

### Abstract

By interpreting exporters' dynamics as a complex learning process, this paper constitutes the first attempt to investigate the effectiveness of different Machine Learning (ML) techniques in predicting firms' trade status. 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 individual treatment effect 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 20*p.p.* in April 2020. Finally, we use a Classification Analysis (CA) to uncover the exporters' characteristics determining higher COVID-19 effects.

### Research Questions

- How to estimate a counterfactual for Colombian exporters under a worldwide shock?
- Is the COVID-19 effect heterogeneous? With respect to which firm's characteristics?

### Estimation Problem

1. We cannot observe the outcome under a pure non COVID-19 scenario for any firm during 2020: no control group is available
2. The economy-wide impact of the shock is coupled with complex interdependences between firms and products belonging to different sectors and countries: no variable measuring the extent of the firm-level exposure to COVID-19 is readily available.

### Data

We use monthly product level transactions data on exports and imports, and firms' characteristics reported at the Colombian Customs Office (DIAN) for 2018, 2019, and 2020. For 2020, we combine this data with four indexes [4] representing the strength of the measures taken by countries to contain the COVID-19 outbreak. Our final data set is composed by 1,975 explanatory variables.

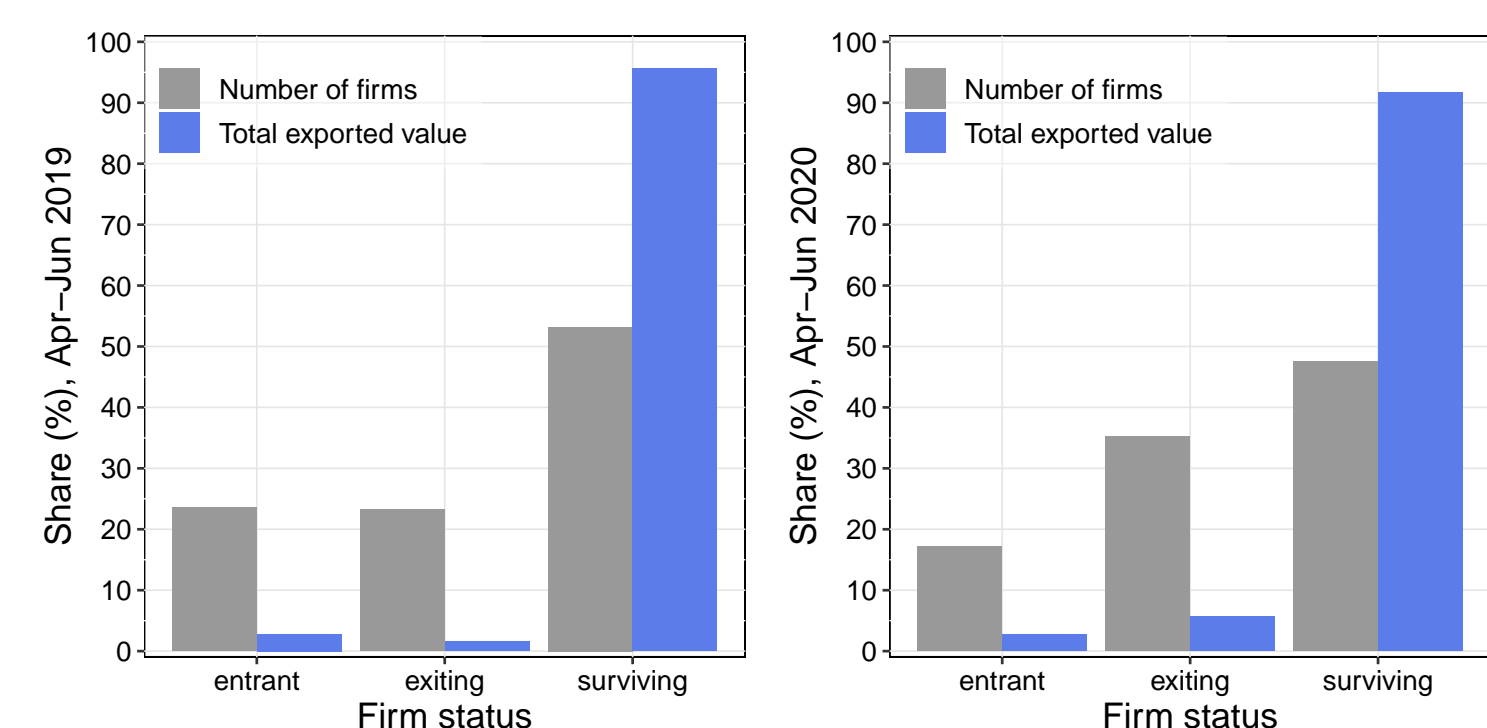


Fig. 1: Entry-exit dynamics and total export value by firms status, in the second quarter of 2019 and 2020.

### Model

We predict the outcome, whether a firm that was exporting in a given month in 2019 will export again in the same month of 2020, under two different scenarios:

1. SAM (Shock Aware Machine) → Observed prediction ( $\hat{Y}_{it}^{SAM}$ )
2. SUM (Shock Unaware Machine) → Counterfactual predictions ( $\hat{Y}_{it}^{SUM}$ )

Therefore, we define COVID-19 effect on individual Colombian exporters as:

$$\hat{\alpha}_{it} = \hat{Y}_{it}^{SAM} - \hat{Y}_{it}^{SUM}, \quad \forall \text{ months in } t. \quad (1)$$

We estimate the models relying on Machine Learning (ML) techniques (see also [2]), because they achieve high accuracy levels [1]. We compare the GoF (RMSE) of 4 models evaluating the SUM performance in 2018/19: *Logit* (0.64), *LASSO* (0.41), *RIDGE* (0.45), *RF* (0.41).

### Results

On March 25 2020, the Colombian government implemented a complete and mandatory lockdown.

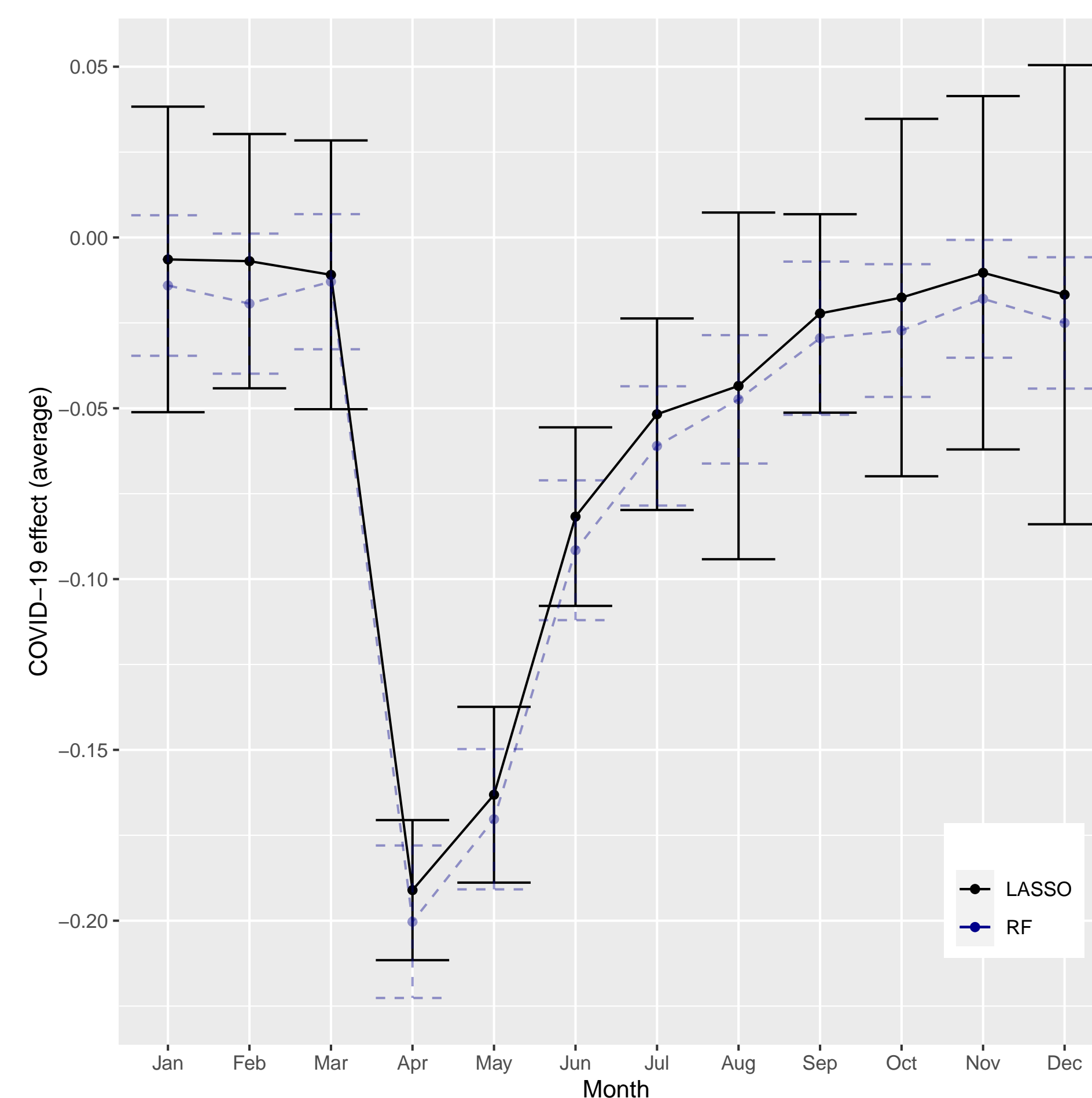
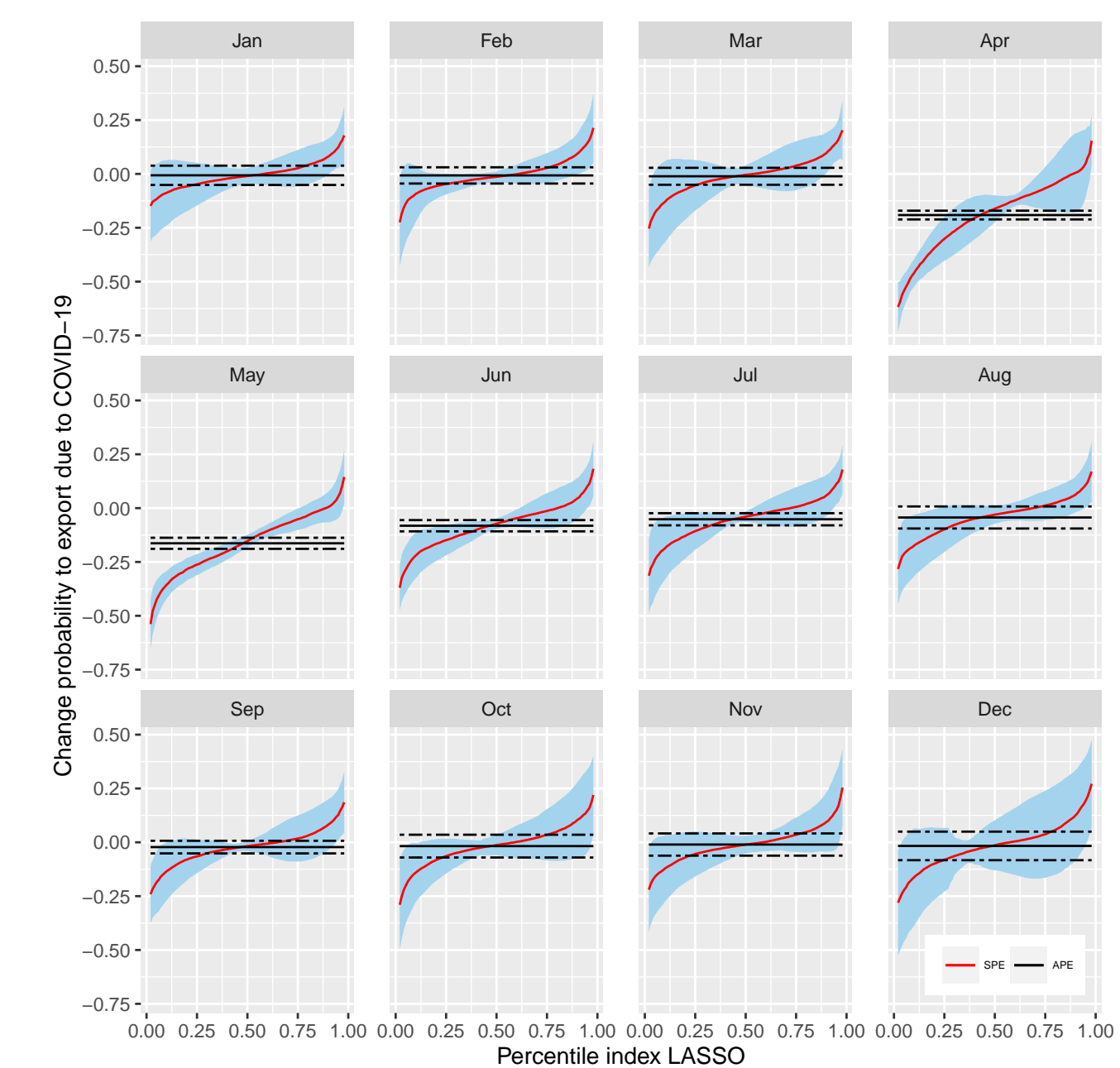


Fig. 2: Average Individual Treatment Effect, by months, comparison between Logit-LASSO and RF. Standard errors obtained with 100 bootstrap replications. Confidence intervals for a 5% significance level.

### Heterogeneity

We study along which firms' characteristics  $\hat{\alpha}$  displays the more relevant heterogeneity. Specifically, we adapt and apply to our setting the Sorted Partial Effects (SPE) and Classification Analysis (CA) method introduced in [3].



Variable	(1)	(2)	(3)
TE	0.313***	0.306***	0.279**
Air	-0.203*	-0.168***	-0.204***
Land	-0.0340	-0.0249	-0.0170
Sea	0.236***	0.192***	0.220***
Export Stringency Index	-19.36***	-19.51***	-7.180*
Import Stringency Index	-19.11***	-20.80***	-7.249***
Value Exported (log)	0.511***	0.449	0.570*
Value Imported (log)	1.816***	2.202***	2.686***
Deviation from sectoral mean		✓	✓
Deviation from monthly mean			✓

Fig. 3: SPE and CA differences.

### Conclusions

We show the potential of using ML to successfully construct a counterfactual scenario when all potential controls are treated by COVID-19. We find that the most affected Colombian exporters tend to have lower levels of imports and exports value in 2019, and higher levels of stringency both at destination and origin levels in 2020.

### References

[1] Falco J Bargagli-Stoffi, Jan Niederreiter, and Massimo Riccaboni. "Supervised learning for the prediction of firm dynamics". In: *arXiv preprint arXiv:2009.06413* (2020).

[2] Augusto Cerqua and Marco Letta. "Local economies amidst the COVID-19 crisis in Italy: a tale of diverging trajectories". In: *COVID Economics* 60 (2020), pp. 142–171.

[3] Victor Chernozhukov, Iván Fernández-Val, and Ye Luo. "The sorted effects method: discovering heterogeneous effects beyond their averages". In: *Econometrica* 86.6 (2018), pp. 1911–1938.

[4] Thomas Hale et al. "Variation in government responses to COVID-19". In: *Blavatnik school of government working paper* 31 (2020), pp. 2020–11.