Intertemporal propensity score matching for casual inference: an application to covid-19 lockdowns and air pollution in Northern Italy

Propensity score matching intertemporale per inferenza causale: un'applicazione su covid-19 lockdown e inquinamento atmosferico nel Nord Italia

Daniele Bondonio and Paolo Chirico

Abstract This paper develops an intertemporal propensity score matching (PSM) approach for estimating the impact of covid-19 lockdowns on air pollution. While PSM has been exclusively applied in the context of matching cross-sectional units, this paper shows that, under specific circumstances, PSM can be also applied for estimating causal inference by means of matching across different temporal units in the context of multivariate time series data. We apply our intertemporal PSM model to the data collected from a large number of air-pollution-measurement stations in Northern Italy, estimating the casual effect of the March-May-2020 lockdown on air-pollution without resorting to the more stringent functional form assumptions of the existing literature.

Abstract *Questo paper sviluppa una procedura intertemporale di propensity score matching (PSM) per stimare l'impatto dei lockdown covid-19 sull'inquinamento atmosferico. Sebbene il PSM sia stato applicato esclusivamente nel contesto dell'abbinamento di unità cross-sezionali, questo paper mostra che, in circostanze specifiche, il PSM può essere applicato anche per stimare l'inferenza causale mediante l'abbinamento tra diverse unità temporali nel contesto di serie temporali multivariate. Applichiamo il nostro modello di PSM intertemporale ai dati raccolti da un ampio numero di stazioni di rilevamento dell'inquinamento atmosferico nel Nord Italia, stimando l'effetto casuale del lockdown di marzo-maggio-2020 senza ricorrere alle più stringenti ipotesi di forma funzionale dei modelli regressivi utilizzati nella letteratura esistente.*

Key words: propensity score matching, air pollution, coronavirus lockdown

Daniele Bondonio
University of Piemonte Orientale (UPO)
e-mail: daniele.bondonio@uniupo.it

Paolo Chirico
University of Piemonte Orientale (UPO)
e-mail: paolo.chirico@uniupo.it

1 Introduction

Statistical matching is widely used as tool for estimating casual inference in the context of quasi-experimental designs in which cross-sectional units are sorted into treated and untreated units based on the exposure of the program intervention of interest. For all known statistical matching estimations, including propensity score matching (PSM), the temporal periods in which the outcome variables are measured have to remain equal across all treated and untreated units. This is a strict requirement for all quasi-experimental designs, in which the existence of the comparison group of untreated units is aimed at controlling for the secular trends that may affect Y independently from the treatment. For this reason, intertemporal PSM matching (i.e. matching across units with Y measured at different times) is not usually considered a viable impact identification strategy. The stringent covid-19 lockdowns of the first half of 2020 provide an exceptional source of an abrupt exogenous decrease on human activities (transportations and productions) that offers a unique opportunity to quantify the effect of human activities on air pollution in urbanized areas. What is unique about the impact identification conditions related to the covid-19 lockdowns, however, is also the fact that the available data are in terms of a number of multivariate time-series, one for each air-pollution measurement station, with units of observations represented by the days in which air pollution and weather characteristic are measured. These features of the data represent a scenario in which the abrupt change in the secular trend (i.e. the lockdown-induced sharp decrease in traffic and production activities) is the treatment of interest, while all the major confounding factors to be controlled for (i.e. the weather characteristics) are observable in the data and they can be safely assumed of not being subject to unobserved secular trend (within the short pre- post-treatment periods of times considered in the analysis). This paper aims at showing that in such a data scenario, an intertemporal statistical matching in which, separately for each air-pollution measurement station, the treated days (corresponding to the lockdown periods) are matched to previous non-lockdown days (untreated units) that share the same weather features, is a viable empirical option that overcomes the limitations of the recently-emerging literature on the air-quality impact of the covid-19 lockdowns.

2 Existing Literature and Impact Identification Properties of Intertemporal PSM

Since spring 2020, an increasingly large number of studies have been using covid-19 lockdowns data on a variety of locations around the world in order to estimate the reduction of air pollution caused by an abrupt decrease of traffic and human activities. A number of these papers fail to properly address the challenges of a reliable causal inference that requires, above all, a suitable controlling of the possible differences in the distribution of the weather characteristics of the pre-lockdown days (untreated units) and the lockdown days (treated units) [1, 3, 4, 14]. These papers identify the causal effect of the lockdowns only under the very stringent assumption of a perfectly equal distribution of the weather characteristics between the lockdown and the pre-lockdown periods. Other studies make use of panel-data regression approaches that have to face the issue of error autocorrelation, and that require strong functional form assumptions with respect to the way in which the observable co-founding factors affect air-pollution [7, 5, 15, 16].

The intertemporal PSM matching approach developed in this paper can be described as follows. Let $\{Y_t\}$ be a stochastic process on which an intervention lasting for a time interval L_1 produces an effect. The causal effect of the intervention should be determined as the difference:

$$\mathrm{E}[Y_t^{(1)}|t \in L_1] - \mathrm{E}[Y_t^{(0)}|t \in L_1] \tag{1}$$

where $Y_t^{(1)}$ is the observable outcome at time t if time t were affected by the intervention, and $Y_t^{(0)}$ if not affected. In the case of the lockdown studies, the temporal units t are represented by the days in which air-pollution and weather characteristics are measured. The difference (1) can be considered as an intertemporal version of the standard *Average Treatment effects on the Treated* (ATT). As $\mathrm{E}[Y_t^{(0)}|t \in L_1]$ is not observable, the intertemporal ATT (1) can be identified in terms of

$$E_{X_1}\left[\mathrm{E}[Y_t|t \in L_1, X_1] - \mathrm{E}[Y_t|t \in L_0, X_1]\right] \quad X_1 = X_t|t \in L_1 \tag{2}$$

when two conditions apply: i) the period of observation $L = L_0 \cup L_1$, that includes also a previous time-interval L_0, is short enough so that the process Y_t can be assumed to have no underlying trend (besides the break in the observed outcome caused by lockdown period that represents the treatment under consideration); ii) the existing multivariate process $\{X_t\}$ of observable covariates (i.e. the weather controls) makes the expected value of $Y_t^{(0)}$ independent of $t \in L$, i.e.:

$$\mathrm{E}[Y_t^{(0)}|t \in L_1, X_t] = \mathrm{E}[Y_t^{(0)}|t \in L_0, X_t] \tag{3}$$

Similarly to cross-sectional PSM, if assumption (3) is satisfied, the intertemporal ATT can be also identified when conditioning on propensity score such as $\varphi_t = P[t \in L_1|X_t]$:

$$E_{\varphi_1}\left[\mathrm{E}[Y_t|t \in L_1, \varphi_1] - \mathrm{E}[Y_t|t \in L_0, \varphi_1]\right] \quad \varphi_1 = \varphi_t|t \in L_1 \tag{4}$$

This leads to the following intertemporal PSM estimation procedure: (i) treated times are matched with the untreated ones that have an equal (similar) propensity score; (ii) the sample mean of Y_t for the matched treated and control times ($M_t = 1$) is computed; (iii) the ATT is estimated as the difference of the sample means ii):

$$(\bar{y}_t|t \in L_1, M_i = 1) - (\bar{y}_t|t \in L_0, M_i = 1) \tag{5}$$

iv) the estimated ATT is validated if the balancing property of PSM is checked [13]; v) in the case of data containing multiple multivariate time-series, the procedure (i)-

(iv) is repeated separately for each time-series, yielding a series of local ATTs that are then averaged-out to a global impact estimate.

Compared to the panel-data regression approaches adopted elsewhere in the literature, this intertemporal PSM approach share the same well-known advantage [2, 8] of the standard cross-sectional PSM in strongly diminishing the sensibility of the estimated impacts to the choice of functional form that links the effects of the controls on the outcome variable.

3 The impact of covid-19 lockdowns on air pollution in northern Italy

We apply the intertemporal PSM model described in the previous section to the air-pollution data from five regions in Northern Italy located along the Po-river Valley, i.e. Piedmont, Lombardy, Emilia-Romagna, Veneto, Friuli Venezia Giulia. These regions represent the most industrialized and densely-populated area of Italy (with an average density of 252 inhab. per km2). Due to these characteristics and the specific orographic features of the Po Valley (that determine a systemic lack of sustained ventilation), this area has some of the highest levels of air-pollution in Europe. For the analysis we collected data on the average daily levels of nitrogen dioxide (NO2) and particulate matter (PM10) detected at 83 ARPA air-pollution measurement stations, located in the main urban areas of the five regions. The observation period covers both the entire length of the nationwide covid-19 lockdown period of March 10 - May 17, 2020 (that posed a very strict stay-at-home mandate and closure of all schools, universities and non-essential services) and the corresponding (untreated) period of the previous year (March 10 – May 17, 2020). The assembled database also includes as controls (X_t) the main weather characteristics that are known in the literature to influence air pollution: average daily temperature (TEMP), rainfall (RAIN), maximum daily wind speed (WIND) [7, 5, 15, 16].

Table 1 shows the average values of the air-pollutants and weather variables during the 2020 lockdown days (treatment) and the corresponding (untreated) period of 2019. These descriptive statistics do not provide a reliable indication of causal effect of the lockdown on air-pollution because the comparison between the treated and non-treated days does not control for the possible differences in the distribution of the weather variables in the two periods. Indeed, the data from Table 1 show, for example, that the lockdown (L_1) period of 2020 experienced less than half of the rain than the corresponding non-treated period of 2019 (L_0). For this reason, any casual inference drawn from a descriptive-statistics comparison between the two periods would be highly biased, particularly for the PM10 outcomes that has a high potential of being influenced by rain and wind conditions.

In order to grant a perfect balancing also of the locational characteristic of the ARPA measuring station (i.e. distance from major roads, highways, airports, surrounding production activities, etc.), we apply to the analysis our intertemporal PSM approach in terms of comparisons between the 2020 lockdown and the 2019 (untreated) period holding constant the same ARPA station. This empirical strategy entails estimating 83 different local impact parameters obtained by applying the intertemporal PSM model (implemented in terms of caliper radius matching) separately to the single multivariate time-series data from each the ARPA stations. The final (global) estimated impact is then obtained as the average of the different local impacts for which the balancing property was successfully tested by means of the Rubin's R and B indices [13].

Table 1 Descriptive statistics of air-pollution and weather characteristics in the 2020 lockdown and corresponding untreated period of 2019

period	N 83x69	NO2 (μg/m^3)	PM10 (μg/m^3)	WIND (m/s)	RAIN (mm)	TEMP (C^o)
10/3 - 17/5, 2019	5727	25.45 (16.08)[a]	20.27 (12.05)	2.28 (1.47)	2.85 (7.73)	12.55 (2.80)
10/3 - 17/5, 2020	5727	15.26 (11.12)	23.90 (17.41)	2.20 (1.49)	1.35 (5.11)	13.74 (4.15)

[a] Std.dev. in parentheses

The main results from the intertemporal PSM analysis are summarized in Table 2. Restricting the focus on the working days only, the covid-19 lockdown is shown to cause a reduction of about 13.1 ($\mu g/m^3$) in the daily average level of NO2 and a reduction of about 3.3 ($\mu g/m^3$) in the level of PM10, compared a counterfactual scenario of no-lockdown (estimated from a comparison group composed by the corresponding calendar days of 2019 with similar weather characteristics). These estimates correspond to 51.3% and a 16.2% decrease in the daily average level of NO2 and PM10, respectively. Taking into account also the week-ends and the festivities ("All days" column), the estimated impacts, albeit slightly smaller, remain of a quite large magnitude: about -12.1 ($\mu g/m^3$), corresponding to -47.7%, and -2.9 ($\mu g/m^3$), corresponding to -14.1%, for the NO2 and the PM10, respectively.

Table 2 Locdown effect on NO2 and PM10 estimated by means of intertemporal PSM

polluttant	Aggregate ATT estimates working days	all week days	Descriptive-statistics Difference $L_1 - L_0$
NO2	-13.07 *** (.221)[a]	-12.12 *** (.206)	-10.27 *** (.745)
PM10	-3.29 *** (.378)	-2.86 *** (.364)	3.60 *** (.220)

Statistical significance: *=at 10% level; **= at 5% level; ***= at 1% level
[a] Std.dev. in parentheses

Due to a higher incidence of unfavourable weather conditions in the lockdown days, compared to previous-year period, the results from a mere descriptive-statistics comparison ($L_1 - L_0$, last column in Table 2) are indeed of a lower magnitude than the intertemporal PSM estimates. This is particularly relevant for the case of PM10, for which the descriptive-statistics analysis is suggestive of a worsening impact of the lockdown (+3.6$\mu g/m^3$), a result that highlights how misleading could be the evidence produced in the absence of adequate statistical tools for a reliable causal inference.

Acknowledgements This work was supported by funding from the University of Piemonte Orientale. We thank Giuseppe Gruttad'Auria, Andrea Nigido, Giulia Pasculli and Michela Rosselli for their contribution in the data acquisition process.

References

1. Anil, I., Alagha, O.: The impact of COVID-19 lockdown on the air quality of Eastern Province, Saudi Arabia. Air Qual Atmos Health 14: 117–128 (2020)
2. Angrist, J.: Estimating the labor market impact of voluntary military service using Social Security data on military applicants. Econometrica 66, 249–88 (1998)
3. Arregocés, H. A., Rojano, R., Restrepo, G.: Impact of lockdown on particulate matter concentrations in Colombia during the COVID-19 pandemic. Science of The Total Environment, 764, 142874 (2021)
4. Collivignarelli, M. C., Abbà, A., Bertanza, G., Pedrazzani, R., Ricciardi, P., Miino, M. C.: Lockdown for CoViD-2019 in Milan: What are the effects on air quality?. Science of the total environment, 732, 139280 (2020)
5. Dang H.A.H., Trinh, T.A.: Does the covid-19 lockdown improve global air quality? new cross-national evidence on its unintended consequences. Journal of Environmental Economics and Management 105:102–126 (2021)
6. Dehejia, R.H., Wahba, S.: Propensity score-matching methods for nonexperimental causal studies. Review of Economics and statistics, 84(1), 151-161 (2002)
7. He G., Pan Y., Tanaka, T.: The short-term impacts of covid-19 lockdown on urban air pollution in china. Nature Sustainability 3(12):1005–1011 (2020)
8. Heckman, J.J., Ichimura, H., Todd, P.: Matching as an econometric evaluation estimator. The review of economic studies, 65(2), 261-294 (1998)
9. Holland, P.W.: Statistics and causal inference. Journal of the American statistical Association 81(396):945–960 (1986)
10. Imbens, G.W., Rubin D.B.: Causal inference in statistics, social, and biomedical sciences. Cambridge University Press (2015)
11. Rosenbaum, P.R., Rubin, D.B.: The central role of the propensity score in observational studies for causal effects. Biometrika 70(1):41–55 (1983)
12. Rubin, D.B.: Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of educational Psychology 66(5):688 (1974)
13. Rubin D.B.: Using propensity scores to help design observational studies: application to the tobacco litigation. Health Services and Outcomes Research Methodology 2(3):169–188 (2001)
14. Seo, J.H., Jeon, H.W., Sung, U.J., Sohn, J.R. (2020). Impact of the COVID-19 outbreak on air quality in Korea. Atmosphere, 11(10), 1137.
15. Song, Y., Li, Z., Liu, J., Yang, T., Zhang, M., Pang, J.: The effect of environmental regulation on air quality in china: A natural experiment during the covid-19 pandemic. Atmospheric Pollution Research 12(4):21–30 (2021)
16. Wang, J., Xu, X., Wang, S., He, S., He, P.: Heterogeneous effects of covid-19 lockdown measures on air quality in northern china. Applied Energy 282:116–179 (2021)