# Assessing the Impact of COVID-19 on Trade: a Machine Learning Counterfactual Analysis

Marco Dueñas[1], Federico Nutarelli[2], Víctor Ortiz*[2], Massimo Riccaboni[2], and Francesco Serti[2]

[1]Universidad de Bogotá Jorge Tadeo Lozano, Bogotá (Colombia)
[2]IMT School for Advanced Studies, Lucca (Italy)

**Abstract**

By interpreting exporters' dynamics as a complex learning process in a context of a generalized shock (COVID-19) and building on recent development in (causal) Machine Learning (ML), this paper investigates the effectiveness of different ML techniques in predicting firms' trade status and in using these predictions to reconstruct the counterfactual distribution of firm's outcomes in absence of the COVID. We focus on the probability of Colombian firms surviving in the export market under two different scenarios: a COVID-19 setting and a non-COVID-19 counterfactual situation. By comparing the resulting predictions, we estimate the distribution of the treatment effects of the COVID-19 shock on firms' outcomes. On average, we find that the COVID-19 shock decreased a firm's probability of surviving in the export market by about 20p.p. in April 2020. We study treatment effect heterogeneity by employing a Classification Analysis of the Differences (CADiff) that compares the characteristics of the firms on the tails of the estimated distribution of the individual treatment effects. We find evidence that focusing on the tails of the distribution of the treatment effects is critical to correct for the estimation error arising from the necessarily imperfect reconstruction of the unobservable counterfactual.

**Keywords:** Machine Learning; International Trade; COVID-19; Counterfactual
**JEL Codes:** F14; F17; D22; L25

## 1 Introduction

The COVID-19 outbreak has affected the world economy, generating unprecedented health, human, and economic crises. To face the health crisis, governments implemented social distancing and lockdown policies, exacerbating supply and demand shocks (World Bank,

*Corresponding author: victor.ortizgimenez@imtlucca.it. Laboratory for the Analysis of Complex Economic Systems, IMT School for Advanced Studies, Piazza San Francesco 19 - 55100 Lucca, Italy

2020). In a highly interconnected world, the impact of the pandemic on international trade has generated great attention (Felbermayr and Görg, 2020; Antràs et al., 2020). International trade is being affected by national lockdowns, trade and trade-related measures adopted by countries, and by the temporal disruption of global value chains (Bonadio et al., 2020; Evenett, 2020). Global trade, which is typically more volatile than output and tends to fall particularly sharply during a crisis, has shown the biggest fall since the 2009 global financial crisis. From the beginning of the COVID-19 epidemics, scholars underlined that, though its impact on international trade could have been comparable to the Great Trade Collapse of 2008-2009, this time, the demand side shock is accompanied by a supply-side shock (Baldwin and Tomiura, 2020). Moreover, this supply-side effect could be reinforced by a supply-side contagion via importing/supply chains, which have grown in relevance during the last decade. In other words, supply disruptions in the countries providing intermediate inputs to a given country are likely to hurt also its export performance.

This paper aims to estimate the causal effect of the COVID-19 shock on a firm's probability of survival in the export markets, and to study the heterogeneity of this effect. The main hurdles for this evaluation task are related to the pervasiveness of the COVID-19 shock. Indeed, the fact that all firms are directly and/or indirectly exposed to the effects of COVID-19 crisis makes it hardly possible to find a control group of firms to be used to build a counterfactual non-COVID-19 scenario. Moreover, identifying the main patterns through which the COVID-19 shock has affected firm-level trade is a demanding task because the economy-wide impact of the shock is coupled with complex interdependencies between firms and products belonging to different sectors and countries, as underlined above.

By interpreting exporters' dynamics as a complex learning process,[1] this paper's first contribution is exploring and comparing the effectiveness of different Machine Learning (ML) techniques in predicting firms' trade status in two different scenarios, a COVID-19 and a non-COVID-19 setting. ML techniques have been successfully applied to predict firm performances and help companies (and public agencies) in their decision-making in complex environments. The accumulated literature shows that ML techniques' ability to classify companies is high and reliable in such high-dimensional contexts (Bargagli-Stoffi et al., 2020). Our paper fits into a nascent literature that is applying ML techniques to study international trade patterns (Breinlich et al., 2021) and, up to what we know, in our study for the first time that ML techniques are used to predict firm-level international trade performance and to estimate causal parameters.

This paper's second contribution is to use these predictions to estimate the causal effect of the COVID-19 shock at the individual firm level. We use the estimated ML

[1] Firms have heterogeneous and incomplete information about the trade opportunities. This is true both on the exporting and the importing side of firm activities. For example, in Albornoz et al. (2012) and Eslava et al. (2015) exporting firms are uncertain and learn about the appeal of their products and, more in general, about the profitability of exporting their products on the international markets. By searching for clients and observing their realized profitability, firms update their beliefs about their capabilities in international markets.

model with the best performance in predicting the 2019 export status of firms exporting in 2018 to build a 2020 non-COVID-19 counterfactual outcome for firms exporting in 2019. Then, we compare these counterfactual non-COVID-19 firm-level export probabilities with the predicted probabilities of the best performing ML model using the characteristics of 2019 exporters to predict their export status in 2020. The latter estimated probabilities summarize the information on the observed COVID-19 scenario and express it in a metric that is comparable with the estimated counterfactual non-COVID-19 outcomes. In the literature using ML counterfactuals (Cerqua and Letta, 2020; Fabra et al., 2020), it is instead common to estimate causal effects by comparing them with the observed outcome in case of treatment, following the so called the "Consistency Assumption": if the outcome in case of treatment is observed, than it also represents the potential outcome under treatment.

Finally, we employ ML techniques to study the heterogeneity of the estimated COVID-19 effects according to firms' characteristics. ML has been proved to be helpful in such high dimensional settings to individuate subgroups, which are particularly responsive to the treatment and, therefore, to identify the most relevant dimensions of the heterogeneity of a treatment. Different ML tools have been used in the literature with a trade-off between precision and interpretability: decision-tree based algorithms, ensemble of trees, Bayesian ensemble of trees, doubly robust approaches, LASSO-based approaches, or meta-learners (Athey and Imbens, 2017; Dominici et al., 2020). We interpret the estimated effects stemming from our ML counterfactual empirical model by using the Classification Analysis of the Differences (CADiff) proposed by Chernozhukov et al. (2018). We show that estimating the treatment effect by using our proposed methodology is fundamental to meaningfully compare the characteristics of units that are more affected with those that are less affected by the treatment. We suggest that to focus on the tails of the distribution of the treatment effects is important to correct for the estimation error arising from the necessarily imperfect reconstruction of the unobservable counterfactual.

We focus on Colombian exporters because of the availability of Colombian Customs data for 2020 and previous years. Similar to many other countries, in 2020, Colombia has witnessed domestic supply and demand shocks related to factory closures, cessation of some public services, and disruptions in the supply chain at home and abroad. de Lucio et al. (2020) found that Spanish exports decreased more in destinations that introduced strict policies to contain COVID-19, particularly between March and May 2020, showing how in Spain export performance during the pandemic depends on COVID-19 induced demand shocks in export markets. Using a sector-level gravity model, Espitia et al. (2021) show that, during the COVID-19 crisis, sectors that tend to be relatively less internationally integrated suffered less from foreign shocks but were more vulnerable to domestic shocks.

The paper is organized as follows. Section 2 briefly describes the Colombian context. Section 3 presents the firm-level data, variables employed in the analysis, and descriptive statistics. Section 4 explains the empirical strategy. Section 5 reports the main estimation results, and section 6 summarizes the findings and discusses both interpretation and limitations of the analysis.

## 2 The Colombian economy amidst the COVID-19 crisis

Colombia is a country that exports little compared to other countries in Latin America with similar development levels. In recent years, the share of total exports of Colombian GDP has oscillated around 15%, well below other countries in the region that practically double this measure, such as Chile and Mexico.

Although the Colombian economy was relatively closed during most of the twentieth century (Ocampo and Tovar, 2000), it has been strongly affected by international crises, as the global financial crisis in 2008-2009 (Zuluaga et al., 2009). The Colombian openness started in the 1990s with several market-oriented reforms aiming at liberalizing financial and capital markets. Nowadays, Colombia has 16 bilateral trade agreements in force. Even though Colombia increased the number of trade partners and the value and volume of trade, the integration into world trade markets is still modest (Cepeda-López et al., 2019).

An essential reason behind Colombia's poor performance is that its export basket exhibits a low diversification level, with a prevalence of primary products, because of the relative abundance of natural resources and low-skilled labor. Besides, the emergence of raw products derived from mining has gained a larger share in total exports, reducing the importance of other products that have been successful, such as coffee, bananas, flowers, some labor-intensive manufactures, and petrochemicals. Bruno et al. (2018) analyzed the export diversification patterns of Colombian manufacturing firms using a product-firm approach (bipartite network analysis). They show that manufacturing firms can be grouped in clusters with a modular structure, meaning that the groups of firms reveal specialization in products that require similar capabilities. Interestingly, these clusters are characterized by a hierarchical structure so that some firms can export a wide range of products, exploiting their economies of scope. On the other side, most of the firms are more specialized, exporting a limited number of products.

Since the outbreak of the COVID-19 pandemic, Colombia implemented early measures to contain the spread of COVID-19 and prepare the health system and mitigate the economic and social impact. The Colombian government issued non-compulsory requests for remote working to private companies on February 24; schools and universities were closed on March 16. On March 25, when there were less than a dozen deaths, the government implemented a complete and mandatory lockdown until April 13. During this period, only a few essential activities – such as health services, public services, communications, banking and financial services, food production, pharmaceuticals, and cleaning and disinfection products – were excluded.

The partial lockdown implementation–between April 27 and May 11–allowed a gradual restoration of mobility, enabling a set of non-essential activities under security guidelines and protocols to guarantee social distancing. Most manufacturing activities were gradually allowed at this stage, while non-authorized activities were restricted to market their products through electronic commerce platforms. Finally, from May 28, restrictions to the services sector have been lifted, and on September 1, the government announced the end of confinement, and airports were opened.

To better cope with the emergency, Colombian authorities have introduced transitory provisions to secure international trade of essential products. Along with the lockdown measures, medicines, supplies, and equipment in the health sector had zero-tariff for six months. Besides, the export and re-export of these products were forbidden. There was a zero-tariff from April 7 to June 30 for raw materials such as maize, sorghum, soybeans, and soybean cake.

The impact of lockdown policies on individuals' behavior and firms' activities is likely to be affected by their endogenous responses to the legal restrictions and to be highly heterogeneous, depending on workers' and firms' characteristics. For instance, Dueñas et al. (2021) find that the responses to lockdown policies largely depend on socio-economic conditions, with the part of the population with worse socio-economic conditions showing lower mobility flows decreases. Regarding business activities, for instance, the lockdown could have led to a more significant impact on formal activities than on informal ones, and some industries could have better adapted than others to remote working. More in general, as mentioned in the introduction, the firm-specific exposure to the COVID-19 shock might depend on multiple factors such as the nature of its final products (de Lucio et al., 2020), its size, the importance of economies of scale and scope, the identity of the destination countries of its shipments, and the origins of its intermediate inputs.

# 3 Data and Descriptive Statistics

## 3.1 Data

To investigate the impact of the COVID-19 pandemic on Colombian firms we use monthly export transactions data reported at the Colombian Customs Office (Dirección de Impuestos y Aduanas Nacionales, DIAN) for 2018, 2019, and 2020. For each transaction, we consider the exporter ID as the firm identifier; the date; a 10-digit Harmonized System code (HS) characterizing the product; the product origin within Colombia (department level); the means of transportation of the shipment; the country of destination; and, the free on board value of the export transaction in US dollars. This data set also contains information related to the value and origin country from which a given exporter is importing from. We remove all transactions related to re-exports of products elaborated in other countries. As a result, we end up with 386,132 customs reports in 2018, 402,140 in 2019, and 365,626 in 2020.

In our analysis, we classify products at the six-digit level of the HS-code. We consider different features of exporters, according to their monthly exports: the total export (and import) value, the number of products ($NP$), the number of export destinations ($ND$), the number of import origin countries ($NO$), the Herfindahl-Hirschman indexes at the product level ($HH_p$) and the destination level ($HH_d$), and a set of dummies for the destinations and origin countries and continents. We create a set of dummies according the Colombian-department from which the product comes from, a set of dummies for the means of transportation used, a set of the dummies classifying the product sector (HS-chapter), and the product industry (HS-section). Moreover, we build two sets of dummy variables indicating whether a firm has experience exporting in specific destinations and product sector in this given month of last year. We also account for the accumulated exporting (importing) experience by summing up the total value exported (imported) during the last twelve months.Furthermore, we create four *size* dummies classifying firms according to the quartiles of the firm-level distribution of the total monthly log-value of exports.

To measure the COVID-19 demand and supply shock, we use the information on government contention measures coming from Hale et al. (2020), which consists of four indexes (ranging from 0 to 100) representing the strength of the measures taken by countries to contain the COVID-19 outbreak. The authors provide an economic index summarizing economic policies ($E$), a health index summarizing health policies ($H$), a government index describing the strictness of 'lockdown style' policies ($G$) and an overall government response index called stringency index ($S$). The value of these indexes ranges from 0 to 100.[2] We build two variables at the firm level for each the four indexes, one at the export and one at the import side, by taking a weighted average of the country level scores according to the proportion of the total monthly value of exports (imports) that a firm ships (source) in each country in 2019. We call these firm-level indexes for firm $i$ "Containment Index$_{i,j,z}$", with $j = \{E, H, G, S\}$ and $z = \{\text{Imp}, \text{Exp}\}$.[3]

Our final data set is composed of 1,975 features. For a summary of all features see Table Appx.1 in Appendix A.

## 3.2 Descriptive Statistics

The left panel in Figure 1 shows the evolution of total monthly exports during 2019 and 2020. The total monthly value of exports in 2020 is significantly lower than the one observed for the corresponding month in 2019, except for January and February. The lockdown measures implemented to contain the COVID-19 outbreak in Colombia and abroad had a severe impact between April and June–the value in April 2020 is half than the one observed in April 2019 (47%).

[2]These indexes are released daily. We average this information at the monthly level.

[3]The value of the Containment Stringency Index Import for firms that are not importing corresponds to the value of the Containment Stringency Index for Colombia (as firms are sourcing all their inputs domestically).

In a typical month, large firms get a lion's share of the total exports. A regular pattern in looking at customs data is that more prominent exporters trade during many months and ship more frequently than smaller firms, which make only a few shipments. The right panel in Figure 1 shows the proportion of surviving exporting firms at year $t$ among those exporting at year $t-1$, by size classes defined at $t-1$. Comparing the figures for 2020 with those for 2019, it seems that the COVID-19 outbreak affected all firms regardless of their size. However, the effect looks proportionally stronger for small firms (Q1 and Q2 of the distribution). In contrast, larger firms are less affected and recover faster to the survival rates observed in 2019.

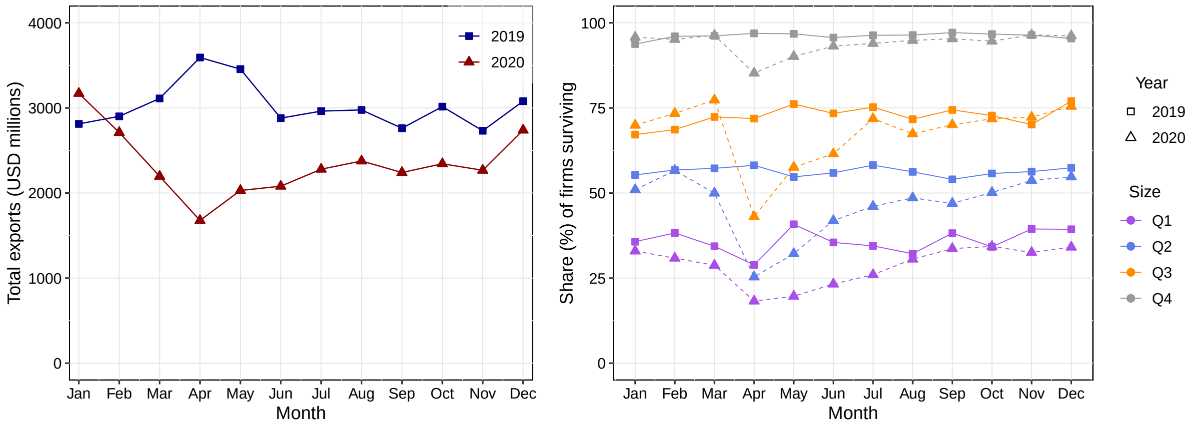


Figure 1: The evolution of total exports (left) and the proportion of surviving exporting firms at year $t$ among those exporting at year $t-1$ within size class at $t-1$ (right). Firm size class derives from the firms' exports (in ln) distribution quartiles in a given month.

Figure 2 shows, separately for the first and second quarter of a year, the percentage of firms that survive, enter, or exit the export market and their corresponding shares of total exports. Thus, for a given quarter in 2019 and the corresponding quarter in 2020, we label each firm as *exiting* when it is present in 2019 and absent in 2020, *entrant* when it is absent in 2019 and present in 2020, and *surviving* when it is present in both years. We average the total value exported by each firm during the same quarter of two different years. Then, we sum the individual average value exported according to the firms' status. It turns that surviving firms play an essential role in explaining total exports: they are around half of the total number of firms in both quarters and account for about 90% of the total export value. The volume lost, during the second quarter of 2020, due to exiting firms is around 5% (assuming they would have exported in 2020 similar export volumes as observed in 2019). Entrant firms almost made up this 5% loss. Despite this, the firms' composition that participates in exports is very different. The number of exiting firms in the second quarter of 2020 is much higher than the share of the first quarter of 2020 and the share of 2019 in the same period of the year.

Figures 3 and 4 show the growth of the total number of exporters and the growth of the total volume of exports between 2019 and 2020, by country of destination and product

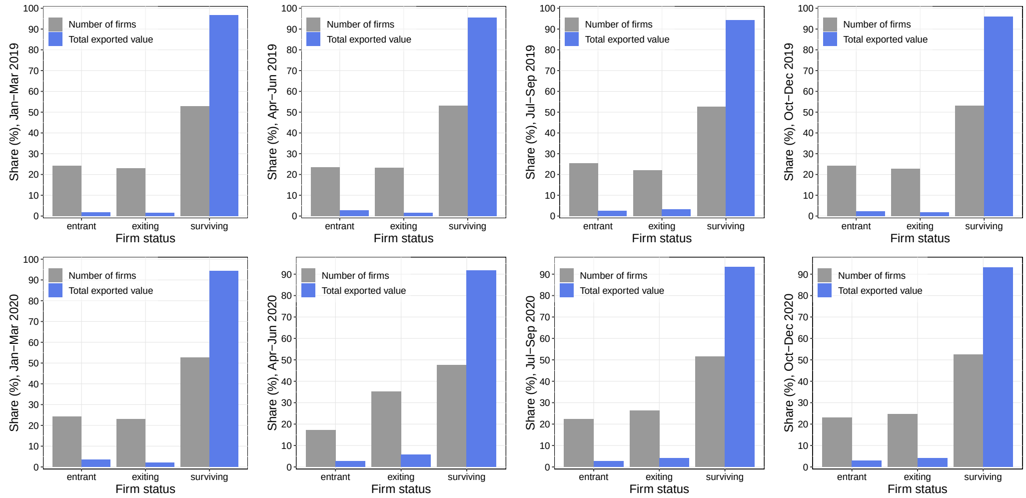

Figure 2: Entry-exit dynamics of firms and total export value by firms that drop, enter or stay active, in 2019 (upper part of the figure) and in 2020 (bottom part of the figure) by quarters. Firm status is defined by looking at the previous year.

sector. We consider the first and the second quarter separately, and we select destinations and product sectors that account for 80% of the total exporters in 2019. In both figures, the product sectors and the destinations are arranged by importance from top to bottom.

Figure 3 shows that the second quarter of 2020 is characterized by a severe and pervasive drop of the number of exporting firms and the volume of exports in almost all the destinations reported. Note that compared to the second quarter, the first quarter export growth exhibits a similar heterogeneity pattern. During the third and fourth quarters the value exported experienced more volatility than the number of firms growth. Nevertheless, the later did not recover to the growth rates of the first quarter of the year. However, growth rates tend to be less extreme and, on average, more stable in the number of exporters and trade volumes.[4]

Exports by product sectors in the second quarter of 2020 (see Figure 4) reveal a generalized decrease in the number of exporting firms and trade values, while the first quarter exhibits very heterogeneous patterns. The sectors that appear to be more severely affected in the second quarter are Footwear (HS64), Leather Articles (HS42), Furniture (HS94), Books (HS49), Articles of Metal (HS83), Knitted and Not-Knitted Accessories (HS61-62), Vehicles (HS87) and Articles of Iron or Steel (HS73). Interestingly, these sectors are relatively more labor-intensive in Colombia, and therefore they could be susceptible to disruptions connected to social distancing. Finally, only for Coffee and Tea (HS08), Other textiles (HS63) and Jewelries (HS71) exports in value significantly grew in the second quarter. Instead, in terms of the number of exporting firms, no product sectors exhibit notable positive dynamics. During the third and the fourth quarter of 2020 there is a rapid back to normality in both the growth of value exported and in the number of exporters growth rate by sector. Figure Appx.2 in the Appendix shows the growth for 2019, suggesting that in periods without strict quarantine – such as the ones of the first quarter of 2020 – the changes in exports are

[4]This evidence matched the observed growth rates for 2019 (see: Figure Appx.1 in the Appendix).

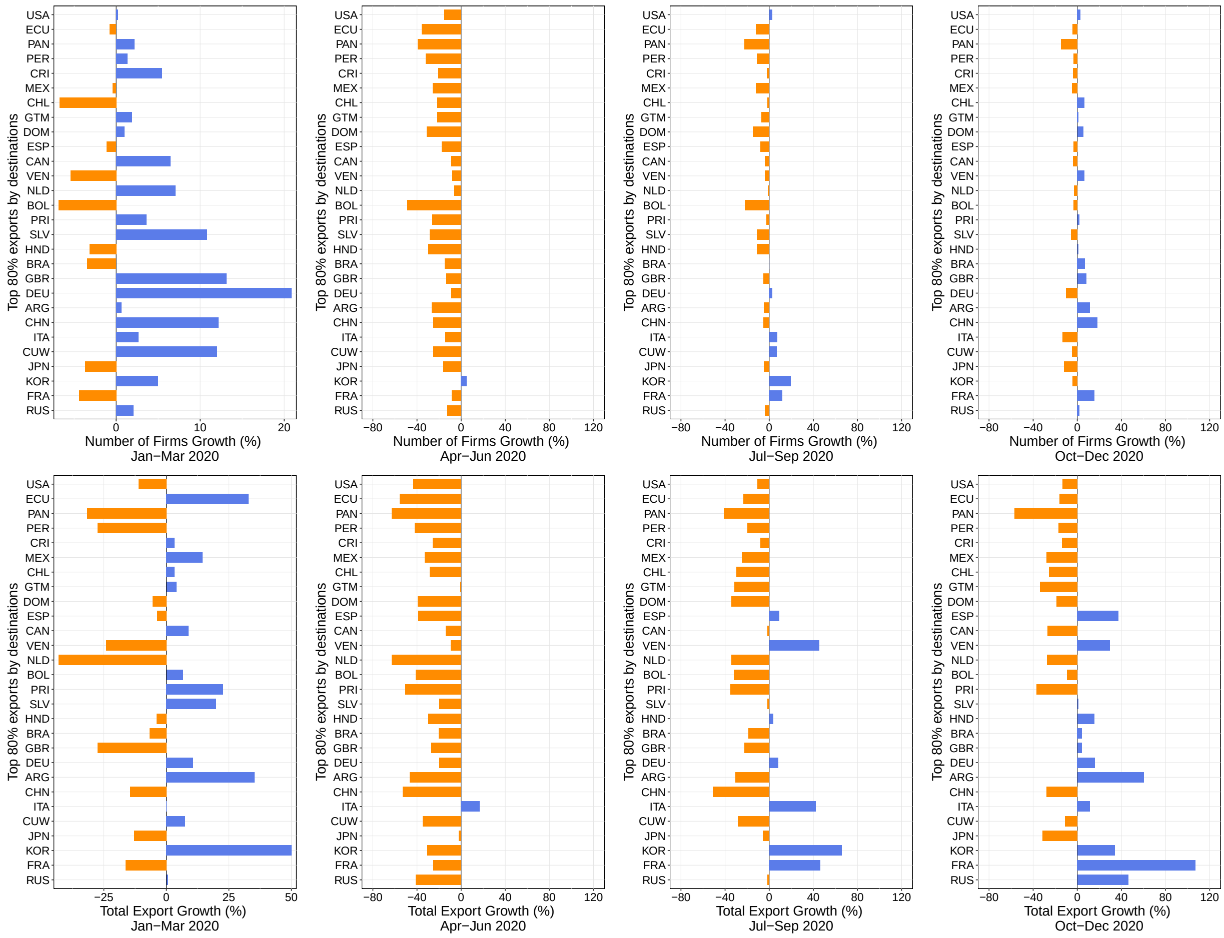


Figure 3: The growth of the total number of exporters and the total value of exports by destination country for the four quarters of 2020. Orange bars represent negative growth and blue bars positive growth. Destination countries are sorted from top to bottom accordingly with their importance in the share of number of exporters in 2019.

also very heterogeneous, but there are not such extreme changes.

In summary, this preliminary evidence suggests that the impact of the COVID-19 shock on Colombian firms' export has been extremely heterogeneous across sectors and destinations.

# 4 Empirical Strategy

This section illustrates our empirical strategy to estimate the effect of the COVID-19 shock on firms' probability of surviving in the export markets, and to study its heterogeneity by firms' observable characteristics.

As in any other evaluation study, the primary identification task is to build a counterfactual outcome, which is not observed, for the treated units. Unfortunately, in considering the effect of the COVID-19 shock, one cannot select any subset of untreated Colombian firms (or if they were available firms of other countries) as a control group because this treatment is affecting, at least indirectly, all firms during 2020. Furthermore, even an identification strategy based on comparing individual firms subject to different intensities of the treatment appears

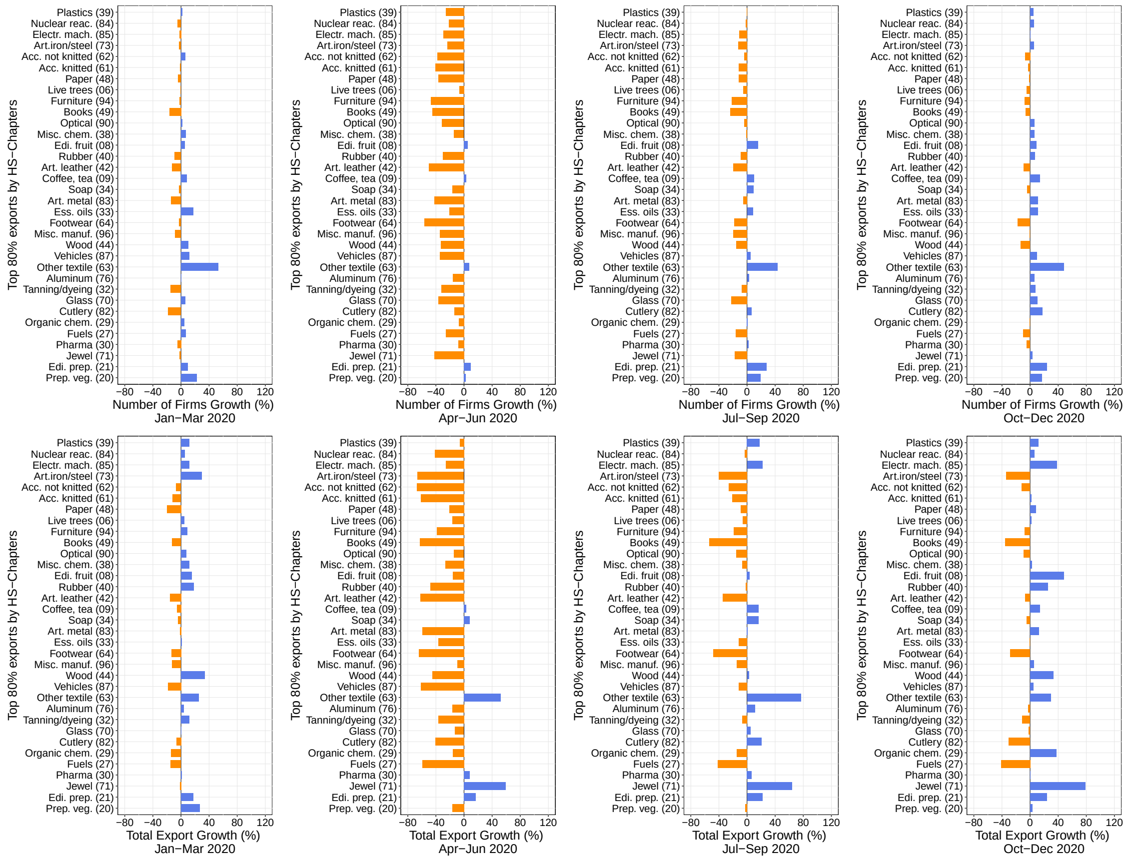


Figure 4: The growth of the total number of exporters and the total value of exports by sector for the four quarters of 2020. Orange bars represent export reductions and blue bars positive export growth. Product sectors are sorted from top to bottom accordingly with their importance in the share of number of exporters in 2019. Product sectors correspond to the chapters of the HS-code in parenthesis, the full name of the chapters is shortened to improve readability.

infeasible due to the complex and ex-ante unknown paths through which firms are potentially exposed to the treatment.[5] In other words, the intensity of treatment might depend on firms' characteristics, such as the identity of suppliers and clients, the characteristics of the traded final product, among many others.

Therefore, as standard in the literature studying the effect of COVID-19, we must resort to using the information on firms' exporting behavior available for periods before the crisis. Following the intuition of Varian (2016), and similarly to the applications of Cerqua and Letta (2020), and Fabra et al. (2020), we use the prediction capabilities of ML techniques to build the counterfactual scenario for the 2020 firms' level outcomes by using pre-pandemic information on firms' export behavior and firms' characteristics. In particular, the outcome (*success*) that we want to predict is whether a company that was exporting in a given month in 2019 will export again in the same month of 2020. The empirical analysis described below is carried out for each month separately to allow the importance of the explanatory variables

[5] A possible identification strategy would be to consider a before-after estimator by comparing firm export behavior during the first quarter of 2020, when presumably firms are still not exposed to the COVID-19 shock, to that of the following quarters. However, this strategy would not properly take into account the strong seasonality of firms' exporting activity.

(e.g., the hypothesized determinants of firm export status) to differ along the year.[6]

Specifically, we denote the potential outcome under the scenario $d \in \{0,1\}$ for firm $i$ at time $t$ as $Y_{it}^d$, where $d$ is an indicator variable for the presence of COVID-19. Also the regressors, $X_{it}^d$, in principle might depend on the presence (absence) of COVID-19. The first step of the analysis is to estimate the counterfactual outcome in 2020: $Y_{i,2020}^0$. We call "Shock Unaware Machine" (SUM) the model that we use to reconstruct this counterfactual (and the counterfactual itself), where the term "Machine" refers to the fact that the counterfactual has been constructed through machine learning techniques. In particular, we will use the outcomes and covariates observed in 2018 and 2019 to reconstruct $Y_{2020}^0$ under the following assumptions (to simplify the notation we will omit the subscript $i$):

(i) Both covariates and outcomes of 2018 and 2019 are not affected by the pandemic:

$$Y_t = Y_t^0 = Y_t^1, \quad X_t = X_t^0 = X_t^1 \quad \text{for} \quad t = 2018, 2019. \tag{1}$$

(ii) Define $Y_t^0 = f_t^0(X_{t-1}^0) + u_t^0$, where $f_t^0(\cdot)$ is a generic model or function representing the relationship between explanatory variables and the outcome in absence of the pandemic such that $\mathbf{E}[Y_t^0|X_{t-1}^0] = f_t^0(X_{t-1}^0)$. Under (i), for $t = 2019$ we have that $Y_{2019} = f_{2019}^0(X_{2018}) + u_{2019}^0$ such that $\mathbf{E}[Y_{2019}|X_{2018}] = f_{2019}^0(X_{2018})$. The second assumption states that the function $f_t^0$ does not depend on $t$, i.e. it is stable over the two considered years:

$$f_{2019}^0 = f_{2020}^0 = f^0 \tag{2}$$

Therefore, under the above assumptions, we can write $Y_{2020}^0 = f^0(X_{2019}) + u_{2020}^0$, such that $\mathbf{E}[Y_{2020}^0|X_{2019}] = f^0(X_{2019})$, and we can use data on 2018 and 2019 to estimate $Y_{2019}^0 = f^0(X_{2018}) + u_{2019}^0$ and retrieve $\hat{f}^0$. By applying this invariant estimated function to the covariates of 2019 we can obtain the predictions for the counterfactual (without COVID-19) outcome in 2020:

$$\hat{Y}_{2020}^0 = \hat{f}^0(X_{2019}) = Y_{2020}^0 - \overbrace{\mathcal{E}_{2020}^0(X_{2019})}^{\textit{Prediction error}} - \overbrace{u_{2020}^0}^{\textit{Orthogonal error}} \tag{3}$$

In general, the estimated counterfactual outcome in 2020, $\hat{Y}_{2020}^0$, will not be a perfect estimate for $Y_{2020}^0$ because $\hat{f}^0$ will not be a perfect estimate of $f^0$ thus producing a prediction error, which in the formula above we have denoted with $\mathcal{E}_{2020}^0(X_{2019}) = f^0(X_{2019}) - \hat{f}^0(X_{2019})$, and because of the existence of other determinants of the outcome that are orthogonal to the covariates, which in the formula above are contained in $u_{2020}^0$. The inaccuracy coming from the estimation of $f^0$, that can vary according to a firm's characteristics $X_{2019}$, will be reduced by experimenting with different ML techniques and using the one associated with

[6] A yearly model with month fixed effects would just allow for different levels of exporting activity for each month.

the best out-of-sample performance.[7] To discriminate between the considered ML techniques we rely on the "K-fold" cross-validation method (with K=5). We divide randomly the set of exporters observed in 2018 (with the exporting success during the same month in 2019 as the outcome) in 5 equally sized groups and obtain the predictions for the firms belonging to a group by estimating $Y_{2019} = f^0(X_{2018}) + u^0_{2019}$ with different ML models on the firms belonging to the other groups. Then we compute the accuracy of the different models for each month and choose the model with the best average performance across months. Notice that this comparison is entirely based on the pre-pandemic accuracy of the ML models by comparing the predictions $\hat{Y}_{2019}$ with the observed $Y_{2019}$, not on its merits in predicting the firms' outcomes in 2020. Finally, we obtain the $\hat{Y}^0_{2020}$ by estimating $Y_{2019} = f^0(X_{2018}) + u^0_{2019}$ on entire set of 2018 exporters (also in this case month by month) and, as shown in (3), applying the estimated function $\hat{f}^0$ to the set of 2019 exporters. Given that during the first three months of 2020 Colombia was in practice not exposed to COVID-19 (and therefore $Y_{2020} = Y^0_{2020}$), if assumption (2) holds we expect that in those months the accuracy of the predictions $\hat{Y}_{2019}$ obtained in the cross-validation step for 2019 will be very similar to that of $\hat{Y}^0_{2020}$ for 2020.

Following Cerqua and Letta (2020) and Fabra et al. (2020), we define as an estimator of the individual-specific COVID-19 effect $\alpha$ the simple comparison of the observed outcome under COVID-19 in 2020 with the estimated counterfactual outcome for a given firm:

$$\hat{\hat{\alpha}} = Y_{2020} - \hat{Y}^0_{2020}. \tag{4}$$

Eq. (4) provides the full distribution of treatment effects. All the parameters of interest of the paper are obtained by computing (conditional) averages and quantiles of such distribution.

Starting from (4), by taking the expected value of the individual treatment effect $\hat{\hat{\alpha}}$ for those units with $X_{2019} = x_{2019}$, we can define the following estimator of the conditional average treatment effect (CATE; the average effect for those units with $X_{2019} = x_{2019}$)

$$\begin{aligned}\mathbf{E}[\hat{\hat{\alpha}}|X_{2019} = x_{2019}] &= \mathbf{E}[(Y_{2020} - Y^0_{2020}) - \mathcal{E}^0_{2020} - u^0_{2020}|X_{2019} = x_{2019}] = \\ &= \underbrace{\Delta(X_{2019} = x_{2019})}_{CATE} - \mathbf{E}[\mathcal{E}^0_{2020}|X_{2019} = x_{2019}] - \underbrace{\mathbf{E}[u^0_{2020}|X_{2019} = x_{2019}]}_{=0 \text{ by assumption}},\end{aligned} \tag{5}$$

where,

$$\Delta(X_{2019} = x_{2019}) = \mathbf{E}[Y_{2020} - Y^0_{2020}|X_{2019} = x_{2019}].$$

Therefore $\mathbf{E}[\hat{\hat{\alpha}}_i]$ will identify the unconditional average treatment effect, $\mathbf{E}[\Delta(X_{2019})] = \Delta$, if on average the prediction error is zero: $\mathbf{E}[\mathcal{E}^0_{2020}] = 0$. The conditional average treatment effect, $\Delta(X_{2019} = x_{2019})$, will be identified by $\mathbf{E}[\hat{\hat{\alpha}}_i|X_{2019} = x_{2019}]$ if on average the prediction error will be zero in the relevant sub-sample: $\mathbf{E}[\mathcal{E}^0_{2020}|X_{2019} = x_{2019}] = 0$.

Now let's decompose the outcome observed in 2020 in presence of the pandemic, $Y^1_{2020}$, in a generic model or function $f^1(X^1_{2019})$, which represents the relationship between explanatory

[7] In order to simplify the notation, from now on, we will denote $\mathcal{E}^d_{2020}(X_{2019})$ as $\mathcal{E}^d_{2020}$.

variables and the outcome during the pandemic, and other determinants of the outcome, $u^1_{2020}$, that are orthogonal to the covariates

$$Y^1_{2020} = f^1(X^1_{2019}) + u^1_{2020}, \quad s.t.\ \mathbf{E}[Y^1_{2020}|X^1_{2019}] = f^1(X^1_{2019}). \tag{6}$$

Given that $Y^1_{2020} = Y_{2020}$ and $X^1_{2019} = X_{2019}$, then

$$Y_{2020} = f^1(X_{2019}) + u^1_{2020}, \quad s.t.\ \mathbf{E}[Y_{2020}|X_{2019}] = f^1(X_{2019}). \tag{7}$$

At this point, we can define an alternative estimator of the individual-specific COVID-19 effect $\alpha$ as the comparison of the predicted outcome under COVID-19 in 2020 with the estimated counterfactual outcome for a given firm:

$$\hat{\alpha} = \hat{Y}_{2020} - \hat{Y}^0_{2020}, \tag{8}$$

where $\hat{Y}_{2020} = \hat{f}^1(X_{2019}) = Y_{2020} - \mathcal{E}^1_{2020} - u^1_{2020}$. We call "Shock Aware Machine" (SAM) the model that we use to predict $Y_{2020}$ (and the predictions $\hat{Y}_{2020}$ themselves), where the term "Machine" refers to the fact that the predictions are constructed through machine learning techniques and "Shock Aware" indicates that they are based the information on the observed COVID scenario and are expressed in a metric that is comparable to the estimated non-COVID counterfactual outcomes deriving from the SUM.[8] The SAM expresses the outcome in 2020 of exporters operating the market in 2019 as a function of their characteristics in 2019 and the information related to governments' COVID-19 related stringency measures all over the world coming from Hale et al. (2020).[9] In order to obtain one 2020 prediction for each firm that exported in 2019, similarly to what we have done to select the best performing SUM, we rely on a 5-folds cross-validation strategy. We randomly group the 2019 exporters in five equally sized subsets and we predict the 2020 outcomes of the firms contained in one subset by using the information of firms contained in the remaining four subsets. In other words, we train the models on a random 80% of the data and tests them on the remaining 20% and we repeat the process five times for each different 20% subset, thus obtaining a 2020 prediction for each 2019 exporter.

Starting from Eq. (8), by taking the expected value of the individual treatment effect $\hat{\alpha}$ for those units with $X_{2019} = x_{2019}$, we can define the following alternative estimator of the

[8]Notice that with $\hat{\hat{\alpha}}$ we are comparing a probability (counterfactual) with a binary value (observed outcome), while with $\hat{\alpha}$ we are comparing two estimated probabilities.

[9]We use the firm-level variables "Containment Stringency Index Export" and "Containment Stringency Index Import" which combine the country level information on stringency measures with the exposure of a given firm to the different markets at the export and at the import side in 2019, as explained in more detail in section 3. We do not introduce these variable explicitly as an argument of $f^1()$ to simplify notation.

conditional average treatment effect (for those units with $X_{2019} = x_{2019}$)

$$\begin{aligned}\mathbf{E}[\hat{\alpha}_i | X_{2019} = x_{2019}] =& \mathbf{E}[(Y_{2020} - Y^0_{2020}) - (\mathcal{E}^1_{2020} - \mathcal{E}^0_{2020}) - (u^1_{2020} - u^0_{2020}) | X_{2019} = x_{2019}] \\ =& \underbrace{\Delta(X_{2019} = x_{2019})}_{CATE} - \mathbf{E}[\underbrace{(\mathcal{E}^1_{2020} - \mathcal{E}^0_{2020})}_{\Delta\mathcal{E}} | X_{2019} = x_{2019}] - \\ & \mathbf{E}[u^1_{2020} - u^0_{2020} | X_{2019} = x_{2019}].\end{aligned} \tag{9}$$

Therefore, $\mathbf{E}[\hat{\alpha}_i]$ will identify the unconditional average treatment effect, $\mathbf{E}[\Delta(X_{2019})] = \Delta$, if on average the difference in prediction errors is zero: $\mathbf{E}[\Delta\mathcal{E}] = 0$. The conditional average treatment effect, $\Delta(X_{2019} = x_{2019})$, will be identified by $\mathbf{E}[\hat{\alpha}_i | X_{2019} = x_{2019}]$ if on average the difference in prediction errors is zero in the relevant sub-sample: $\mathbf{E}[\Delta\mathcal{E} | X_{2019} = x_{2019}] = 0$.

Given the definitions of SUM and SAM, to simplify the reasoning in the following we will refer to Equations (4) and (8) respectively as

$$\hat{\hat{\alpha}} = Y - \hat{Y}_{SUM} = Y - SUM. \tag{10}$$

$$\hat{\alpha} = \hat{Y}_{SAM} - \hat{Y}_{SUM} = SAM - SUM. \tag{11}$$

The assumptions behind these identification results are not directly testable as they are expressed in terms of the expected values of the prediction error $\mathcal{E}^0_{2020}$ that is a function of the unobservable counterfactual $Y^0_{2020}$. Table 1 distinguishes the five different scenarios concerning the values of $\mathcal{E}^0_{2020}$ and $\mathcal{E}^1_{2020}$ that are relevant in determining whether applying the statistic $\mathbf{T}$ to $Y - SUM$ and $SAM - SUM$ is able to recover the corresponding treatment effect estimand (e.g., whether averaging the estimated individual treatment effects would recover the average treatment effect).

| | $\mathbf{T}(SAM - SUM)$ | $\mathbf{T}(Y - SUM)$ |
|---|---|---|
| $\mathbf{T}[\mathcal{E}^1_{2020}] \neq 0$ and $\mathbf{T}[\mathcal{E}^0_{2020}] = 0$ | X | ✓ |
| $\mathbf{T}[\mathcal{E}^1_{2020}] = \mathbf{T}[\mathcal{E}^0_{2020}] = 0$ | ✓ | ✓ |
| $\mathbf{T}[\mathcal{E}^1_{2020}] = 0$ and $\mathbf{T}[\mathcal{E}^0_{2020}] \neq 0$ | X | X |
| $\mathbf{T}[\mathcal{E}^1_{2020}] = \mathbf{T}[\mathcal{E}^0_{2020}] \neq 0$ | ✓ | X |
| $\mathbf{T}[\mathcal{E}^1_{2020}] \neq \mathbf{T}[\mathcal{E}^0_{2020}] \neq 0$ | X | X |

Table 1: Identification of generic functions of the individual treatment effects, $\mathbf{T}$, according to the corresponding value taken by the prediction errors.

The estimators based on $Y - SUM$ identify the population parameters when $\mathbf{T}[\mathcal{E}^0_{2020}] = 0$. The estimators based on $SAM - SUM$ are unbiased whenever $\mathbf{T}[\mathcal{E}^1_{2020}] = \mathbf{T}[\mathcal{E}^0_{2020}]$.
Under the assumption, that strength of the COVID-19 effect on export propensity was at most very limited during the first trimester of 2020, we will use the out of sample prediction errors for the first trimester of 2020 as a proxy for the unobservable behavior of $\mathcal{E}^0_{2020}$ in the following months. As explained in detail in Section 5.2, the distribution of the estimated

treatment effects during the first trimester will be used to check the credibility of the above assumptions for the set of all 2019 exporters and for different subsets of 2019 exporters defined according to their characteristics $X_{2019}$ or to their position in the distribution of such effects.

As a final step, we perform the heterogeneity analysis by adapting the Sorted Partial Effects (SPE) method introduced in Chernozhukov et al. (2018) to our setting. Formally, the SPE are defined as percentiles of the Treatment Effects (TE) and can supply a more detailed summary of the distribution of TE than the Average Treatment Effects (ATE), commonly employed in econometric analysis.

$$\alpha^*(u) = u^{th} - \text{ percentile of } \alpha \tag{12}$$

In our setting $\alpha^*(u)$ is therefore a function of $X_{2019}$ defined over its distribution in the population of 2019 exporters.

The SPE are used to do a classification analysis (CA) that allocate the 2019 exporters in two groups, the most and least affected by the COVID-19 shock, according to whether their $\alpha$ are lower than $\alpha^*(25)$ or greater than $\alpha^*(75)$, respectively. Notice that, since the effect of COVID-19 shock is negative, we have defined as the most affected units those whose $\alpha$ lie in the left tail of the sorted distribution of treatment effects. Finally, to study the determinants of treatment effect heterogeneity we observe the mean of the $X_{2019}$ across the most and least affected groups by looking at the difference in means ($CADiff$). In the estimation we use sample analogs of $\alpha^*(u)$ and $CADiff$. We calculate standard errors of $\alpha^*(u)$ and $CADiff$ by bootstrapping the entire estimation process, starting from the initial $\alpha$ estimation step.

The application of the SPE technique presents several advantages in our setting. First, the estimation of the $\alpha^*(u)$s allows a detailed summary of the estimated distribution of the $\alpha$s. Second, it identifies a subgroup of the population which is more affected and guarantees an easy interpretation of how the heterogeneity of the treatment depends on observables without imposing (additional) functional form assumptions. Third, it provides p-values adjustments to account for joint testing of all the covariates considered in the $CADiff$ final step used to detect which observables are associated to greater treatment heterogeneity. In other words, the main idea is to test the null hypothesis of no difference between the value of the covariates in the most and the least affected groups, considering the fact that we conduct simultaneous inference on multiple variables.[10]

[10] Starting from $B$ bootstrap replications of all the estimation steps (including the prediction stage), we calculate the CADiff $B$ times. To determine the significance of the CADiffs we perform a two tailed test. The p-values are constructed as follows:

$$2 \cdot min\{Pr(S \geq t|H_0), Pr(S \leq t|H_0)\}$$

being $t$ the observed $t$ test statistic, $t = \frac{\widetilde{CADiff_{original}}}{\tilde{\sigma}}$, drawn from the unknown distribution $S$. $\tilde{\sigma}$ represents the standard deviation of the bootstrapped CADiffs. To adjust the p-value and obtain the joint p-values taking into account that we are testing jointly hypotheses on many covariates, we reproduce the "single-step" method employ in Chernozhukov et al. (2018) to control for the family-wise error rate.

# 5 Results

## 5.1 Selection of the machine learning algorithm

Once we have defined the methodology to build the SUM and the SAM, we select the best model in terms of prediction performance among a set of standard ML techniques and compare them with a benchmark logistic regression.

The prediction performance out of sample of our empirical models is of fundamental importance because our identification strategy is based on our ability to reconstruct a counterfactual that is practice out of sample, because it is unobserved. Our ML counterfactual approach recognizes that this is a complex task because it is high dimensional (i.e., we have a very high number of potential explanatory variables to take into account) and because of the existence complex interdependencies between firms and products belonging to different sectors and countries which are difficult to know ex-ante. In such a situation, an approach that is based on the maximization of the accuracy of in-sample predictions will be prone to overfitting. Instead, ML techniques are have been shown to constitute the best way to choose the optimal positioning on bias-variance tradeoff for (out of sample) predictive tasks. For machine learning, hyprparameter tuning and cross-validation should be employed in order to avoid overfitting.

We compare four different models: *Logit, Logit-LASSO, Logit-Ridge, and Random Forest (RF)*. Logit represents a natural model that is usually chosen when dealing with a binary dependent variable context, with a binary outcome. Therefore, though in other contexts different from the prediction of firm trade status the literature strongly suggests that in presence of many predictors ML outperforms outperforms it in terms of prediction accuracy, we report also the results from teh Logit model that we consider a natural benchmark for our analysis. Logit-LASSO is used for model selection, i.e. reducing the dimensionality of the matrix of predictors Ahrens et al. (2020). To select the most relevant predictors, the model shrinks the coefficients of some variables to zero. It is powerful when only a bunch of predictors have a lot of predictor power. In particular, we use the *plugin* version of LASSO (also called "Rigorous-LASSO" in Chernozhukov et al. (2016)) because it guarantees the optimal rate of convergence for estimation of the parameters and for the predictions (see Ahrens et al. (2020)). Logit-Ridge shrinks the coefficients with minor contribution to the outcome close to zero but doesn't exclude regressors. It has good predictive performance when many variables of the model are relevant. In the Random Forest version we do not include any interactions as the models creates all possible interactions that are useful to accurately predict the outcome.[11] The prediction analysis is repeated for all months between

[11]Note that it is important to optimize (tune) the hyper-parameters of Logit-Ridge and Random Forest for an accurate predicting exercise. The hyperparameter to optimize in Logit-Ridge is $\lambda$, which controls the impact of the penalty or shrinkage on parameters (when $lambda = 0$ we are in a Logit scenario, when $lambda$ increases the penalty impact grows). We find the optimal hyper-parameter for Logit-Ridge by choosing the

January-December 2020. For Logit, Logit-Ridge, and Logit-LASSO models we include interactions between *size* of the company and some of the main product characteristics, *industry*, *sector*, *means of transportation* as well as with *destination country* dummies.[12]

Table 2 compares the model's prediction power (accuracy) by presenting two commonly used performance measures for classification problems: Area Under the receiver operating Curve (AUC), and Root-Mean-Square Error (RMSE). AUC takes value 0.5 when the model predicts randomly and takes value 1 when the model perfectly classifies the outcome. RMSE ranges from 0 (most accurate model) to 1 (a model that is not able to predict accurately). This table estimates the model using the universe of exporters in 2018. In particular, it measures the the out-of sample performance by cross-validating the sample (K-folds), in the absence of COVID-19. This setting shows that both Logit-LASSO and RF models are the best performers. Models in Table 3 are trained with exporters characteristics in 2018 and their observed outcome in 2019. However, these models are tested using the set of exporters of 2019 and their observed outcome in 2020, the COVID-19 year. Therefore we are in the SUM context. During the months of January, February, and March, the accuracy of Logit-LASSO and RF remains unchanged, as expected, compared to the accuracy (AUC/RMSE) obtained in 2. After April, the accuracy obtained in Table 3 decreases slightly because it does not use any COVID-19 information to predict under a COVID-19 shock scenario. Models in Table 4 are trained and tested with the universe of exporters in 2019 and their observed outcomes in 2020. Using these models we construct SAM predictions. The accuracy of the predictions is very similar to the ones obtained under no COVID-19 information in Table 2. In this analysis is crucial to have good predictions because the individual treatment effects, $\hat{\alpha}_i$, depends on the quality of the estimation accuracy. Both the SUM and the SAM show acceptable levels of accuracy when predictions are done with Logit-LASSO and Random Forest.

$\lambda$ that minimizes the mean cross-validated error. One of the main RF parameters is the number of random trees used. Because of the RF design, it is very difficult to have over-fitted predictions when using this model. Therefore, we set the number of trees to a large enough number (500). We find this number is large enough after repeating the same Random Forest model with a different number of random trees. Even with a lower than 500 number of random trees the error rate of the model remains unchanged.

[12]For more information about all the features included to build the SUM and SAM see Table Appx.1 in Appendix A.

Table 2: Goodness of Fit for SUM in 2018/19

| | AUC | | | | RMSE | | | |
|---|---|---|---|---|---|---|---|---|
| | Logit-LASSO | Logit-Ridge | Random Forest | Logit | Logit-LASSO | Logit-Ridge | Random Forest | Logit |
| Jan | 0.73 | 0.53 | 0.73 | 0.59 | 0.40 | 0.45 | 0.41 | 0.64 |
| Feb | 0.70 | 0.50 | 0.71 | 0.58 | 0.41 | 0.45 | 0.41 | 0.64 |
| Mar | 0.70 | 0.56 | 0.71 | 0.57 | 0.41 | 0.44 | 0.41 | 0.65 |
| Apr | 0.73 | 0.59 | 0.73 | 0.60 | 0.40 | 0.43 | 0.40 | 0.63 |
| May | 0.72 | 0.52 | 0.71 | 0.59 | 0.40 | 0.44 | 0.41 | 0.64 |
| Jun | 0.71 | 0.50 | 0.72 | 0.59 | 0.40 | 0.45 | 0.41 | 0.64 |
| Jul | 0.73 | 0.50 | 0.73 | 0.55 | 0.40 | 0.45 | 0.40 | 0.66 |
| Aug | 0.70 | 0.51 | 0.72 | 0.58 | 0.41 | 0.45 | 0.40 | 0.64 |
| Sep | 0.72 | 0.50 | 0.71 | 0.58 | 0.41 | 0.45 | 0.40 | 0.64 |
| Oct | 0.73 | 0.58 | 0.74 | 0.58 | 0.40 | 0.44 | 0.41 | 0.64 |
| Nov | 0.71 | 0.51 | 0.72 | 0.57 | 0.41 | 0.45 | 0.41 | 0.64 |
| Dec | 0.70 | 0.50 | 0.71 | 0.58 | 0.41 | 0.45 | 0.41 | 0.64 |

Table 3: Goodness of Fit for SUM in 2019/20

| | AUC | | | | RMSE | | | |
|---|---|---|---|---|---|---|---|---|
| | Logit-LASSO | Logit-Ridge | Random Forest | Logit | Logit-LASSO | Logit-Ridge | Random Forest | Logit |
| Jan | 0.72 | 0.53 | 0.72 | 0.49 | 0.41 | 0.45 | 0.41 | 0.75 |
| Feb | 0.69 | 0.50 | 0.69 | 0.56 | 0.41 | 0.45 | 0.42 | 0.64 |
| Mar | 0.72 | 0.54 | 0.73 | 0.59 | 0.40 | 0.44 | 0.41 | 0.63 |
| Apr | 0.67 | 0.56 | 0.66 | 0.51 | 0.48 | 0.50 | 0.49 | 0.70 |
| May | 0.69 | 0.51 | 0.69 | 0.60 | 0.46 | 0.48 | 0.46 | 0.63 |
| Jun | 0.68 | 0.50 | 0.68 | 0.59 | 0.43 | 0.47 | 0.44 | 0.63 |
| Jul | 0.70 | 0.50 | 0.69 | 0.59 | 0.42 | 0.46 | 0.43 | 0.63 |
| Aug | 0.68 | 0.51 | 0.69 | 0.58 | 0.42 | 0.45 | 0.43 | 0.63 |
| Sep | 0.69 | 0.50 | 0.70 | 0.59 | 0.42 | 0.45 | 0.42 | 0.63 |
| Oct | 0.71 | 0.59 | 0.70 | 0.60 | 0.42 | 0.45 | 0.43 | 0.63 |
| Nov | 0.71 | 0.51 | 0.71 | 0.59 | 0.41 | 0.45 | 0.41 | 0.63 |
| Dec | 0.69 | 0.50 | 0.69 | 0.58 | 0.42 | 0.46 | 0.42 | 0.63 |

Table 4: Goodness of Fit for SAM in 2019/20

| | AUC | | | | RMSE | | | |
|---|---|---|---|---|---|---|---|---|
| | Logit-LASSO | Logit-Ridge | Random Forest | Logit | Logit-LASSO | Logit-Ridge | Random Forest | Logit |
| Jan | 0.73 | 0.58 | 0.74 | 0.50 | 0.41 | 0.45 | 0.41 | 0.71 |
| Feb | 0.70 | 0.50 | 0.70 | 0.49 | 0.41 | 0.46 | 0.42 | 0.70 |
| Mar | 0.73 | 0.50 | 0.73 | 0.50 | 0.40 | 0.46 | 0.40 | 0.71 |
| Apr | 0.74 | 0.66 | 0.73 | 0.52 | 0.42 | 0.47 | 0.42 | 0.69 |
| May | 0.76 | 0.74 | 0.77 | 0.50 | 0.41 | 0.46 | 0.41 | 0.71 |
| Jun | 0.73 | 0.69 | 0.73 | 0.48 | 0.42 | 0.46 | 0.42 | 0.72 |
| Jul | 0.73 | 0.63 | 0.72 | 0.51 | 0.41 | 0.45 | 0.42 | 0.69 |
| Aug | 0.72 | 0.50 | 0.72 | 0.53 | 0.41 | 0.46 | 0.42 | 0.69 |
| Sep | 0.71 | 0.50 | 0.70 | 0.55 | 0.42 | 0.47 | 0.42 | 0.67 |
| Oct | 0.72 | 0.50 | 0.71 | 0.52 | 0.42 | 0.46 | 0.42 | 0.70 |
| Nov | 0.72 | 0.52 | 0.72 | 0.49 | 0.41 | 0.45 | 0.41 | 0.71 |
| Dec | 0.71 | 0.51 | 0.70 | 0.51 | 0.41 | 0.45 | 0.42 | 0.70 |

Given the above results on the prediction accuracy of the considered models, we rely on Logit-LASSO and Random Forest (RF) models. For compactness, we only provide the results for Logit-LASSO.

## 5.2 Evaluation of the COVID-19 effect

We use Logit-LASSO predicted probabilities to estimate the average monthly effect of the COVID-19 shock as the monthly average of $\hat{\alpha}_i$ (the difference between the firm-level predicted probabilities of success in the SUM and the SAM scenarios), which are presented Figure 5. If we assume that, in the first months of 2020, average firms are not affected by the COVID-19 shock, we can consider the estimates comparing the SAM and SUM predictions as a falsification test, similarly to the in-time placebo test routinely used in Synthetic Control Methods-SCM (Abadie et al., 2015). Estimating an economically significant effect of the COVID-19 treatment in the months before the actual economic shock happened would indicate that our model is mechanically predicting a COVID-19 effect even when it is not expected. We will also apply this placebo study conditioning on exogenous firms' characteristics observed in 2019 by estimating COVID-19 effects for selected subsamples of firms according to such characteristics. We interpret these placebo studies as a robustness check on our results on treatment heterogeneity. This Figure also contains the RF predicted probabilities to show that Logit-LASSO and RF results are quantitatively similar.[13]

[13]Note that Random Forest results are available upon request.

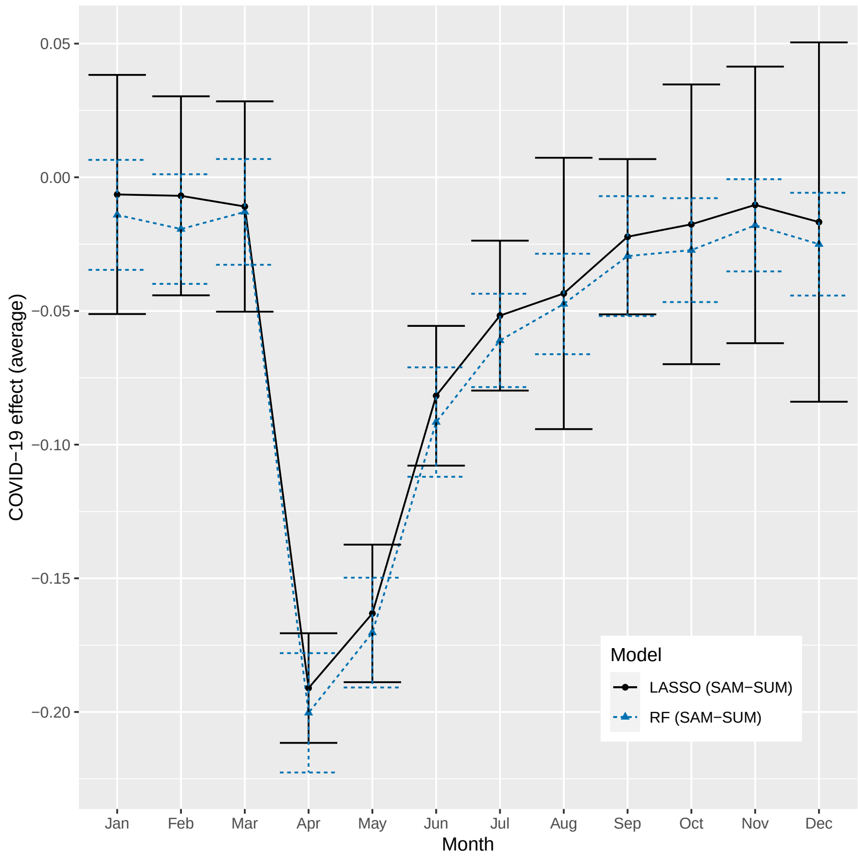


Figure 5: Average Individual Treatment Effect, by months, comparison between Logit-LASSO and RF. Standard errors obtained with 100 bootstrap replications. Confidence intervals for a 5% significance level.

As shown in Figure 5, the probabilities obtained from the SUM and the SAM are almost identical on average for January, February, and March. This result is reassuring since only on March 25, 2020, the Colombian government implemented a complete and mandatory lockdown. More in general, we can conclude that our identification strategy is not mechanically recovering COVID-19 effects for a period with low incidence in Colombia and in the rest of the world.

We find that the peak of the COVID-19 effect is in April 2020, when we find an average difference between the predicted probabilities of exporting of nearly 20 percentage points. In the following months, the estimated average effect is declining with the time.

Figure 6 shows evidence of big variations in the quarterly estimated average individual treatment effect by Industry. On the one hand, during the first, third, and fourth quarters of 2020 there is no evidence of aggregated heterogeneity due to COVID-19 effect on industries, and the aggregated COVID-19 shock is economically small or not significant at all. On the other hand, during the second quarter of 2020, Colombian exporters belonging almost to every industry are predicted to reduce significantly their probabilities to survive in the international markets. Moreover, this reduction in the rates of success is more remarkable for industries like Textile, Footwear or Jewelries. Other exporters in industries like Food

Preparations or Vegetables are estimated to not vary much the likelihood to survive due to COVID-19 shock.

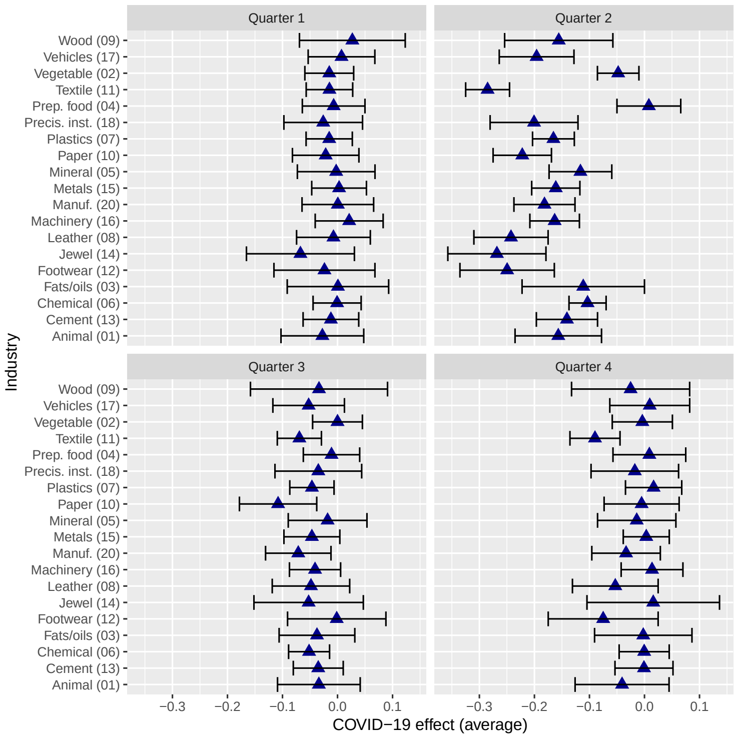


Figure 6: Quarterly mean difference in the predicted probability of success (SAM vs. SUM) by industry, using Logit-LASSO predictions. COVID-19 effect in 2020. Standard errors obtained with 100 bootstrap replications. Confidence intervals for a 5% significance level.

## 5.3 Heterogeneity of the COVID-19 effect on Colombian exporters

This section explores the exporters' characteristics determining higher COVID-19 effects. Notice, indeed, that, though informative, the Average Individual Treatment Effect of Figure 5 alone is not able to disentangle the heterogeneity of the effect of COVID-19 on export probabilities.

For this purpose we first need to sort the treatment effects estimated (see Chernozhukov et al. (2018)) in the previous section.[14] This is done yearly in Figure 7 and monthly in Figure 8. In the following we will focus on the monthly analysis as it gives more hints with respect to its yearly counterpart where pandemic and non-pandemic months are mixed together (hence COVID-19 effects might be compensated by non-pandemic months). In particular, Figure 8 shows the monthly Sorted Partial Effects (SPE) and Average Partial Effects (APE) computed as specified in Section 4, by percentiles. The two mentioned figures also report the 95% confidence intervals with the blue bands (SPE), and with the black dashed lines (APE). APE provides the average range of fluctuation that works as a reference level or benchmark effect to interpret the SPE results. The solid black line in Figure 8 coincides with the

[14]We use estimations coming from Logit-LASSO for this purpose.

black dots in Figure 5.The heterogeneity of the SPE is more evident and more statistically significant during the months of April and May. We also find significant COVID-19 effects during June, July and August. In September, the SPE confidence intervals tend to settle around pre-pandemic levels (though not fully recovering). In general, the monthly SPE (red line) ranges from around $-25p.p.$ to $14p.p$ in pre-pandemic months, while covering a wider range toward negative values in the middle of the pandemic. The treatment effect of COVID-19 reaches negative values of around $-27\%$.

It is interesting to observe how the SPE does almost coincide with the APE of "no COVID-19 effect" in pre-pandemic months, confirming the validity of our results. As detailed above, this is not the case for COVID-19 months where negative and significant values of $\alpha^*(u)$ are reported. The SPE analysis for October, November and December clearly displays the recovery of the export probability evidencing a not significant $\alpha^*(u)$. While the mentioned heterogeneity of the SPE is more pronounced and unbalanced toward negative values of $\alpha^*(u)$ in the middle of the pandemic (April and May), negative and positive values tend to balance in the next months. This evidence enforces the discussed recovery starting from autumn. Figure 8 highlights also the seasonal trend of the effect of COVID-19 on export probabilities. Moreover, from June to September, when contagion and, accordingly, restrictions slow down, the SPE suddenly flattens out breaking the "negative trend" of April and May.

These last considerations can be condensed as follows: *"no COVID-19 effect - significant COVID-19 effect - no COVID-19 effect"* respectively in *"pre-pandemic months - pandemic months - autumn and winter months"*.

In addition to being very interesting per se in that they reflect the distribution of heterogeneous effects, the annual SPE are further employed to perform the Classification Analysis reported in Table 5.

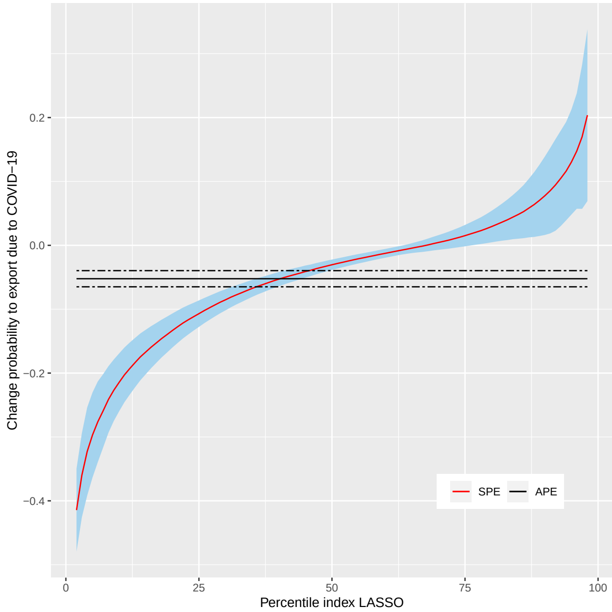


Figure 7: Annual Sorted Partial Effects (SPE) and Average Partial Effects (APE) of COVID-19 on Colombian firm export's status. TE is calculated as a difference between SAM and SUM predictions. Standard errors obtained with 100 bootstrap replications. Confidence intervals for a 5% significance level.

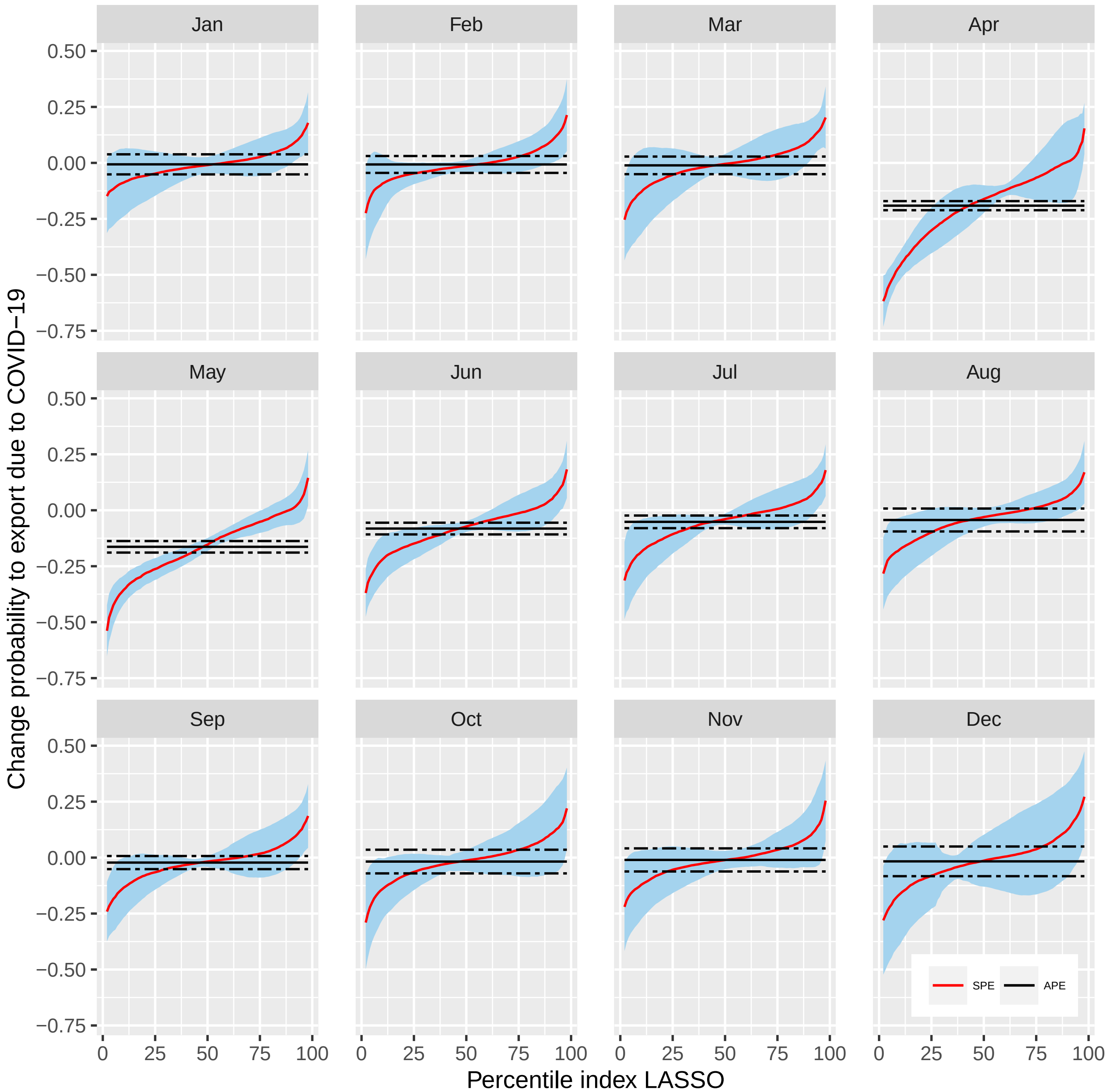


Figure 8: Monthly Sorted Partial Effects (SPE) and Average Partial Effects (APE) of COVID-19 on Colombian firm export's status. TE is calculated as a difference between SAM and SUM predictions. Standard errors obtained with 100 bootstrap replications. Confidence intervals for a 5% significance level.

In Table 5 we report the variables of interest including firms' characteristics. With respect to Section 4, we slightly modify the CADiff as follows. First, most and least affected firms are defined according to the first and last quartiles. Notice that since the Treatment Effects (TE) are predominantly negative, we define the most affected as those firms whose estimated TE are lower than the first quartile of the effect (below percentile 25). Similarly, the least affected are defined as those firms whose estimated effect exceeds the third quartile of the effect (over the percentile 75).

We compute the difference between most and least affected firms via regression, i.e. regressing the variables of interest over a dummy $q = \mathbf{1}_{\{i\in\alpha^*(75)\}}$. The dummy $q$ is 0 for $\alpha^*(25)$ in the first quartile.[15] This regression allows to recover the same results as the ones obtained by making the difference between the most and the least affected firms but the regression also offers the possibility of controlling for time and sector variables. Controlling for sectors and months allows to perform a *ceteris paribus* analysis, i.e. to dig into the effects of COVID-19 within specific sectors and specific months. Therefore it facilitates the differentiation in the Treatment Effect (TE) of variables for which the COVID-19 effect depends on the sector and/or time patterns.

In this respect, Table 5 is divided into 3 columns according to the controls included in the mentioned regression. In particular, denoting as $v$ the different variables of interest (the estimated individual treatment effect and the selected firm characteristics) and as $k$ other controls, we estimate the following general form:

$$v_f = \beta^m_{0,f} + \beta^m_{1,f} q + \beta^m_{2,f} k + \varepsilon^m_f$$

where $m \in \{1, 2, 3\}$ are the models depending on the different controls used, and $f$ corresponds to the f-th outcome variable selected.[16] $\beta^1_{1,f}$ represents the average difference in $\hat{\alpha}$, for a given dependent variable, between the most and the least affected firms, that in section 4 is defined as $CADiff$; $\beta^2_{1,f}$ is the conditional $CADiff$ obtained controlling for the firm sector dummies; $\beta^3_{1,f}$ is the conditional $CADiff$ obtained controlling for the firm sector and month of the year dummies. Adopting this notation, the columns of Table 5 differ according to the regressors included in $k$.

Considering the estimated individual Treatment Effects as dependent variable we find a negative and significant $\beta^m_{1,f}$ in all the three specifications. These results show that the most affected exporters (those located in the first SPE quartile distribution) experienced a decrease in the probabilities to export between 27.9$p.p.$ and 31.3$p.p.$ lower than the one experienced by the least affected firms (those located in the third SPE quartile distribution). Table 5 shows no large significant effect by industries (aggregated). Focusing on treatment effect heterogeneity with respect to the aggregate sector to which firms belong, we detect the share of Textile firms among the most affected 2019 exporters is 16$p.p.$ higher with respect to the one estimated for the group of the least affected firms. Similarly, we find the presence of 2.9$p.p.$ more wood exporters among the most affected than among the least affected firms. Therefore, textile and wood industries were the most negatively affected aggregate sectors by the COVID-19. We also detect the existence of treatment heterogeneity associated to the means of transportation used by exporters in 2019. On the one hand, we observe that there are 16.8$p.p.$ to 20.4$p.p.$ more exporters using the air among the most affected than among

[15] To perform CADiff analysis, we only take those estimations that are located in the first or last quartile of the SPE distribution.

[16] Note that the rows in Table 5 represent the different outcome variables we use.

the least affected firms. However, there are 19.2*p.p.* to 23.6*p.p.* less Colombian exporters using the sea to transport among the most affected than among the least affected firms. Looking into the heterogeneity by months, the first pattern we notice is that only the months from April to August have a positive estimated parameter. However, only April and May are statistically significant. There are 18.6*p.p.* to 19.5*p.p.* (17.7*p.p.* to 18.2*p.p.*) more firms exporting in April (May) among the most affected than among the least affected firms. The coefficients turn negative and significant from September to November signaling the start of the recovery.

The $\beta^m_{1,f}$ estimated when using the Export (Import) Containment Stringency Index as dependent variables provide insightful hints on the difficulties of Colombian firms in exporting (importing) to (from) countries adopting severe stringency measures. In particular, the most affected Colombian exporters face on average an higher Export (Import) Containment Stringency Index with respect to the one faced by least affected firms by 7.18 to 19.51 (7.25 to 20.8) basic points. [17]

Finally, in 2019 the least affected firms exported (imported) 156.7% to 176.83% (614.7% to 1467.3%) more value than the most affected firms. Therefore, Colombian exporters trading in larger volumes (in value) are more resilient under a COVID-19 scenario.[18]

[17]Remember that the Index ranges from 0 to 100.

[18]Table 5 results commented above are robust to the inclusion of sectoral and sectoral-monthly information (second and third columns).

| **Outcome variable** | $\beta^1_{1,f}$ | $\beta^2_{1,f}$ | $\beta^3_{1,f}$ |
|---|---|---|---|
| TE | $-0.3130^{***}$ | $-0.3060^{***}$ | $-0.2790^{**}$ |
| Agriculture | -0.1940 | | |
| Chemicals | -0.0057 | | |
| Manufacturing | -0.0092 | | |
| Metals | 0.0134 | | |
| Special | $0.0056^{***}$ | | |
| Textile | $0.1600^{***}$ | | |
| Wood | $0.0292^{***}$ | | |
| Air | $0.2030^{*}$ | $0.1680^{***}$ | $0.2040^{***}$ |
| Land | 0.0340 | 0.0249 | 0.0170 |
| Sea | $-0.2360^{***}$ | $-0.1920^{***}$ | $-0.2200^{***}$ |
| Jan | -0.0738 | $-0.0766^{***}$ | |
| Feb | -0.0710 | $-0.0768^{***}$ | |
| Mar | -0.0751 | $-0.0773^{***}$ | |
| Apr | $0.1860^{***}$ | $0.1950^{***}$ | |
| May | $0.1770^{***}$ | $0.1820^{***}$ | |
| Jun | 0.0754 | $0.0784^{***}$ | |
| Jul | 0.0132 | 0.0159 | |
| Aug | 0.0021 | 0.0008 | |
| Sep | $-0.0412^{***}$ | $-0.0406^{**}$ | |
| Oct | $-0.0604^{***}$ | $-0.0609^{**}$ | |
| Nov | $-0.0723^{***}$ | $-0.0763^{**}$ | |
| Dec | -0.0557 | $-0.0621^{**}$ | |
| ND | -0.1990 | -0.1640 | -0.2480 |
| NO | -1.7470 | $-1.9820^{***}$ | $-2.4440^{**}$ |
| NP | 0.2400 | -0.2570 | -0.3440 |
| Containment Index Stringency Export | $19.3600^{***}$ | $19.5100^{***}$ | $7.1800^{*}$ |
| Containment Index Stringency Import | $19.1100^{***}$ | $20.8000^{***}$ | $7.2490^{***}$ |
| Value Exported (log) | $-0.5110^{***}$ | -0.4490 | $-0.5700^{*}$ |
| Value Imported (log) | $-1.8160^{***}$ | $-2.2020^{***}$ | $-2.6860^{***}$ |
| Deviation from sectoral mean | | ✓ | ✓ |
| Deviation from monthly mean | | | ✓ |

Table 5: Estimated differences in means of the estimated treatment effect and other covariates between the group of more affected and the group of less affected firms ($CADiff$) applying the classification Analysis to the $SAM - SUM$ estimates. The considered variable varies by row. Column 1 does not include sector nor month variables in the regression. Column 2 includes sector in the regression. Column 3 includes both the sector and month variables. *** means significant at 1%, ** at 5%, * at 10%. Note that standard errors are obtained by boostrapping the whole estimation process and joint p-values are adjusted to take into account the simultaneous testing of all the variables as described in the text.

## 5.4 Estimations based on $(Y - SUM)$

Following Fabra et al. (2020) and Cerqua and Letta (2020) we use the estimators based on Eq. (10). These estimators capture the differences between the observed outcome, Y, (binary variable accounting for the success of a Colombian exporter in 2020) and its counterfactual predictions (SUM). Figure 9 shows that the average individual treatment effect for COVID-19 is very similar when the individual treatment effect is estimated as the average of $\hat{\alpha}$ (black line, $SAM - SUM$) or as the average of $\hat{\hat{\alpha}}$ (yellow line, $Y - SUM$). As shown in Figure 9, when the interest lies in estimating the average treatment effects (by months in this case) the results based on $Y - SUM$ do not differ from those obtained by using $SUM - SAM$. We obtain similar results for the two methodologies also in terms of conditional treatment effects based on subgroups defined on firm characteristics (e.g., by industry or main export destination country).[19]

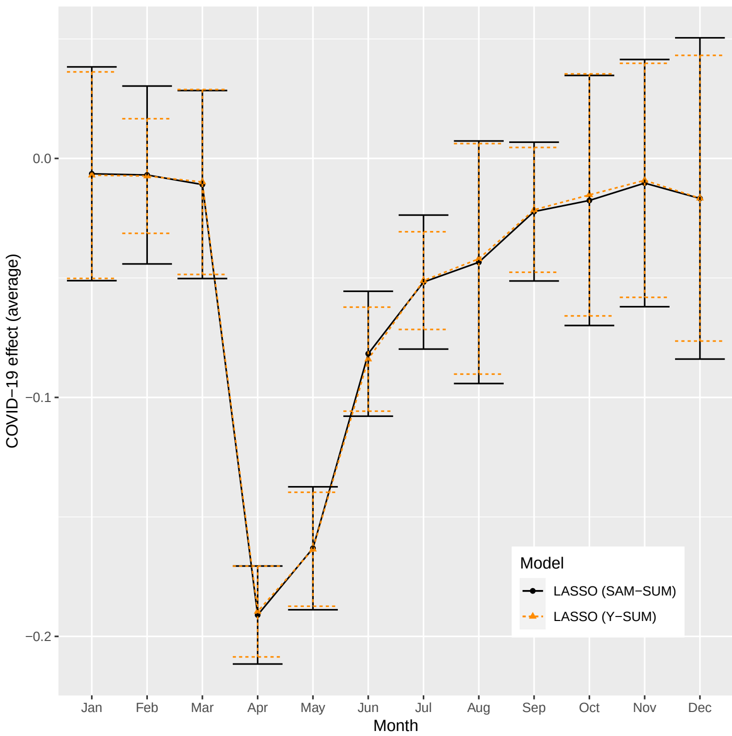


Figure 9: Mean difference in the predicted probability of success (SAM vs. SUM / Y vs. SUM) by month, using Logit-LASSO predictions and (SAM vs. SUM). COVID-19 effect in 2020. Standard errors obtained with 100 bootstrap replications. Confidence intervals for a 5% significance level.

The fact that the two estimators consistently find zero estimated effects for all 2019 exporters (and subgroups based on the values of individual observables) during the first quarter suggests that the estimation error of both SUM and SAM, $\mathcal{E}_0$ and $\mathcal{E}_1$ respectively, goes to zero when we average the individual treatment effects across the whole distribution of 2019 exporters or in subgroups defined by one of the possible dimension of treatment heterogeneity defined by observables (e.g., by industry or main export destination country).

[19] Results available upon request.

However, our estimation strategy to identify the main dimensions of treatment heterogeneity is based on classifying units in two groups having the highest and the lowest treatment estimated effects, therefore we are also interested in the performance of our alternative estimation strategies in identifying treatment effects in the tails of the distribution. Figure 10 shows the estimated average treatment effects obtained with the two estimators for intervals defined by the estimated percentiles of $Y - SUM$. On the one hand, it is apparent that the estimator based on $Y - SUM$ is identifying significant treatment effect heterogeneity also in the first quarter suggesting that the estimation error, $\mathcal{E}_0$, of the distribution is not zero on average in the tails. Moreover, the shape of the $Y - SUM$ curve is practically constant across quarters, suggesting that this estimation method will be prone to misclassify units when using the Sorted Effects strategy suggested above. On the other hand, in the first quarter the shape of the $SAM - SUM$ curve is flat showing a constant average estimated effect that is zero along the whole distribution of the $Y - SUM$ estimated effects, suggests that on average $\mathcal{E}_1 = \mathcal{E}_0$ is zero because by using the $SAM$ we are able to wash out the estimation error of the $SUM$. This behavior of the estimators based on $SAM - SUM$ is clearly consistent with the results shown in Figure 8 for the Sorted Effects analysis. In Figure 11, we show that the intuition on the inadequacy of the $Y - SUM$-based estimators to identify treatment effects on the tail of the distribution is confirmed also by the Sorted Effects analysis based on this estimation strategy. When using the $Y - SUM$ individual level estimates to feed the SPE methodology, in the first quarter we find economically and statistically significant effects of the COVID-19 shock all along the percentile distribution. This indicates again that, though on average the $\mathcal{E}_0$ goes to zero when considering all the observations, when concentrating on local portions of the treatment effect distribution this does not happen.

Table 6 shows the results of the heterogeneity analysis estimating the $CADiffs$ by using the $(Y - SUM)$ approach. We obtain no significant results for basically all the firm characteristic that we consider. This is consistent with the inability of the $Y - SUM$ approach to consistently estimate treatment effects in the tails of the distribution of the $\alpha$s and, therefore, to identify the groups of more affected and less affected firms. In other words, such groups will be contaminated by the inclusion of firms wrongly classified due to the estimation error $\mathcal{E}_0$.

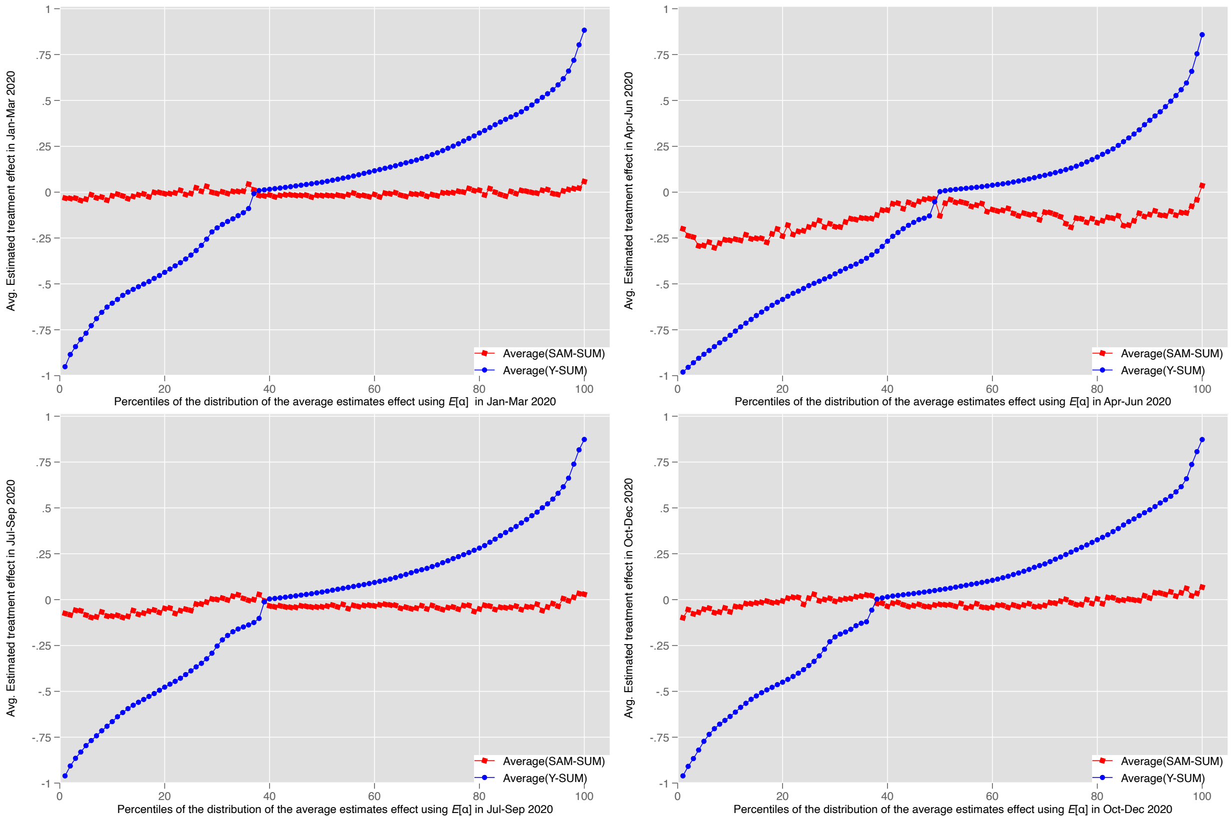


Figure 10: Estimated average treatment effects for intervals defined by the estimated percentiles of $Y - SUM$, $SAM - SUM$ (red line) and $Y - SUM$ (blue line), by quarters.

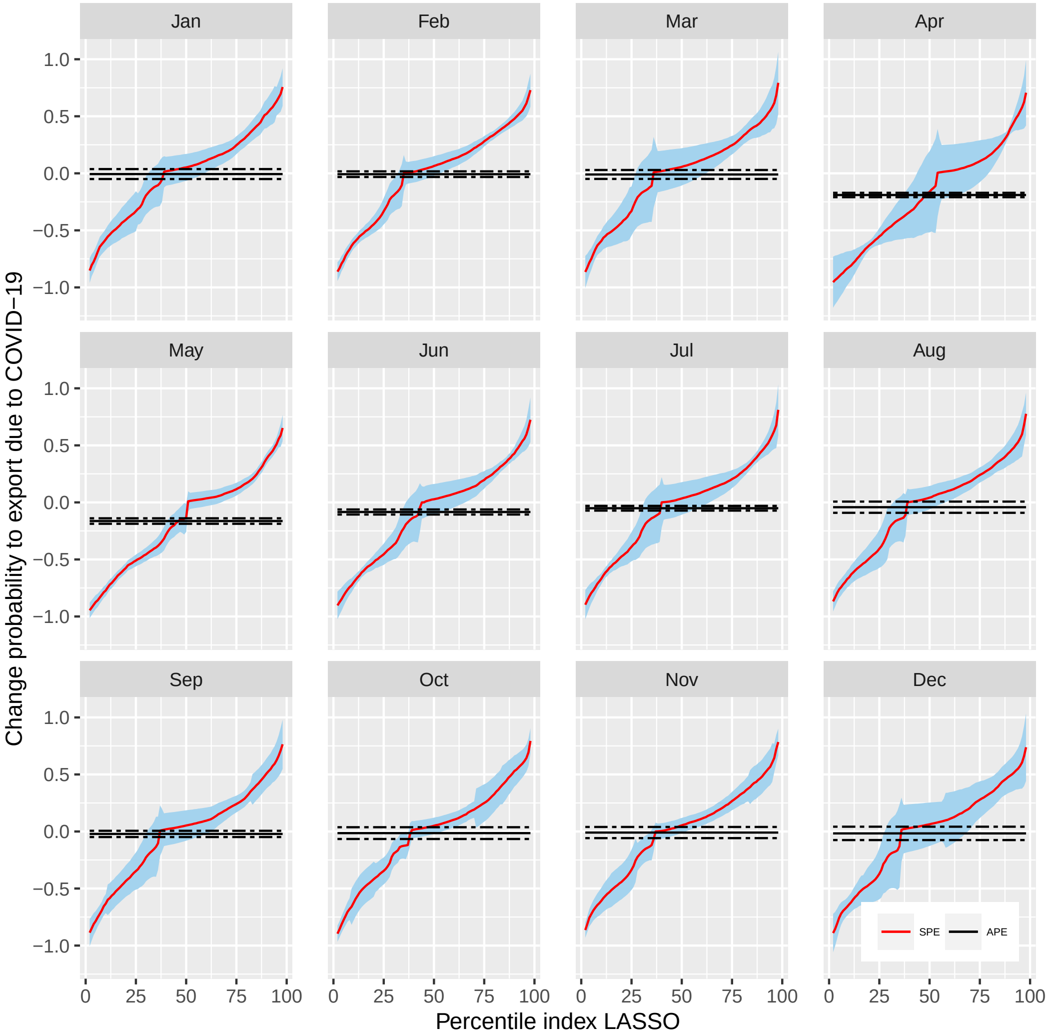


Figure 11: Monthly Sorted Partial Effects (SPE) and Average Partial Effects (APE) of COVID-19 on Colombian firm export's status. TE is calculated as a difference between the observed outcome (Y) and SUM predictions. Standard errors obtained with 100 bootstrap replications. Confidence intervals for a 5% significance level.

| **Outcome variable** | $\beta_{1,f}^{(1)}$ | $\beta_{1,f}^{(2)}$ | $\beta_{1,f}^{(3)}$ |
|---|---|---|---|
| TE | -1.0910 | $-1.0930^{***}$ | -1.0710 |
| Agriculture | -0.0616 | | |
| Chemicals | -0.0192 | | |
| Manufacturing | 0.0112 | | |
| Metals | 0.0109 | | |
| Special | 0.0059 | | |
| Textile | 0.0486 | | |
| Wood | 0.0041 | | |
| Air | 0.0411 | 0.0271 | |
| Land | 0.0086 | 0.0062 | |
| Sea | -0.0482 | -0.0321 | |
| Jan | -0.0190 | -0.0189 | |
| Feb | -0.0242 | -0.0237 | |
| Mar | -0.0181 | -0.0181 | |
| Apr | 0.0631 | 0.0630 | |
| May | 0.0620 | 0.0612 | |
| Jun | 0.0166 | 0.0167 | |
| Jul | 0.0033 | 0.0028 | |
| Aug | -0.0050 | -0.0053 | |
| Sep | -0.0169 | -0.0167 | |
| Oct | -0.0216 | -0.0208 | |
| Nov | -0.0218 | -0.0222 | |
| Dec | -0.0183 | -0.0181 | |
| ND | 0.3310 | 0.3470 | |
| NO | 0.0350 | -0.0595 | |
| NP | 0.6050 | 0.4670 | |
| Containment Index Stringency Export | -0.2280 | -0.0264 | 0.9690 |
| Containment Index Stringency Import | -4.2180 | -4.4910 | -0.0520 |
| Value Exported (log) | -0.2700 | -0.2760 | -0.1800 |
| Value Imported (log) | -0.0910 | 0.0296 | 0.0040 |
| Deviation from sectoral mean | | ✓ | ✓ |
| Deviation from monthly mean | | | ✓ |

Table 6: Estimated differences in means of the estimated treatment effect and other covariates between the group of more affected and the group of less affected firms ($CADiff$) applying the classification Analysis to the $Y - SUM$ estimates. The considered variable varies by row. Column 1 does not include sector nor month variables in the regression. Column 2 includes sector in the regression. Column 3 includes both the sector and month variables. $^{***}$ means significant at 1%, $^{**}$ at 5%, $^{*}$ at 10%. Note that standard errors are obtained by boostrapping the whole estimation process and joint p-values are adjusted to take into account the simultaneous testing of all the variables as described in the text.

# 6 Final discussion

We estimate by using counterfactuals built with ML techniques that the COVID-19 shock decreased a firm's probability of surviving in the export market by about 15$p.p.$ to 20$p.p$ in April and May and by approximately 5$p.p.$ to 8$p.p.$ in June and July. By studying the distribution of the estimated individual treatment effect, we also show that these average results hide considerable heterogeneity. We do it by integrating the Sorted Partial Effects methodology with our causal ML approach. We show that when the focus lies on the tails of the distribution of the treatment effects, it is critical to correct for the estimation error arising from the necessarily imperfect reconstruction of the unobservable counterfactual.

This paper also shows how ML methods can be applied successfully to predict firms' trade potential. We consider this method and its application promising avenues of research to assist firms and public agencies in their decision-making processes. The bulk of countries possess export promotion agencies whose objective is to sustain firms' internationalization activities by lowering the costs of information acquisition (Broocks and Van Biesebroeck, 2017; Munch and Schaur, 2018). Given that exporter dynamics can be understood as a complex learning process dense of interdependencies (complementarity or substitutability) between products and destination markets (from the perspective of technology/knowledge, local tastes, legal requirements, and marketing and distribution costs)[20] and that ML techniques have been successfully applied to predict firm performances in such settings, we think that these techniques and firm-level data can be fuitfully used to build recommendation systems to help firms learning their latent comparative advantages and providing export diversification/differentiation recommendations.

[20]In the core competence model by Eckel and Neary (2010) the focus is on process innovation and productivity: each firm has a core product in which its productivity is the highest and bears adaptation costs to produce different products. In Montinari et al. (2021), instead, product innovation and cumulative growth are central: to enlarge the portfolio of produced and exported goods, firms have to invest in R&D and "the more products a firm has, the more resources it can devote to research and develop new products." Hidalgo et al. (2007) and Jun et al. (2020) study bilateral trade at the product-country level and represent the portfolio of goods shipped by countries as a network, the so-called product space. They define a measure of relatedness between products based on co-exporting patterns to capture common capabilities and knowledge flows between products. Finally, find that the entry costs a firm has to bear to start exporting to a new market depend on the similarity of the new destination with respect to those of the portfolio of markets already reached by the firm (in terms of geographic location, language, and income per capita) and profit shocks in a market affects firms' exports in other markets.

# References


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

Ahrens, A., C. B. Hansen, and M. E. Schaffer (2020). lassopack: Model selection and prediction with regularized regression in stata. *The Stata Journal 20*(1), 176–235.

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

Antràs, P., S. Redding, and E. Rossi-Hansberg (2020). Globalization and Pandemics. (No. w27840). National Bureau of Economic Research.

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

Baldwin, R. and E. Tomiura (2020). Thinking ahead about the trade impact of COVID-19. *Economics in the Time of COVID-19 59*.

Bargagli-Stoffi, F. J., J. Niederreiter, and M. Riccaboni (2020). Supervised learning for the prediction of firm dynamics. *arXiv preprint arXiv:2009.06413*.

Bonadio, B., Z. Huo, A. A. Levchenko, and N. Pandalai-Nayar (2020). Global supply chains in the pandemic. National Bureau of Economic Research, No. w27224.

Breinlich, H., V. Corradi, N. Rocha, M. Ruta, J. Santos Silva, and T. Zylkin (2021). Machine learning in international trade research.

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

Bruno, M., F. Saracco, T. Squartini, and M. Dueñas (2018). Colombian export capabilities: building the firms-products network. *Entropy 20*(10), 785.

Cepeda-López, F., F. Gamboa-Estrada, C. León-Rincón, and H. Rincón-Castro (2019). Colombian liberalization and integration to world trade markets: Much ado about nothing. *Borradores de Economía; No. 1065*.

Cerqua, A. and M. Letta (2020). Local economies amidst the covid-19 crisis in italy: a tale of diverging trajectories. *COVID Economics* (60), 142–171.

Chernozhukov, V., I. Fernández-Val, and Y. Luo (2018). The sorted effects method: discovering heterogeneous effects beyond their averages. *Econometrica 86*(6), 1911–1938.

Chernozhukov, V., C. Hansen, and M. Spindler (2016). hdm: High-dimensional metrics. *arXiv preprint arXiv:1608.00354*.

de Lucio, J., R. Mínguez, A. Minondo, and A. F. Requena (2020, December). Impact of Covid-19 containment measures on trade. Working Papers 2101, Department of Applied Economics II, Universidad de Valencia.

Dominici, F., F. J. Bargagli-Stoffi, and F. Mealli (2020). From controlled to undisciplined data: estimating causal effects in the era of data science using a potential outcome framework. *arXiv preprint arXiv:2012.06865*.

Dueñas, M., M. Campi, and L. Olmos (2021). Changes in mobility and socioeconomic conditions during the covid-19 outbreak. *Humanities and Social Sciences Communications 8*(1), 1–10.

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

Eslava, M., J. Tybout, D. Jinkins, C. Krizan, J. Eaton, et al. (2015). A search and learning model of export dynamics. In *2015 Meeting Papers*, Number 1535. Society for Economic Dynamics.

Espitia, A., A. Mattoo, N. Rocha, M. Ruta, and D. Winkler (2021). Pandemic trade: COVID-19, remote work and global value chains. *The World Economy*.

Evenett, S. (2020). Sicken thy neighbour: The initial trade policy response to COVID-19. *The World Economy 43*(4), 828–839.

Fabra, N., A. Lacuesta, and M. Souza (2020). Degrowth versus Decoupling: Competing strategies for carbon abatement? Available at: https://eel107.s3.us-east-2.amazonaws.com/EEL_109.pdf.

Felbermayr, G. and H. Görg (2020). Implications of covid-19 for globalization. *KIELER BEITRÄGE ZUR*, 3.

Hale, T., A. Petherick, T. Phillips, and S. Webster (2020). Variation in government responses to covid-19. *Blavatnik school of government working paper 31*, 2020–11.

Hidalgo, C. A., B. Klinger, A.-L. Barabási, and R. Hausmann (2007). The product space conditions the development of nations. *Science 317*(5837), 482–487.

Jun, B., A. Alshamsi, J. Gao, and C. A. Hidalgo (2020). Bilateral relatedness: knowledge diffusion and the evolution of bilateral trade. *Journal of Evolutionary Economics* (30), 247–277.

Montinari, L., M. Riccaboni, and S. Schiavo (2021). Innovation, trade and multi-product firms. *Canadian Journal of Economics/Revue canadienne d'économique*.

Munch, J. and G. Schaur (2018). The effect of export promotion on firm-level performance. *American Economic Journal: Economic Policy 10*(1), 357–87.

Ocampo, J. A. and C. Tovar (2000). Colombia in the classical era of 'inward-looking development', 1930–74. In E. Cárdenas, J. Ocampo, and R. Thorp (Eds.), *An economic history of twentieth-century Latin America*, pp. 239–281. Springer.

Pierce, J. R. and P. K. Schott (2012). A concordance between ten-digit US Harmonized System Codes and SIC/NAICS product classes and industries. *Journal of Economic and Social Measurement 37*(1-2), 61–96.

Varian, H. R. (2016). Causal inference in economics and marketing. *Proceedings of the National Academy of Sciences 113*(27), 7310–7315.

World Bank (2020). Global Economic Prospects, June 2020. Washington, DC: World Bank. Available at: https://openknowledge.worldbank.org/handle/10986/33748.

Zuluaga, Ó. I., C. G. Cano, J. J. Echavarría, F. Tenjo, J. P. Zárate, and J. D. Uribe (2009). Informe de la Junta Directiva al Congreso de la República – Marzo de 2009. Banco de la República de Colombia.

# A Appendix - Data

Table Appx.1: Predictors for exporters success

| Variable | Description | Source |
|---|---|---|
| | *Models: SUM and SAM* | |
| $NP$, $ND$, $NO$ | Number of products exported by, number of destinations where a company exports, and number of import origin countries for an exporter in a given month, respectively. | Authors' own elaboration from Colombian Customs Office (DIAN). |
| $HH_p$, $HH_d$ | Product-Herfindahl Index, and Destination-Herfindahl Index. Measure the concentration of products at 6-digits HS, and the concentration at destination by company-month, respectively. | Authors' own elaboration from Colombian Customs Office (DIAN). |
| Total value (exports) | Free on board value of the export transaction in US dollars for each company-month. | Colombian Customs Office (DIAN) |
| Total value (imports) | Free on board value of the import transaction in US dollars for each company-month. | Colombian Customs Office (DIAN) |
| Size | 4 class dummies classifying firms according to the quartiles of the firm-level (Q1, Q2, Q3 and Q4) distribution of the total monthly value of exports (in ln). | Authors' own elaboration |
| Destination | Factor variable with one level (dummy variable) for each destination country where Colombian exporters operate by month. | Colombian Customs Office (DIAN) |
| Origin | Factor variable with one level (dummy variable) for each import origin country where Colombian exporters operate by month. | Colombian Customs Office (DIAN) |
| Continent | Factor variable with one level (dummy variable) for each continent where Colombian exporters operate. | Authors' own elaboration |
| Department | Factor variable with one level (dummy variable) for each department (region) in Colombia from which companies operate. | Colombian Customs Office (DIAN) |
| Means of Transportation | 4 class dummies indicating the means of transportation a company use to perform a transaction (land, sea, air, others). | Colombian Customs Office (DIAN) |
| Sector | 99 class dummies classifying company products at 2-digit HS code (corresponding to a HS-chapter). | Authors' own elaboration |
| Industry | 22 class dummies indicating the industries (HS-sections) where companies operate. | Authors' own elaboration from Colombian Customs Office (DIAN). |
| Sector Experience | Factor variable with one level (dummy variable) for each sector. Takes value 1 in all periods after a company exports for first time in a given sector (reflecting past experience in a sector). | Authors' own elaboration from Colombian Customs Office (DIAN). |
| Destination Experience | Factor variable with one level (dummy variable) for each destination. Takes value 1 in all periods after a company exports for first time in a given destination (reflecting past experience in a destination). | Authors' own elaboration from Colombian Customs Office (DIAN). |
| Exporter (importer) Experience | Variable counting the accumulated value exported (imported) in the last twelve months. | Authors' own elaboration from Colombian Customs Office (DIAN). |
| | *Models: SAM (COVID-19 variables)* | |
| Containment Economic Index | We consider the Economic Index from Hale et al. (2020) that provides a measure of the strength of the economic policies set in place to deal with the pandemic (such as income support and debt relief) for each country in the world. It ranges from 0 to 100. At the firm level we define two variables, one at the export and one at the import side, by taking a weighted average of these country level scores according to the monthly value of exports(imports) that a company ships(source) in every country.[1] | Hale et al. (2020) and Colombian Customs Office (DIAN). |
| Containment Government Index | We consider the Government Index from Hale et al. (2020) that measures the strictness of 'lockdown' style policies that primarily restrict people's behavior. It ranges from 0 to 100. At the firm level we define two variables, one at the export and one at the import side, by taking a weighted average of these country level scores according to the monthly value of exports(imports) that a company ships(source) in every country. | Hale et al. (2020) and Colombian Customs Office (DIAN). |
| Containment Health Index | We consider the Health Index from Hale et al. (2020) that combines 'lockdown' restrictions and closures with measures such as testing policy and contact tracing, short term investment in healthcare, as well investments in vaccine). Ranges from 0 to 100. At the firm level we define two variables, one at the export and one at the import side, by taking a weighted average of these country level scores according to the monthly value of exports(imports) that a company ships(source) in every country. | Hale et al. (2020) and Colombian Customs Office (DIAN). |
| Containment Stringency Index | We consider the Stringency Index from Hale et al. (2020) that records how the response of governments has varied over all indicators, becoming stronger or weaker over the course of the outbreak. Ranges from 0 to 100. At the firm level we define two variables, one at the export and one at the import side, by taking a weighted average of these country level scores according to the monthly value of exports(imports) that a company ships(source) in every country. | Hale et al. (2020) and Colombian Customs Office (DIAN). |
| | *Models: SUM and SAM (variables only for Logit, Logit-LASSO, and Logit-Ridge)* | |
| Size*Industry | Factor variables with 5 levels for each industry. Takes value 1 when the company size is Q1, value 2 when company size is Q2, value 3 when the size is Q3 and 4 when the size is Q4 while operating in a given industry. However, it takes value 0 if a company is not operating in this industry (for any size level). | Authors' own elaboration from Colombian Customs Office (DIAN). |
| Size*Sector | Factor variables with 5 levels for each sector. Takes value 1 when the company size is Q1, value 2 when company size is Q2, value 3 when the size is Q3 and value 4 when the size is Q4 while operating in a given sector. However, it takes value 0 if a company is not operating in this sector (for any size level). | Authors' own elaboration from Colombian Customs Office (DIAN). |
| Size*Means of Transportation | Factor variables with 5 levels for each sector. Takes value 1 when the company size is Q1, value 2 when company size is Q2, value 3 when the size is Q3 and value 4 when the size is Q4 while operating using a given means of transportation. However, it takes value 0 if a company is not operating using this means of transportation (for any size level). | Authors' own elaboration from Colombian Customs Office (DIAN). |
| Size*Destination | Factor variables with 5 levels for each sector. Takes value 1 when the company size is Q1, value 2 when company size is Q2, value 3 when the size is Q3 and value 4 when the size is Q4 while operating in a given destination. However, it takes value 0 if a company is not operating in this destination (for any size level). | Authors' own elaboration from Colombian Customs Office (DIAN). |

* https://data.europa.eu/euodp/en/data/dataset/covid-19-coronavirus-data

[1] When an exporter does not import we impute the corresponding internal Index (Economic, Government, Health, and Stringency) of Colombia to create the corresponding import side Index.

Table Appx.2: Sector-Industry mapping

| Section (Industry) | Industry Name | HS-Chapter (Sector) |
|---|---|---|
| 1 | Live Animals/ Animal Products | 1-5 |
| 2 | Vegetable Products | 6-14 |
| 3 | Animal or Vegatable Fats/Oils | 15 |
| 4 | Prepared Foodstuffs | 16-24 |
| 5 | Mineral Products | 25-27 |
| 6 | Products of Chemical Industries | 28-38 |
| 7 | Plastics, Rubber | 39-40 |
| 8 | Raw Hides, Skins and Leather | 41-43 |
| 9 | Wood | 44-46 |
| 10 | Paper | 47-49 |
| 11 | Textile | 50-63 |
| 12 | Footwear | 64-67 |
| 13 | Art. of Stone, Cement | 68-70 |
| 14 | Jewelries | 71 |
| 15 | Base Metals | 72-83 |
| 16 | Machinery Equipment | 84-85 |
| 17 | Vehicles | 86-89 |
| 18 | Precision Instruments | 90-92 |
| 19 | Arms | 93 |
| 20 | Misc. Manuf. Art. | 94-96 |
| 21 | Works of Art | 97 |
| 22 | Special Classification Provisions | 98-99 |

Source: Author's elaboration using Pierce and Schott (2012) tables.

# B Appendix - Growth of exports in 2019

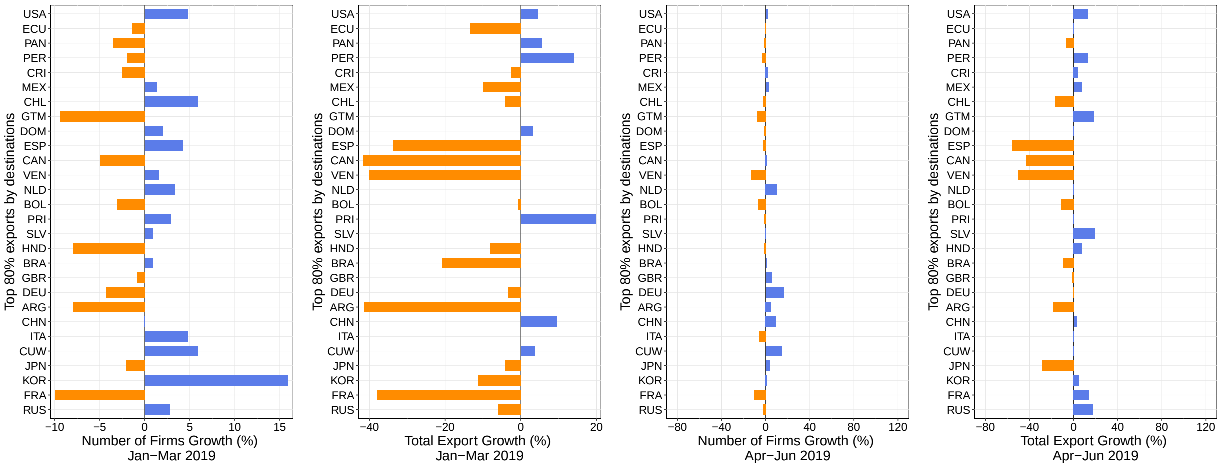


Figure Appx.1: The growth of the total number of exporters and the total value of exports by destination country for the first and the second quarters of 2019. Orange bars represent negative growth and blue bars positive growth. Destination countries are sorted from top to bottom accordingly with their importance in the share of number of exporters in 2019.

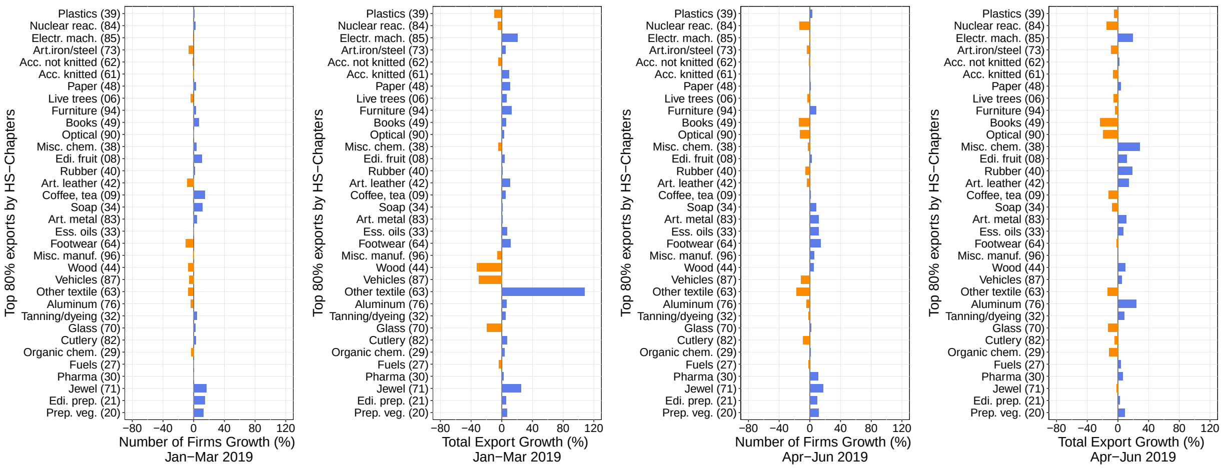


Figure Appx.2: The growth of the total number of exporters and the total value of exports by sector for the first and the second quarters of 2019. Orange bars represent export reductions and blue bars positive export growth. Product sectors are sorted from top to bottom accordingly with their importance in the share of number of exporters in 2019. Product sectors correspond to the chapters of the HS-code in parenthesis, the full name of the chapters is shortened to improve readability.

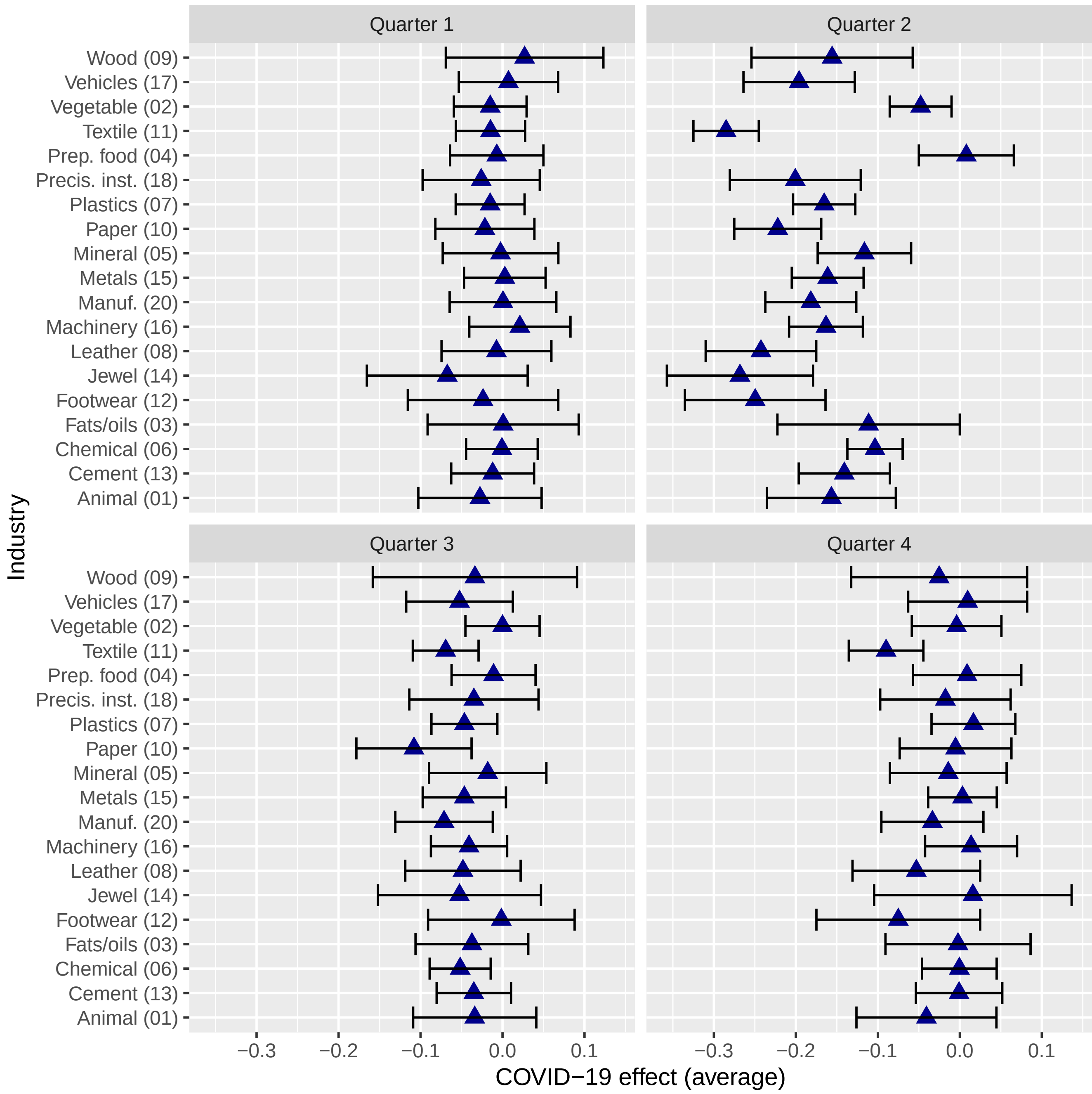


Figure Appx.3: Confidence intervals for the Treatment Effect of COVID-19, by export's industry. Intervals created using bootstrapping techniques.

# C Appendix - Construction of joint p-values

In the following is described the single-step algorithm. We will indicate the bootstrap version of a variable, $v$, as $\tilde{v}$ and its estimated version (on the original data) as $\hat{v}$.

The single step algorithm proceeds as follows: 1) for each variable $x \in X_t$, compute the CADiff for most (denoted by $(.)^{+u}$) and least (denoted by $(.)^{-u}$) affected firms as $\tilde{\Lambda}(x)^{+u}$ and $\tilde{\Lambda}(x)^{-u}$ respectively. We want to test the null hypothesis, $H_0$, that $\Lambda^u(x) = 0$, for $\Lambda^u(x) = [\Lambda(x)^{-u}, \Lambda(x)^{+u}]$. 2) Construct a bootstrap draw of the distribution of $(\hat{\Lambda}^{+u}(x) - \Lambda^{-u}(x))$, $Z^u_\infty(x)$. The latter is

obtained by exploiting the bootstrap version of $\Lambda^{+u}(x)$ and $\Lambda^{-u}(x)$, namely: $\tilde{Z}_\infty(x) = \sqrt{n}(\tilde{\Lambda}^u(x) - \hat{\Lambda}^u(x))$ where $\tilde{\Lambda}^u(x) = [\tilde{\Lambda}(x)^{-u}, \tilde{\Lambda}(x)^{+u}]$[21]. 3) Repeat steps 1) and 2) $B$ times; 4) compute a bootstrap estimator of the variance of $Z_\infty$ as $\hat{\Sigma}^u(x) = \frac{q^u_{0.75}(x) - q^u_{0.25}(x)}{z_{0.75} - z_{0.25}}$ being $q^u_p(x)$ the $p^{th}$ sample quantile of $\tilde{Z}_\infty(x)$ and $z_p$ the $p^{th}$ quantile of a standard normal distribution. 5) Use the latter to construct the test statistic $\tilde{\tau}(X_t) = sup_{x \in X_t} |\tilde{Z}_\infty(x)| \cdot |\hat{\Sigma}^u(x)|^{-1/2}$. A p-value for the null $H_0$ that $\Lambda^u(x) = 0$ for all $x \in X_t$ of the realization of the estimated statistic, $sup_{x \in X_t} |\hat{\Lambda}^u(x)| \cdot |\hat{\Sigma}(x)|^{-1/2} = s$, is given by the average number of times that $\tilde{\tau}(X_t)$ is greater than $s$.
We can relate the above algorithm to the setting of the present work, by noting that $s = \frac{\widetilde{CADiff_{original}}}{\hat{\Sigma}(x)}$, where $CADiff_{original}$ is employed in the main text instead of the more general $\hat{\Lambda}^u$ derived by Chernozhukov et al. (2018). Moreover, the numerator of $\hat{\Sigma}(x)$[22] is composed by the first and last quartiles of the distribution of the differences between the bootstrapped CADiffs, $CADiff_{boot}$ (equivalent of $\tilde{\Lambda}^u$ in the algorithm) and the resized original CADiffs, $CADiff_{original}$, i.e. $\tilde{Z}_\infty$ in the algorithm.

[21] Similarly, $\hat{\Lambda}^u(x) = [\hat{\Lambda}(x)^{-u}, \hat{\Lambda}(x)^{+u}]$

[22] As detailed in the procedure, the denominator is simply the difference between standard normal quantiles