selected destinations are less relevant for French exporters. Notably, none of the alternative matrix configurations[25] achieved the same level of prediction performance as our baseline, as shown in Appendix Table B1.7. For this reason, we prefer to keep our baseline matrices. A third concern is that results are driven by the specific choice of a matrix completion algorithm. As we discussed in Section 2.4, the main difference between the algorithm that we adapt from Athey, Bayati, et al., 2021 and standard proposals in computer science literature (E. Candes & Recht, 2012; E. J. Candes & Plan, 2010) is the inclusion of vectors of fixed effects before proceeding with the singular value decomposition. In our case, we remove the vector of firm-level fixed effects, and we find that the prediction performance slightly worsens. We do not see a fundamental change in the results, but we prefer to keep our baseline results.

[25]The list of alternative destinations by each selection strategy is reported in Appendix Table B1.8.

Finally, we investigate what happens when we change the definition of treated firms. In our baseline, a multiproduct firm is treated when it exports at least two products in Canada and, among them, at least one is enlisted by the CETA. Briefly, by our definition, we have some treated firms with a portfolio that encompasses both products that have seen a regime change and products that have not. If we change our definition and consider as treated only those firms that export at least two products all enlisted by the CETA, what we observe is that the sample shrinks dramatically to the point that it is not representative anymore. In fact, we have that 41% of multiproduct firms usually have in their portfolio both product types; they are usually bigger exporters, and we would introduce a major sample selection. For this reason, we conclude that results with a different definition of treated firms cannot be trusted.

2.7 Conclusions

The present work proposes a novel approach to evaluating the impact of trade agreements using a causal machine learning framework. The aim is to provide a robust empirical strategy capable of handling the complexities and heterogeneity of trade effects at both the product and firm levels while mitigating concerns about endogenous selections into trade agreements. As a case study, we consider the entry into force of the EU-Canada Comprehensive Economic and Trade Agreement (CETA) and adapt an algorithm proposed by Athey, Bayati, et al. (2021) to the case of French customs data. The main advantage is that we can predict multidimensional counterfactuals at the firm, product, and destination levels and, thus, obtain consistent estimates of causal effects.

Findings reveal an average small albeit statistically significant positive impact of the CETA on the product-level intensive margin in the year after the CETA. The Weighted Average Treatment Effects on the Treated (WATET) is 1.28%. Yet, product-level heterogeneity of the impact is relevant, and we show how the full distribution of treatment effects needs to be evaluated. Notably, we find that the impact is higher on those products for which France showed a comparative advantage before the Treaty. On the extensive margin, we record a product churning due to the treaty, which goes beyond the numbers of regular entry-exit dynamics. Due to the CETA agreement, there is a 13.1% of products not exported before that substitute 11.9% of products that are no longer exported. Interestingly, entering and exiting products are those that are more responsive to trade cost changes, i.e., whose trade elasticity is higher. At the firm level, we test the case of multiproduct firms. Consistent with the mechanism of portfolio adjustment predicted by Mayer et al. (2014), we observe that multiproduct exporters reallocate shares towards their first and most exported product, possibly due to an increasing local market competition after trade liberalization. Finally, our empirical strategy is suitable for capturing general equilibrium effects. Indeed, when we look at alternative destinations, we show that CETA induces trade diversion. As the trade treaty makes destinations different from Canada relatively more costly, product flows are partly redirected from other destinations towards Canada.

In conclusion, we believe we showed the validity of a matrix completion approach in evaluating changing trade policies. We believe the same approach can be adapted in the evaluation of other trade policy actions. The main advantage is the possibility of predicting multidimensional counterfactuals as cells of a well-designed matrix, thus returning a more complete picture of the heterogeneity of the impact of trade regime changes, including general equilibrium effects from different policies in destinations that are not parties to trade agreements.

Chapter 3

A dose-response function for learning-by-exporting

3.1 Introduction

Previous literature has extensively studied the relationship between productivity and exporting status. The main challenge was to unravel reverse causality and check which mechanism prevails. On the one hand, there is a self-selection mechanism into the exporting status, by which only the most productive firms can reach foreign markets because beachhead costs are relevant (Bernard & Jensen, 1999; Bernard et al., 2007, 2012; Melitz, 2003; Melitz & Ottaviano, 2008; Roberts & Tybout, 1997). On the other hand, there is a mechanism of learning by exporting, by which a firm's productivity improves after entering a foreign market thanks to knowledge spillovers coming directly from buyers or through increased competition from foreign producers (Atkin et al., 2017; J. R. Baldwin & Gu, 2003; Clerides et al., 1998; Crespi et al., 2008; De Loecker, 2013; Liang et al., 2024).

Our perspective is different. Our aim is to isolate the effect of export intensity on a firm's performance from the self-selection mechanism of exporting due to firm heterogeneity. We adopt a potential outcome framework and estimate a dose-response function by assuming that exporting intensity represents a continuous treatment that has an impact on the firm's productivity. Briefly, we test whether firms react heterogeneously to different levels of export intensity after they already decided to export.

Our main intuition is that firms' productivity may benefit from exporting only after reaching some capacity. When export intensity is low, a firm still needs to establish the necessary absorptive capacity and the logistical organization needed to reap productivity gains from foreign markets. After export activity increases, firms streamline production processes to remain competitive in foreign markets, eventually registering efficiency gains. Consequently, the full benefits from exporting, i.e., the channel of *learning by exporting*, is activated after a minimum threshold of export intensity.

Our hypothesis is confirmed after we investigate exports and firm-level outcomes of French firms in the time interval 2010-2018. In particular, we estimate a dose-response curve following Cerulli (2015), where export intensity is administered as a dose of treatment to firms that have already decided to export.

As expected, the typical dose-response curve shows that the relationship between export intensity and firms' productivity is nonlinear. Notably, its shape and the corresponding shapes of the impact of export intensity on firms' costs and revenues unveil some interesting dynamics that help explain the rationale for an minimum threshold. We find that a firm can appreciate the benefits of learning by exporting only with an intensity at least equal to 60%.

We argue our findings point to exporting as a crucial source of additional productivity gains, on top of self-selection into exporting, only when a firm builds efficient logistics to maintain beachheads in export destinations, on the one hand, and absorptive capacity to internalize knowledge spillovers, on the other hand. Both activities require additional efforts that exporters may find inconvenient below a critical mass.

Indeed, our results corroborate the idea that exporting is inconvenient when export activity falls in an unstable intensity interval, in which firms either prefer to transit to a higher level of exporting or drop from active exporting. We define the latter interval *low productivity trap*.

Finally, we delve into industry heterogeneity by focusing on the vertical linkages and trajectories of technological change that exporting firms may have. For our purpose, we adopt the seminal Pavitt, 1984 taxonomy, and we find that firms in the category of *Scale and Information Intensive* exhibit clear productivity gains when export intensity is relatively higher, as they have a type of technology for which economies of scale are relevant. Conversely, firms that have *Specialised suppliers* and that are *Science-based* show negligible effects of export intensity on productivity, consistent with the idea that they already operate at relatively high productivity in more competitive markets. Finally, we find that firms active in more traditional manufacturing sectors are heavily influenced by their suppliers, and their benefits from an increased export intensity are more limited.

The rest of the paper is structured as follows. We begin with an introduction to data in Section 3.2. In Section 3.3, we outline our estimation strategy. Results are discussed in Section 3.4, with a focus on the exporting dynamics of the firms in Subsection 3.4.1. Robustness and sensitivity to the technological trajectories are presented respectively in Section 3.5 and 3.5.1. Finally, policy implications and conclusions are sketched in Section 3.6.

3.2 Data and descriptive statistics

We source firm-level information for French exporters in the time interval 2010-2018 from Orbis, by Bureau Van Dijk[1]. In particular, we focus on France as it is a well-explored case study for firm-level trade data, providing a foundation for building upon and confronting previous literature. See, among others, Crozet et al., 2012 and Fontagné et al. (2018).

Our primary variable of interest is a firm's export intensity, which we derive from information about export revenues[2] on the total revenues. Firms' outcomes include Total Factor Productivity (TFP), estimated following Ackerberg et al. (2015), along with sales and costs . See Appendix Table C1.1 for more details on firm-level accounts.

For our purpose, we consider only firms that have engaged at least once in exporting in our analysis period.

Moreover, to remove noise in the relationship between export intensity and firm performance, we keep in our sample permanent exporters, i.e., firms that do not engage in temporary trade once in a while without commitment to foreign markets. Following the definition provided by Békés and Muraközy (2012), a firm needs to export for at least four consecutive years to be considered a permanent exporter.

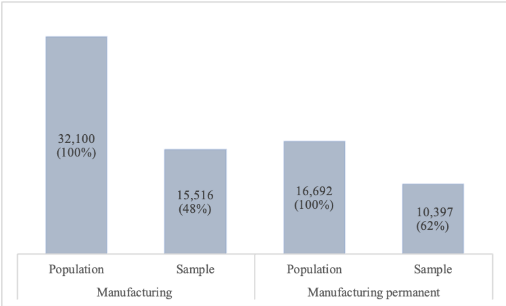
Finally, we eliminate from our sample firms that belong to sectors different from manufacturing. In this way, we do not consider intermediaries in trade, as these are firms that professionally trade on behalf of other firms.

Our final sample encompasses 13,542 manufacturing exporters in the period 2010-2018, distributed heterogeneously throughout our time interval. In Figure 18, we report a snapshot of the sample coverage in 2015, in which we also show the share of permanent exporters. Despite rep-

[1]The Orbis database is a recognized global source for firm-level financial accounts and has been used in previous studies, including G. Gopinath et al., 2017, Cravino and Levchenko, 2016, Del Prete and Rungi, 2017, and Del Prete and Rungi, 2018, Micocci and Rungi (2023).

[2]Interestingly enough, French firms must report revenues from exports separately, as from the subsequently amended *Règlement n. 99-03 du Comité de la réglementation comptable.*

Figure 18: Sample coverage of the number of exporters in the manufacturing sector



Note: The Figure reports the number of exporters in the Manufacturing sector according to INSEE, and those of our sample, split by total and permanent exporters. The data on the population of French exporters are sourced from INSEE (2023).

resenting only the 32% of total exporters in manufacturing, our sample represents the 62% of the population of 'permanent' exporters, i.e., those firms that exported all years from 2011 to 2015.

Once we explore potential heterogeneity, we will make use of a classification that provides an idea about firms' technological trajectories. We follow Pavitt, 1984's classification [3]. In Appendix Table C1.2, we report our sample distribution in the four main classes offered by the latter classification.

3.3 Empirical strategy

In this paper, we aim to investigate the causal impact of export intensity on overall firm performance. To achieve this, we draw on the econometric literature on treatment effects estimation (Imbens & Wooldridge, 2009), with a particular emphasis on dose-response models (Bia et al., 2014; D'Haultfœuille et al., 2023; Hirano & Imbens, 2004; Kluve et al.,

[3] To map the seminal Pavitt, 1984 classification on Nace Rev.2, we use the mapping by Bogliacino and Pianta, 2016

2012).

Dose-response models are particularly well-suited for socio-economic contexts like ours where it is crucial to consider not just the binary treatment status (treated vs. untreated) but also the degree of exposure (or 'dose') experienced. These models allow us to:

1. Go beyond estimating a single average effect by providing an effect as a function (the dose-response function) across different levels of the dose variable.

2. Present results in a clear and intuitive graphical format via the dose-response function plot, making the pattern of the causal relationship more apparent.

3. Examine the entire distribution of the causal effect, thereby improving the precision of the observed treatment effect pattern.

In what follows, we briefly present the model and the notation based on the econometric model developed by Cerulli (2015).

The dose-response framework is based on Rubin's potential outcome equation:

$$y_i = y_{0,i} + w_i(y_{1,i} - y_{0,i}) \tag{3.1}$$

Here $y_{0,i}$ represents the potential outcome for unit i if untreated, $y_{1,i}$ is the potential outcome when treated, and w_i is a dummy variable indicating treatment status. In our framework, the treatment consists of having positive export revenues in $t-1$.

Expanding this equation into a continuous framework, we define t_i as a continuous treatment indicator ranging from 0 to 100. t_i is our measure of export intensity, and it is computed as the ratio between export revenues and total revenues.

The relationship between the export intensity and the outcome of interest depends on $h(t_i)$, a general differentiable function of the export intensity t_i, a function $g(x_i)$ of the M confounders $\mathbf{x}_i = [x_{1,i}, x_{2,i}, \ldots, x_{M,i}]$, and and a set of components depending on the treatment status w_i. Notably, μ_1 and μ_0 are scalars, and e_1 and e_0 are error terms corresponding to random variables with an unconditional mean 0 and constant variance. The population-generating process for the two potential outcomes is expressed as follows (for conciseness, we get rid of index it):

$$\begin{cases} w = 1: & y_1 = \mu_1 + g(\mathbf{x}) + h(t) + e_1 \\ w = 0: & y_0 = \mu_0 + g(\mathbf{x}) + e_0 \end{cases} \tag{3.2}$$

with the function $h(t)$ being nonzero only when a unit is in the treated status. Using this model and defining the treatment effect as $TE = (y_1 - y_0)$, we can define the causal parameters of interest, i.e., the population Average Treatment Effect conditional on $\mathbf{x}$ and t:

$$ATE(\mathbf{x}, t) = E(y_1 - y_0 | \mathbf{x}, t) \tag{3.3}$$

By the law of iterated expectation, the corresponding population unconditional ATE can be obtained as:

$$ATE = E_{(\mathbf{x},t)}\{ATE(\mathbf{x}, t)\} \tag{3.4}$$

We now assume a linear-in-parameters form for $g(\mathbf{x}) = \mathbf{x}\delta$. The *ATE* conditional on $\mathbf{x}$, t, and w becomes:

$$ATE(\mathbf{x}, t, w) = w \times \{\mu + h(t)\} + (1 - w) \times \{\mu\} \tag{3.5}$$

where $\mu = (\mu_1 - \mu_0)$. The corresponding unconditional *ATE* will be:

$$ATE = p(w = 1) \times (\mu + \overline{h}_{t>0}) + p(w = 0) \times (\mu) \tag{3.6}$$

where $p(w = 1)$ is the probability of treatment status, and $h_{t>0}$ is the average of the response function taken over $t > 0$. In this model, the dose-response function is equal to the conditional *Average Treatment Ef-*

fect, given the level of treatment t.

Substituting the potential outcomes in model (3.2) into Rubin's potential outcome equation (3.1), we obtain the following model:

$$\begin{aligned}
y &= y_0 + w(y_1 - y_0) && (3.7)\\
&= \mu_0 + \mathbf{x}\delta + \epsilon_0 + w[(\mu_1 + \mathbf{x}\delta + h(t) + \epsilon_1) - (\mu_0 + \mathbf{x}\delta + \epsilon_0)] && (3.8)\\
&= \mu_0 + \mathbf{x}\delta + w(\mu_1 - \mu_0) + w(h(t)) + \epsilon_0 + w(\epsilon_1 - \epsilon_0) + \mathbf{w}\overline{\mathbf{h}} - \mathbf{w}\overline{\mathbf{h}} && (3.9)\\
&= \mu_0 + \mathbf{x}\delta + w(\mu_1 - \mu_0 + \overline{h}) + w(h(t) - \overline{h}) + \epsilon_0 + w(\epsilon_1 - \epsilon_0) && (3.10)\\
&= \mu_0 + \mathbf{x}\delta + wATE + w(h(t) - \overline{h}) + \eta && (3.11)
\end{aligned}$$

To estimate this model, we assume a three-degree polynomial form for the function $h(t_i)$ and use the fixed effect coefficient regression to estimate:

$$\ddot{y}_{it} = \alpha_0 + \ddot{\mathbf{x}}_{it}\delta_0 + \ddot{w}_{it}ATE + \ddot{w}_{it}[a\ddot{T}_{1it} + b\ddot{T}_{2it} + c\ddot{T}_{3it}] + \ddot{\eta}_i \qquad (3.12)$$

Here, each variable $\ddot{v}$ is computed as $v_{it} - \overline{v}_i$, i.e. as a deviation form the individual mean of the period $\overline{v}_i = \sum_t v_{it}/t$. Note that, in this framework, $\ddot{w}$ represents difference between the exporting behaviour of the unit at time t and its propensity to export in the entire period. For firms exporting all years, $\ddot{w} = 0$, thus meaning the only effect on productivity is given by the variation in export intensity. On the other hand, the productivity of intermittent exporters will be affected both by the exporting activity as such, and by the export intensity. Finally α_0 is the average value of the fixed effects, i.e., the grand average of y across all units and all periods, while $T_j = t^j - E(t^j)$ for $j = 1, 2, 3$.

Under the hypothesis of *Conditional Mean Independence*, an OLS estimation of equation (3.12) produces consistent estimates of the parameters, that is: $\hat{\delta}_0$, $A\hat{T}E$, $\hat{a}$, $\hat{b}$, $\hat{c}$. With these parameters at hand, we can finally estimate the dose-response function as:

$$\begin{aligned}
A\hat{T}E(\ddot{t}_{it}) = & w\left[A\hat{T}E_{t>0} + \hat{a}\left(\ddot{t}_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{i=1}^{T}\ddot{t}_{it}\right) + \hat{b}\left(\ddot{t}_{it}^{2} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{i=1}^{T}\ddot{t}_{it}^{2}\right)\right. \\
& \left. + \hat{c}\left(\ddot{t}_{it}^{3} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{i=1}^{T}\ddot{t}_{it}^{3}\right)\right] + (1-w)A\hat{T}E_{t=0}
\end{aligned} \tag{3.13}$$

A simple plot of the curve $A\hat{T}E(t)_{t>0}$ over the support of t returns the pattern of the dose-response function.

3.4 Results

We now present the results of our analysis, where we estimate the model specified in Equation (3.12). Our primary interest lies in studying the impact of export intensity on Total Factor Productivity (TFP), as computed following Ackerberg et al., 2015.

In our vector of controls $\mathbf{x}$, we include (a) the logarithm of the number of employees to adjust for changes in firm size, (b) the size-age indicator by Hadlock and Pierce (2010) as a proxy for financial constraints, and (c) the presence of patents owned by the firm as a measure of innovation. It is important to emphasize that firm, industry, and regional-specific characteristics are controlled through firm and year-fixed effects. Therefore, the remaining predictors capture variations relative to individual means over the period.

Column (1) of Table 13 shows that the variation in a firm's exporting status from the previous year does not significantly impact productivity. Our sample consists of permanent exporters who have exported for at least four consecutive years within our time frame. Moreover, we employ a fixed-effect model to estimate the marginal contribution of exporting to productivity, which means that variations in export participation are only observed among firms that do not export every year.[4] Being such variations minimal, the lack of significant impact on productivity is to be expected.

Increased financial constraints are linked to reduced TFP, consistent with previous research showing that financial and liquidity constraints hinder R&D investment decisions (Butler & Cornaggia, 2011) and lead firms to forgo profitable investment opportunities (Almeida & Campello, 2007), lowering productivity.

A more surprising result is that an increase in size is associated with decreased TFP. However, TFP results from allocating capital and labor in the production function; an increase in labor alone does not automatically translate into an increase in TFP unless capital is adjusted accordingly. Moreover, the newly hired labor force might need some learning time before becoming fully operative, meaning their impact on TFP might be initially negative, as the increase in sales does not compensate for the rise in total costs.

Finally, the presence of patents owned by the firm positively correlates with its productivity. This finding corroborates similar literature showing how innovative firms are more productive than their non-innovative counterparts.[5]

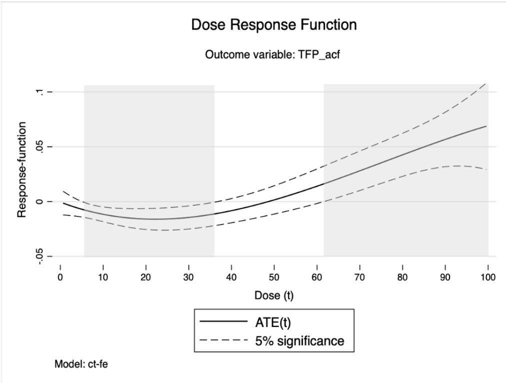
Central to our analysis is the dose-response function illustrated in Figure 19, which is obtained by plugging the coefficients from Table 13 into Equation (3.13) and plotting the resulting curve over the support t.

For export intensities lower than 5%, the impact on firm performance is insignificant. This low export intensity may indicate that the firm is a passive exporter, engaging in one-time shipments in response to foreign orders without establishing a permanent logistical infrastructure. The engagement of these firms in international markets is minimal and does not provide sufficient opportunities for gaining new knowledge

[4]For these firms, we observe a positive value of $\ddot{w} > 0$ when they export and a negative one when they do not.

[5]See Zhao and Jin (2020) for a recent review on the effects of innovation and globalization on productivity.

Figure 19: Dose-response function of export intensity on TFP



Note: The figure reports the DRFs obtained by plugging-in the estimated coefficients in Table 13 in equation 3.13 and plotting it against the support t.The figure shows the relationship between export intensity Total Factor Productivity. The grey highlighted areas identify intervals of export intensity where the DRF is statistically different from zero using a significance level of 5%.

and skills. Additionally, such low export levels are insufficient for firms to benefit from economies of scale, explaining the lack of observable effects on their overall performance.

Export intensities between 5% and 35% significantly negatively affect total factor productivity. This suggests that firms transition into active exporters within this range. Exporting shifts from a one-time event to a structured strategy, distributing sales across international markets. As firms become active exporters, they incur the costs associated with stable entry into foreign markets, such as packaging, upgrading product quality, establishing marketing channels, and accumulating information on demand sources (Roberts & Tybout, 1995). These investments initially have a negative impact on TFP, but are progressively compensated by increased production efficiency as export intensity increases.

Export intensity exceeding 60% marks the point where a firm begins to fully benefit from exporting. At such high levels, exporting becomes a critical driver of profitability, while inducing economies of scale, enhancing a firm's productivity as operational scale expands.

Overall, the shape of the dose-response curve shows that it is after a firm has reached a critical mass of exports that LBE mechanisms start operating. This result aligns with other empirical firm-level studies showing that low to medium levels of export intensity can have either no effect or even a negative impact on firm productivity (Fryges & Wagner, 2008; López, 2005; Van Biesebroeck, 2005).

However, in contrast to these studies, we have identified two distinct export intervals within the lower end of the export intensity distribution. The first interval comprises *pure passive exporters* who do not invest in exporting infrastructure, exporting volumes too small to benefit from economies of scale. For these firms, exporting activities have a negligible impact on productivity.

The second interval, termed the "low-productivity trap," includes firms that begin investing in foreign activities but struggle to reallocate resources efficiently from their core domestic operations to support exports. The additional managerial and operational complexities associated with exporting detract from productivity gains, negatively impacting firm performance.

3.4.1 The low-productivity trap

For firms caught in the "low-productivity trap", it may be challenging to bear the costs of active exporting. In response to a productivity setback, they may gradually reduce their international presence or increase export intensity to capitalize on expanded market demand. Similar to a poverty trap, the "low-productivity trap" explains why firms with low-export intensity continue to export only a small portion of their sales abroad and why the distribution of firms over export intensity remains concentrated at lower levels (see Figure C3.1a).

These findings align with Van den Berg et al. (2022), who suggest that firms must not only meet the productivity threshold to enter export markets, as predicted by Melitz (2003), but also exceed this threshold to remain competitive. Thus, firms must surpass two productivity thresholds to sustain their presence in foreign markets.

Figures 20a-20d corroborate this idea of a second productivity threshold affecting the firm's export intensity. Here, we categorized firms into four groups based on their export behaviors: (a) *low-export*, exporting less than 5% of their sales abroad at time t; (b) *low-productivity trap*, exporting between 5% and 35% of their sales abroad; (c) *high-export*, exporting between 35% and 75% of their sales abroad; and (d) *very high-export*, exporting more than 75% of their sales abroad. Using these categories, we track firms' export behavior over subsequent years.

Firms initially in the low-productivity trap at time t either maintain their export intensity or exhibit divergent behaviors: some decrease to lower levels and eventually exit the foreign markets. In contrast, others significantly increase their exports in subsequent years. Conversely, firms initially in the low-export class exit foreign markets the following years or increase exports by up to 35%. Sustaining export intensities above 35% proves challenging for all but a few. Figures 20a and 20b further illustrate how many firms shift between the low-productivity trap and low-export intensity classes, underscoring the difficulty of achieving the critical export mass required to move beyond this trap. Theoretically, these dynamics align with the model of sequential exporting proposed by Albornoz et al. (2012). According to their framework, firms' internationalization processes are gradual and experiential. Firms use initial export activities as experiments to reduce uncertainty and assess their competitiveness in international markets. Positive export experiences encourage firms to expand their exporting activities, while a poor performance make them more prone to reduce or cease them.

Instead, firms with high export intensity typically sustain or boost their exports over time. Even more so for exporters with very-high export intensity, which demonstrate remarkable resilience to their export intensity: 70% of them remain in the same export intensity class and only 12% export less than 35% of their sales abroad after four years. Solely a fraction of the high-export intensity class initially falls back to the low-productivity class, reducing their export intensity and eventually exiting foreign markets.

What is most interesting is that an export intensity of 35% marks a threshold identifying groups of exporters who tend to persist at their exporting levels: either always below or consistently above the threshold.

This evidence suggests that expanding their international presence may not be sustainable for firms below the threshold due to limited resources for sustaining internationalization processes (Bernard et al., 2011). Consequently, they often experience lower export performance and higher exit rates from international markets, as confirmed by Kneller and Pisu (2007).

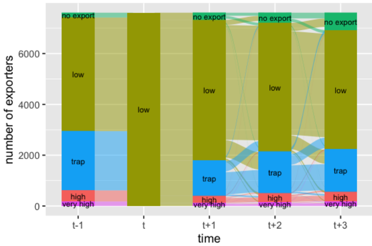
Conversely, the stability observed above this export intensity threshold indicates that once a certain critical mass of foreign activities is reached, maintaining or even increasing export levels becomes profitable. Firms with higher export intensities benefit from economies of scale and scope, leading to increased efficiency and competitiveness in foreign markets (Wang et al., 2022). Furthermore, they are often more productive and innovative, enabling them to adapt to changing market conditions and sustain their international operations (Albornoz et al., 2012).

3.4.2 Channels of productivity increase

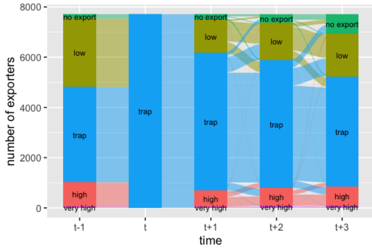
The production process improvement resulting from expanded export intensity can be explained through two mechanisms. On the one hand, competition in international markets drives knowledge spillovers, enabling firms to catch up with technological frontiers. On the other hand, economies of scale from foreign markets enhance productivity by optimizing the sales-to-production cost relationship. Both mechanisms have a visible effect on the relationship between the firm's sales and costs. To

Figure 20: Exporter dynamics for firms in different classes of export intensity

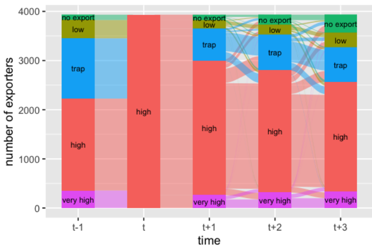
(a) Low-intensity exporters



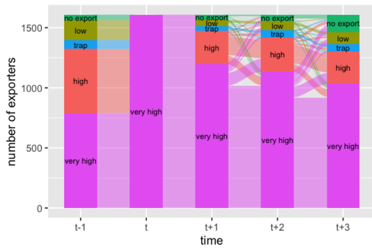
(b) Exporters in low-productivity trap



(c) High-intensity exporters



(d) Very high-intensity exporters



Note: The Figures shows the export dynamics of exporters, which, in time t, were exporting within a certain export intensity class. Figure (a) shows the export dynamics of exporters in the low-exporting class (0-5%); figure (b) show the behaviour of exporters in the export growth trap (5%-35%); figure (c) represents the exporters on the high-export intensity class (35%-75%); figure (d) encompasses firms in the very-high exporting class (75%-100%).

investigate this further, we estimate our model using the firm's total sales and total costs as outcome variables.

Table 13: Regression models

	TFP	Sales	Total Costs
	(1)	(2)	(3)
Variation in export status in (t-1)	0.00032	0.01524**	0.01584**
	(0.00767)	(0.00512)	(0.00492)
size-age	-0.0761***	0.17559***	0.17313***
	(0.00917)	(0.00598)	(0.00574)
log(n. of employees)	-0.38705***	0.32863***	0.35569***
	(0.00661)	(0.00437)	(0.00418)
patents	0.07278**	0.05782***	0.04965***
	(0.02212)	(0.01521)	(0.01462)
T_1	-0.00149	-0.00068	-0.00119*
	(0.00083)	(0.00055)	(0.00053)
T_2	0.00004	0.00004*	0.00006***
	(0.00002)	(0.00001)	(0.00001)
T_3	0.00000	0.00000*	0.00000***
	(0)	(0)	(0)
Constant	11.48258***	15.17938***	14.64527***
	(0.04024)	(0.02638)	(0.02527)
Firm FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
	45,074	49,728	49,731

Note: The table reports the estimates of equation (3.12), using as dependent variables respectively TFP, Sales, Total Costs and profitability computed as *EBITDA/Total Assets*. Sales and Total Costs are in real terms and in logarithmic form.

Columns (2) and (3) of Table 13 present estimates where sales and costs are considered as dependent variables.

The dose-response functions for sales and costs reported in Figure 21 indicate that exporting significantly impacts a firm's operations only when export intensity exceeds 30%. Below this threshold, firms do not experience substantial changes in their activity levels. This finding aligns with the work of Bernard et al. (2012), who demonstrated that firms need to reach a critical mass in export activities to see notable benefits.

For firms where more than 30% of activities are destined for foreign markets, there is a marked increase in both sales and operational volumes. This aligns with the theory of economies of scale, as suggested by Melitz (2003), where increased production for exports leads to higher output.

Interestingly, for export intensities above 75%, we observe a steady decrease in costs while sales levels remain stable. This pattern suggests that firms achieve greater efficiency and cost savings at higher export intensities, possibly due to better optimization of supply chains and production processes tailored for large-scale exports. This phenomenon is supported by Helpman et al. (2008), who found that firms focusing extensively on exports can exploit advanced production techniques and more efficient logistics.

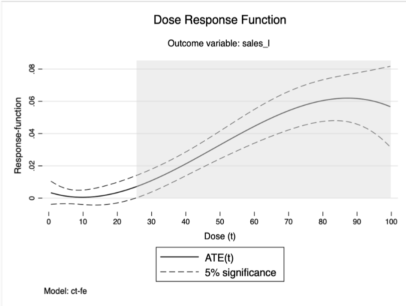
In summary, the observed patterns confirm that significant export activities (above 30% intensity) are necessary for firms to experience substantial growth in sales and activity levels. Moreover, at very high export intensities (above 75%), firms benefit from decreasing costs and stable sales, highlighting the efficiency gains and competitive advantages of a strong export focus.

3.5 Robustness checks

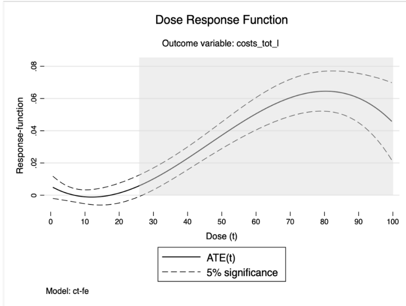
In our baseline estimation, we restricted our sample to permanent exporters to isolate the effect of export intensity from export status. We then checked the robustness of our results by including temporary exporters. Column (1) of Table 14 shows that the estimated coefficients remain robust with the inclusion of temporary exporters, as does the shape of the dose-response function in Figure 22a. However, when we consider only temporary exporters, the effect of export intensity on a firm's productivity completely disappears (see Figure 22b). This finding is crucial as it confirms that the intensive margin of exporting matters only for perma-

Figure 21: Dose-Response Functions

(a)



(b)



Note: The figures report the DRFs obtained by plugging-in the estimated coefficients in Table 13 in equation 3.13 and plotting it against the support t. Figure (a), (b) show the relationship between export intensity and respectively Log of real sales, and Log of Real Total costs. The grey highlighted areas identify intervals of export intensity where the DRF is statistically different from zero using a significance level of 5%.

nent exporters. Temporary exporters, who respond to foreign demand without investing in the necessary infrastructure for stable foreign market entry, do not experience productivity gains from export intensity.

We also examined the duration of the effect of export intensity on a firm's productivity. We ran our models considering further lags in the firm's exporting activity. Figure 22c display the dose-response function for exporting activity in $(t-3)$, where we can see no significant effect after three years. This suggests that most learning-by-exporting occurs immediately after an increase in export intensity. In fact, figures 20c-20d show that once a certain level of export intensity is reached, firms tend to maintain similar levels. The productivity improvements are driven by the process required to reach such levels, with firms reaping most of the associated rewards immediately and then maintaining the reached level of productivity.

Finally, we investigated heterogeneous treatment effects due to changes in the controls of the treated population. We were concerned that firms with varying export statuses might differ in size, growth, and innovation paths compared to those consistently exporting. We interacted the controls with the treatment status and reported the estimated coefficients in Column (5) of Table 14. Excluding financial constraints, the remaining interactions are not statistically significant, and the dose-response function shape remains unaffected by the new controls (see Figure 22d). This confirms that among permanent exporters, there are no significant differences between firms that export every year and those that do not. Therefore, we keep our baseline specification.

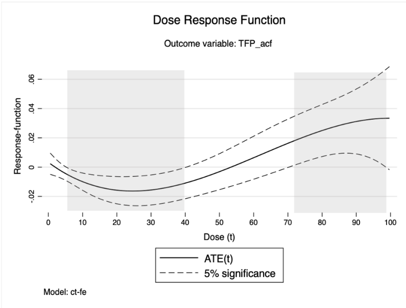
Table 14: Alternative Specifications

	TFP (1)	TFP (2)	TFP (3)	TFP (4)	TFP (5)
Δ export status	-0.001	-0.001	0.008	0	-0.004
	(0.005)	(0.006)	(0.01)	(0.011)	(0.01)
size-age	-0.075***	-0.076***	-0.089***	-0.083***	-0.111***
	(0.008)	(0.016)	(0.011)	(0.012)	(0.014)
log(n. of employees)	-0.383***	-0.367***	-0.425***	-0.46***	-0.371***
	(0.006)	(0.011)	(0.007)	(0.008)	(0.011)
patents	0.077***	0.165***	0.066*	0.081*	0.081*
	(0.021)	(0.046)	(0.027)	(0.03)	(0.034)
Δ export status#size-age					0.036***
					(0.01)
Δ export status#log(n. of employees)					-0.021
					(0.019)
Δ export status#patents					-0.026
					(0.026)
T_1	-0.002*	-0.001	0	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
T_2	0*	0	0	0	0
	(0)	(0)	(0)	(0)	(0)
T_3	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)
Constant	11.356***	11.001***	11.63***	11.813***	11.359***
	(0.035)	(0.064)	(0.046)	(0.05)	(0.066)
Firm FE	YES	YES	YES	YES	YES
year FE	YES	YES	YES	YES	YES
Lag export	t-1	t-1	t-2	t-3	t-1
Eporters	All exporters	Only temporary	Only permanent	Only permanent	Only permanent
(N)	60,183	20,818	34,878	30,682	39,365

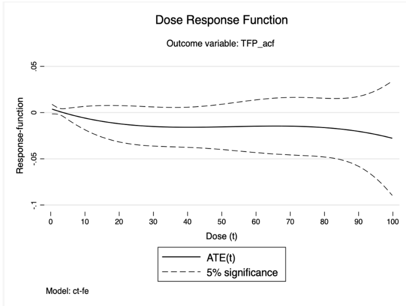
Note: The table reports the estimated coefficients we obtain in different specifications. Column (1) reports the estimated model when we include all exporters, while Column (2) includes the estimates for temporary exporters only. In Columns (3) and (4) we used as treatment the variation in export status and export intensity in (t-2) and (t-3), respectively. Finally, Column (5) reports the estimates we obtain when we control for heterogeneous effects among exporters exporting all years and the remaining exporters.

Figure 22: Dose-Response Functions - Alternative specifications

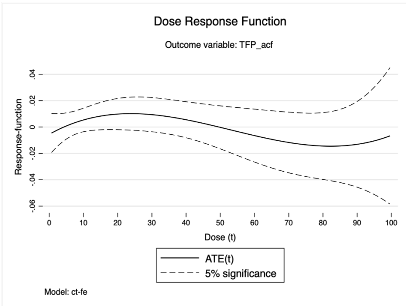
(a) All exporters



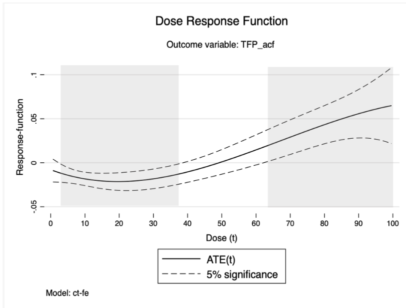
(b) Temporary exporters only



(c) Treatment in t-3



(d) Including heterogeneous treatments



Note: The figures report the DRFs obtained by plugging-in the estimated coefficients in Table 14 in equation 3.13 and plotting it against the support t. Figure (a) reports the estimated model when we include all exporters, while Figure (b) includes the estimates for temporary exporters only. In Figure (c) we use as treatment the variation in export status and export intensity in (t-3). Finally, Figure (d) reports the estimated DRF we obtain when we control for heterogeneous effects among exporters exporting all years and the remaining exporters. The grey highlighted areas identify intervals of export intensity where the DRF is statistically different from zero using a significance level of 5%.

3.5.1 Heterogeneity across technological trajectories

We now want to dig deeper into the sensitivity of our results to the technological trajectory of the firm. We classify the firms according to the Pavitt Taxonomy while following the mapping to the Nace Rev.2 classification by Bogliacino and Pianta (2016) and repeat the previous analysis within each of the Pavitts' classes. Such an exercise allows us to investigate the possible patterns of LBE further.

Pavitt Taxonomy categorizes industrial firms according to sources of technology, requirements of the users, and appropriability regime (Pavitt, 1984). It consists of four categories of industrial firms: (a) *Supplier dominated*, the most traditional manufacturing industries relying on sources of innovation external to the firm (ex. textiles, footwear, food and beverages, paper and printing, and wood); (b) *Scale-intensive*, mainly characterized by large firms producing basic materials and consumer durables for which sources of innovation may be both internal and external to the firm, with a medium level of appropriability (ex. basic metals, motor vehicles, trailers, and semi-trailers); (c) *Specialized suppliers*, which are smaller, more specialized firms producing technology to be sold to other firms. Here, there is a high level of appropriability due to the tacit nature of the knowledge (ex., machinery and equipment, office, accounting, and computing machinery, and medical, precision, and optical instruments); (d) *Science-based* are high-tech firms which rely on R&D from both in-house sources and university research. They have a high degree of appropriability from patents, secrecy, and tacit know-how (ex., chemicals, pharmaceuticals, and electronics).

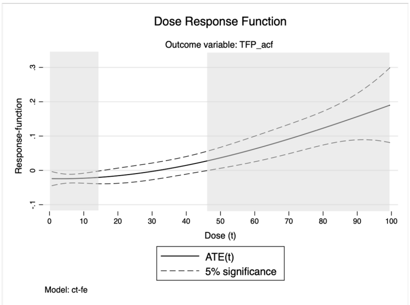
When we replicate our analysis isolating firms within the same Pavitt's class, the results on TFP are quite heterogeneous. The regression table is reported in Table 23.

Results for *Scale and Information Intensive* firms align perfectly with those we discussed in section 3.4. Their productivity steeply increase for high values of export intensity, now for export intensity values above 45%. Here, the mechanism behind the LBE is clearly one of economies of scale. For these firms, an expanded presence in the international markets implies opportunities for cost-cutting technical change, reflected in increased factor productivity.

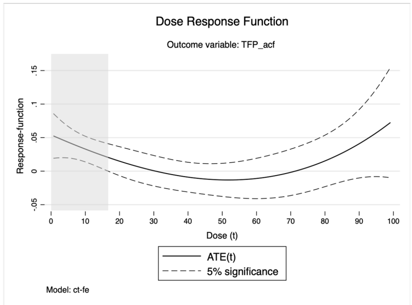
Specialised suppliers seem to benefit from exporting only for low-medium

Figure 23: Dose-Response Functions for TFP in different Pavitt's classes

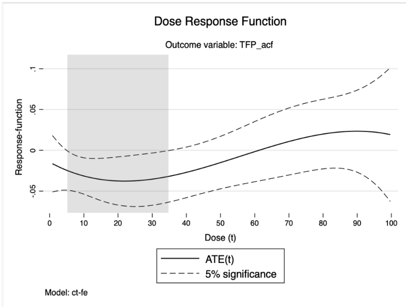
(a) Scale-intensive



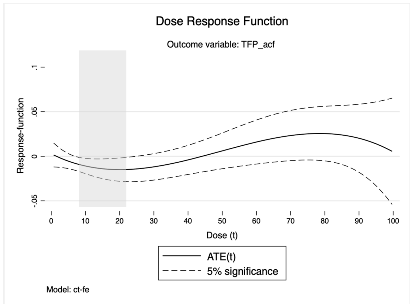
(b) Specialised Suppliers



(c) Science-Based



(d) Supplier Dominated



Note: The figures show the relationship between export intensity and Total Factor Productivity (TFP) for firms in different Pavitt's classes.

export intensities. However, it is essential to note that the relationship between export intensity and firm performance for these firms is theoretically less straightforward than for other exporters. A critical aspect to consider is the nature of the products they offer. Specialized suppliers primarily deal in capital goods, which inherently possess higher values and often require customization or specific configurations to meet the importing businesses' needs. This customization journey can be laborious and resource-intensive, imposing practical limitations on the volume of feasible exports. The technological trajectories for specialized suppliers are more strongly oriented towards performance-increasing product innovation rather than cost-reducing process innovation (Pavitt, 1984). Moreover, these firms often target niche markets where their unique expertise and product offerings are highly valued. While these markets may offer lucrative opportunities, they may not support high export volumes due to their specialized nature and limited demand. These considerations explain why the productivity of specialized suppliers is less affected by export intensity: the gains acquired by entering foreign markets and facing foreign competition are independent of the export share.

For *science-based* firms, the positive effect of high export intensity on TFP is rather small and for export intensity above 80%. At the same time, there is a consistent negative effect in the whole export growth trap. According to Pla-Barber and Alegre (2007), traditional internationalization theories may not apply to science-based industries. Instead of evolving through a series of international stages, firms in science-based industries will likely encounter global pressures much earlier. Moreover, specific features of these firms point to a lack of critical mass that cancels out the benefits of economies of scale (Khilji et al., 2006). Whatever innovation, which is the main source of a firm's productivity in these sectors, generally takes place well before the firm enters foreign markets. Science-based industries are highly globalized, with research teams having a scientific reputation and frequenting international conferences and scientific meetings (Elmes & Kasouf, 1995). It should then come with no surprise that the export intensity plays almost no role in firm productivity.

Firms in more traditional manufacturing sectors, whose technological trajectory is strongly influenced by their suppliers, do not benefit in terms of either profit margins or Total Factor Productivity (TFP) from increasing export intensity. On the contrary, the low-productivity trap seem to be a major concern for this subset of firms. In industries where suppliers dominate, design and productive efficiency investments are the primary channels to increase firm productivity. However, such industries are populated mainly by Small and Medium Enterprises (SMEs), as evidenced by our sample, where they represent 72% of firms in this category (See Appendix Table C1.3). According to Love and Roper (2015), SMEs often encounter particular behavioral, cultural, and resource-related challenges that impede their ability or willingness to engage with design as part of their innovation activity. Additionally, they may fail to grasp the potential value of design for innovation success. Furthermore, as highlighted by Gkypali et al. (2021), once SMEs have surpassed the productivity threshold necessary to enter foreign markets and aim to preserve and enhance their competitive position, they must leverage knowledge flows from learning-by-exporting. This process enables them to upgrade and diversify the quality and variety of their products to align with the needs of both domestic and foreign customers. Nonetheless, this is not a straightforward process. Particularly from a short-term perspective, the adjustment and marketing costs required to promote both old and new products may disrupt current and future business planning, thus resulting in a decrease in productivity.

In conclusion, by categorizing firms according to the Pavitt Taxonomy, we highlight significant heterogeneity in the impact of export intensity on key performance indicators. This underscores the need to recognize and account for the specific characteristics of each firm when assessing the consequences of heightened export activity.

In particular, our results suggest that a one-size-fits-all approach is inadequate when studying the implications of export intensity on firm performance. Recognizing the diversity in technological trajectories and corresponding strategies is essential for policymakers, industry practitioners, and researchers aiming to formulate targeted interventions and strategies to foster economic growth and competitiveness.

3.6 Conclusions and policy implications

The present work studies the impact of a firm's export intensity (the proportion of sales exported) on its performance metrics. We utilize a dose-response model to estimate how various levels of export intensity affect firm' productivity in subsequent years.

Our findings reveal a nonlinear relationship between export intensity and firm productivity. Exporting firms do not immediately experience benefits from increased export intensity; significant rewards are only observed when export intensity surpasses 60%. Beyond this threshold, profits and productivity notably escalate as exporting becomes a primary revenue source. Conversely, very low export intensities below 5% exhibit minimal impact on production processes, likely reflecting passive exporting behaviors. Additionally, we identify a "low-productivity trap" within the range of 5-35% export intensity. Within this export intensity interval, exports negatively affect productivity, as firms allocate resources towards exporting infrastructure without corresponding increased returns.

When delving into the exporting behaviour in firms in different intervals of export intensities, we show that the threshold of 35% identifies groups of exporters who then stick to the same levels of foreign activities in the following years. Firms exporting below this threshold hardly mange surpass such crucial exporting mass in the following years, thus maintaining a low-medium level of export intensity. On the other hand, firms above this threshold consistently keep exporting with higher export intensities in the following years. This evidence supports the notion that LBE mechanisms only take place after a certain level of foreign activities has been reached, while exporting can be irrelevant or even detrimental to firm productivity before.

The analysis of the relationship between costs and sales along the export intensity distribution further identifies, for export intensity values higher than 75%, the specific benefits coming from economies of scale.

When categorizing firms according to Pavitt's Taxonomy, we elucidate how the impacts of export intensity on firm performance vary significantly based on a firm's sector and technological trajectory. The results of our analysis suggest several crucial policy implications.

Firstly, there exists a strong nonlinear relationship between export intensity and firm performance. Exporting firms do not immediately experience benefits from increased export intensity, and significant rewards are only observed when export intensity surpasses 60%. Consequently, firms need to invest consistently in accumulating exporting capabilities to be able to sustain increased export intensity. This is crucial, especially for lower-intensity firms, which may struggle to attain the necessary export intensity to benefit from the learning-by-exporting (LBE) phenomenon. [6]

Secondly, export intensity is not a relevant source of productivity for all firms. Specialised suppliers and firms with science-based technology, do not primarily innovate through exports; the innovation process typically concludes before entering foreign markets. In the case of specialised suppliers, the innovation is rooted in the human know-how and the learning process happens at the stage of the specific customization planning, rather than the exporting activity. On the other hand, the global nature of science-based firms encourages international collaborations during product development, and there is generally minimal need for customization.

These outcomes underscore the inadequacy of a "one-size-fits-all" policy approach and emphasize the necessity for targeted support tailored to firms' technological trajectories and export strategies.

[6] In Appendix C we show the results of a common support analysis, which, indeed, reveals structural disparities between very-low exporting firms and their more export-oriented counterparts.

Conclusions

This thesis investigates the application of novel empirical models, leveraging machine-learning techniques and dose-response models, to address key challenges in international economics. Specifically, it examines three crucial issues: predicting export potential, assessing the impact of trade agreements, and understanding productivity gains from exporting. Each chapter not only delves into these topics but also contributes to a broader narrative on how advanced analytical methods can enhance our understanding of complex economic phenomena and guide effective policy-making.

The first chapter delves into predicting firms' exporting ability, highlighting the power of financial data and sophisticated statistical learning models in identifying potential exporters. Among the algorithms tested, the Bayesian Additive Regression Tree with Missingness In Attributes (BART-MIA) model achieved the highest accuracy, up to 90%, particularly adept at handling missing data from smaller firms. Importantly, the endogeneity of the predictors, which would pose challenges in traditional econometric approaches, actually enhances our models by revealing how closely a firm resembles a successful exporter. Moreover, among the algorithms tested, tree-based models proved most precise, highlighting the complex non-linear interactions among firm characteristics

Our predictions hold robustly across different definitions of exporters and training strategies, offering valuable insights for trade promotion programs, trade credit assessments, and firms' trade potential evaluations. Indeed, governments often formulate trade policies aimed at promoting exports to boost economic growth. Notable examples include Germany's Euler Hermes and France's Bpifrance Assurance Export, which provide export credit insurance and guarantees, and UK Trade & Investment (UKTI) and Enterprise Ireland, which offer support services to expand businesses' exports. Predicting which firms are likely to become exporters can significantly help these programs to target their support more effectively, while better allocating their resources and reducing wastage.

For instance, our study revealed significant regional heterogeneity in trade potential across France, indicating that targeted policy interventions could effectively enhance trade promotion efforts.

The second chapter employs a causal machine learning framework to evaluate the heterogeneous effects of the EU-Canada Comprehensive Economic and Trade Agreement (CETA) on France's trade. We adapt a Matrix Completion model, initially proposed by Athey, Bayati, et al. (2021), to the context of French customs data, predicting multidimensional counterfactuals at the firm, product, and destination level.

This approach exposes the heterogeneity in the impacts of a trade agreement, thus emphasizing the importance of evaluating the entire distribution of treatment effects. We identify both positive and negative effects, notably observing that products in which France held a comparative advantage before the treaty experienced a more pronounced positive impact. Additionally, significant product churning was observed due to the CETA provisions, with new products entering the export market as others phase out. These diverse treatment effects highlight the limitations of analyses focusing solely on average effects, where instances of positive and negative treatment effects may cancel each other out.

Furthermore, our methodology allows for the evaluation of spillover effects. In our analysis, these manifest as classical Vinerian trade diversion effects induced by CETA, with trade flows redirecting toward

Canada from other destinations, especially for products with a higher elasticity of substitution.

Importantly, the matrix completion approach discussed in this chapter is adaptable to other trade policy evaluations, offering the potential for a more comprehensive understanding of trade regime impacts. For instance, continuously updating trade matrices with new data could enable ongoing assessment of evolving trade patterns and policy impacts over time. This dynamic analysis would help policymakers adjust strategies and interventions effectively in response to changing economic conditions. Moreover, extending matrix completion to conduct sector- or regional-specific analyses of trade policy impacts allows for tailored interventions and support measures to maximize sectoral benefits and address specific challenges arising from trade policy changes. Similarly, a regional perspective could reveal disparities in economic outcomes and guide the allocation of resources to promote balanced development and regional integration.

An additional compelling application lies in using matrix completion to analyze global supply chain dynamics. Representing supply chain networks as matrices—where rows denote suppliers or manufacturers, and columns represent customers or distribution points—enables the inclusion of metrics like transportation costs, lead times, inventory levels, and transaction volumes between nodes. By completing these matrices, it becomes possible to predict demand patterns, optimize inventory levels, and minimize excess inventory costs and stockouts across the supply chain network. Moreover, leveraging insights derived from the completed matrices can facilitate the implementation of contingency plans and resilience strategies, mitigating risks associated with potential disruptions such as supplier failures, natural disasters, or geopolitical events.

The third chapter advances the literature on the "learning-by-exporting" hypothesis, which posits that participating in international trade can enhance firm productivity. We use the dose-response model by Cerulli (2015) to quantify productivity gains associated with varying levels of export intensity while accounting for self-selection into exporting.

Unlike traditional studies that use a binary treatment for a firm's exporting status, our methodology isolates the effects of varying export intensity on firm performance. By treating exporting as a continuous variable, we estimate a dose-response function that captures both the direction and magnitude of effects across different export intensity levels. The rationale is that treatment effects could be heterogeneous and non-linear, with exporting becoming profitable only after a critical mass of exports is reached.

Our findings reveal, indeed, that the relationship between export intensity and firm productivity is non-linear. Significant productivity improvements occur when export intensity exceeds 60%, while lower levels can be either ineffective or even detrimental to productivity. Specifically, we identify a "low-productivity trap" within the 5-35% export intensity range, where exports negatively affect productivity due to resource allocation without corresponding returns. Moreover, we show that firms exporting below the 35% threshold struggle to surpass it in subsequent years, maintaining low-medium export levels, while those above it consistently sustain higher export intensities.

These findings suggest that learning-by-exporting mechanisms become effective only after reaching a certain level of foreign engagement, explaining mixed results observed in models using a binary export status indicator. Moreover, these results have several policy implications. For instance, firms with lower export intensity mostly need capacity-building initiatives to stay profitably in foreign markets. Conversely, high-intensity exporters benefit especially from logistics optimization and market diversification, which capitalize further on economies of scale. Moreover, classifying firms using Pavitt's Taxonomy shows considerable variability in productivity outcomes across different sectors and technological capabilities, highlighting diverse impacts of exporting on firm performance and the need for sector-specific policy interventions.

Together, these essays illustrate how advanced empirical analysis, utilizing machine learning and dose-response methods, offers valuable tools for international trade analysis. They underscore the significance of data-driven approaches in exploring economic outcomes, uncovering insights that traditional methods may overlook. Historically, challenges such as limited computational power, data availability, and analytical techniques constrained the incorporation of large, complex datasets in economic research. However, advancements in technology, data infrastructure, and methodologies have now facilitated the full utilization of these datasets.

Through the application of sophisticated analytical techniques, the three essays demonstrate how leveraging detailed data can unveil the intricate mechanisms shaping trade patterns, firm performance, and the effects of trade policies. This integration not only enhances our comprehension of international trade dynamics but also highlight the potential for innovative methodologies to inform policy decisions and stimulate economic growth.

Appendix A

Supplementary materials for Chapter 1

This Appendix is based on Micocci and Rungi (2023), "Predicting Exporters with Machine Learning", World Trade Review. 2023;22(5):584-607. Available at https://doi.org/10.1017/S1474745623000265.

Appendix A1: Data

Table A1.1: List of predictors

Variable	Description
Corporate Control	A binary variable equal to one if a firm belongs to a corporate group.
Dummy Patents	equal to 1 if the firm issued any patent, and 0 otherwise.

Continued on next page

Table A1.1 – continued from previous page

Variable	**Description**
Consolidated Accounts	A binary variable equal to one if the firm consolidates accounts of subsidiaries
NACE rev. 2	A 2-digit industry affiliation following the European Classification
NUTS 2-digit	The region in which the company is located following the European classification.
Productive Capacity	It is an indicator of investment in productive capacity computed as *Fixed Assets*$_t$/(*Fixed Assets*$_{t-1}$+*Depreciation*$_{t-1}$)
Capital Intensity	It is a ratio between fixed assets and number of employees for the choice of factors of production.
Labour Productivity	It is a ratio between value added and number of employees for the average productivity of labor services.

Continued on next page

Table A1.1 – continued from previous page

Variable	Description
Value Added, Depreciation, Creditors, Current Assets, Current liabilities, Non-current liabilities, Current ratio, Debtors, Operating Revenue Turnover, Material Costs, Costs of Employees, Taxation, Financial Revenues, Financial Expenses, Interest Paid, Number of Employees, Cash Flow, EBITDA, Total Assets, Fixed Assets, Intangible Fixed Assets, Tangible Fixed Assets, Shareholders' Funds, Long-Term Debt, Loans, Sales, Solvency Ratio, Working Capital	Original financial accounts expressed in euro.
Interest Coverage Ratio (ICR)	It is a ratio between EBIT and Interest Expenses, as yet another proxy of financial constraints as in Caballero et al., 2008.
TFP	It is the Total Factor Productivity of a firm computed as in Ackerberg et al. (2015).
Financial Constraints	It is a proxy of financial constraints as in Nickell and Nicolitsas, 1999, calculated as a ratio between interest payments and cash flow
Markup	It an estimate of a firm's markup following De Loecker and Warzynski, 2012.
ROA	It is a ratio of EBITDA on Total Assets for returns on assets.

Continued on next page

Table A1.1 – continued from previous page

Variable	Description
Financial Sustainability	It is a ratio between Financial Expenses and Operating Revenues.
Size-Age	It is a synthetic indicator proposed by Hadlock and Pierce (2010), computed as (-0.737$\cdot$ *log*(*total assets*))+(0.043 $\cdot$ *log*(*total assets*))^2 -(0.040 $\cdot$ *age* to catch the non-linear relationship between financial constraints, size and age.
Capital Adequacy Ratio	It is a ratio of Shareholders' Funds over Short and Long Term Debts.
Liquidity Ratio	A ratio between Current Assets minus Stocks and Current Liabilities.
Liquidity Returns	It is a ratio between Cash Flow and Total Assets
Regional Spillovers	It is a proxy proposed by Bernard and Jensen, 2004 computed as a share of exporting plants out of total plants in a region.
Industrial spillovers	It is a proxy proposed by Bernard and Jensen, 2004 computed as a share of exporting plants on total plants in a 2-digit industry.

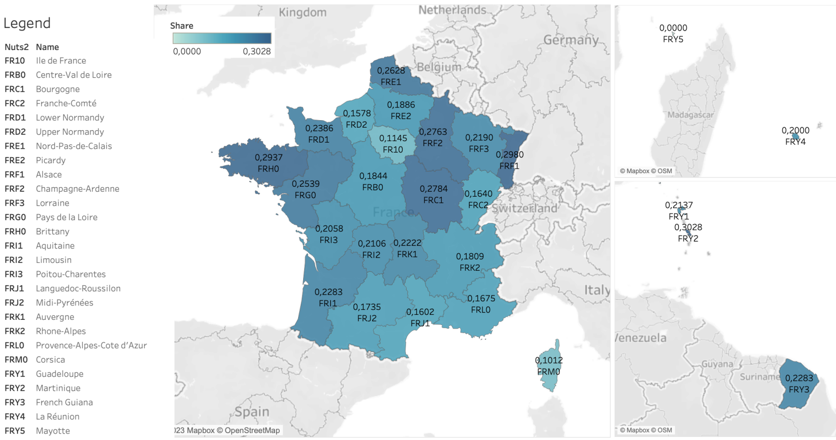
Continued on next page

Table A1.1 – continued from previous page

Variable	Description
External Economies of Scale	It is a proxy proposed by Bernard and Jensen, 2004 computed as a share of exporting plants out of the total in an industry-region cell.
Size	Measure of firm size computed as (log of) number of employees.
Average Wage Bill	It is computed as (log of) costs of employees divided by number of employees.
Inward FDI	It is a binary variable with value 1 if the firm has foreign headquarters and 0 otherwise.
Outward FDI	It is a binary variable with value 1 if the firm has subsidiaries abroad and 0 otherwise.

Appendix A2: Figures and Tables

Figure A2.1: Sample coverage: exporters by region



Note: Unitary shares indicate exporters on total firms in NUTS 2-digit regions.

Table A2.1: Sample coverage by industry

NACE rev. 2	code	Sample				Population			
		non-exporters	exporters	total	(%)	non-exporters	exporters	total	(%)
		(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Food products	10	13,057	1,429	14,486	0.254	49,153	2,135	51,288	0.293
Beverages	11	1,176	395	1,571	0.028	3,028	825	3,853	0.022
Textiles	13	919	389	1,308	0.023	4,278	798	5,076	0.029
Wearing apparel	14	1,060	336	1,396	0.024	8,813	881	9,694	0.055
Leather and related products	15	374	142	516	0.009	2,930	313	3,243	0.019
Wood and products of wood and cork	16	2,203	509	2,712	0.048	8,920	1,036	9,956	0.057
Paper and paper products	17	455	362	817	0.014	823	469	1,292	0.007
Printing and reproduction of recorded media	18	2,995	584	3,579	0.063	14,347	969	15,316	0.088
Coke and refined petroleum	19	17	14	31	0.001	-	-	25	0.0001
Chemicals and chemical products	20	958	705	1,663	0.029	1,388	1,127	2,515	0.014
Pharmaceutical products	21	151	148	299	0.005	93	159	252	0.001
Rubber and plastic products	22	1,436	931	2,367	0.042	1,780	1,425	3,205	0.018
Other non-metallic products	23	1,929	393	2,322	0.041	7,026	777	7,803	0.045
Basic metals	24	354	267	621	0.011	295	304	599	0.003
Fabricated metal prod., except machinery and equipment	25	8,135	2,540	10,675	0.187	14,557	3,903	18,460	0.106
Computer, electronic and optical products	26	965	605	1,570	0.028	1,304	991	2,295	0.013
Electrical equipment	27	789	495	1,284	0.023	1,321	727	2,048	0.012
Machinery and equipment	28	1,938	1,194	3,132	0.055	2,567	1,967	4,534	0.026
Motor vehicle, trailers and semi-trailers	29	748	424	1,172	0.021	1,119	516	1,635	0.009
Other transport equipment	30	330	186	516	0.009	847	260	1,107	0.006
Furniture	31	1,416	249	1,665	0.029	8,758	598	9,356	0.053
Other manufacturing	32	2,796	518	3,314	0.058	19,960	1,378	21,338	0.122
Total		44,201	12,815	57,016	1.00	153,307	21,558	174,890	1.00

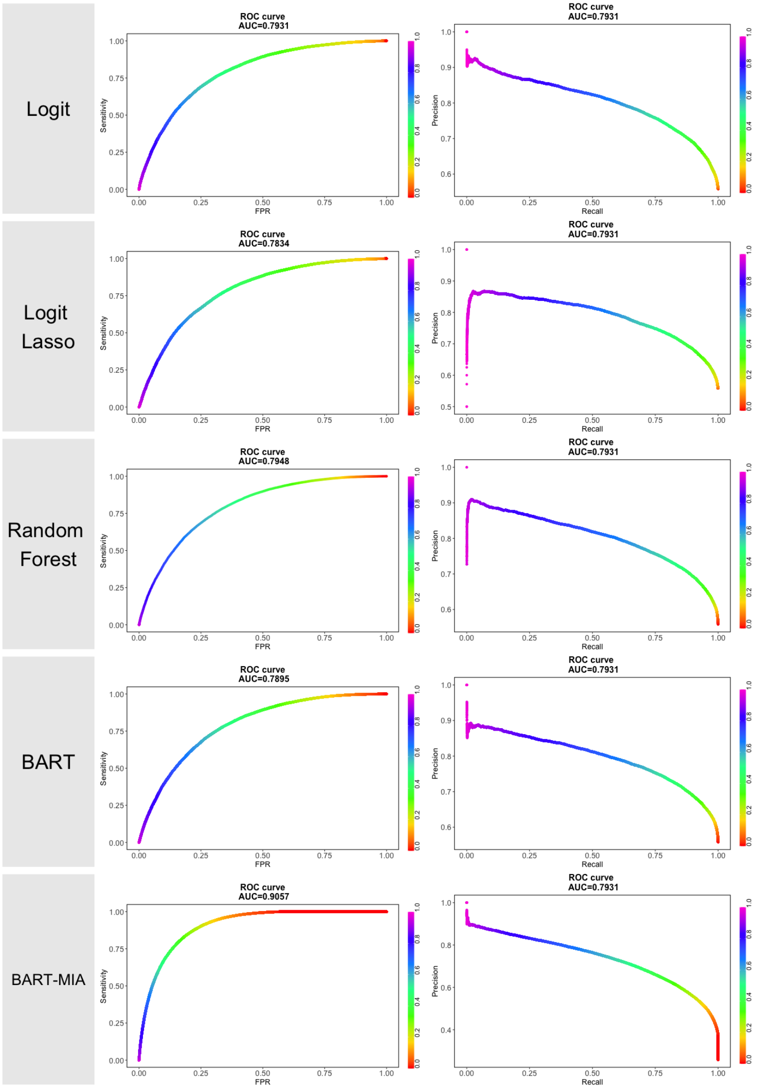
Note: French manufacturing firms are sourced from Orbis, by Bureau Van Dijk. On columns 3 and 4, we separate exporters and non-exporters in our sample. On column 5 we report the total number of manufacturing firms by NACE rev.2. On columns 7-9 a comparison with Eurostat census. When we look at shares on columns 6 and 10, we find our sample is well balanced by industry if compared with the population.

Table A2.2: Sample coverage - size classes

NACE	Sample - N. employees						Population - N. employees					
rev.2	0-9	10-19	20-49	50-249	250+	Total	0-9	10-19	20-49	50-249	250+	Total
10	1,649	711	611	488	172	3,631	45,798	3,225	1,382	679	204	51,288
11	233	105	93	59	21	511	3,397	205	147	76	28	3,853
13	93	76	107	80	7	363	4,586	209	151	113	17	5,076
14	117	51	49	47	22	286	9,391	140	89	57	16	9,694
15	43	24	36	47	16	166	3,038	70	69	45	21	3,243
16	274	182	178	93	8	735	8,869	560	337	168	21	9,956
17	48	64	105	129	39	385	865	123	121	120	62	1,292
18	381	144	167	86	6	784	14,455	445	277	123	17	15,316
19	1	3	4	6	5	19	NA	NA	3	3	7	25
20	134	109	177	223	87	730	NA	NA	190	219	99	2,515
21	16	18	36	58	61	189	NA	NA	31	50	55	252
22	192	173	274	279	53	971	1,963	405	431	319	86	3,205
23	348	135	161	136	59	839	7,094	266	234	136	72	7,803
24	39	33	53	122	51	298	377	60	56	70	35	599
25	988	792	869	571	75	3,295	13,917	2,174	1,498	734	136	18,460
26	134	113	136	154	70	607	1,700	219	157	171	49	2,295
27	106	83	120	123	64	496	1512	169	168	136	63	2,048
28	281	171	320	319	101	1,192	2,983	455	536	399	160	4,534
29	84	62	103	157	98	504	1,092	156	160	152	75	1,635
30	36	22	30	70	41	199	838	57	63	95	55	1,107
31	148	55	78	66	9	356	8,976	164	134	68	13	9,356
32	311	121	108	102	26	668	20,551	394	217	133	44	21,338
Total	5,656	3,248	3,816	1, 091	3,415	17,226	151,402	9,496,	6,451	4,066	1,335	174,898

Note: French manufacturing firms are sourced from Orbis, by Bureau Van Dijk. Sample coverage by number of employees in 2017 (left panel) is compared with information on population sourced from EUROSTAT Structural Business Statistics. Please note that number of employees may report missing values from sample data, thus number of observations do not sum up to sample totals.

Figure A2.2: Out-of-sample Goodness-of-Fit



Note: We report the ROC Curves and Precision-Recall curves of the models. See Appendix A for the details on the construction of the curves and their interpretation.

Table A2.3: Prediction accuracies after cross-validating training and testing sets

Measure	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Sensitivity	0.649	0.647	0.654	0.65	0.648
Specificity	0.911	0.904	0.905	0.905	0.907
Balanced Accuracy	0.780	0.775	0.780	0.778	0.778
ROC	0.909	0.903	0.907	0.903	0.908
PR	0.739	0.738	0.742	0.732	0.739
N.Obs	103,540	102,748	102,169	102,028	101,712

Note: We report prediction accuracies of BART-MIA after cross-validating the algorithm on five different random training and testing sets. Our aim is to check whether predictions are robust against data sampling.

Table A2.4: Prediction accuracies with optimal thresholds (Liu, 2012)

Model	Sensitivity	Specificity	Balanced Accuracy	ROC	PR	Threshold
Logit-Lasso	0.786	0.676	0.716	0.785	0.789	0.513
Logit	0.760	0.688	0.724	0.794	0.805	0.517
Random forest	0.760	0.686	0.723	0.795	0.801	0.560
BART	0.730	0.708	0.719	0.791	0.800	0.569
BART-MIA	0.863	0.791	0.827	0.905	0.738	0.280

Note: We report prediction accuracies when we select the optimal prediction threshold following Liu, 2012.

Table A2.5: Prediction accuracies with a subset of predictors

Model	Sensitivity	Specificity	Balanced Accuracy	ROC	PR
Logit-Lasso	0.668	0.768	0.718	0.786	0.785
CART	0.512	0.907	0.710	-	-
Random forest	0.810	0.627	0.719	0.791	0.793
BART	0.807	0.629	0.718	0.790	0.791
BART-MIA	0.623	0.914	0.768	0.902	0.725

Note: We report prediction accuracies after reducing the battery of predictors from 52 to 23 variables selected by a robust LASSO (Ahrens et al., 2020).

Table A2.6: Prediction accuracies after training and testing on separate years

Measure	2011	2012	2013	2014	2015	2016	2017	2018
Sensitivity	0.907	0.896	0.885	0.896	0.901	0.918	0.924	0.928
Specificity	0.637	0.632	0.641	0.627	0.639	0.651	0.652	0.654
Balanced Accuracy	0.772	0.764	0.763	0.761	0.770	0.784	0.788	0.791
ROC	0.903	0.889	0.886	0.888	0.894	0.910	0.919	0.930
PR	0.759	0.718	0.725	0.723	0.722	0.729	0.734	0.727
N.Obs	11,375	11,377	11,378	11,383	11,386	11,392	11,388	11,387

Note: We report prediction accuracies of BART-MIA after training and testing on separate years. Our aim is to check whether predictions are robust along the timeline.

Table A2.7: Prediction accuracies of exporters defined *á la* Békés and Muraközy, 2012

Exporter Class	Sensitivity	Specificity	Balanced Accuracy	ROC	PR	Num. Obs.
Permanent Exporters	0.723	0.779	0.751	0.849	0.934	76,185
Temporary Exporters	0.421	0.820	0.621	0.755	0.447	73,647
Non-Exporters		0.949				158,625
Total	0.650	0.9066	0.7783	0.9048	0.7383	232,272

Note: We report prediction accuracies after BART-MIA for firms classified 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 A2.8: Prediction accuracies after an exporters' definition based on thresholds of the share of export revenues over total revenues

Measure	1^{st} Percentile	2^{nd} Percentile	5^{th} Percentile	Benchmark
Sensitivity	0.652	0.641	0.625	0.658
Specificity	0.835	0.837	0.852	0.833
Balanced Accuracy	0.744	0.739	0.738	0.745
ROC	0.836	0.835	0.836	0.836
PR	0.737	0.731	0.724	0.738
N.Obs	41,911	41,911	41,911	41,911

Note: We report prediction accuracies of BART-MIA after defining as exporters the firms with share of export revenues over total revenues above some specific thresholds, at the 1^{st},2^{nd}, and 5^{th} percentiles of the distribution of the share of export revenues over total revenues.

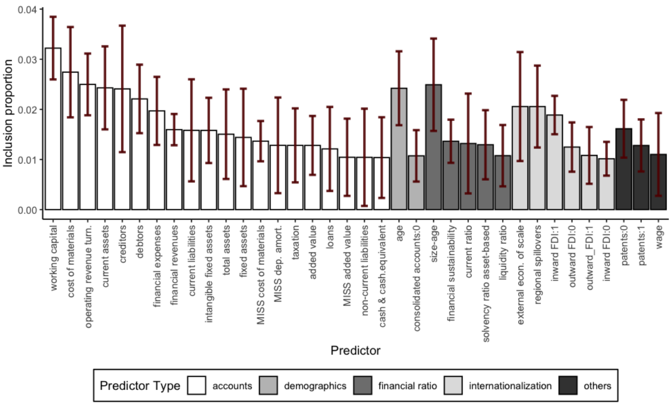
Table A2.9: Prediction accuracies - Imputation of missing values

	Specificity	Sensitivity	Balanced Accuracy	ROC	PR	N. obs.
LOGIT	0.817	0.751	0.784	0.784	0.528	382,606
LOGIT-LASSO	0.913	0.541	0.727	0.880	0.682	382,606
CART	0.893	0.617	0.755			382,606
Random Forest	0.910	0.647	0.778	0.907	0.738	382,606
BART	0.910	0.635	0.772	0.905	0.731	382,606

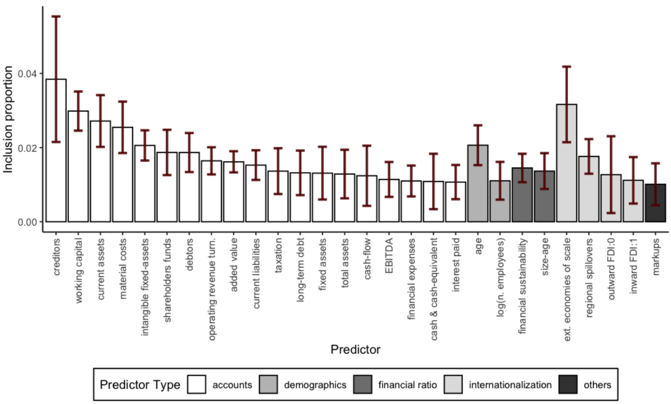
Note: For a robustness check, we report prediction accuracies after an imputation of missing values based on median values, while adding a predictor indicating the number of missing entries by observation (number of missing values by row).

Figure A2.3: Variable inclusion proportions in Île-de-France *versus* the rest of France

(a) VIPs Île-de-France

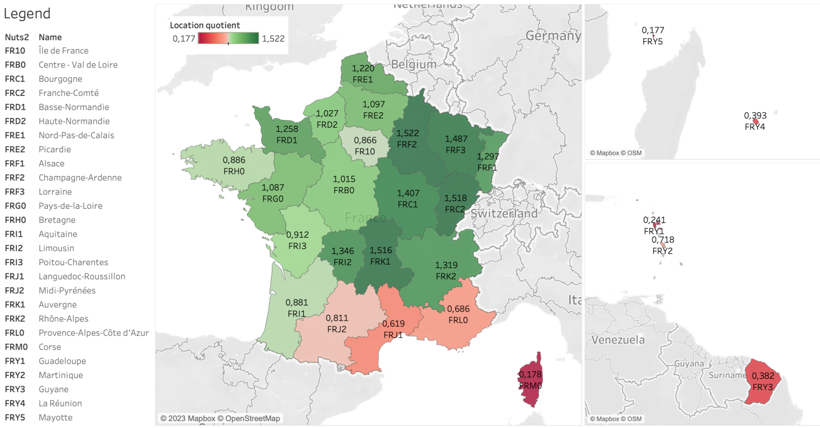


(b) VIPs France, excluding Île-de-France



Note: We report Variable Inclusion Proportions (VIPs) in (a)Île-de-France, (b) in all France *excluding*Île-de-France. Of all the predictors in baseline, we visualize only those with a VIP higher than 1%. The bars represent standard deviations obtained by replicating five different times the BART-MIA on the same random training set.

Figure A2.4: The potential for extensive margin across France



Note: We report location quotients of non-exporters whose score is above the median in the national distribution. Regions with location quotients greater than one (lower than one) are those where potential exporters are more (less) concentrated than what one would expect given manufacturing density. See Appendix D for details on the computation of location quotients.

Appendix A3: Evaluation of prediction accuracy

Different metrics are used to evaluate the prediction accuracy of machine learning algorithms. Briefly, prediction accuracy metrics compare the classes predicted by the algorithm with the actual ones. In the case of a binary outcome, the comparison generates four classes of results:

- **True Positives:** cases when the actual class of the data point is 1 (Positive) and the predicted is also 1 (Positive);
- **False Positives:** cases when the actual class of the data point is 0 (Negative) and the predicted is 1 (Positive);

- **False Negatives:** cases when the actual class of the data point is 1 (Positive) and the predicted is 0 (Negative);

- **True Negatives:** cases when the actual class of the data point is 0 (Negative) and the predicted is also 0 (Negative);

In an ideal scenario, we want to minimize the number of False Positives and False Negatives.

Table A3.1: Confusion Matrix

		Actual	
		$Positives$ (1)	$Negatives$ (0)
Predicted	$Positives$ (1)	True Positives (TP)	False Positives (FP)
	$Negatives$ (0)	False Negatives (FN)	True Negatives (TN)

The metrics we use to evaluate prediction accuracy in our exercises are based on the relationship between the sizes of the above classes.

Sensitivity (or Recall) Sensitivity (or Recall) is a measure of the proportion of correctly Predicted Positives out of the total Actual Positives.

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

Specificity Specificity is a measure that catches the proportion of correctly Predicted Negatives, out of total Actual Negatives.

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives}$$

Balanced Accuracy (BACC) Balanced Accuracy (BACC) is a combination of Sensitivity and Specificity. It is particularly useful when classes are imbalanced, i.e., when a class appears much more often than the other. It is computed as the average between the True Positives rate and

True Negatives rate.

$$BACC = \frac{Sensitivity + Specificity}{2}$$

Receiving Operating Characteristics (ROC) The ROC curve is a graph showing the performance in classification at different thresholds, expressed in terms of the relationship between True Positive Rate (TPR) and False Positive Rate (FPR), defined as follows:

$$True\ Positive\ Rate = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

$$FalsePositiveRate = \frac{False\ Positives}{False\ Positives + True\ Negatives}$$

The Area Under the Curve (AUC) of ROC is then useful to evaluate performance in a bounded range between 0 and 1, where 0 indicates complete misclassification, 0.5 corresponds to an uninformative classifier, and 1 indicates perfect prediction.

Precision-Recall (PR) The PR curve is a graph showing the trade-off between Precision and Recall at different thresholds. Note that Precision and Recall are defined as follows:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

As for the ROC curve, the PR AUC is used to evaluate the classifier performance. A High AUC represents both high recall and high precision, thus meaning the classifier is returning accurate results (high precision), as well as returning a majority of all the positive results (high recall).

Appendix A4: Location Quotients

Let us define $\mathcal{I} = \{1, \ldots, n\}$ the set of non-exporting firms and $\mathcal{R} = \{1, \ldots, r\}$ the set of regions (NUTS 2-digit). The r partitions of $\mathcal{I}$ by region $j \in \mathcal{R}$ are defined as:

$$I_j \subset \mathcal{I}, j = 1, \ldots, r \quad s.t. \quad \bigcup_{j=1}^{r} I_j = \mathcal{I}$$

Let $\mathcal{P}$ be the set of non-exporting firms whose exporting score e is above the one of the median firm in the total distribution of non-exporters, i.e.:

$$\mathcal{P} \subset \mathcal{I} = \{i \in \mathcal{I} : e_i > median(e)\}$$

Again we can define the r partitions of $\mathcal{P}$ by region $j \in \mathcal{R}$ as

$$P_j \subset \mathcal{P}, j = 1, \ldots, r \quad s.t. \quad \bigcup_{j=1}^{r} P_j = \mathcal{P}$$

The location quotient, for each region $j = 1, \ldots, r$ is computed as

$$LQ_j = \frac{\#P_j/\#I_j}{\#\mathcal{P}/\#\mathcal{I}}$$

In our case, location quotients (LQ) detect the concentration of potential exporters in excess of what one would expect from the national distribution. If, for example, region j has $LQ_j = 1.5$, it implies that firms with a high trade potential are 1.5 times more concentrated in such a region than the average.

Appendix B

Supplementary materials for Chapter 2

This Appendix is based on Fontagné et al. (2024), "The heterogeneous impact of the EU-Canada agreement with causal machine learning", Papers 2407.07652, arXiv.org, revised July 2024. Preprint at https://doi.org/10.48550/arXiv.2407.07652.

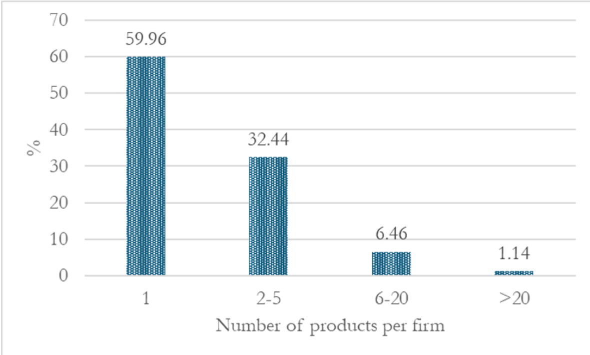
Appendix B1: Tables and graphs

Table B1.1: Distribution of tariff changes in the Canada-EU Comprehensive Trade Agreement (CETA)

Tariff decrease (%)	N. products	% products
0.3 - 5	1,871	51.04
6 - 10	1,290	35.19
11 - 20	479	13.06
>20	26	0.71
Total	3,666	100.00

Note: The table shows the distribution of tariff changes by HS 6-digit products as it has been negotiated in the CETA. The simple average tariff decrease has been 5.8% with a 4.3 standard deviation.

Figure B1.1: Products per exporter in Canada in 2016



Note: The figure shows the distribution of product portfolios by exporters to Canada before the entry into force of the CETA. On the left, the first bar indicates exporters with one product delivered to Canada. Then, the following bars refer to product portfolios sold to Canada by multiproduct firms.

Table B1.2: Which products in the intensive margin

Case	Traded in 2015	Traded in 2016	Traded in 2017	Intensive margin	Note:
1)	Yes	Yes	Yes	Yes	Always traded
2)	Yes	Yes	No	No	Not traded after CETA
3)	Yes	No	Yes	Yes	Intermittently traded
4)	No	Yes	Yes	Yes	Intermittently traded
5)	Yes	No	No	No	Intermittently traded
6)	No	Yes	No	Yes	Intermittently traded
7)	No	No	Yes	No	Traded only after CETA
8)	No	No	No	No	Never traded

Note: The table separates cases of intensive margins from different trade patterns in the original data. For each of them, we report in column (4) whether the corresponding product is included in the analyses on intensive margins.

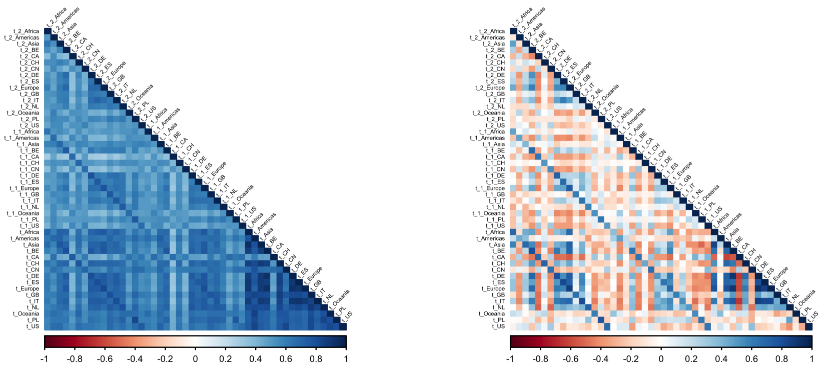
Table B1.3: Ranking export destination by trade volumes and number of products

Destination	Export volume (in mln €) (1)	# Products (2)	Rank by Values (3)	Rank by #Products (4)	Combined Rank (5)
Germany	73.134	4,816	1	4	2.5
Italy	36.084	4,842	4	2	3
Spain	36.692	4,825	3	3	3
Belgium	30.752	4,857	6	1	3.5
USA	38.771	4,091	2	9	5.5
United Kingdom	35.721	4,594	5	7	6
Netherlands	16.350	4,775	8	5	6.5
Switzerland	15.922	4,691	9	6	7.5
China	19.489	3,836	7	10	8.5
Poland	8.356	4,193	10	8	9
Canada	4.217	3,812	26	18	22
Rest of Asia	74.958	4,890			
Rest of Europe	55.077	4,999			
Africa	29.825	4,912			
Rest of Americas	16.377	4,245			
Oceania	5.449	4,106			

Note: Countries in this table are included in the trade matrix at the product level introduced in Section 2.5.1. The decision is based on two criteria: in column (1), we report the average trade values exported by the French Exporters from 2015-2016; in column (2), we report the average number of products exported to each destination in 2015-2016. Columns (3) and (4) report the ranking position of each country by average trade values and average number of exported products, respectively. Column (5) reports an average of rankings in columns (3) and (4). The (rest of the) continents at the bottom of the table are also included in the analyses to close and balance the trade matrix.

Figure B1.2: Time-destination correlation matrices on the extensive margins: original data matrix and the predicted low-rank matrix

(a) Original data matrix $\mathbf{Y}$ **(b)** Predicted matrix $\tilde{\mathbf{Y}}$



Note: he figure on the left displays the correlation matrix of the columns of the original matrix $\mathbf{Y}$, which records the exports of products (at the HS6 level) to country d at time t. The figure on the right shows the correlation matrix of the columns of the corresponding predicted low-rank matrix $\tilde{\mathbf{Y}}$. This predicted matrix accounts for the residual correlation between the rows and columns of the original matrix after removing the row and column fixed effects ($\tilde{\gamma}_i$ and $\tilde{\delta}_j$ in equation 1.2). Figure (a) illustrates significant and consistent correlation patterns between destinations over time. Figure (b) demonstrates that these correlation patterns are effectively learned and captured by the predicted low-rank matrix $\tilde{\mathbf{Y}}$, enhancing the accuracy of the predicted outcomes.

Table B1.4: A placebo test for the intensive margin to Canada

Product class	Class name	WATET (1)	weighted st. dev. (2)	N. products (3)
01-97	All products	-1.038	11.664	2,219
01-05	Live animals & Animal products	0.932	85.550	44
06-15	Vegetable products	5.380	0.696	122
16-24	Foodstuffs	0.415	4.262	120
25-27	Mineral products	-32.675	232.346	23
28-38	Chemicals & Allied industries	-1.613	12.084	244
39-40	Plastics / Rubbers	-1.289	9.967	129
41-43	Raw Hides, Skins, Leather & Furs	-1.021	5.609	31
44-49	Wood & Wood products	0.578	9.423	31
50-63	Textiles	15.36	13.84	458
64-67	Footwear / Headgear	3.189	26.784	30
68-71	Stone / Glass	3.388	33.419	74
72-83	Metals	2.216	3.766	234
84-85	Machinery / Electrical	-1.655	4.607	418
86-89	Transportation	-9.612	6.021	66
90-97	Miscellaneous	1.253	3.382	195

Note: The table reports the Weighted Average Treatment Effects on the Treated (WATET) exports to Canada after a placebo test, considering the same definitions of treatment but in the period September 2012-August 2015. TET^*_{pdt}, are weighted for the relevance each product had in the year before the treaty signature to obtain the unique $WATET$. The weighted standard deviations are computed as $\sqrt{\frac{\sum_{i=1}^{N} s_{pdt}\left(TET^*_{pdt}-WATET\right)^2}{(\mathcal{L}-1)\backslash\mathcal{L}\sum_{i=1}^{N} s_{pdt}}}$, where $\mathcal{L}$ is the number of counterfactuals in the trade matrix for Canada. *, **, *** stand, respectively, for $p < 0.05$, $p < 0.01$, $p < 0.001$.

Table B1.5: Changing alternative destinations in the trade matrix

Model	WATET (1)	weighted std. dev. (2)	N. products (3)
Baseline	1.278***	0.524	2,165
Number of exporters	1,217***	0.423	2,165
Import structure similarity	1.006***	0.431	2,167
Import market size	0.939***	0.429	2,165

Note: The table reports the Weighted Average Treatment Effects on the Treated (WATET) exports to Canada after changing the set of alternative destinations in the trade matrix. TET^*_{pdt} are weighted for the relevance each product had in the year before the treaty signature to obtain the unique $WATET$. The weighted standard deviations are computed as $\sqrt{\frac{\sum_{i=1}^{N} s_{pdt}\left(TET^*_{pdt}-WATET\right)^2}{(\mathcal{L}-1)\backslash\mathcal{L}\sum_{i=1}^{N} s_{pdt}}}$, where $\mathcal{L}$ is the number of counterfactuals in the trade matrix for Canada. *, **, *** stand, respectively, for $p < 0.05$, $p < 0.01$, $p < 0.001$.

Table B1.6: CETA and alternative destinations - general equilibrium trade effects - Robustness checks

Dependent variable TET_{pdt}	(1)	(2)	(3)
$TET_{CA,pt}$	-0.927*	-1.951	-1.740*
	(0.332)	(1.250)	(0.648)
$Value_{pdt-1}$	1.755***	1.661***	1.381***
	(0.189)	(0.233)	(0.276)
Constant	-5,379,224.0**	-7,379,045.8*	-4,700,216.0*
	(1,534,383.2)	(3,395,757.2)	(1,947,475.7)
N. obs.	32,505	32,505	32,505
R squared	0.773	0.693	0.602
Clusters by country	Yes	Yes	Yes
Clusters by product class	Yes	Yes	Yes
Model	Number of exporters	Import structure similarity	Import market size

Note: The Table shows results after a linear regression model whose dependent variable includes the treatment effects on the treated in monetary values, TET_{pdt}, where destination d is different from Canada. Each column corresponds to a different set of destinations, as reported in Table B1.8. The main regressor of interest is the vector of treatment effects on the treated in monetary values, TET_{pdt}, where destination d is instead Canada. The unique control variable is the value of the product p export flow in destination d different from Canada in the period before the CETA, $t-1$. Errors are double-clustered by country and product class. **, *** stand, respectively, for $p < 0.05$, $p < 0.01, p < 0.001$.

Table B1.7: Prediction accuracy at the product level intensive margin - Robustness checks

Model	min RMSE	$\overline{Y}$	SI	NRMSE
Baseline	7.12126	7,060,711	0.000100858	0.00027172
No fixed effects	7.328702	7,060,711	0.000103796	0.00027963
Number of exporters	8.322443	7,037,844	0.000118253	0.00034071
Import structure similarity	9.581219	7,204,660	0.000132986	0.00049488
Import market size	11.518196	7,041,990	0.000163565	0.00053770

Note: The table reports the statistics of the prediction accuracy that we obtain when we train the model while removing the Fixed Effects, or on matrices where we used different matrix structure strategies.

Appendix B2: Difference-in-difference

We consider the simple difference-in-difference as a conventional empirical method for benchmarking against our preferred empirical strategy.

Table B1.8: Choice of destinations using different selection criteria

Selection Criterion	Individual Destinations	Aggregates
Baseline	Belgium, Canada, Switzerland, China, Germany, Spain, the United Kingdom, Italy, The Netherlands, Poland, the United States of America	Africa, Americas, Asia, Europe, Oceania
Number of exporters	Belgium, Canada, Switzerland, China, Germany, Spain, the United Kingdom, Italy, Japan, Morocco, the United States of America	Africa, Americas, Asia, Europe, Oceania
Import structure similarity	Austria, Australia, Canada, Germany, Spain, Finland, United Kingdom, New Zealand, Poland, Sweden, The United States of America	Africa, Americas, Asia, Europe, Oceania
Import market size	China, Germany, the United Kingdom, Hong Kong, India, Italy, Japan, Korea, the Netherlands, the United States of America	Africa, Americas, Asia, Europe, Oceania

Note: The table reports, for each destination selection criterion, the list of partner countries included in the trade matrix.

Following our definitions, a treated product is a product that is enlisted in the CETA, while a treated firm is a firm that exports to Canada at least one product under CETA. Basic formulations are, for the intensive margins:

$$Y_{ut} = c_u + \gamma_t + \beta_D \cdot D_{ut} + \epsilon_{ut} \tag{B.1}$$

the extensive margin for products:

$$Pr(Q_{pt} = 1 | X_{pt} = 1) = c_u + \gamma_t + \beta_D \cdot D_{ut} + \epsilon_{ut} \tag{B.2}$$

and the extensive margin for firms:

$$Q_{it} = \exp(c_i + \gamma_t + \beta_D D_{it} + \epsilon_{it}) \tag{B.3}$$

where Y_{ut} represents the total exports of the u-th unit of observation where $u = (p, i)$ is either a p-th product or an i-th-firm observed at time t in Canada. Product fixed effects, c_p, and time fixed effects, γ_t, are included. The binary variable D_{pt} is the treatment indicator, while the error term ϵ_{pt} captures stochastic variation. In eq. B.2, we examine the impact of CETA on the product's extensive margin of trade with either a linear probability model (LPM) or a logit, whose dependent variable, Q_{pt} is equal to one if the product was exported and zero otherwise. In eq. B.3, instead, we study the impact of CETA on the firms' extensive margin with either a simple OLS or a Pseudo-Poisson, where Q_{it} is the number of products exported in Canada by a firm i at time t.

Table B2.1: Difference-in difference for products and firms

	Product-level			Firm-level		
	Intensive Margin	Extensive Margin		Intensive Margin	Extensive Margin	
	(OLS)	(LPM)	(Logit)	(OLS)	(OLS)	(Poisson)
	Y_{pt}	$P(Q_{pt}=1)$	$OR(Q_{pt}=1)$	Y_{it}	$Q_{it,CA}$	$Q_{it,CA}$
	(1)	(2)	(3)	(4)	(5)	(6)
ATT	91.63	0.017	1.244	-32.03	0.324***	0.0842***
	(115.1)	(0.010)	(0.161)	(48.52)	(0.075)	(0.028)
Year fixed effects:						
t-4	12.95	0.005	1.067	-5.452	-0.16	-0.008
	(46.53)	(0.005)	(0.085)	(16.65)	(0.030)	(0.010)
t-3	79.97	0.016**	1.228**	21.00	0.127***	0.040*
	(58.09)	(0.005)	(0.085)	(16.65)	(0.049)	(0.016)
t-2	15.83	0.014*	1.201*	-10.05	0.245 ***	0.077***
	(62.16)	(0.006)	(0.087)	(25.52)	(0.061)	(0.020)
t-1	89.88	0.033***	1.526***	18.91	0.420***	0.133***
	(62.61)	(0.006)	(0.113)	(22.83)	(0.070)	(0.021)
t	82.20	0.015	1.219	76.37	0.232***	0.085*
	(128.9)	(0.010)	(0.150)	(50.76)	(0.051)	(0.023)
constant	1,063.6***	0.550***		291.1***	2.407***	
	(45.67)	(0.004)		(16.54)	(0.043)	
Product fixed effect	YES	YES	YES	NO	NO	NO
Firm fixed effect	NO	NO	NO	YES	YES	YES
N. obs.	15,763	31,236	10,980	53,338	53,338	45,729

Note: We report product-level results in columns 1-3. Column (1) reports results on the intensive margin expressed in thousands of euros. Columns (2) and (3) report results on the extensive margin either computed using a Linear Probability model (LPM) or a logit (Logit). We report firm-level results in columns 4-6. Column (4) reports the results on the intensive margin expressed in thousands of euros. Column (5) reports the results on the extensive margin computed using a Linear Model, while column (6) reports results after using a Pseudo-Poisson Model. Robust standard errors in parentheses. *, **, *** stand, respectively, for $p < 0.05$, $p < 0.01$, $p < 0.001$.

Appendix B3: Prediction accuracy

Different metrics are used to evaluate the prediction accuracy of machine learning algorithms. Briefly, prediction accuracy metrics compare the classes predicted by the algorithm with the actual ones.

In the case of continuous outcomes, we can use the following measures:

- **Root-Mean-Square Error (RMSE)**, which is computed as

$$RMSE = \sqrt{\sum_{i=1}^{NRD} (\hat{y}_{ird} - y_{ird})^2 / NRD} \tag{B.4}$$

- **Scatter Index (SI)**, computed as

$$SI = RMSE/\overline{y}_{ird} * 100 \tag{B.5}$$

  It gives the percentage of expected error for the parameter of interest

- **Normalised Root-Mean-Square Error (NRMSE)**, computed as

$$NRMSE = RMSE/(Q3 - y_{min}) * 100 \tag{B.6}$$

  it relates the RMSE to the observed range of the variable, thus allowing comparisons with other models

Appendix C

Supplementary materials for Chapter 3

Appendix C1: Data

Table C1.1: List of variables

Variable	Description
Sales, Number of employees, Profit Margins, P/L after tax, Operating revenue turnover, Working capital, Long-term debt, Debtors, Tangible fixed assets, Intangible fixed assets, Financial Expenditure	Original financial accounts expressed in euro.
Export intensity	Indicator computed as *Export revenues/ Total revenues*

Continued on next page

Table C1.1 – continued from previous page

Variable	Description
Total Costs	Total costs of production, computed as *Real Cost of materials +Real Cost of employees*
Profitability	Measure of profitability expressing how much earnings are generated by the firm's assets. It is computed as *EBITDA/Total Assets*
NACE rev. 2	A 2-digit industry affiliation following the European Classification
NUTS 2-digit	The region in which the company is located following the European classification.
TFP	It is the Total Factor Productivity of a firm computed as in Ackerberg et al. (2015).
Size-Age	It is a synthetic indicator proposed by Hadlock and Pierce (2010), computed as (-0.737$\cdot$ $log(total\ assets)$)+(0.043 $\cdot$ $log(total\ assets))^2$ - (0.040 $\cdot$ *age* to catch the non-linear relationship between financial constraints, size and age.
patents	It is a binary variable with value 1 if the firm possess at least one patent at time *t*
D(export in t-1)	It is a binary variable with value 1 if the firm reported positive export revenues in *t-1*
Pavitt Class	Taxonomy which describes a firm's patterns of technical change. The classification follows the methodology of Bogliacino and Pianta (2016), which is based on Nace Rev.2 classification.
Inward FDI	It is a binary variable with value 1 if the firm has foreign headquarters
Outward FDI	It is a binary variable with value 1 if the firm has subsidiaries abroad

Continued on next page

Table C1.1 – continued from previous page

Variable	Description
N.patents	Total number of patents owned by the firm at time t
Corporate Control	A binary variable equal to one if a firm belongs to a corporate group.
Labour Productivity	It is a ratio between value added and number of employees for the average productivity of labor services.
Productive Capacity	It is an indicator of investment in productive capacity computed as $\textit{Fixed Assets}_t/(\textit{Fixed Assets}_{t-1}+\textit{Depreciation}_{t-1})$
Capital Adequacy Ratio	It is a ratio of Shareholders' Funds over Short and Long Term Debts.
Financial Sustainability	It is a ratio between Financial Expenses and Operating Revenues.
Capital Intensity	It is a ratio between fixed assets and number of employees for the choice of factors of production.

Continued on next page

Table C1.1 – continued from previous page

Variable	Description
Firm Size	Size classification sourced from Orbis: • *Very Large*: they match at least one of the following conditions: – Op. revenue $\geq$100 million € – Total assets $\geq$ 200 million € – Employees $\geq$ 1,000 – Listed • *Large:* they match at least one of the following conditions: – Op. revenue $\geq$ 10 million € – Total assets $\geq$ 20 million € – Employees $\geq$ 150 – Not very large • *Medium:* when they match at least one of the following conditions: – Ope. revenue $\geq$ 1 million € – Total assets $\geq$ 2 million € – Employees $\geq$ 15 – Not very large or large • *Small:* Residual Class

Table C1.2: Distribution of firms and export across Pavitt's classes - Averages 2010-2018

Pavitt class	N. firms	N. exporters	Export value (in mln €)
	(1)	(2)	(3)
Scale and information intensive	10,893	3,389	39,535.8
	(19.13%)	(23.05%)	(24.36%)
Science based	3,529	1,575	48,976.90
	(6.2%)	(10.71%)	(30.18%)
Specialised Suppliers	4,924	2,112	39,938.10
	(8.65%)	(14.36%)	(24.61%)
Suppliers dominated	37,609	7,625	33,822.70
	(66.03%)	(51.87%)	(20.84%)
Total	56,954	14,701	162,273.50
	(100%)	(100%)	(100%)

Note: We report in column (1) the distribution of the number of firms in our sample over the Pavitt classes, while column (2) reports the corresponding number of exporters. Column (3) shows the export value generated by the exporters in each Pavitt's class. Note that all numbers are means over the period 2010-2018.

Table C1.3: Firms' distribution in our sample, accross firm size and Pavitt's class

Size Class	Pavitt's Class: Scale and Inform. Intensive	Science-based	Specialized Suppliers	Suppliers Dominated	Total
Small	729	289	381	1,926	3,325
	(22.61%)	(18.11%)	(17.85%)	(26.68%)	(23.46%)
Medium	1,392	636	987	3,453	6,468
	(43.18%)	(39.85%)	(46.23%)	(47.83%)	(45.63%)
Large	919	519	652	1,580	3,670
	(28.50%)	(32.52%)	(30.54%)	(21.88%)	(25.89%)
Very Large	184	152	1,5	261	712
	(5.71%)	(9.52%)	(5.39%)	(3.61%)	(5.02%)
Total	3,224	1,596	2,135	7,220	14,175
	(100%)	(100%)	(100%)	(100%)	(100%)

Note: The table reports the number of observations in our sample by firm size and Pavitt's Class. Each observation refers to a firm i in a time t. Please note that the sample of firms includes only permanent exporting firms.

Appendix C2: Alternative Specifications

Table C2.1: Regression models for TFP in each Pavitt's class

	Pavitt's class			
Dep. Variable: TFP	Scale and information intensive	Science-based	Specialised Suppliers	Suppliers dominated
	(1)	(2)	(3)	(4)
Export status variation in (t-1)	0.00271	-0.01827	0.02064	-0.00192
	(0.01458)	(0.02415)	(0.0207)	(0.01052)
size-age	-0.09234***	-0.15481***	-0.07524**	-0.03059*
	(0.01666)	(0.02657)	(0.0243)	(0.01386)
log(n. employees)	-0.43643***	-0.16015***	-0.47774***	-0.39757***
	(0.01363)	(0.01958)	(0.0168)	(0.00924)
patents	0.10137*	0.17533**	0.07226	-0.01774
	(0.04235)	(0.05533)	(0.04519)	(0.0405)
T_1	-0.00021	-0.00182	-0.00224	-0.00193
	(0.00184)	(0.00257)	(0.002)	(0.00114)
T_2	0.00003	0.00005	0.00001	0.00006
	(0.00005)	(0.00007)	(0.00005)	(0.00003)
T_3	0	0	0	0
	(0)	(0)	(0)	(0)
Constant	11.3563***	11.30115***	12.31453***	11.40184***
	(0.07917)	(0.11442)	(0.1074)	(0.05853)
Firm FE	YES	YES	YES	YES
Time FE	YES	YES	YES	YES
(N)	11,189	5,624	6,999	21,262

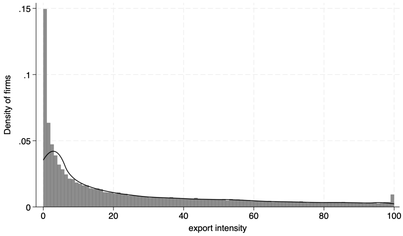
Note: In this table we report the results of the model estimated considering only TFP as dependent variables for firms in different Pavitt's classes. Column (1) reports the results when we restrict the sample on firms in Information Intensive Industries. Column (2) considers firms in Science-based industries. Column (3) encompasses Specialised suppliers. Column (4) considers firms in Supplier-dominated industries.

Appendix C3: Export Intensity and Common Support

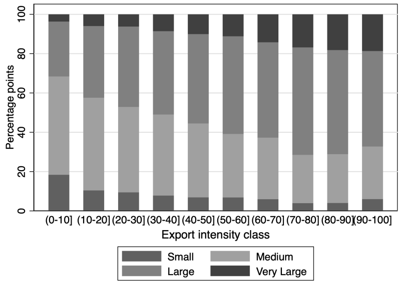
In this appendix we verify the presence of a common support for firms in our sample. This is an essential measure to ensure the validity of our results and their policy implications. We start by studying the relationship between a firm's size and its export intensity. Figure C3.1a illustrates the distribution of export intensity in our sample, while Figure C3.1b displays how firms of different sizes are distributed across various classes of export intensity. Notably, the distribution of export intensity is skewed towards the left, with the majority of firms exporting less than 10% of their sales abroad. A long right tail of firms strongly committed to export is observed, with a peak corresponding to an export intensity of 100%. This peak identifies firms exclusively exporting their products abroad, without competing in their domestic market. Examining Figure C3.1b, a positive correlation emerges between firm size and export intensity. This evidence hints at structural differences between firms exporting different intensities, suggesting that smaller firms may be less capable of sustaining larger export intensities.

To further investigate this hypothesis, we verify the presence of significant correlations between relevant firm dimensions and the firm's export intensity. Figure C3.2 demonstrates that export intensity is significantly positively correlated with firm size, measured by assets or the number of employees, innovation, international participation, and several measures of productivity and financial sustainability. This evidence indicates that more export-intensive firms tend to be larger, more innovative, and more productive than their less export-intensive counterparts. To assess whether the correlation is so strong that comparisons between firms with different export intensities become problematic, we conduct two analyses of common support. On one hand, we employ Mahalanobis weights based on the relevant firm dimensions selected in the correlation

Figure C3.1: Descriptive statistics on firms' distributions



(a) Distribution of export intensities in our sample



(b) Distribution firms of different sized within export intensities

Note: Figure (a) reports the distribution of firms in our sample over the export intensities: Figure (b) shows how firms of different sizes distrubute within export intensity intervals.

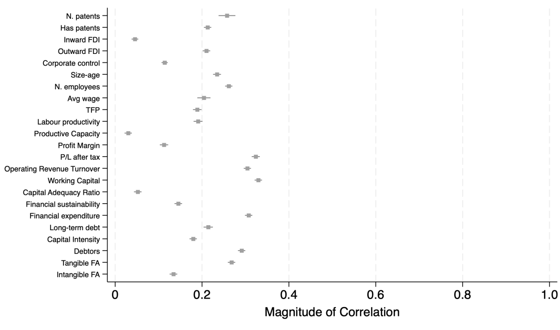
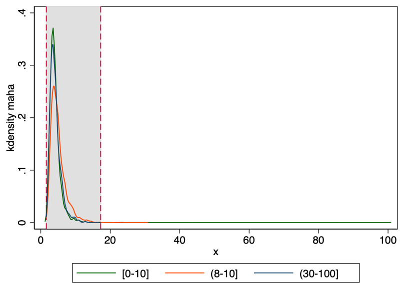


Figure C3.2: Pairwise correlation between export intensity and relevant firms' characteristics analysis. On the other hand, we select the first component obtained from a PCA on the same variables to compare populations of firms in different export intensity classes. Figures C3.3a and C3.3b clearly show that a few firms with export intensity lower than 10% lie outside the common support. This suggests that such firms possess characteristics quite different from their more export-intensive counterparts. Specifically, based on the results of the correlation analysis, these firms appear to be smaller and less productive, indicating structural limitations in their ability to increase export intensity.

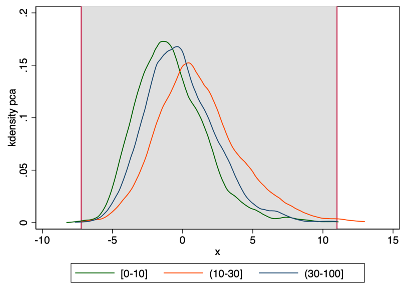
Finally, we utilize a propensity score matching exercise to compare firms with export intensity below 10% to those with intensity above 10%. The aim is to confirm that the characteristics of firms with export intensity smaller than 10% differ from those of more intensive exporters. Once again, we use the set of variables identified in the correlation analysis. However, the propensity score matching exercise fails to satisfy the balancing property, confirming that the characteristics of firms with export intensity less than 10% are structurally different from other firms.

The failure of the common support hypothesis for firms characterised by very low export intensity is crucial for policy considerations.

Figure C3.3: Common support analysis



(a) Distribution of Mahalanobis weights across firm-intensity classes



(b) Distribution of the first component across firm-intensity classes

Note: Figure (a) reports the distribution of firms in our sample over the export intensities: Figure (b) shows how firms of different sizes distribute within export intensity intervals.

Bibliography

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, *105*(490), 493–505.

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, *59*(2), 495–510.

Abrams, R. K. (1980). International trade flows under flexible exchange rates. *Economic Review*, *65*(3), 3–10.

Ackerberg, D. A., Caves, K., & Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, *83*(6), 2411–2451.

Ahrens, A., Hansen, C. B., & Schaffer, M. E. (2020). Lassopack: Model selection and prediction with regularized regression in stata. *arXiv preprint arXiv:1901.05397*.

Aitken, N. D. (1973). The effect of the eec and efta on european trade: A temporal cross-section analysis. *The American Economic Review*, *63*(5), 881–892.

Albornoz, F., Pardo, H. F. C., Corcos, G., & Ornelas, E. (2012). Sequential exporting. *Journal of International Economics*, *88*(1), 17–31.

Alessi, L., & Detken, C. (2018). Identifying excessive credit growth and leverage. *Journal of financial stability*, *35*, 215–225.

Alfaro, L., Charlton, A., & Kanczuk, F. (2009). Plant size distribution and cross-country income differences. *NBER International seminar on macroeconomics*, *5*(1), 243–272.

Almeida, H., & Campello, M. (2007). Financial constraints, asset tangibility, and corporate investment. *The Review of Financial Studies*, *20*(5), 1429–1460.

Almeida, H., Campello, M., & Weisbach, M. S. (2004). The cash flow sensitivity of cash. *The Journal of Finance*, *59*(4), 1777–1804.

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

Altman, E. I., et al. (2000). Predicting financial distress of companies: Revisiting the z-score and zeta models. *Stern School of Business, New York University*, 9–12.

Anderson, J. E., & Yotov, Y. V. (2016). Terms of trade and global efficiency effects of free trade agreements, 1990–2002. *Journal of International Economics*, *99*, 279–298.

Angrist, J. D., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, *24*(2), 3–30.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2019). *Synthetic difference in differences* (tech. rep.). National Bureau of Economic Research.

Arkolakis, C., Ganapati, S., & Muendler, M.-A. (2021). The extensive margin of exporting products: A firm-level analysis. *American Economic Journal: Macroeconomics*, *13*(4), 182–245.

Arkolakis, C., & Muendler, M.-A. (2013). Exporters and their products: A collection of empirical regularities. *CESifo Economic Studies*, *59*(2), 223–248.

Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda* (pp. 507–547). University of Chicago Press.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., & Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 1–15.

Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353–7360.

Athey, S., Imbens, G., Metzger, J., & Munro, E. (2021). Using Wasserstein generative adversarial networks for the design of Monte Carlo simulations. *Journal of Econometrics*.

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic perspectives*, *31*(2), 3–32.

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*(1), 685–725.

Atkeson, A., & Burstein, A. T. (2010). Innovation, firm dynamics, and international trade. *Journal of political economy*, *118*(3), 433–484.

Atkin, D., Khandelwal, A. K., & Osman, A. (2017). Exporting and Firm Performance: Evidence from a Randomized Experiment*. *The Quarterly Journal of Economics*, *132*(2), 551–615.

Aw, B. Y., Lee, Y., & Vandenbussche, H. (2023). Quantifying consumer taste in trade: Evidence from the food industry. *CESifo Working Paper No. 10234*.

Baier, S. L., & Bergstrand, J. H. (2002). *On the endogeneity of international trade flows and free trade agreements* (tech. rep.). mimeo New York.

Baier, S. L., & Bergstrand, J. H. (2004). Economic determinants of free trade agreements. *Journal of International Economics*, *64*(1), 29–63.

Baier, S. L., & Bergstrand, J. H. (2007). Do free trade agreements actually increase members' international trade? *Journal of international Economics*, *71*(1), 72–95.

Baier, S. L., & Bergstrand, J. H. (2009). Estimating the effects of free trade agreements on international trade flows using matching econometrics. *Journal of international Economics*, *77*(1), 63–76.

Baier, S. L., Bergstrand, J. H., & Mariutto, R. (2014). Economic determinants of free trade agreements revisited: Distinguishing sources of interdependence. *Review of International Economics*, *22*(1), 31–58.

Baier, S. L., Yotov, Y. V., & Zylkin, T. (2019). On the widely differing effects of free trade agreements: Lessons from twenty years of trade integration. *Journal of International Economics*, *116*, 206–226.

Bajari, P., Nekipelov, D., Ryan, S. P., & Yang, M. (2015). Machine learning methods for demand estimation. *American Economic Review*, *105*(5), 481–85. https://doi.org/10.1257/aer.p20151021

Baldwin, J., & Gu, W. (2009). The impact of trade on plant scale, production-run length and diversification. In *Producer dynamics: New evidence from micro data* (pp. 557–592). University of Chicago Press.

Baldwin, J. R., & Gu, W. (2003). Export-market participation and productivity performance in canadian manufacturing. *Canadian Journal of Economics/Revue canadienne d'économique*, *36*(3), 634–657.

Bandiera, O., Prat, A., Hansen, S., & Sadun, R. (2020). Ceo behavior and firm performance. *Journal of Political Economy*, *128*(4), 1325–1369.

Bargagli-Stoffi, F., Riccaboni, M., & Rungi, A. (2020). *Machine learning for zombie hunting. firms' failures and financial constraints* (EIC Working Paper Series No. 1/2020). IMT School for Advanced Studies.

Barro, R. J., & Sala-i-Martin, X. (1997). Technological diffusion, convergence, and growth. *Journal of Economic Growth*, *2*(1), 1–26.

Bartelme, D., Lan, T., & Levchenko, A. A. (2024). Specialization, market access and real income. *Journal of International Economics*, *150*, 103923.

Bas, M., & Bombarda, P. (2013). Chinese trade reforms, market access and foreign competition: The patterns of french exporters. *the world bank economic review*, *27*(1), 80–108.

Békés, G., & Muraközy, B. (2012). Temporary trade and heterogeneous firms. *Journal of International Economics*, *87*(2), 232–246.

Belloni, A., Chernozhukov, V., et al. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, *19*(2), 521–547.

Belloni, A., Chernozhukov, V., Fernández-Val, I., & Hansen, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica*, *85*(1), 233–298.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, *28*(2), 29–50.

Belloni, A., Chernozhukov, V., Hansen, C., & Kozbur, D. (2016). Inference in high-dimensional panel models with an application to gun control. *Journal of Business & Economic Statistics*, *34*(4), 590–605.

Bergstrand, J. H. (1985). The gravity equation in international trade: Some microeconomic foundations and empirical evidence. *The review of economics and statistics*, 474–481.

Bernard, A. B., & Jensen, J. B. (1999). Exceptional exporter performance: Cause, effect, or both? *Journal of international economics*, *47*(1), 1–25.

Bernard, A. B., & Jensen, J. B. (2004). Why some firms export. *Review of economics and Statistics*, *86*(2), 561–569.

Bernard, A. B., Jensen, J. B., & Lawrence, R. Z. (1995). Exporters, jobs, and wages in us manufacturing: 1976-1987. *Brookings papers on economic activity. Microeconomics*, *1995*, 67–119.

Bernard, A. B., Jensen, J. B., Redding, S. J., & Schott, P. K. (2007). Firms in international trade. *Journal of Economic perspectives*, *21*(3), 105–130.

Bernard, A. B., Jensen, J. B., Redding, S. J., & Schott, P. K. (2012). The empirics of firm heterogeneity and international trade. *Annual Review of Economics*, *4*(1), 283–313.

Bernard, A. B., Redding, S. J., & Schott, P. K. (2010). Multiple-product firms and product switching. *American economic review*, *100*(1), 70–97.

Bernard, A. B., Redding, S. J., & Schott, P. K. (2011). Multiproduct firms and trade liberalization. *The Quarterly journal of economics*, *126*(3), 1271–1318.

Bia, M., Flores, C. A., Flores-Lagunes, A., & Mattei, A. (2014). A stata package for the application of semiparametric estimators of dose–response functions. *The Stata Journal*, *14*(3), 580–604.

Bleich, J., Kapelner, A., George, E. I., & Jensen, S. T. (2014). Variable selection for bart: An application to gene regulation. *The Annals of Applied Statistics*, *8*(3), 1750–1781.

Bluwstein, K., Buckmann, M., Joseph, A., Kapadia, S., & Şimşek, Ö. (2023). Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. *Journal of International Economics*, *145*, 103773. https://doi.org/https://doi.org/10.1016/j.jinteco.2023.103773

Bogliacino, F., & Pianta, M. (2016). The pavitt taxonomy, revisited: Patterns of innovation in manufacturing and services. *Economia Politica*, *33*, 153–180.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. *Belmont, CA: Wadsworth & Brooks*.

Breinlich, H., Corradi, V., Rocha, N., Ruta, M., Santos Silva, J., & Zylkin, T. (2022). Machine learning in international trade research-evaluating the impact of trade agreements.

Breinlich, H., Corradi, V., Rocha, N., Ruta, M., Santos Silva, J., & Zylkin, T. (2021). *Machine learning in gravity models: An application to agricultural trade* (Research Working Paper No. 9629). World Bank.

Butler, A. W., & Cornaggia, J. (2011). Does access to external finance improve productivity? evidence from a natural experiment. *Journal of Financial Economics*, *99*(1), 184–203.

Caballero, R. J., Hoshi, T., & Kashyap, A. K. (2008). Zombie lending and depressed restructuring in japan. *American Economic Review*, *98*(5), 1943–77.

Candes, E., & Recht, B. (2012). Exact matrix completion via convex optimization. *Communications of the ACM*, *55*(6), 111–119.

Candes, E. J., & Plan, Y. (2010). Matrix completion with noise. *Proceedings of the IEEE*, *98*(6), 925–936.

Cerulli, G. (2015). Ctreatreg: Command for fitting dose–response models under exogenous and endogenous treatment. *The Stata Journal*, *15*(4), 1019–1045.

Chalfin, A., Danieli, O., Hillis, A., Jelveh, Z., Luca, M., Ludwig, J., & Mullainathan, S. (2016). Productivity and selection of human capital with machine learning. *American Economic Review*, *106*(5), 124–27. https://doi.org/10.1257/aer.p20161029

Chandra, A., Flack, E., & Obermeyer, Z. (2024). The health costs of cost-sharing. *The Quarterly Journal of Economics*, qjae015.

Chen, H. (, & Chen, S. ((2012). Investment-cash flow sensitivity cannot be a good measure of financial constraints: Evidence from the time series. *Journal of Financial Economics*, *103*(2), 393–410.

Chen, J., & Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, *95*(3), 759–771.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.

Chernozhukov, V., Wüthrich, K., & Zhu, Y. (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. *Journal of the American Statistical Association*, *116*(536), 1849–1864.

Chipman, H. A., George, E. I., McCulloch, R. E., et al. (2010). Bart: Bayesian additive regression trees. *The Annals of Applied Statistics*, *4*(1), 266–298.

Chor, D., & Manova, K. (2012). Off the cliff and back? credit conditions and international trade during the global financial crisis [Sympo-

sium on the Global Dimensions of the Financial Crisis]. *Journal of International Economics*, *87*(1), 117–133.
Clerides, S. K., Lach, S., & Tybout, J. R. (1998). Is learning by exporting important? micro-dynamic evidence from colombia, mexico, and morocco. *The Quarterly Journal of Economics*, *113*(3), 903–947.
Cravino, J., & Levchenko, A. A. (2016). Multinational Firms and International Business Cycle Transmission*. *The Quarterly Journal of Economics*, *132*(2), 921–962.
Crespi, G., Criscuolo, C., & Haskel, J. (2008). Productivity, exporting, and the learning-by-exporting hypothesis: Direct evidence from uk firms. *The Canadian Journal of Economics / Revue canadienne d'Economique*, *41*(2), 619–638.
Crowley, M., & Han, L. (2022, July). The pro-competitive effects of trade agreements.
Crozet, M., Head, K., & Mayer, T. (2012). Quality sorting and trade: Firm-level evidence for french wine. *The Review of Economic Studies*, *79*(2), 609–644.
De Chaisemartin, C., & d'Haultfoeuille, X. (2023). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The Econometrics Journal*, *26*(3), C1–C30.
De Loecker, J. (2007). Do exports generate higher productivity? evidence from slovenia. *Journal of international economics*, *73*(1), 69–98.
De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics*, *5*(3), 1–21.
De Loecker, J., & Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, *102*(6), 2437–71.
Del Prete, D., & Rungi, A. (2017). Organizing the global value chain: A firm-level test. *Journal of International Economics*, *109*, 16–30.
Del Prete, D., & Rungi, A. (2018). The smile curve at the firm level: Where value is added along supply chains. *Economics Letters*, *164*, 38–42.
Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., & Reif, J. (2019). The mortality and medical costs of air pollution: Evidence from

changes in wind direction. *American Economic Review*, *109*(12), 4178–4219. https://doi.org/10.1257/aer.20180279

D'Haultfœuille, X., Hoderlein, S., & Sasaki, Y. (2023). Nonparametric difference-in-differences in repeated cross-sections with continuous treatments. *Journal of Econometrics*, *234*(2), 664–690.

Dhingra, S. (2013). Trading away wide brands for cheap brands. *American Economic Review*, *103*(6), 2554–84.

Dhyne, E., Ludwig, P., & Vandenbussche, H. (2023). *Export entry and network interactions: Evidence from the belgian production network* (tech. rep.). National Bank of Belgium.

Dobbie, W., Liberman, A., Paravisini, D., & Pathania, V. (2021). Measuring bias in consumer lending. *The Review of Economic Studies*, *88*(6), 2799–2832.

Dubé, J.-P., & Misra, S. (2023). Personalized pricing and consumer welfare. *Journal of Political Economy*, *131*(1), 131–189.

Eaton, J., & Fieler, A. C. (2019). *The margins of trade* (tech. rep.). National Bureau of Economic Research.

Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, *70*(5), 1741–1779.

Eaton, J., Kortum, S., & Kramarz, F. (2004). Dissecting trade: Firms, industries, and export destinations. *American Economic Review*, *94*(2), 150–154.

Eaton, J., Kortum, S., & Kramarz, F. (2011). An anatomy of international trade: Evidence from french firms. *Econometrica*, *79*(5), 1453–1498.

Eaton, J., Kortum, S. S., & Sotelo, S. (2012). *International trade: Linking micro and macro* (tech. rep.). National bureau of economic research.

Eckel, C., & Neary, J. P. (2010). Multi-product firms and flexible manufacturing in the global economy. *The Review of Economic Studies*, *77*(1), 188–217.

Egger, P. H., & Tarlea, F. (2021). Comparing apples to apples: Estimating consistent partial effects of preferential economic integration agreements. *Economica*, *88*(350), 456–473.

Ellison, G., & Glaeser, E. (1997). Geographic concentration in u.s. manufacturing industries: A dartboard approach. *Journal of Political Economy*, *105*(5), 889–927.

Elmes, M. B., & Kasouf, C. j. (1995). Knowledge workers and organizational learning: Narratives from biotechnology. *Management Learning*, *26*(4), 403–422.

Fazzari, S. M., Hubbard, R. G., & Petersen, B. C. (1988). Financing Constraints and Corporate Investment. *Brookings Papers on Economic Activity*, *19*(1), 141–206.

Feenstra, R., & Ma, H. (2007). *Optimal choice of product scope for multiproduct firms under monopolistic competition* (tech. rep.). National Bureau of Economic Research.

Feenstra, R. C., Markusen, J. R., & Rose, A. K. (2001). Using the gravity equation to differentiate among alternative theories of trade. *Canadian Journal of Economics/Revue canadienne d'économique*, *34*(2), 430–447.

Fontagné, L., Guimbard, H., & Orefice, G. (2022). Tariff-based product-level trade elasticities. *Journal of International Economics*, *137*, 103593.

Fontagné, L., Micocci, F., & Rungi, A. (2024). The heterogeneous impact of the EU-Canada agreement with causal machine learning. *arXiv preprint arXiv:2407.07652*.

Fontagné, L., Secchi, A., & Tomasi, C. (2018). Exporters' product vectors across markets. *European Economic Review*, *110*, 150–180.

Fryges, H., & Wagner, J. (2008). Exports and productivity growth: First evidence from a continuous treatment approach. *Review of World Economics*, *144*, 695–722.

Gaulier, G., Santoni, G., Taglioni, D., & Zignago, S. (2013). *In the wake of the global crisis. evidence from a new quarterly database of export competitiveness* (Policy Research Working Paper No. 9629). World Bank.

Geishecker, I., Schröder, P. J., & Sörensen, A. (2019). One-off export events. *Canadian Journal of Economics/Revue canadienne d'économique*, *52*(1), 93–131.

Geman, S., & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6), 721–741.

Gentzkow, M., Shapiro, J. M., & Taddy, M. (2019). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica*, *87*(4), 1307–1340.

Gkypali, A., Love, J. H., & Roper, S. (2021). Export status and sme productivity: Learning-to-export versus learning-by-exporting. *Journal of Business Research*, *128*, 486–498.

Gnecco, G., Nutarelli, F., & Riccaboni, M. (2023). Matrix completion of world trade: An analysis of interpretability through shapley values. *The World Economy*, *46*(9), 2707–2731.

Goldberg, P. K., & Pavcnik, N. (2016). The effects of trade policy. In *Handbook of commercial policy* (pp. 161–206, Vol. 1). Elsevier.

Gopinath, G., Kalemli-Özcan, Ş., Karabarbounis, L., & Villegas-Sanchez, C. (2017). Capital allocation and productivity in south europe. *The Quarterly Journal of Economics*, *132*(4), 1915–1967.

Gopinath, M., Batarseh, F. A., & Beckman, J. (2020). *Machine learning in gravity models: An application to agricultural trade* (Working Paper No. 27151). National Bureau of Economic Research.

Gordeev, S., & Steinbach, S. (2024). Determinants of pta design: Insights from machine learning. *International Economics*, *178*, 100504.

Greenaway, D., & Kneller, R. (2008). Exporting, productivity and agglomeration. *European economic review*, *52*(5), 919–939.

Grossman, G. M., & Helpman, E. (1990). Comparative advantage and long-run growth. *The American Economic Review*, *80*(4), 796–815.

Grossman, G. M., & Helpman, E. (1991). Trade, knowledge spillovers, and growth. *European economic review*, *35*(2-3), 517–526.

Hadlock, C. J., & Pierce, J. R. (2010). New evidence on measuring financial constraints: Moving beyond the kz index. *The review of financial studies*, *23*(5), 1909–1940.

Handel, B., & Kolstad, J. (2017). Wearable technologies and health behaviors: New data and new methods to understand population health. *American Economic Review*, *107*(5), 481–85. https://doi.org/10.1257/aer.p20171085

Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, *57*, 97–109.

Head, K., & Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In *Handbook of international economics* (pp. 131–195, Vol. 4). Elsevier.

Head, K., & Ries, J. (1998). Immigration and trade creation: Econometric evidence from canada. *Canadian journal of economics*, 47–62.

Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *The quarterly journal of economics*, *123*(2), 441–487.

Helpman, E., Melitz, M. J., & Yeaple, S. R. (2004). Export versus fdi with heterogeneous firms. *American economic review*, *94*(1), 300–316.

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

Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, *226164*, 73–84.

Hottman, C. J., Redding, S. J., & Weinstein, D. E. (2016). Quantifying the Sources of Firm Heterogeneity *. *The Quarterly Journal of Economics*, *131*(3), 1291–1364.

Hübner, K., Deman, A.-S., & Balik, T. (2017). Eu and trade policy-making: The contentious case of ceta. *Journal of European Integration*, *39*(7), 843–857.

Iacovone, L., & Javorcik, B. S. (2010). Multi-product exporters: Product churning, uncertainty and export discoveries. *The Economic Journal*, *120*(544), 481–499.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of economic literature*, *47*(1), 5–86.

INSEE. (2023). *Les opérateurs du commerce extérieur*.

Jaax, A., Mourougane, A., & Gonzales, F. (2024). Nowcasting services trade for the g7 economies. *The World Economy*, *47*(4), 1336–1386.

Joseph, A. (2020). *Parametric inference with universal function approximators* [Bank of England Staff Working Paper no. 784].

Joy, M., Rusnák, M., Šmídková, K., & Vašíček, B. (2017). Banking and currency crises: Differential diagnostics for developed countries. *International Journal of Finance & Economics*, *22*(1), 44–67.

Kapelner, A., & Bleich, J. (2013). Bartmachine: Machine learning with bayesian additive regression trees. *arXiv preprint arXiv:1312.2171*.

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

Khilji, S. E., Mroczkowski, T., & Bernstein, B. (2006). From invention to innovation: Toward developing an integrated innovation model for biotech firms. *Journal of product innovation management*, *23*(6), 528–540.

Kim, D., & Steinbach, S. (2023). Preferential trading in agriculture: New insights from a structural gravity analysis and machine learning. *Available at SSRN 4412389*.

Kluve, J., Schneider, H., Uhlendorff, A., & Zhao, Z. (2012). Evaluating continuous training programmes by using the generalized propensity score. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *175*(2), 587–617.

Kneller, R., & Pisu, M. (2007). Industrial linkages and export spillovers from fdi. *World Economy*, *30*(1), 105–134.

Larch, M., & Yotov, Y. V. (2023, February). Estimating the effects of trade agreements: Lessons from 60 years of methods and data.

Leamer, E., & Stern, R. (1970). Quantitative international economics. Allin; Bacon, Boston.

Liang, Y., Shi, K., Tao, H., & Xu, J. (2024). Learning by exporting: Evidence from patent citations in china. *Journal of International Economics*, *150*, 103933.

Limão, N. (2016). Preferential trade agreements. In *Handbook of commercial policy* (pp. 279–367, Vol. 1). Elsevier.

Limodio, N. (2022). Terrorism financing, recruitment, and attacks. *Econometrica*, *90*(4), 1711–1742.

Lin, F. (2015). Learning by exporting effect in china revisited: An instrumental approach. *China Economic Review*, *36*, 1–13.

Lin, F. (2017). Credit constraints, export mode and firm performance: An investigation of china's private enterprises. *Pacific Economic Review*, *22*(1), 123–143.

Liu, X. (2012). Classification accuracy and cut point selection. *Statistics in medicine*, *31*(23), 2676–2686.

López, R. A. (2005). Trade and growth: Reconciling the macroeconomic and microeconomic evidence. *Journal of Economic surveys*, *19*(4), 623–648.

Love, J. H., & Roper, S. (2015). Sme innovation, exporting and growth: A review of existing evidence. *International small business journal*, *33*(1), 28–48.

Magee, C. S. (2003). Endogenous preferential trade agreements: An empirical analysis. *Contributions in Economic Analysis & Policy*, *2*(1), 1–17.

Manova, K. (2012). Credit Constraints, Heterogeneous Firms, and International Trade. *The Review of Economic Studies*, *80*(2), 711–744.

Mayer, T., Melitz, M. J., & Ottaviano, G. I. (2014). Market size, competition, and the product mix of exporters. *American Economic Review*, *104*(2), 495–536.

Mayer, T., Melitz, M. J., & Ottaviano, G. I. (2021). Product mix and firm productivity responses to trade competition. *Review of Economics and Statistics*, *103*(5), 874–891.

Mazumder, R., Hastie, T., & Tibshirani, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *The Journal of Machine Learning Research*, *11*, 2287–2322.

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

Melitz, M. J., & Ottaviano, G. I. P. (2008). Market Size, Trade, and Productivity. *The Review of Economic Studies*, *75*(1), 295–316.

Melitz, M. J., & Redding, S. J. (2014). Chapter 1 - heterogeneous firms and trade. In G. Gopinath, E. Helpman, & K. Rogoff (Eds.), *Handbook of international economics* (pp. 1–54, Vol. 4). Elsevier.

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

Micocci, F., & Rungi, A. (2023). Predicting exporters with machine learning. *World Trade Review*, *22*(5), 584–607.

Mrázová, M., & Neary, J. P. (2017). Not so demanding: Demand structure and firm behavior. *American Economic Review*, *107*(12), 3835–3874.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Nagengast, A., & Yotov, Y. V. (2024). Staggered difference-in-differences in gravity settings: Revisiting the effects of trade agreements. *American Economic Journal: Applied Economics*.

Nickell, S., & Nicolitsas, D. (1999). How does financial pressure affect firms? *European Economic Review*, *43*(8), 1435–1456.

Pavitt, K. (1984). Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research policy*, *13*(6), 343–373.

Pla-Barber, J., & Alegre, J. (2007). Analysing the link between export intensity, innovation and firm size in a science-based industry. *International business review*, *16*(3), 275–293.

Qiu, L. D., & Zhou, W. (2013). Multiproduct firms and scope adjustment in globalization. *Journal of International Economics*, *91*(1), 142–153.

Reis, J. G., Wagle, S., & Farole, T. (2010). *Analyzing trade competitiveness : A diagnostics approach*. The World Bank.

Richardson, J. (1971a). Constant-market-shares analysis of export growth. *Journal of International Economics*, *1*(2), 227–239.

Richardson, J. (1971b). Some sensitivity tests for a "constant-market-shares" analysis of export growth. *The Review of Economics and Statistics*, *53*(3), 300–304.

Rivera-Batiz, L. A., & Romer, P. M. (1991). International trade with endogenous technological change. *European Economic Review*, *35*(4), 971–1001.

Roberts, M. J., & Tybout, J. R. (1995). *An empirical model of sunk costs and the decision to export* (Vol. 1436). World Bank Publications.

Roberts, M. J., & Tybout, J. R. (1997). The decision to export in colombia: An empirical model of entry with sunk costs. *The American Economic Review*, *87*(4), 545–564.

Romer, P. (1994). New goods, old theory, and the welfare costs of trade restrictions. *Journal of Development Economics*, *43*(1), 5–38.

Rubin, D. B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association*, *100*(469), 322–331.

Soloaga, I., & Wintersb, L. A. (2001). Regionalism in the nineties: What effect on trade? *The North American Journal of Economics and Finance*, *12*(1), 1–29.

Tybout, J. R. (2000). Manufacturing firms in developing countries: How well do they do, and why? *Journal of Economic literature*, *38*(1), 11–44.

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

Van Biesebroeck, J. (2005). Exporting raises productivity in sub-saharan african manufacturing firms. *Journal of International economics*, *67*(2), 373–391.

Van Biesebroeck, J., Konings, J., & Volpe Martincus, C. (2016). Did export promotion help firms weather the crisis? *Economic Policy*, *31*(88), 653–702.

Van den Berg, M., Boutorat, A., Franssen, L., & Mounir, A. (2022). Intermittent exporting: Unusual business or business as usual? *Review of World Economics*, *158*(4), 1173–1198.

Viner, J. (1950). *The customs union issue*. Carnegie Endowment for International Peace, New York.

Volpe Martincus, C., & Carballo, J. (2008). Is export promotion effective in developing countries? firm-level evidence on the intensive and the extensive margins of exports. *Journal of International Economics*, *76*(1), 89–106.

Volpe Martincus, C., & Carballo, J. (2010a). Export promotion: Bundled services work better. *The World Economy*, *33*(12), 1718–1756.

Volpe Martincus, C., & Carballo, J. (2010b). Entering new country and product markets: does export promotion help? *Review of World Economics*, *146*(3), 437–467.

Volpe Martincus, C., Estevadeordal, A., Gallo, A., & Luna, J. (2010). Information barriers, export promotion institutions, and the extensive margin of trade. *Review of World Economics*, *146*(1), 91–111.

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242.

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

Wang, F., Milner, C., & Scheffel, J. (2022). Export destination and the skill premium: Evidence from chinese manufacturing industries. *Canadian Journal of Economics/Revue canadienne d'économique*, *55*(2), 1057–1094.

Westerlund, J., & Wilhelmsson, F. (2011). Estimating the gravity model without gravity using panel data. *Applied Economics*, *43*(6), 641–649.

Yang, S., & Martinez-Zarzoso, I. (2014). A panel data analysis of trade creation and trade diversion effects: The case of asean–china free trade area. *China Economic Review*, *29*, 138–151.

Zhao, S., & Jin, M. (2020). Globalization, innovation, and productivity. *Handbook of Production Economics*, 1–19.