A Synthetic Control Method Analysis of Schools Opening and Covid-19 Outbreak in Italy

Vincenzo Alfano, Salvatore Ercolano, Lorenzo Cicatiello

A Synthetic Control Method Analysis of Schools Opening and Covid-19 Outbreak in Italy

Abstract

Schools have been very central in the debate about COVID-19. On one hand, there have been many supporters of the argument to keep them open, for their importance for the youngsters, and for the effort many countries played in establishing protocols to keep them safe. On the other hand, a counter-argument supported by several other stakeholders, has been that being a major occasion of aggregation between teenager and adults accompanying their children, and being a major occasion of congestion per the public transportation, keeping the schools open favours the spread of the virus. In this article, by the means of a quantitative analysis that exploits the quasi experimental setting offered by the scattered opening that schools have had in Italy, we aim to shed some light on the subject. More precisely, a synthetic control method approach suggests that Bolzano, the first province in Italy to have opened the schools after the summer break, has way more cases than its synthetic counterfactual, built from a donor pool constituted by the other Italian provinces. Results seem to confirm that that opening the schools causes and increase in the infections, and this has to be taken into account by the policymakers.

Vincenzo Alfano
Department of Structures for Engineering and Architecture
University of Naples Federico II / Italy
vincenzo.alfano@unina.ot

Salvatore Ercolano
Department of Mathematics, Information
Sciences and Economics
University of Basilicata / Italy
salvatore.ercolano@unibas.it

Lorenzo Cicatiello
Department of Human and Social Science
University of Napoli L'Orientale
Naples / Italy
lcicatiello@unior.it

Introduction

A growing debate about the efficacy of non-pharmaceutical measures in containing COVID-19 outbreak (Lau et al. 2020; Sardar et al. 2020; Alfano and Ercolano, 2020a; Wong et al., 2020) has emerged over the last year. Differently from classical policies aimed to strengthen the health system, non-pharmaceutical measures, with their long history of being put in place to fight pandemics (Alfano and Sgobbi, 2020), tried to reduce the probability to contract the virus (Alfano and Ercolano, 2020a; Piguillem and Shi, 2020). Following Hale et al. (2020) the mix of measures adopted has characterized in term of heterogeneity the responses of national governments. In building a composite indicator, the cited authors make a distinction among eight different measures: i) School closing, ii) Workplace closing, iii) Cancel public events, iv) Restrictions on gathering size, v) Stop public transportation, vi) Stay at home requirements, vii) Restrictions on internal movement, viii) Restrictions on international travel. This work aims to focus on the first family of measures, assessing its importance on COVID-19 spread by the means of a quantitative analysis, exploiting a quasi-experimental setting occurred in Italy during September 2020.

We consider this as a very relevant issue, since the second wave of the pandemic, begun in European countries by the end of summer, has pushed this debate toward the possible effect of the school opening on the outbreak of the COVID pandemic (Lai et al., 2020; Tian et al., 2020; Viner et al., 2020, Alfano and Ercolano, 2020a). It is worth noting that during the first wave, schools closure has represented one of the first non-pharmaceutical measures adopted by the most of the countries in facing with the pandemic, and the most widely adopted in the world: no fewer than 107 countries implemented national school closures by 18 March 2020 (Viner et al., 2020). In fact, despite the importance of face to face teaching activities, keeping schools open could increase the diffusion of the virus via two mechanism: a catalyst of infection spread could be both the presence of numerous people in classrooms and the increase of people using local public transport (in order to reach schools) (Alfano and Ercolano, 2020b). Nevertheless, according to some opinions (Marino and Grassi, 2020), at the moment schools closure seems to have undefined benefits but more defined costs. All this makes the closure of schools a peculiar case among non-pharmaceutical intervention, especially since, despite a vast debate, its efficacy is still under scrutiny.

The evidence of the efficacy of closure of schools in order to fight the spread of an infection, comes mainly from previous evidence regarding other viruses. The literature especially focused on influenza (Bayham and Fenichel, 2020). With regard to this, there are also nuanced evidences: as a matter of fact, some studies suggest that to close schools is neither necessary, nor useful to reduce the spread of infection, if certain alternative policies are adopted (Melnick and Darling-Hammond, 2020). On the

other hand, Rashid et al. (2015) have suggested that school closures might be effective not because they prevent children to meet among them, but because forcing the parents to work at home, this policy manages to reduce the spread of the virus in workplaces.

It remains unclear whether closure of schools is effective in reducing the spread of COVID-19. The literature presents some modelling studies based on SARS, that have produced conflicting results (Becker et al., 2005), while COVID-19 modelling predicts that school closures prevent 2-4% of deaths (Viner et al., 2020). This reduction rate seems anyhow much lower than other social distancing measures. The main challenge, remains to understand how much closing the schools contributes in avoiding other, not specifically related to schools, contacts, such as the ones on public transportation to reach the schools

As pointed out by Alfano and Ercolano (2020a), Italy represents an interesting case study with regard to COVID-19, per several reasons. Among them, there is also the fact that at the beginning of September the spread of the virus started to increase in different provinces, but also due to the final authority on school opening being left by the national government to the different regions. In fact, while central government is responsible for the definition of the whole numbers of the days in the school year, local governments are responsible for the planning of the calendar: this means that each and any of the 19 regions in which Italy is divided, and of the 2 autonomous provinces of Trento and Bolzano, may decide when to open the schools on the territory they govern.

For this reason, looking at the Italian case, it is possible to observe different starting days for schools opening. In 2020, the autonomous province of Bolzano was the first one to open the schools on the 7th of September, while the most of the regions, following a governmental advice, opened the schools on 14th September. Some other regions delayed the opening to 16th September (Friuli), 22nd of September (Sardinia) or up to 24th of September (Abruzzo, Basilicata, Calabria, Campania and Puglia).

Thanks to this specificity of the Italian case, the present paper aims to contribute to the literature evaluating the efficacy, efficiency and the impact of policies to fight COVID-19 spread, by means of a synthetic control method (SCM from now on) design on coronavirus impact in Italy, based on provincial level data. SCM allows to evaluate the evolution of the pandemic against a synthetic counterfactual scenario (Abadie et al., 2010). In more detail, with this design we aim to analyse the effect of opening the schools in the province of Bolzano, a province that opened more than a week before the neighbouring provinces (and two weeks before the late openers) its schools. With this process, we may derive the effect of having the schools opened on the spread of COVID-19 infection. In order to do so, we build a synthetic control unit that mimics the trend observed in Bolzano before

the schools opened. Therefore, any difference observed after the schools opened in Bolzano can be imputed to the early opening of schools in this province.

Differently from previous contributions interested in the effect of school opening, to the best of our knowledge this paper represents the first attempt to use this methodology in order to measure the impact of such measure on pandemic containment, while this methodology has already been used to test the effect of policies against the spread of COVID-19, namely lockdown measures (Born et al., 2020) and the use of face masks (Sang-Wook, 2020). This attempt could mitigate some limitations of other methodological approaches based on cross regions comparisons. In fact, previous results could simply describe, on average, the correlation between school opening and the spread of pandemic. But in absence of a counterfactual, these results could derive also from unobservable characteristics within regions. Instead, SCM allows to assess a causal effect of the opening of school on the evolution of the contagion, and therefore provide very valuable information to policy-makers.

The main limitation of SCM is its external consistency (or lack thereof). In fact, the generalization of the results of this study to different contexts should be done with extreme caution, given the intrinsic limitation on external validity of the procedure. However, the evidence presented in this article could nonetheless help solving a part of the puzzle about the spread of the pandemic, and to suggest the impact of schools opening on the unfolding of the pandemic, at least in Italy.

Nevertheless in our opinion, the measurement issue in this kind of measures, suggests the need of a deeper investigation using local data. In fact, as pointed out by Hale et al. (2020), the measurement issue derives from the heterogeneity in the Governments' responses to COVID 19. Looking at the school case, the authors state that in each country "in some places, all schools have been shut; in other places, universities closed on a different timescale than primary schools; in other places still, schools remain open only for the children of essential workers." For this reason we think that our approach could represent one of the possibility to overcome this kind of limitations.

The rest of the paper is structured as follow: section 2 describes the data and methodology, section 3 comments the main results of the empirical estimations while section 4 conclude the paper.

2. Data and methodology

SCM is designed to evaluate the effects of treatments performed on a small number of units (Abadie, Diamond, and Hainmueller, 2010; Abadie, 2019). Our empirical setting is coherent with this scenario,

as we exploit the decision by the province of Bolzano of opening the schools earlier than any other province in Italy. Therefore, the treated unit in our analysis is the province of Bolzano, and the treatment is the opening of schools, that occurred the 7th of September 2020. The rationale behind SCM is that all the units (the provinces in our empirical setting) behave similarly before the treatment, and that just one of them differs from the others. However, since all the provinces are different from one another, the idea is to build a counterfactual (synthetic) unit by estimating a weighted average of the non-treated units (known as donor pool). The weights are computed in order to have a synthetic control unit that matches the treated unit on the relevant variables before the treatment. Therefore, differences between treated and synthetic control unit after the treatment depend on the treatment itself (Abadie, 2010).

Our analysis focuses on the log of positive cases, in order to take into account the exponential nature of the evolution of the COVID-19 pandemic. However, since the pandemic hit rather heterogeneously during the first wave in Italy (lasted about from the end of February to the beginning of June), we focus on the total of new positive cases observed from the first of August. In fact, by using the total number of cases observed from the beginning of the pandemic we would underestimate the variation of positive cases in regions where the pandemic hit more severely during the first wave, and overestimate the variation of positive cases where the first wave was less acute. For this reason, we track the evolution of the pandemic from the first of August, when the curve of contagion became flat (as presented in Figure 1), and take as outcome variable the log of the total number of cases observed after that day (for the sake of clarity we will call this variable total new cases).

To build a counterfactual, we require the synthetic control unit to track as close as possible the log of total new cases each day up to seven days before the treatment. We also require the synthetic control unit to be as close as possible to a number of variables that potentially predict the spread of the virus within a province. We include the total number of cases the day before the treatment, because a greater number of cases during the first wave could impact on herd immunity and on citizens behaviour, a proxy for income per capita in the province, total population, the share of population in schooling age, population density, and the share of population in the province living in municipalities with less than 30 thousand inhabitants. Data about COVID-19 infections are gathered from the Italian Ministry of Health's dataset, which reports official data for each province and day; income per capita in the province is computed by dividing the total taxable income of the province per the population of the same province (data about taxable income taken from the Italian Ministry of Economic and Finance, MEF, and about population from ISTAT); all the other data is extracted from the Italian National Statistics Institute (ISTAT) warehouse.

At a first step of the analysis we opt for including all the provinces different from Bolzano in the donor pool. This choice allows to perform the computation of weights on a large number of provinces, therefore exploiting the potential of SCM. Then, we run a set of placebo tests performing the same analysis on all the provinces in the donor pool. According to Galiani and Quistorff (2017, p. 836), "if the distribution of placebo effects yields many effects as large as the main estimate, then it is likely that the estimated effect was observed by chance". This allows to estimate p-values of the estimated effect. However, the placebo effect may be imprecisely matched in the pre-treatment period, which would result in p-values being too conservative. For this reason, we weight placebo effects using their pre-treatment match quality (measured as the root mean squared prediction error, or RMSPE) to get standardized p-values (Galiani and Quistorff, 2017).

Table 1 summarizes the six predictors for Bolzano and both the operationalization of synthetic Bolzano, where the latter is constructed with positive weights assigned to Reggio nell'Emilia, Aosta, Crotone and Caserta, in descending order.

3. Results

Figure 2 graphically presents the infection dynamic in Bolzano's province (solid line) compared with a counterfactual outcome of Bolzano (dashed line) built from all the other Italian provinces (that, as explained, at the time had closed schools, given that Bolzano was the first province to open them). The vertical axis measures cumulative infections in logs, while the horizontal one represents calendar days, starting from August the 26th. As it is possible to see, the period before the vertical line (posed at September the 7th, day in which Bolzano opened the schools) suggest that our choice of variables created a counterfactual that follows a dynamic pretty similar to the one of Bolzano, given the strong overlapping among the two lines.

After the schools opened, the dynamic of COVID's incidence in the real Bolzano grows more than the one in its counterfactual. The incidence of COVID in the former gets bigger than in the latter on September the 12th, 5 days after the opening of schools, and keeps growing more than the counter factual. Table 2 summarizes analytically the data from which the graph per Bolzano and synthetic Bolzano is built. It shows how 11 days after the opening of schools, the gap between the two becomes very high, jumping from an incidence equal to 0.08 to one of 0.19. This time lag is coherent with the time needed to suspect being infected and getting tested, suggesting an impact of schools into this.

Hence, these findings suggest that indeed there has been a strong impact on the COVID-19 infections due to the opening of the schools.

Figure 3 reports the difference between each province in the donor pool and its estimated synthetic counterfactual, showing the placebo effects estimated for all the provinces in the donor pool contrasted to the effect estimated for Bolzano, represented by the solid black line. The trend of Bolzano starts to increase around 10 days after the opening of schools, becoming higher than the majority of the placebo effects. This correspond to lower standardized p-values, as shown in Table 2: the 18th of September the effect estimated for Bolzano is larger than the 97% of the placebo effects (i.e. 1 – the standardized p-value). all these suggests that the incidence measured in Bolzano is due to what differentiate this province from the others, hence it is due to the opening of schools.

Please also notice that this finding, in terms of lag of days from the treatment to the registration of an increase in incidence, is in line with the average amount of time the literature suggest is needed in order to present symptoms after a COVID-19 infection; as a matter of fact, 97.5% of those who develop symptoms do so within 11.5 days of infection, with a 95% confidence interval of between 8.2 and 15.6 days (Lauer et al., 2020). The cases in Bolzano keep being higher and growing more than in the synthetic Bolzano for all the period analysed, going from a delta of 0.192649 in the incidence, observed on September the 18th, ten days after the opening of schools, up to a delta of 0.387235 observed in the end of the month. All these suggest a major and very significant role in the spread of the infection played by the opening of the schools.

Conclusion

Schools are a very important part of the formation of each human being. Children and teenagers obtain much more from the hours spent in school than just what they learn in classes (which, nonetheless, is a very important thing in building grown-up women and man and forging the citizens of tomorrow). Nonetheless, schools plays a major role in determine occasion of meeting between students and their relatives. During a pandemic, this means to increase the occasion of circulation of the virus.

Even if schools are very controlled, and several protocols are put in place to avoid having un-detected infections between their walls, students have to reach somehow the school, and it is hard to check their behaviour once the classes are over. Furthermore, to accompany children at school means to increase the circulation within the city, when many people at the same time have to reach specific points in the city, and also to fill the public transportation system, for parents that chose to, or that have to, do not reach their children school with private means of transportation.

All this suggest that there is indeed an effect of schools opening on the spread of the infection. This is what our analysis suggests: the first province in Italy to open the schools, has had an incidence on its COVID-19 cases much higher two weeks after the opening than a synthetic build counterfactual, from data of other Italian provinces, suggests it would have been.

While the external validity of our results is questionable, and of course it has to be taken with caution, our findings suggest that to open schools causes an increase in COVID-19 cases. While this do not necessary means that the cost in terms of spread of the infection is lower than the cost of imposing distance learning to a generation of students, it is an important result that the policymaker has to take into account while deciding how to face the pandemic.

Accordingly, we may suggest to the policymaker that any decision to open or re-open the schools should be considered very carefully. As a matter of fact, a rushed re-opening of the schools may condemn hard-won results due to social-distance policies or other sacrifices to waste.

References

Abadie, A., Diamond, A. and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. Journal of the American Statistical Association 105 (490), 493–505.

Abadie, A. (2019). Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. Article prepared for the Journal of Economic Literature.

Alfano, V. and Ercolano, S. (2020a). The Efficacy of Lockdown Against COVID-19: A Cross-Country Panel Analysis. Appl Health Econ Health Policy 18, 509–517 (2020).

Alfano, V. & Ercolano, S. (2020b). In fila per tre. Apertura delle scuole e nuova ondata di COVID, Economia e Politica, 12-2/2020.

Alfano, V. and Sgobbi, M. (2020). A Peste, Fame et Bello libera Nos, Domine. An analysis of the black death in Chioggia in 1630. Journal of Family History, forthcoming.

Bayham, J. and Fenichel, E. P. (2020). Impact of school closures for COVID-19 on the US health-care workforce and net mortality: a modelling study. The Lancet Public Health, 5 (5): 271-278.

Becker, N. G., Glass, K., Li, Z., Aldis, G.K. (2005) Controlling emerging infectious diseases like SARS. Math Biosci 2005; 193: 205–21.

Born, B., Dietrich, A., Müller, G. (2020). The lockdown effect: A counterfactual for Sweden, CEPR Discussion Paper 14744.

Galiani, S., & Quistorff, B. (2017). The Synth_Runner Package: Utilities to Automate Synthetic Control Estimation Using Synth. The Stata Journal: Promoting Communications on Statistics and Stata, 17(4), 834–849. https://doi.org/10.1177/1536867X1801700404.

Hale, T., Angrist, N., Cameron-Blake, E., Hallas, L., Kira, B., Majumdar, S., Petherick, A., Phillips, T., Tatlow, H., Webster, S. (2020). Oxford COVID-19 Government Response Tracker, Blavatnik School of Government.

Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Schubert, J., Bania, J. and Khosrawipour, T. (2020). The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. Journal of Travel Medicine.

Lauer S. A., Grantz K. H., Bi Q., Jones, F. K., Zheng, Q., Meredith, H., Azman, A. S., Reich, N. G. and Lessler, J. (2020). The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Ann Intern Med. 2020

Marino, I. and Grassi, I. (2020). Dalle scuole chiuse benefici incerti, ma costi certi. Lavoce.info, https://www.lavoce.info/archives/70803/dalle-scuole-chiuse-benefici-incerti-ma-costi-certi/

Melnik, H. and Darling-Hammond, L. D. (2020). Reopening Schools in the Context of COVID-19: Health and Safety Guidelines From Other Countries. Learning policy Institute, Policy Brief.

Piguillem, F. and Shi, L. (2020). The optimal covid-19 quarantine and testing policies (No. 2004). Einaudi Institute for Economics and Finance (EIEF).

Rashid, H., Ridda, I., King, C., Begun, M., Tekin, H., Wood, J. G., Booy, R. (2015). Evidence compendium and advice on social distancing and other related measures for response to an influenza pandemic. Paediatr Respir Rev; 16: 119–26.

Sang-Wook (Stanley), C. (2020). Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: The case of Sweden, The Econometrics Journal 23 (3): 323–344.

Sardar, T., Nadim, S. S. and Chattopadhyay, J. (2020). Assessment of 21 Days Lockdown Effect in Some States and Overall India: A Predictive Mathematical Study on COVID-19 Outbreak. arXiv preprint arXiv:2004.03487.

Viner R. M., Russell S. J., Croker H., Packer J., Ward J., Stansfield C., et al. (2020). School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. Lancet Child Adolesc Health.; 4(5): 397-404. https://doi.org/10.1016/S2352-4642(20)30095-X PMID: 32272089

Wong, M., Huang, J., Teoh, J., & Wong, S. (2020). Evaluation on different non-pharmaceutical interventions during COVID-19 pandemic: an analysis of 139 countries. The Journal of Infection.

Table 1 – Profile of synthetic Bolzano

Variables	Bolzano	Synthetic Control
Log new COVID-19 cases	5.537668	5.538229
Total COVID-19 cases	3003	4890.966
Income	22067.89	21426.52
Total population	532080	496281.9
Population between 6 and 18 (percent)	0.138203	0.130536
Population density	71.9027	216.1658
Population in municipalities under 30,000 inhabitants (percent)	0.720762	0.678091
RMSPE	-	0.270721

Table 2 – Incidence of COVID-19 in Bolzano and synthetic Bolzano (donors' pool composed by all Italian provinces), per each day

Time	Bolzano	Synthetic Control	Effect	Standardized p-values
Donors and weights: Aosta (.057), Caserta (.011), Crotone (.046), Reggio nell'Emilia (.886)				
31 August	5.3844953	5.4061125	-0.02162	-
1 September	5.4424176	5.4387505	0.003667	-
2 September	5.484797	5.4573819	0.027415	-
3 September	5.5012584	5.4940976	0.007161	-
4 September	5.5451775	5.5626968	-0.01752	-
5 September	5.6312118	5.5911222	0.04009	-
6 September	5.655992	5.6384523	0.01754	-
7 September	5.655992	5.7172209	-0.06123	0.009434
8 September	5.6698809	5.7554905	-0.08561	0.009434
9 September	5.7037826	5.7664534	-0.06267	0.0849057
10 September	5.768321	5.8060448	-0.03772	0.1981132
11 September	5.8464389	5.8584832	-0.01204	0.754717
12 September	5.9162021	5.8789369	0.037265	0.245283
13 September	6.0185933	5.93553	0.083063	0.1037736
14 September	6.0354815	5.9727754	0.062706	0.1320755
15 September	6.0844994	6.023238	0.061261	0.1509434
16 September	6.1114674	6.0393867	0.072081	0.1415094
17 September	6.1441855	6.0571279	0.087058	0.1415094
18 September	6.2878585	6.0952099	0.192649	0.0283019
19 September	6.3350544	6.1164861	0.218568	0.0188679
20 September	6.3699012	6.1489584	0.220943	0.0188679
21 September	6.403574	6.1642845	0.23929	0.0188679
22 September	6.4167323	6.1836441	0.233088	0.0377358
23 September	6.4754329	6.2073806	0.268052	0.0283019
24 September	6.505784	6.2227948	0.282989	0.009434
25 September	6.5806394	6.2411897	0.33945	0.009434
26 September	6.6200733	6.256578	0.363495	0.0188679
27 September	6.6489844	6.2787581	0.370226	0.0188679

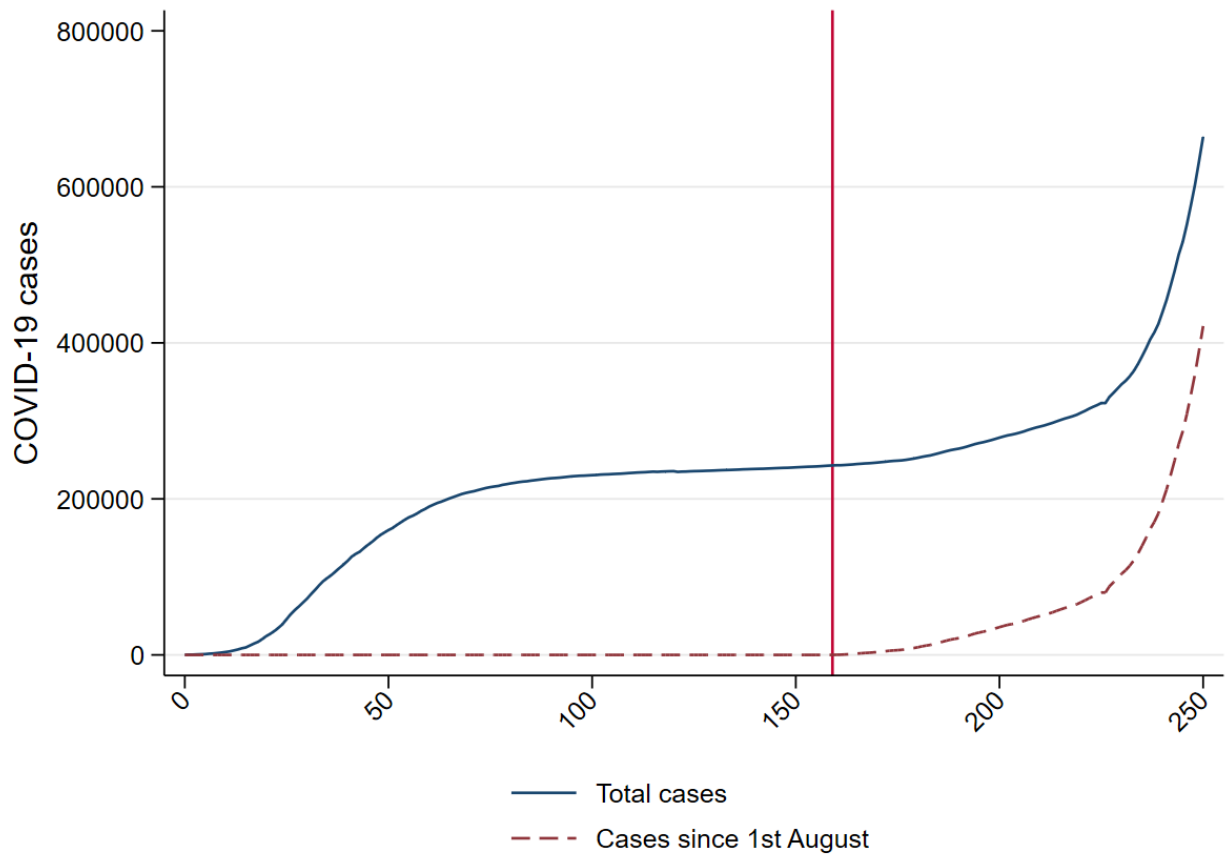|				
28 September	6.6808548	6.2913282	0.389527	0.009434
29 September	6.6908422	6.3129962	0.377846	0.0188679
30 September	6.7117405	6.3245054	0.387235	0.0188679

Figure 1 – Total COVID-19 cases in Italy, from the beginning of the pandemic (22nd of February) and from August the 1st (second wave). Vertical line represents 1st of August.
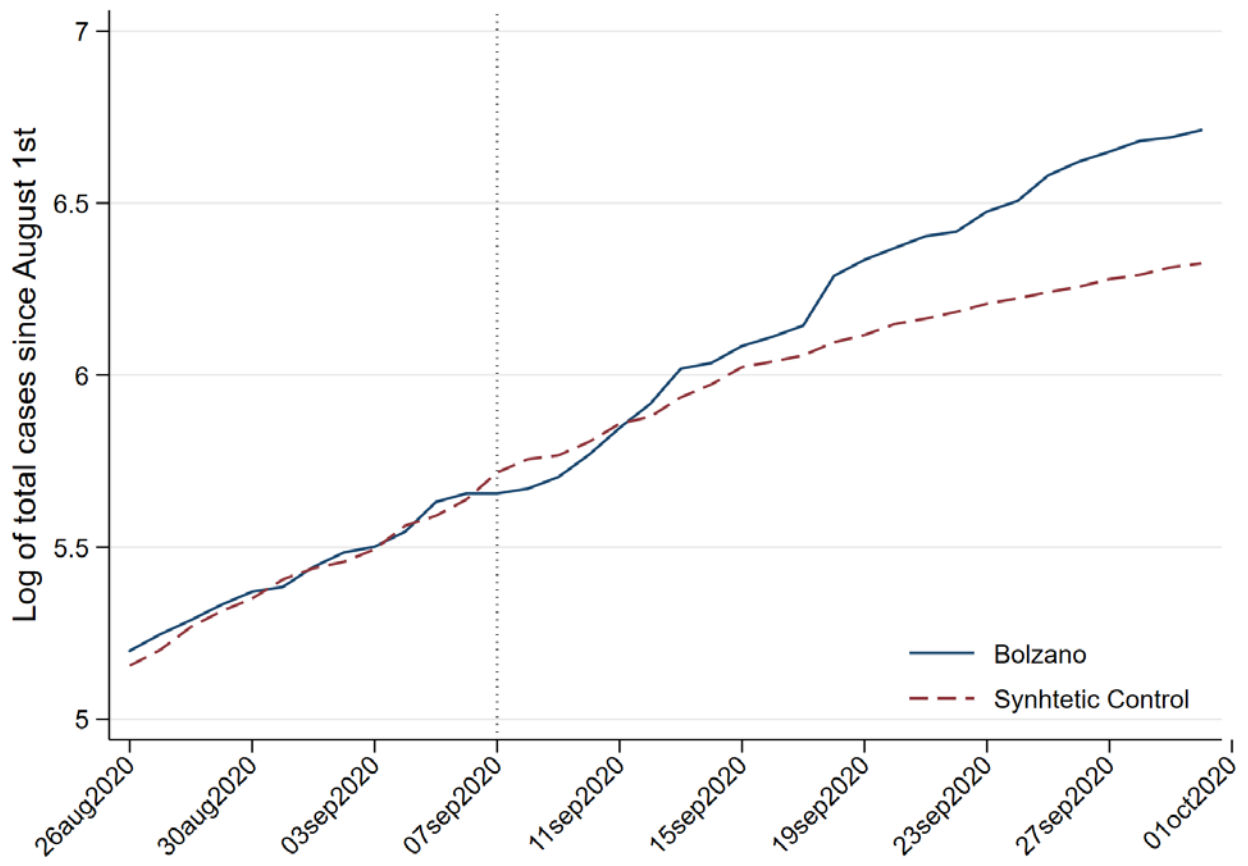
Figure 2 – SCM with synthetic Bolzano build from a donor pool composed by all Italian provinces.
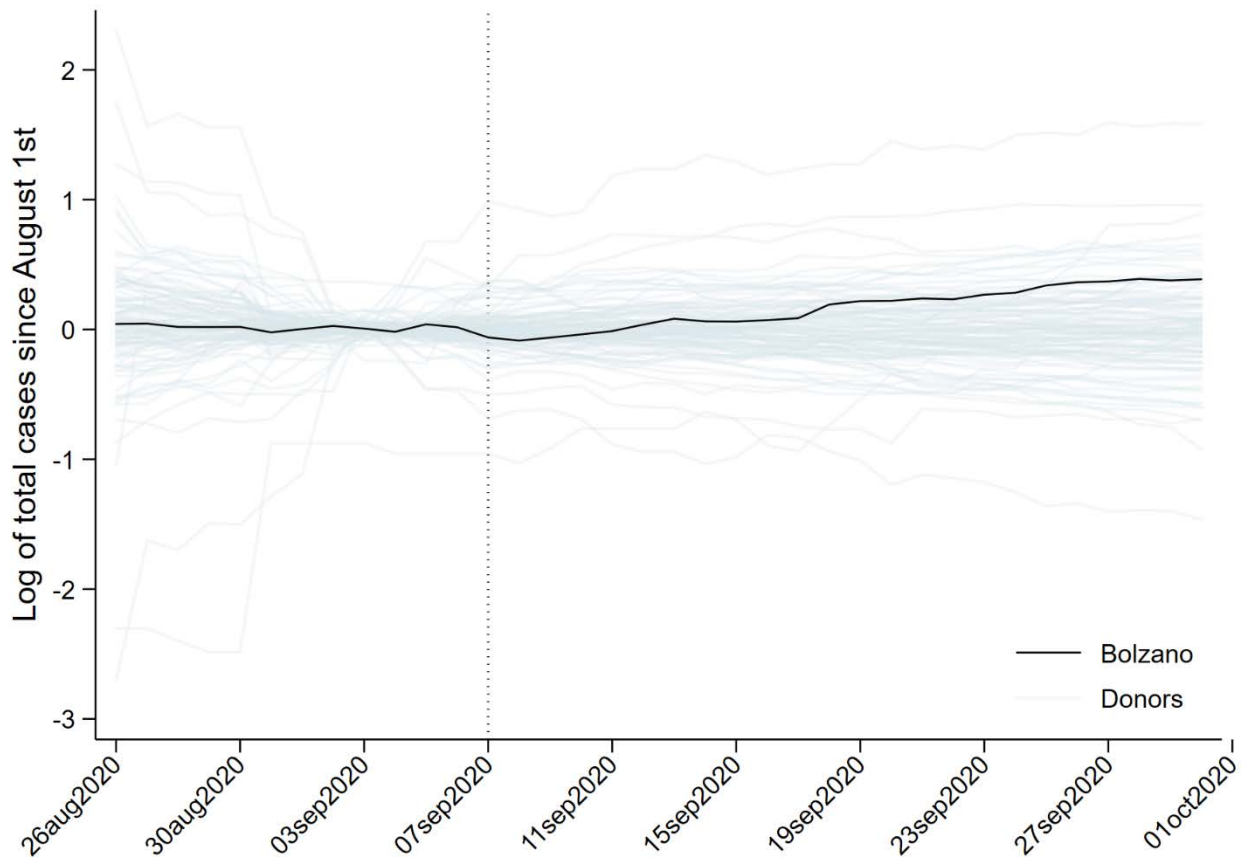
Figure 3 – Effects estimated for all the provinces in the donor pool for the same treatment period.