The impact of the COVID-19 lockdown on air pollution in Lombardy

Francesco Granella [1,2] Lara Aleluia Reis [2] Valentina Bosetti [1,2] Massimo Tavoni [2,3]

[1]Bocconi University

[2]RFF-CMCC European Institute on Economics and the Environment

[3]Politecnico of Milan

June 30, 2020

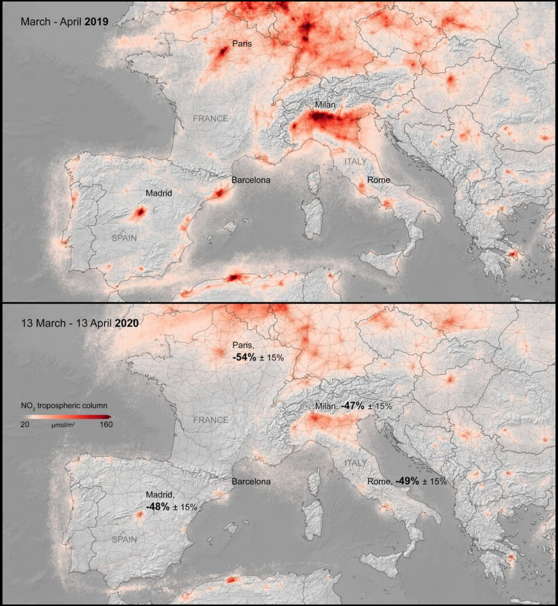
- $PM_{2.5}$ fifth leading risk factor for death worldwide, 4.2 million premature deaths in 2015.
  GBD, Cohen et al. (2017).
  Lower physical, cognitive productivity e.g. Graff Zivin and Neidell (2012); Künn et al. (2019).
- 71 000 premature deaths in Europe attributed to NO_2 in 2016 European Environment Agency (2019)

- Uncertainty over Policy $\mapsto$ Emissions $\mapsto$ Concentrations

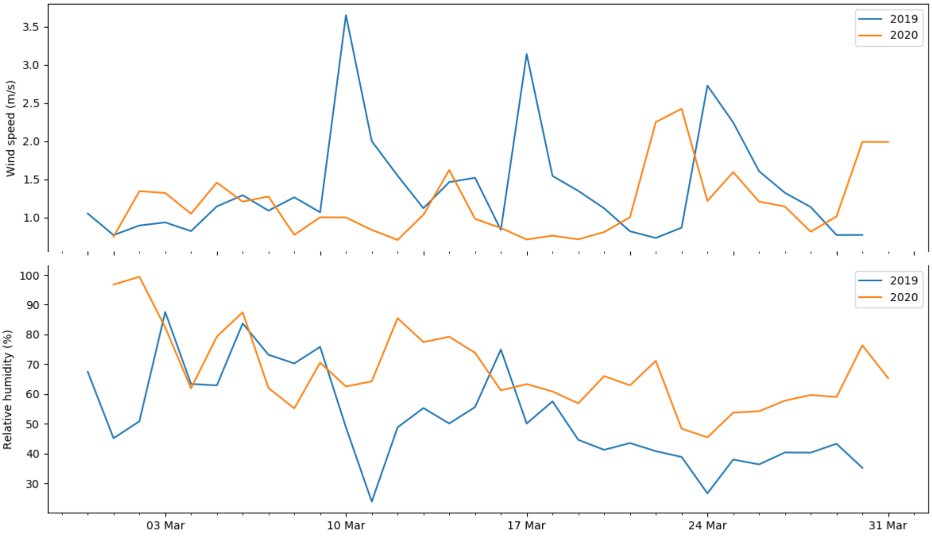
- $PM_{2.5}$ fifth leading risk factor for death worldwide, 4.2 million premature deaths in 2015. GBD, Cohen et al. (2017).
  Lower physical, cognitive productivity e.g. Graff Zivin and Neidell (2012); Künn et al. (2019).
- 71 000 premature deaths in Europe attributed to NO_2 in 2016 European Environment Agency (2019)
- Uncertainty over Policy $\mapsto$ Emissions $\mapsto$ Concentrations

This paper

1. Estimating effect of lockdown on concentrations of $PM_{2.5}$ and NO_2 in Lombardy.
2. ML: Do not require comparable unit assumption $\rightarrow$ scalable
3. Compare years of life saved by air quality improvement to years of life lost to COVID-19


March - April 2019
Paris
Milan
FRANCE
ITALY
Barcelona
Rome
Madrid
SPAIN
13 March - 13 April 2020
Paris,
-54% ± 15%
NO_2 tropospheric column
20
μmol/m²
160
Milan, -47% ± 15%
FRANCE
ITALY
Barcelona
Rome, -49% ± 15%
Madrid,
-48% ± 15%
SPAIN

Weather in March in Milan, 2020 vs 2019



Similar studies

- Effect of lockdown of Chinese cities on pollution.
  - He et al. (2020) DID, control for weekly average temp, $temp^2$, wind speed, precipitation
  - Cole et al. (2020) Weather-normalized data + augmented synthetic control method
  - **Require comparable but untreated cities**
- ML and pollution
  - Short-term predictions

Turning the problem upside down

1. Train a ML algorithm to predict concentrations as a function of only weather in BAU times
2. Predict concentrations during lockdown: counterfactual

In Lombardy

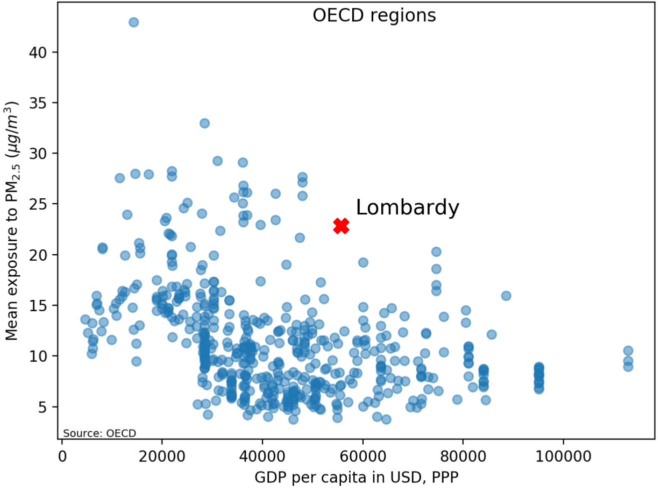
Pop-weighted $PM_{2.5}$ down by 22% (6 $\mu g/m^3$) Pop-weighted NO_2 down by 36% (17 $\mu g/m^3$)
Years of life saved by improved air quality $>$ 10% years of life lost to COVID

Outline

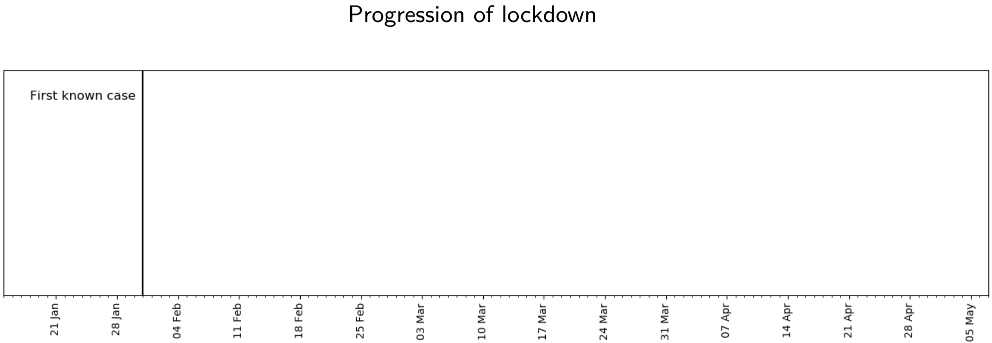
1. Lockdown of Lombardy
2. Empirical strategy
3. Data
4. Validity of counterfactual
5. Results

Why Lombardy?

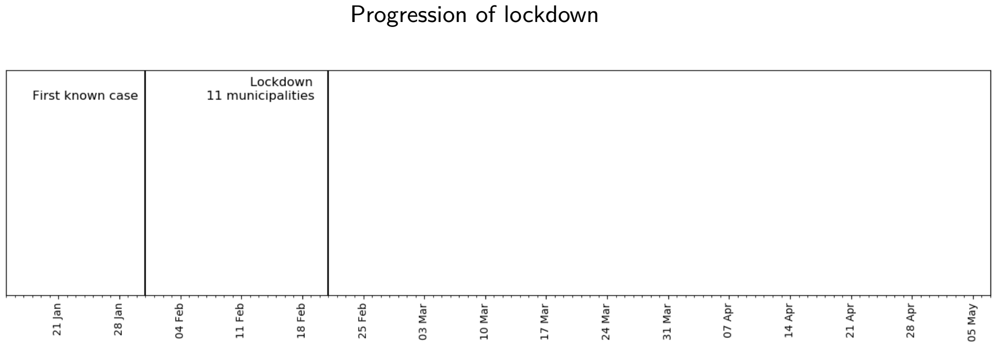
- First major outbreak in Europe, one of first lockdowns
- One of most polluted regions in OECD, given income level



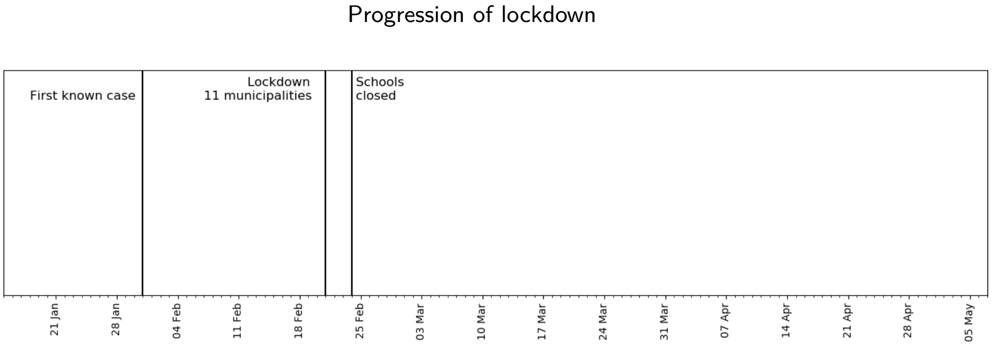
Lockdown of Lombardy



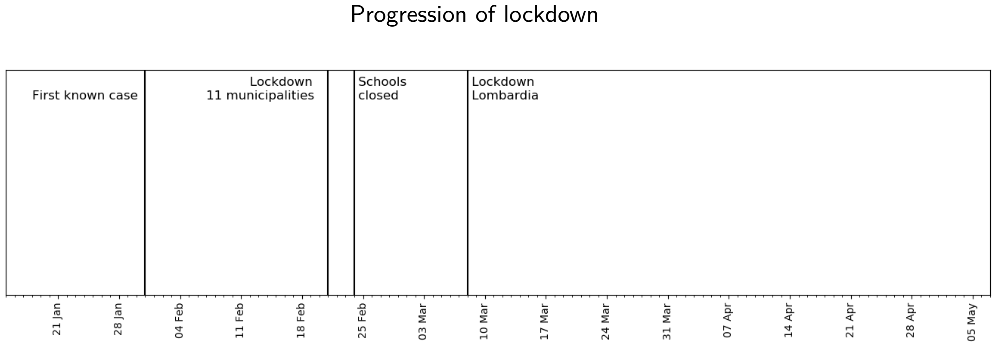
Lockdown of Lombardy



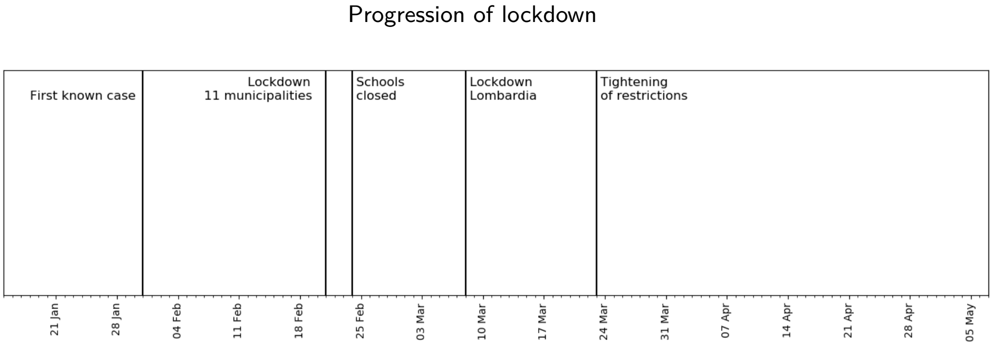
Lockdown of Lombardy



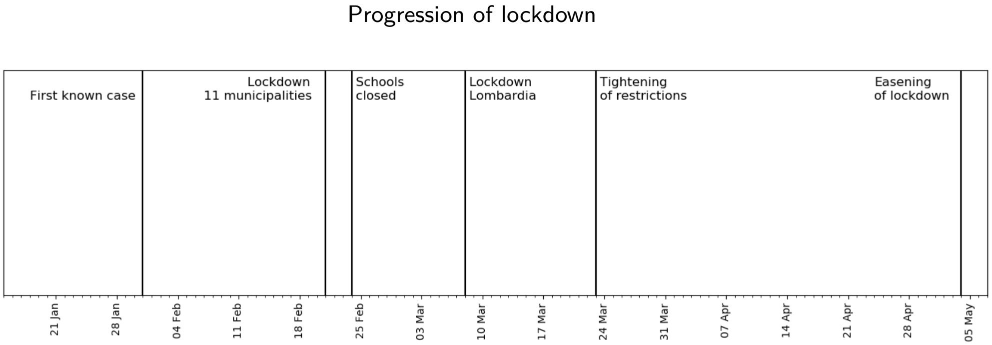
Lockdown of Lombardy



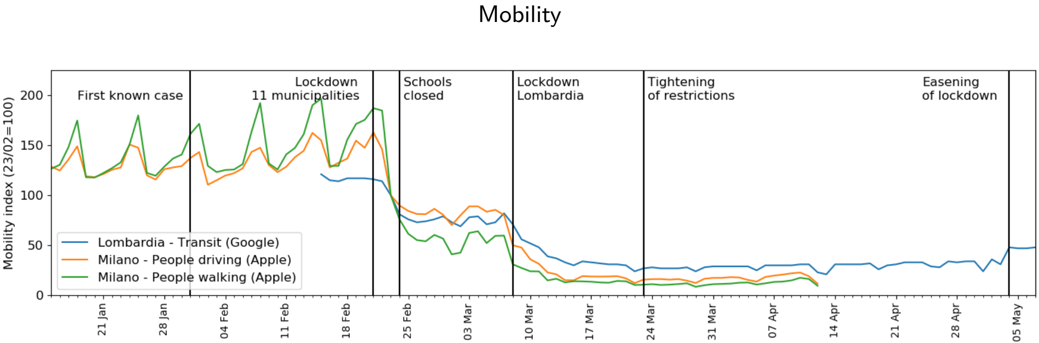
Lockdown of Lombardy



Lockdown of Lombardy

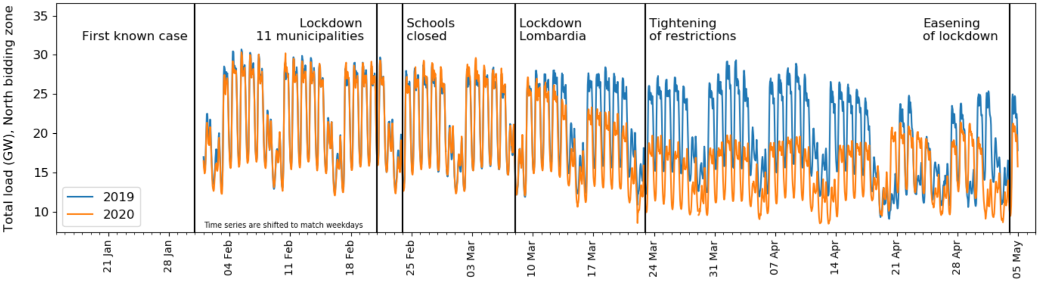


Lockdown of Lombardy



Lockdown of Lombardy

Energy demand



Buildings $\updownarrow$?
Agriculture $\leftrightarrow$

Empirical strategy

For every monitoring station i:

1. Fit* $y_{it} = g(Weather_t) \quad t \in 2012 \ldots 2019$ and learn $\widehat{g}$
2. Predict $\widehat{y}_{it} = \widehat{g}(Weather_t) \quad t \in 2020$
3. Evaluate prediction over January 1 to Febrary 22 (pre-lockdown)
4. Effect of lockdown: $DID = \underbrace{(\overline{y}_{i,post} - \overline{\widehat{y}}_{i,post})}_{\Delta_{Observed\,,\,Counterfactual\,,\,post}} - \underbrace{(\overline{y}_{i,pre} - \overline{\widehat{y}}_{i,pre})}_{\Delta_{Observed\,,\,Counterfactual\,,\,pre}}$

- Requires stable emissions over time (pre-lockdown)
- Does not require comparable and untreated region

*Gradient Boosting Machine cross-validated on 4 folds of data from January to April of 2016, 2017, 2018, 2019, respectively.

Data

$PM_{2.5}$, NO_2 Daily avg at 79 monitors in 64 municipalities

Weather Daily T max, T min, wind speed and direction, relative humidity, precipitation, atmospheric sounding. Contemporaneous & lagged

+ Year, month, week, day-of-week, season
+ Ratio $\frac{PM_{2.5}}{PM_{10}}$ to account for transboundary dust (desert dust)

Map Pollution monitors by type

Accuracy of prediction

Population-weighted average centered R^{2*}

- 72% for $PM_{2.5}$
- 62% for NO_2

How does that look like?

Taylor diagrams

Days exceeding EU limit values for annual average exposure

- Precision of 92% and Recall of 96% for $PM_{2.5}$
- Precision of 81% and Recall of 92% for NO_2

Confusion matrices

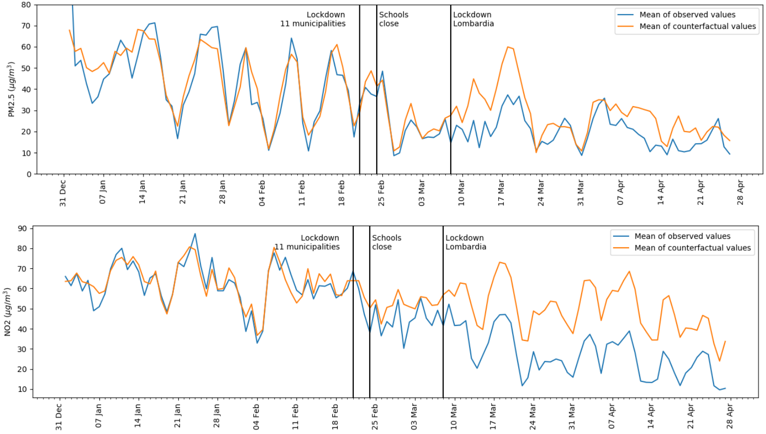
* Centered R^2 is $1 - \frac{\sum_i^n (y_i - \bar{y} - \widehat{y_i} + \widehat{\bar{y}})^2}{\sum_i^n (y_i - \bar{y})^2}$ from January 1 to February 22.

Results

Results

Non-parametric

Population-weighted average of observed and counterfactual values



Results

Parametric

	$\Delta_{Observed,Counterfactual}$	
	(1) PM 2.5	(2) NO2
Lockdown	-6.06***	-16.62***
	(1.45)	(1.30)
Constant	-0.76	-1.14**
	(1.63)	(0.47)
Average baseline concentration	27.89	45.60
Observations	3491	8970

OLS weighted by population of the municipality where the monitor is located. Population is adjusted by the number of monitors in that municipality. The dependent variable is the difference between the observed values and the counterfactual.

Lockdown is a dummy variable equal to 0 from January 1st, 2020 to February 22nd, and equal to 1 after February 22nd, 2020.

Standard errors, in brackets, are clustered by monitor. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Δ vs Population   Unweighted OLS   Map

Results

By type of monitor

Background: Pollution representative of exposure of general population

Industrial: Proximity to industral source

Traffic: Proximity to single major road

	$\Delta_{Observed,Counterfactual}$	
	(1) PM 2.5	(2) NO2
Lockdown=1 (Background)	-4.18***	-12.39***
	(0.39)	(1.37)
Industrial	5.27***	6.27**
	(1.07)	(2.65)
Traffic	4.88***	0.50
	(1.49)	(0.86)
Lockdown=1 × Industrial	-2.86***	-0.39
	(0.95)	(2.16)
Lockdown=1 × Traffic	-4.24**	-7.28***
	(1.83)	(2.08)
Constant	-2.92***	-1.54***
	(0.92)	(0.44)
Average baseline concentration - Background	27.63	36.21
Average baseline concentration - Industrial	27	29
Average baseline concentration - Traffic	28	53
Observations	3491	8970

Unweighted

Results

Population-weighted concentrations at background stations:

$PM_{2.5}$ $24.7 \rightarrow 20.5\ \mu g/m^3$

NO_2 $34.7 \rightarrow 22.3\ \mu g/m^3$

Days exceeding EU annual average limits

Years of life saved in Lombardy

Years of life saved		Years of life lost
$PM_{2.5}$	NO_2	COVID-19
8486	22313	191966
[6887, 10084]	[17402, 27224]	

95% confidence interval in brackets.
YLL for COVID-19 in Lombardy until May 26.
Exposure-response relationship from European Environment Agency (2019) (WHO, Henschel et al. (2013))

Formula

Take-aways

- Properly factoring in **weather** is crucial
- Substantial and yet partial improvement of air quality during lockdown
- Important sectoral emissions (buildings, agriculture) little affected by lockdown
  $\Rightarrow$ The scope of **air pollution policy must include all relevant sectors** (varying with season)
- Improvement in air quality saved $>$**10% of the years of life lost to COVID-19**.

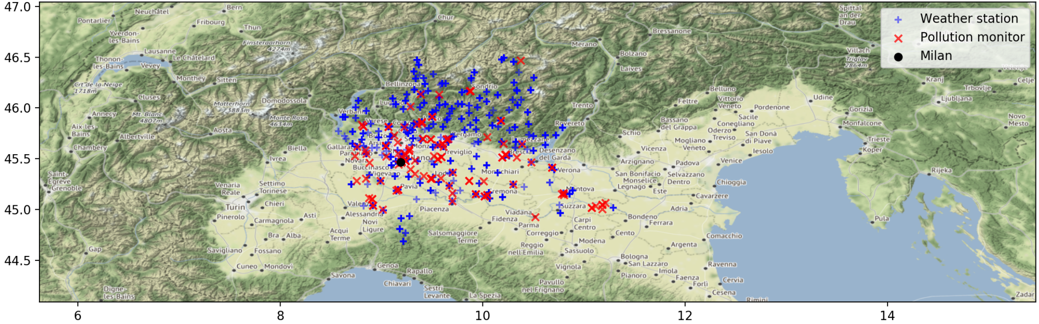
Thank you

francesco.granella@eiee.org

Appendix

Pollution monitors and weather stations

Location of pollution monitors (red) and weather stations (blue) in Lombardy



Back

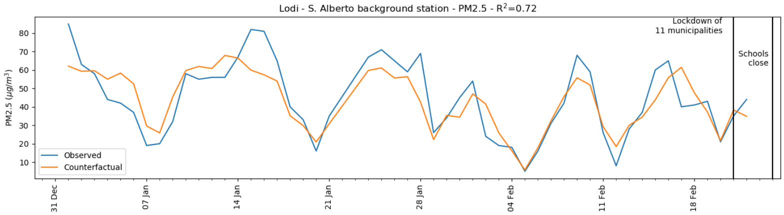
Pollution monitors

Pollutant	Type of monitor	Number of municipalities	Number of monitors
NO2	Background	45	46
NO2	Industrial	8	8
NO2	Traffic	19	23
PM2.5	Background	19	19
PM2.5	Industrial	2	2
PM2.5	Traffic	10	10

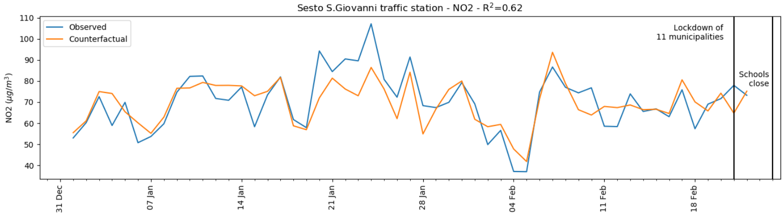
At background monitoring stations pollution levels are representative of the average exposure of the general population or vegetation. Industrial stations are located in close proximity to an industrial area or an industrial source. Traffic stations are located in close proximity to a single major road.

Back

Examples of average R^2



(a) $R^2 = 0.72$



(b) $R^2 = 0.62$

Back

Confusion matrices

PM2.5	Predicted below threshold	Predicted above threshold
Observed below threshold	266	110
Observed above threshold	47	1186

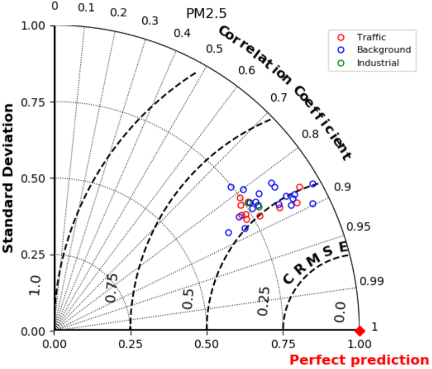
92.0% of days predicted above EU limit value has exceeded the threshold (precision).
96.0% of days above EU limit value is correctly predicted (recall). Values are not weighted.

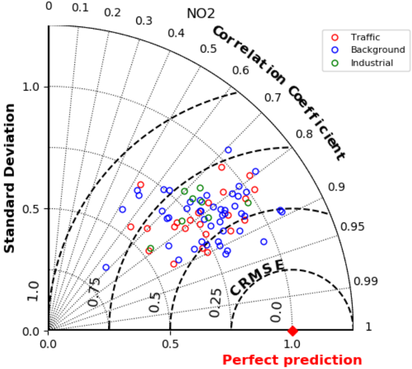
NO2	Predicted below threshold	Predicted above threshold
Observed below threshold	704	160
Observed above threshold	60	685

81.0% of days predicted above EU limit value has exceeded the threshold (precision).
92.0% of days above EU limit value is correctly predicted (recall). Values are not weighted.

Back

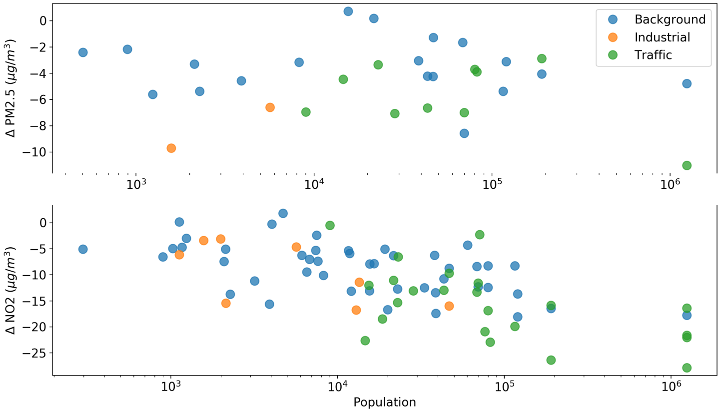
Taylor diagrams





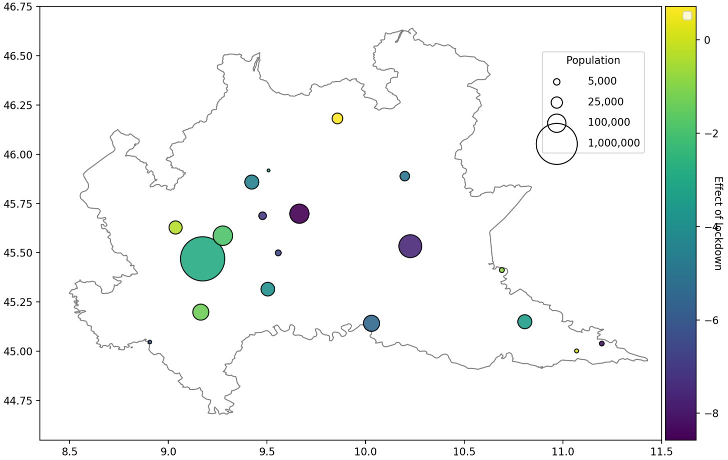
Back

Effect by size of municipality



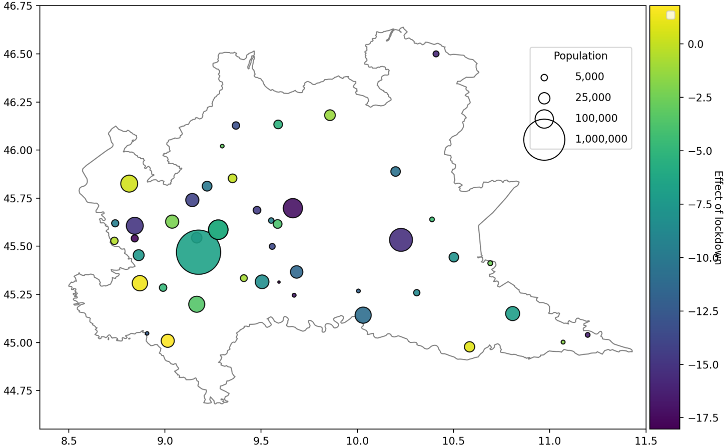
Back

Reduction in PM2.5



Back

Reduction in NO_2



Back

Unweighted OLS

	$\Delta_{Observed,Counterfactual}$	
	(1) PM 2.5	(2) NO2
Lockdown	-4.42***	-11.01***
	(0.48)	(0.72)
Constant	-0.23	-1.13***
	(0.41)	(0.41)
Average baseline concentration	25.79	33.16
Observations	3491	8970

Unweighted OLS. The dependent variable is the difference between the observed values and the counterfactual. Lockdown is a dummy variable equal to 0 from January 1st, 2020 to February 22nd, and equal to 1 after February 22nd, 2020.
Standard errors, in brackets, are clustered by monitor. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Back

Unweighted OLS, by type of monitor

	$\Delta_{Observed, Counterfactual}$	
	(1) PM 2.5	(2) NO2
Lockdown=1	-3.37***	-8.76***
	(0.50)	(0.72)
Industrial	3.83***	3.18
	(0.92)	(2.01)
Traffic	1.22	0.44
	(0.78)	(0.98)
Lockdown=1 × Industrial	-4.67***	-0.81
	(1.37)	(2.17)
Lockdown=1 × Traffic	-2.32**	-7.40***
	(0.91)	(1.44)
Constant	-0.87*	-1.58***
	(0.49)	(0.33)
Average baseline concentration - Background	25.21	29.16
Average baseline concentration - Industrial	28	27
Average baseline concentration - Traffic	26	44
Observations	3491	8970

Unweighted OLS. The dependent variable is the difference between the observed values and the counterfactual. Background monitors is the omitted category.

Lockdown is a dummy variable equal to 0 from January 1st, 2020 to February 22nd, and equal to 1 after February 22nd, 2020.

Standard errors, in brackets, are clustered by monitor. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Back

Days exceeding EU yearly average limits

Pollutant	Type of monitor	Observed number of exceedances	Predicted number of exceedances	Change
NO2	Background	6.8	29.8	-77.2 %
NO2	Industrial	2.6	7.5	-65.2 %
NO2	Traffic	19.3	49	-60.6 %
PM2.5	Background	15.6	35	-55.4 %
PM2.5	Industrial	15.6	31.5	-50.5 %
PM2.5	Traffic	17.1	31.7	-46.1 %

Average number of readings above the EU limit values of 25 and 40 $\mu g/m^3$ for PM 2.5 and NO2, respectively for the 66 days from February 22 to April 27. Days exceeding limit values are counted for each monitor and an average number of exceedances is computed by municipality, pollutant and type of monitor (background, industrial, traffic. This number is then average over all municipalities weighting for population.

Back

Years of life saved

$$YLS_{p,g,a} = \underbrace{\gamma_p \cdot \widehat{\beta}_p \cdot Baseline\ Risk_{g,a}}_{\text{Change in 1-year individual risk}} \cdot \frac{1}{6} \cdot Life\ Expectancy_{g,a} \cdot N_{g,a}$$

$$YLS_p = \sum_g \sum_a YLS_{p,g,a}$$

Where p indexes pollutants, g gender, and a age group.
γ_p summarizes the exposure-mortality function. From European Environment Agency (2019) (WHO, Henschel et al. (2013))
$\gamma_{PM2.5} = 0.0062$ for all ages above 30, for all concentration.
$\gamma_{NO2} = 0.0055$ for all ages above 30, for all concentrations above 20 $\mu g/m^3$.
β_p is the estimated change in concentrations of pollutant p at background stations.
$N_{g,a}$ is number of individuals of gender g in age group a in Lombardy.

Back

References I

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska, L., Pope, C. A., Shin, H., Straif, K., Shaddick, G., Thomas, M., van Dingenen, R., van Donkelaar, A., Vos, T., Murray, C. J. L., and Forouzanfar, M. H. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the global burden of diseases study 2015. *The Lancet*, 389(10082):1907–1918.

Cole, M. A., Elliott, R. J. R., and Liu, B. (2020). The Impact of the Wuhan Covid-19 Lockdown on Air Pollution and Health: A Machine Learning and Augmented Synthetic Control Approach. Technical report, Department of Economics, University of Birmingham.

European Environment Agency (2019). *Air quality in Europe - 2019 report*.

Graff Zivin, J. and Neidell, M. (2012). The Impact of Pollution on Worker Productivity. *American Economic Review*, 102(7):3652–3673.

He, G., Pan, Y., and Tanaka, T. (2020). Covid-19, city lockdowns, and air pollution: Evidence from china. *medRxiv*.

References II

Henschel, S., Chan, G., Organization, W. H., et al. (2013). Health risks of air pollution in europe-hrapie project: New emerging risks to health from air pollution-results from the survey of experts.

Künn, S., Palacios, J., and Pestel, N. (2019). Indoor Air Quality and Cognitive Performance. IZA Discussion Papers 12632, Institute of Labor Economics (IZA).