Innovation & Society

Statistics and Data Science for Evaluation and Quality

Editors: Giovanna Boccuzzo, Enrico Bovo, Marica Manisera, Luigi Salmaso

University of Padova
Casa della Gioventù
Brixen – Bressanone
June 25-27, 2025

cleup

First edition: June 2025
ISBN 978 88 5495 849 4
© IES 2025 - Statistical Methods for Evaluation and Quality
2025 Cleup sc
"Coop. Libraria Editrice Università di Padova"
via G. Belzoni 118/3, Padova (t. +39 049 8753496), Italy
www.cleup.it
www.facebook.com/cleup
All rights reserved.

No part of this book may be translated or copied in whole or in part without written permission of the authors, except for brief excerpts in connection with reviews or scholarly analysis.

IES 2025 - Innovation & Society: Statistics and Data Science for Evaluation and Quality

BOOK OF SHORT PAPERS

Editors: Giovanna Boccuzzo, Enrico Bovo,
Marica Manisera, Luigi Salmaso

Book of Short Papers
2025 Conference of the 12th Scientific Meeting of the Statistics for the Evaluation and Quality of Services Group of the Italian Statistical Society (SVQS)
IES 2025 - Innovation & Society: Statistics and Data Science for Evaluation and Quality University of Padova

Casa della Gioventù, Brixen - Bressanone
June 25-27, 2025

Traffic Restrictions and Air Quality: A Counterfactual Matrix Completion Analysis of Milan's Area C

Roxana Adam[1], Francesco Biancalani[1], Rodolfo Metulini[2]
[1]AXES, IMT School for Advanced Studies Lucca
[2]Department of Economics, University of Bergamo

Traffic Restrictions and Air Quality: A Counterfactual Matrix Completion Analysis of Milan's Area C

Roxana Adam, Francesco Biancalani, Rodolfo Metulini

Abstract Urban traffic is seen as one of the major contributors to air pollution in densely populated areas. Many cities have implemented traffic restriction zones to mitigate the high levels of pollutants within urban areas. Milan's Area C, introduced in January 2012, is a prominent case. While prior studies utilizing classical econometric techniques have produced mixed results, we apply matrix completion, a statistical learning method, to evaluate the causal effect of Area C on air quality. Our analysis reveals a statistically significant reduction in PM_{10} levels due to Area C restrictions within the treated area.
Keywords: air quality, matrix completion, environmental policy, traffic restrictions

1. Introduction and literature review

Urban centers face a problem with air pollution, because of the high dense concentration of vehicles. Various policies have been implemented globally to mitigate these negative impacts, such as traffic restrictions. The previous literature investigated the effectiveness of these strategies in reducing air pollution in high-density urban areas, drawing on empirical evidence from case studies in the United Kingdom [1], China [2], India [3] and Portugal [4]. We focus our attention on the peculiar case of the city of Milan (Italy) since its urban area represents one of the most polluted areas in Western Europe.

The 2012 establishment of Milan's Area C, a traffic control zone in "Cerchia dei Bastioni", spurred research on its air quality impact. A key study by Fassò et al. [5] analyzed traffic restrictions within and around Area C, revealing complex environmental effects. While the policy improved air quality inside Area C and potentially reduced overall pollution, concerns emerged about increased pollution outside due to traffic diversion. Findings suggest both benefits and unintended consequences, highlighting the need for further assessment. Invernizzi et al. [6] assessed Milan's Ecopass zone's impact on air pollution, focusing on black carbon concentrations. Unlike other pollutants, black carbon, mainly traffic-generated, showed a notable decline in restricted areas. This highlights black carbon's sensitivity as a traffic pollution indicator and confirms the effectiveness of traffic restrictions in reducing it. A more recent study by Maranzano et al. [7] examined the same impact on a wider restricted zone in Milan ("Area B"). Using time series intervention analysis, they found initial reductions in pollutant concentrations but the effects on nitrogen oxides (NO_x) and nitrogen dioxide (NO_2) remained inconclusive, underscoring the need for continuous monitoring and analysis to evaluate the long-term effects of such policies accurately.

Given Milan's significance as a case study and the availability of a rich and free database for the Lombardy region, we aim to contribute to the existing literature by applying the statistical learning technique called matrix completion to analyze the impact of "Area C" in Milan. This innovative approach enables a rigorous analysis of the problem using a robust methodology, ensuring an accurate measurement of the causal effects of policy interventions

This short paper is structured as it follows: Section 2 introduces the methodological strategy, Section 3 describes the dataset, Section 4 presents the results and 5 concludes.

2. Methods

In this paper, we employ a novel methodological approach, matrix completion [8] , to assess the impact of environmental policy, as done by Biancalani et al. [9] for the study of the impact of the European union emissions trading system on CO_2. This method is particularly well suited for our analysis, as it relaxes the stringent parallel trend assumption, which is often difficult to satisfy in real-world settings [10]. Traditional methods, such as those based on instrumental variables, regression discontinuity design, and propensity score matching, are often not applicable because they require strong identifying assumptions. However, by resorting to matrix completion, we can effectively estimate causal effects, making this method a powerful tool for our analysis. Specifically, we adopt the Fixed Effects Nuclear Norm Matrix Completion method [10] (from now on, MCFE) via the soft-impute algorithm [11], implemented in the `fect` package in `R` [12]. Unlike other implemented packages for matrix completion, `fect` allows for the inclusion of unit-time varying covariates such as the meteorological ones, making it a preferred choice. MCFE optimization problem can be formulated as follows:

$$\begin{aligned} \underset{\hat{\mathbf{M}}\in\mathbb{R}^{m\times n},\hat{\mathbf{L}}\in\mathbb{R}^{m\times n},\hat{\mathbf{\Gamma}}\in\mathbb{R}^{m\times 1},\hat{\mathbf{\Delta}}\in\mathbb{R}^{n\times 1},\boldsymbol{\beta}\in\mathbb{R}^{p\times 1}}{\text{minimize}} \quad & \left(\frac{1}{|\Omega^{\text{tr}}|}\sum_{(i,j)\in\Omega^{\text{tr}}}\left(M_{i,j}-\hat{M}_{i,j}\right)^2+\lambda\|\hat{\mathbf{L}}\|_*\right), \\ \text{subject to} \quad & \hat{\mathbf{M}}=\hat{\mathbf{L}}+\hat{\mathbf{\Gamma}}\mathbf{1}_n^\top+\mathbf{1}_m\hat{\mathbf{\Delta}}^\top+\mathbf{X}\boldsymbol{\beta}. \end{aligned} \tag{1}$$

where Ω^{tr} is a subset of pairs of indices (i, j) corresponding to the positions of known entries of a matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$, $\mathbf{1}_n$ and $\mathbf{1}_m$ are column vectors consisting of, respectively, n and m entries all equal to 1, $\hat{\mathbf{M}}$ is the completed matrix to be estimated, where $\hat{\mathbf{L}}$, $\hat{\mathbf{\Gamma}}$ and $\hat{\mathbf{\Delta}}$ have to be estimated, $\lambda \geq 0$ is a regularization constant chosen via an internal cross validation that controls the tradeoff between adequately fitting the known entries of the matrix $\mathbf{M}$ and achieving a small nuclear norm of the first term $\hat{\mathbf{L}}$, and $\|\hat{\mathbf{L}}\|_*$ is the nuclear norm of the matrix $\hat{\mathbf{L}}$. The inclusion of time-fixed effects accounts for factors that may have contributed to pollution reduction over the years, such as the adoption of low-emission vehicle fleets and improvements in building heating systems. Simultaneously, unit-fixed effects control for variations in baseline pollution levels across different areas, enhancing the accuracy of the analysis. MCFE does not explicitly specify assumptions on the residual terms; however, by incorporating unit- and time-fixed effects, it isolates idiosyncratic shocks from correlated trends. This ensures robustness even under strong serial and spatial autocorrelation.

We adopt the average of the Average Treatment Effect on the Treated (ATT) as a measure to quantify the causal effect of the traffic policy on the pollutants' level. ATT is the difference between the observed values under treatment and what would have happened in the absence of treatment (i.e. the counterfactual values) and can be defined, for treated period t, as:

$$ATT_t = \frac{1}{N_t}\sum_{i\in\tau}(Y_{it}-\hat{Y}_{it}^0), \tag{2}$$

where Y_{it} is the observed value for treated unit i at time t, $\hat{Y}_{it}^0$ the estimated value in the absence of treatment, τ the set of treated units, and N_t number of treated units at time t.

In our context, the untreated levels of pollutants for the treated Area C in Milan's downtown after January 2012 are unknown and must be estimated. MCFE allows us to compare them to actual levels of pollutants and quantify the impact of the traffic restriction. By considering a matrix where monitoring stations are in rows and time instants in columns, the core idea is to artificially obscure the entries related to monitoring stations located within Area C after January 2012 while considering the remaining matrix entries as observed.

3. Data processing and preliminary evidences

We use data from the Lombardy Regional Environmental Agency (ARPA Lombardia), freely accessible at `https://www.arpalombardia.it/` and retrieved using `ARPALData` package in `R` [13] . The data comprises pollution and weather variables from several air quality monitoring stations in the Lombardy region. These stations can collect high-frequency data, though some may experience intermittent outages or lack certification. Two stations are located within Milan's traffic-restricted Area C, namely, *Milano Verziere* (ID 528) and *Milano, Via Senato* (ID 548). We consider these two stations as the treated group. To exclude potential sources of spillover effects on air quality, we exclude from the analysis the stations within the municipality of Milan that are outside Area C. We use the remaining Lombardy stations as a control group. In this study, we employ PM_{10} monthly data from 2008 to 2019 (expressed in $\mu g/m^3$), obtained by aggregating by month original data. The choice of the period was driven by the potential shocks caused by the onset of COVID-19 in 2020 that might affect pollutants' levels, while the decision to aggregate at the monthly level was made to account for climate variations throughout the year. We incorporate temperature, rainfall, and relative humidity to avoid the possible presence of confounding factors. Furthermore, meteorological data were filtered to retain only values within the following valid ranges: rainfall (0-100 mm), temperature (-30°C to +50°C), and relative humidity (0-100%). Temperature is converted from Celsius to Kelvin by adding 273.15. To maintain consistency when merging air quality and meteorological data, we only include monitoring stations situated within a 10 km radius of a weather station. Following data processing, the analysis uses a matrix of 23 monitoring stations (including two treated) over 144 months.

Table 1 presents summary statistics for PM_{10} levels by group (treatment in Area C, control outside) and period (pre-December 2011, post-January 2012). Considering the full period, the control group displays a lower mean compared to the treatment group (34.14 versus 39.34). Before the treatment, the PM_{10} mean for the control group equals 37.31, while the mean of the treated group is 44.88. After the treatment, the PM_{10} mean for the control group decreases to 32.56 (-4.75), while the mean of the treated group decreases to 36.56, for a reduction of 8.32 points. Standard deviations are relatively similar across groups and periods, suggesting comparable variability in PM_{10} levels. It appears that treatment may have decreased PM_{10} in Area C. The matrix completion approach in Section 4 will allow us to evaluate the causal impact of the restriction on the reduction of pollution.

Table 1: Summary statistics for the PM_{10} pollutant. sd stays for standard deviation.

period	group	mean	sd	min	Q1	Q2	Q3	max
full	control	34.14	15.64	6.95	22.25	30.02	43.57	104.45
	treated	39.34	17.09	13.86	25.77	34.21	50.45	97.46
pre treatment	control	37.31	16.95	11.64	24.44	32.52	47.65	104.45
	treated	44.88	19.15	18.42	29.03	39.37	57.40	97.46
post teatment	control	32.56	14.69	6.95	21.04	28.60	41.75	96.79
	treated	36.56	15.28	13.86	23.64	32.34	47.72	84.69

4. Matrix completion results

Table 2 shows the results of the analysis based on the matrix of PM_{10} values without the inclusion of covariates, indicating a significant reduction on the average PM_{10} concentration due to the traffic restriction policy. The ATT is approximately -3.344, with a standard error of 0.943. The 95% bootstrap Confidence Interval (CI) for this effect ranges from -5.193 to -1.496, and the p-value is 0.000. These statistics suggest a strong and significant reduction in PM_{10} levels associated with the treatment, highlighting its effectiveness in decreasing air pollution concentrations in the studied area. Table 3 shows the results of

Table 2: Average of the Average Treatment Effect on the Treated (ATT), estimated using fixed effects matrix completion (MCFE) in `fect` package. PM_{10} as response variable.

	estimate	standard error	CI lower	CI upper	p-value
Average ATT	-3.344	0.943	-5.193	-1.496	0.000

the analysis with the addition of meteorological variables rainfall, temperature, and relative humidity as covariates, indicating a similar effect of the restriction policy on the average PM_{10} concentrations as compared to the case without covariates. Specifically, the ATT indicates a reduction of approximately 3.446 units in PM_{10} levels, with a standard error of 0.992. The confidence interval for this effect ranges from -5.390 to -1.503, and the p-value is 0.001, indicating strong statistical significance and suggesting that the treatment was effective in reducing PM_{10} concentrations. Rainfall, temperature, and humidity show limited influence on PM_{10} levels, with small coefficients and high p-values (0.670, 0.505, and 0.836), indicating no significant effects. Overall, the main finding is a significant reduction in PM_{10} levels due to the treatment, with estimates remaining consistent across different sets of covariates, which themselves have minimal impact.

5. Conclusions

Our findings suggest that Milan's Area C traffic restriction significantly reduced PM_{10} levels, with an estimated ATT of -3.3 units. Results remain robust across models, indicating the reduction is driven by policy rather than weather conditions. This highlights the effectiveness of traffic restrictions in curbing urban air pollution and provides insights for

Table 3: Average of the Average Treatment Effect on the Treated (ATT), estimated using fixed effects matrix completion (MCFE) in `fect` package. PM_{10} as response variable. Addition of meteorological covariates.

	estimate	standard error	CI lower	CI upper	p-value
Average ATT	-3.346	0.992	-5.390	-1.503	0.001
Rainfall	1.360	3.191	-4.894	7.614	0.670
Temperature	0.049	0.074	-0.096	0.195	0.505
Relative Humidity	0.007	0.036	-0.063	0.078	0.836

policymakers. However, further research is needed to refine the model and assess impacts on other pollutants.

Acknowledgement

The contribution has been developed for the project "Study of mobile phone siGNals for the evalUation of the interconnections between Mobility and the environment in Lombardia (SIGNUM)" CUP: F53D23010910001- PRIN 2022 PNRR M4C2 - financed by the European Union - Next Generation EU (DD MUR n. 1409 del 14/09/2022). The authors thank Alessandro Fassò for his valuable feedback.

References

[1] Alan M. Jones, Roy M. Harrison, Benjamin Barratt, and Gary Fuller. A large reduction in airborne particle number concentrations at the time of the introduction of "sulphur free" diesel and the London Low Emission Zone. Atmospheric Environment, 50:129-138, 2012.

[2] Zheng Pan, Xuxian Zhang, and Chengzheng Li. Does driving restriction affect air quality: evidence from Guangzhou city in China. Applied Economics Letters, 32(3):357-364, 2025.

[3] Asha Chelani. Statistical characteristics of ambient PM2.5 concentration at a traffic site in Delhi: Source identification using persistence analysis and nonparametric wind regression. Aerosol and Air Quality Research, 13(6):1768-1778, 2013.

[4] Raquel Menezes, Helena Piairo, Pilar García-Soidán, and Inês Sousa. Spatial-temporal modellization of the NO2 concentration data through geostatistical tools. Statistical Methods & Applications, 25:107-124, 2016.

[5] Alessandro Fassò. Statistical assessment of air quality interventions. Stochastic Environmental Research and Risk Assessment, 27(7):1651-1660, Oct 2013.

[6] Giovanni Invernizzi, Ario Ruprecht, Roberto Mazza, Cinzia De Marco, Grisa Mocnik, Costantinos Sioutas, and Dane Westerdahl. Measurement of black carbon concentration as an indicator of air quality benefits of traffic restriction policies within the ecopass zone in Milan, Italy. Atmospheric Environment, 45(21):3522-3527, 2011.

[7] Paolo Maranzano, Alessandro Fassò, Matteo Pelegatti, and Manfred Mudelsee. Statistical modeling of the early-stage impact of a new traffic policy in Milan, Italy. International Journal of Environmental Research and Public Health, 17(3):1088, 2020.

[8] Trevor Hastie, Robert Tibshirani, and Martin Wainwright. Statistical learning with sparsity. Monographs on statistics and applied probability, 143(143):8, 2015.

[9] Francesco Biancalani, Giorgio Gnecco, Rodolfo Metulini, and Massimo Riccaboni. The impact of the European Union emissions trading system on carbon dioxide emissions: a matrix completion analysis. Scientific Reports, 14(1):19676, 2024.

[10] Susan Athey, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. Matrix completion methods for causal panel data models. Journal of the American Statistical Association, 116(536):1716-1730, 2021.

[11] Rahul Mazumder, Trevor Hastie, and Robert Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. The Journal of Machine Learning Research, 11:2287-2322, 2010.

[12] Licheng Liu, Ye Wang, and Yiqing Xu. A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. American Journal of Political Science, 68(1):160-176, 2024.

[13] Paolo Maranzano and Andrea Algieri. ARPALData: an r package for retrieving and analyzing air quality and weather data from ARPA Lombardia (Italy). Environmental and Ecological Statistics, pages 1-32, 2024.