Estimating the Causal Effect of Social Mobility on COVID-19 Related Mortality

Kei Irizawa, Adam Oppenheimer, and Jason Yang
University of Chicago

May 26, 2020

Table of Contents

1 Research Question

2 Literature Review

3 Model

4 Data

5 IV Regression

6 Lasso

7 Result

8 Measurement Error Bias

9 Conclusion

10 Questions

Research Question

Research Question

How effective are policies aimed at restricting social mobility on COVID-19 related mortality?

Identification Issue

- Social mobility measures are contaminated by measurement error $\rightarrow$ Attenuation Bias.
- Many potential controls that are correlated with social mobility (endogeneity issue lead to omitted variable bias and reverse causality) $\rightarrow$ Diff-in-Diff.
- Difficulty to find an instrument that satisfied both relevance and exclusion restriction.

Goals

- Estimate the causal effect by using instrument rainfall.
- Quantify the bias from measurement error and endogeneity.

Table of Contents

1 Research Question

2 Literature Review

3 Model

4 Data

5 IV Regression

6 Lasso

7 Result

8 Measurement Error Bias

9 Conclusion

10 Questions

Literature Review

- Adda (2016): social distancing on viral spread in France. Endogeneity, serial correlation $\rightarrow$ weather IV
- Qiu et al. (2020): climate IV to measure social distancing on coronavirus spread in China
- Dave et al. (2020) and Villas-Boas et al. (2020): DiD to measure policy on social distancing

Difference-in-Difference Problems

Figure: Time Series of Mobility Level in Texas

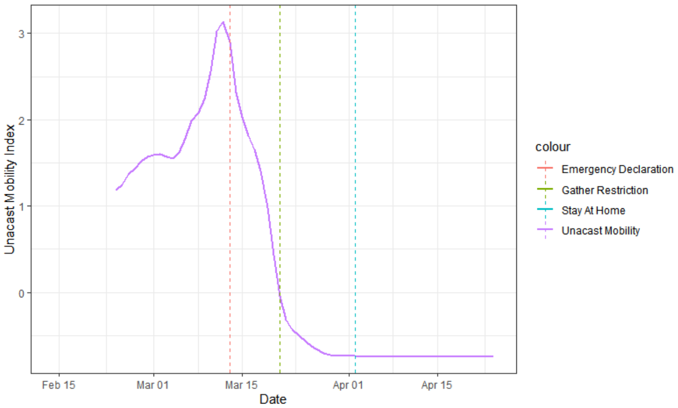


Table of Contents

1 Research Question

2 Literature Review

3 Model

4 Data

5 IV Regression

6 Lasso

7 Result

8 Measurement Error Bias

9 Conclusion

10 Questions

Empirical Model

$$\Delta D_{i,t} = \beta_0 + \sum_{j=1}^{J} \beta_j \Delta SD_{i,t-1+j} + \eta_1 \mathbf{P}_{i,t} + \lambda_i \chi_i + \gamma_i \chi_i \cdot t + \epsilon_{i,t}$$

- i and t are subscripts for state and week.
- $\Delta D_{i,t}$ denotes the change in COVID-19 related deaths per capita.
- $\mathbf{P}_{i,t}$ are indicator variables for whether a particular policy was implemented (Emergency Declaration and Gather Restrictions)
- χ_i is a state indicator variable and $\chi_i \cdot t$ capture linear trends for each state.

Table of Contents

1 Research Question

2 Literature Review

3 Model

4 Data

5 IV Regression

6 Lasso

7 Result

8 Measurement Error Bias

9 Conclusion

10 Questions

Data Sources

- Social Mobility
  1) Descartes Lab (Warren and Skillman (2020)): Daily index of absolute "level" of mobility
  2) Unacast (Unacast (2020)): Daily percent change in mobility
  3) Twitter (Xu et al. (2020)): Weekly index of mobility
- Policy (Fullman et al. (2020)): Date enacted for Emergency Declaration and Gathering Restrictions.
- Rainfall (Community Collaborative Rain, Hail, & Snow Network (n.d.)): Average daily rainfall for each state in inches.
- Mortality (Systems Science and Engineering at Johns Hopkins University (2020)): Daily COVID-19 related cases and mortality for each state.

Table of Contents

1 Research Question

2 Literature Review

3 Model

4 Data

5 IV Regression

6 Lasso

7 Result

8 Measurement Error Bias

9 Conclusion

10 Questions

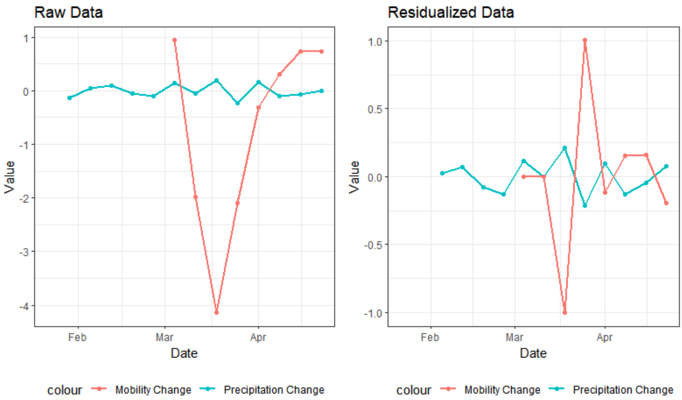
Instrument Relevance

$$\Delta SD_{i,t} = \beta \Delta R_{i,t} + P'_{i,t}\eta + \lambda_i \chi_i + \gamma_i \chi_i t$$

- $P_{i,t}$ refer to whether a policy was implemented at time t
- χ_i is a State indicator
- t is the week number (which we treat as a continuous variable to capture a trend effect)

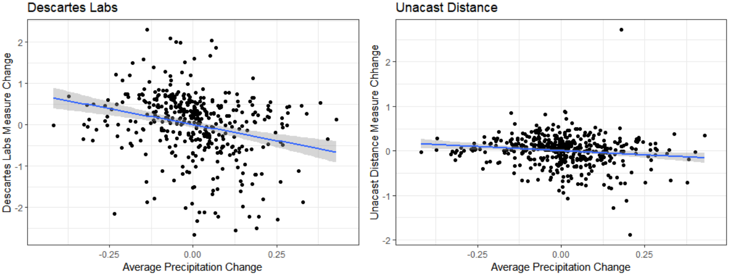
Instrument Relevance

Figure: Time Series of Mobility Change and Rainfall Change in Texas



Instrument Relevance

Figure: First Stage: Residualized Rainfall Change versus Residualized Social Distancing Change



Instrument Relevance

Table: First Stage: Change in Mobility

	Descartes Lab				Unacast				Twitter			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ΔAvg. Precipitation$_t$	−1.804***	−1.805***	−2.066***	−1.626***	−0.465***	−0.470***	−0.511***	−0.445***	−0.318	−0.258	0.100	0.250
	(0.337)	(0.363)	(0.301)	(0.416)	(0.097)	(0.104)	(0.098)	(0.123)	(0.465)	(0.506)	(0.634)	(1.807)
Emergency Declaration$_t$			−1.490***	−3.438***			−0.575***	−0.817***			−0.652***	−0.398
			(0.218)	(0.388)			(0.062)	(0.102)			(0.117)	(0.664)
Gathering Restriction (Any)$_t$			−0.316	−2.541***			0.059	−0.574***			−0.215*	−0.729
			(0.247)	(0.416)			(0.091)	(0.211)			(0.125)	(0.621)
ΔCOVID Mortality Per Capita$_{t-1}$			16461.470**	−3.84e+04**			4,278.941***	−6865.973**			3.04e+05***	2.52e+06
			(6,490.225)	(15019.083)			(1,483.819)	(3,165.341)			(85588.700)	(4.13e+06)
Constant	−0.597***	−0.536***	0.944***	−15.591***	−0.226***	−0.241***	0.151***	0.524	−0.594***	−0.409***	−0.004	8.230
	(0.025)	(0.003)	(0.094)	(5.758)	(0.001)	(0.002)	(0.031)	(1.683)	(0.033)	(0.025)	(0.055)	(15.223)
N	400	400	400	400	450	450	450	450	260	260	200	200
R^2	0.028	0.042	0.201	0.717	0.015	0.015	0.189	0.494	0.002	0.074	0.343	0.705
Adj. R^2	0.025	−0.095	0.079	0.541	0.013	−0.108	0.080	0.232	−0.002	−0.159	0.104	−0.276
State FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
State · Week	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
F-stat	28.7	24.7	47.0	15.3	23.1	20.6	27.5	13.1	.467	.260	.025	0.019

Robust standard errors in parentheses. Standard errors clustered by state.

Exclusion Restriction

1. Tosepu et al. (2020) document empirical relationship between coronavirus transmission and meterological variables (temperature, humidity, and rainfall) in Indonesia. Find a relationship for temperature & humidity but not for rainfall.
2. Pica and Bouvier (2012) review literature covering relationship between respiratory viruses and meterological factors. They find that the relationship between rainfall and virus transmission is very non-robust, it varies significantly with the country (even province of study).
3. Lowen and Steel (2014) provide experimental evidence for and review scientific models concerning the relationship between influenza and temperature & humidity.
4. Although other meterological factors seem important for viral transmission, this does not appear to be true for rainfall.
5. We had trouble collecting other meterological data, a future version of the paper ideally should include temperature & humidity controls.

Table of Contents

1 Research Question

2 Literature Review

3 Model

4 Data

5 IV Regression

6 Lasso

7 Result

8 Measurement Error Bias

9 Conclusion

10 Questions

Variable Selection

Lasso Regression:

$$\min_{\beta} \sum_{i=1}^{N} \left(D_{i,t} - \beta_0 - \sum_{j=1}^{J} \beta_j SD_{i,t-1+j} - C'_{i,t}\eta \right)^2 \quad \text{subject to} \sum_{j=1}^{P} |\beta_j| \leq t$$

Interested in finding the correct lags, while keeping the other controls in the regression model (C_{it}). Apply Frisch-Waugh Theorem:

$$\min_{\beta} \sum_{i=1}^{N} \left(M_C D_{i,t} - \sum_{j=1}^{J} \beta_j M_C SD_{i,t-1+j} \right)^2 + \lambda \sum_{j=1}^{P} |\beta_j|$$

We run the following regression model:

$$M_C D_{i,t} = M_C SD\beta + \lambda|\beta|$$

Lasso Implementation

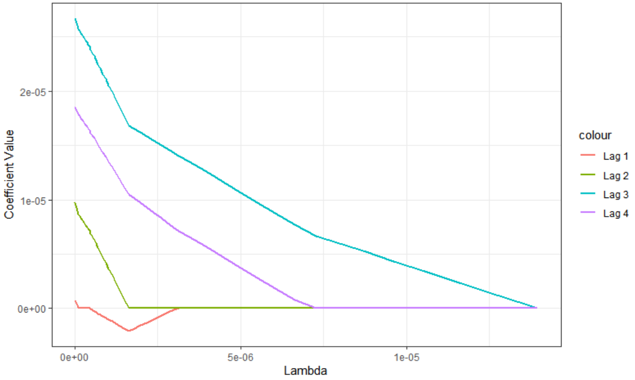
1) Split the data into in-sample and out-of-sample data.
2) Using the in-sample data, generate $M_C D_{i,t}$ and $M_C SD_{i,t}$. These are residuals from an OLS regression of $D_{i,t}$ and $SD_{i,t}$ on $C_{i,t}$ respectively.
3) Using the in-sample data, regress $M_C SD_{i,t}$ on $R_{i,t}$ and store the predicted $\widehat{M_C SD}_{i,t}$. This is equivalent to the predicted values from the first stage of an IV regression.
4) Consider a list of possible λ values that can be used for the Lasso regression.
5) For each λ, using the in-sample data, run the following Lasso Regression:

$$M_C D_{i,t} = \widehat{M_C SD}_{i,t}\beta + \lambda|\beta|$$

6) Using the out-of-sample data, calculate the MSE based on our estimates from (5). Store the MSE value.
7) Repeat (5) - (6) for the range of λ considered in (4). Choose the λ that gives the lowest MSE.
8) Given this optimal λ parameter, find the lagged variables where $\beta_i > 0$.

Lasso Results

Figure: Coefficient Estimates for a range of λ



Lasso Results

Figure: MSE for Range of λ

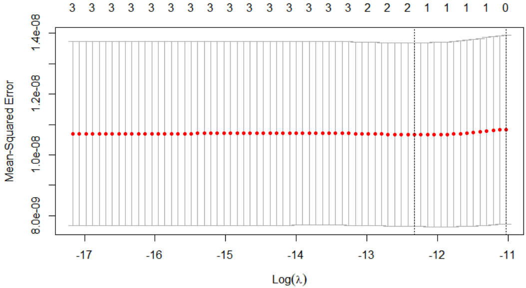


Table of Contents

1 Research Question

2 Literature Review

3 Model

4 Data

5 IV Regression

6 Lasso

7 **Result**

8 Measurement Error Bias

9 Conclusion

10 Questions

Regression Result

Table: Second Stage: Change in COVID Mortality Per Capita (lag 3)

	Descartes Lab				Unacast				Twitter			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ΔMobility$_{t-3}$	6.45e-06***	3.83e-06**	2.76e-06*	1.06e-06	1.80e-05**	1.90e-05**	1.19e-05*	3.05e-06	−1.29e-04	−1.63e-04	1.93e-04	2.47e-05
	(2.36e-06)	(1.60e-06)	(1.58e-06)	(1.34e-06)	(8.33e-06)	(8.87e-06)	(6.63e-06)	(2.40e-06)	(1.94e-04)	(2.86e-04)	(1.05e-03)	(7.49e-05)
Emergency Declaration$_{t-3}$			−1.92e-06	−1.26e-06			9.78e-06***	−6.65e-06**			1.29e-04	6.13e-06
			(4.85e-06)	(4.01e-06)			(3.58e-06)	(3.03e-06)			(6.79e-04)	(3.89e-05)
Gathering Restriction (Any)$_{t-3}$			2.37e-05***	−2.76e-06			2.14e-05***	−1.44e-06			5.68e-05	5.68e-06
			(7.15e-06)	(2.85e-06)			(6.88e-06)	(2.19e-06)			(2.31e-04)	(1.85e-05)
ΔCOVID Mortality Per Capita$_{t-4}$			4.19e-01	−6.81e+00***			8.55e-01**	−5.77e+00***			−5.45e+01	−1.59e+01
			(4.34e-01)	(1.60e+00)			(3.92e-01)	(1.25e+00)			(3.21e+02)	(6.20e+01)
Constant	2.78e-05***	1.28e-05***	−1.30e-06	−7.49e-05***	2.31e-05***	1.34e-05***	−6.49e-06**	−7.73e-05***	−6.86e-05	−6.09e-05	4.41e-07	−8.13e-05*
	(6.33e-06)	(1.90e-06)	(2.26e-06)	(1.58e-05)	(5.84e-06)	(3.05e-06)	(3.07e-06)	(1.63e-05)	(1.15e-04)	(1.13e-04)	(1.22e-05)	(4.55e-05)
N	250	250	250	250	300	300	300	300	260	260	200	200
R^2	.	0.626	0.705	0.970	.	0.412	0.620	0.958	.	.	.	0.662
Adj. R^2	.	0.533	0.625	0.949	.	0.294	0.538	0.935	.	.	.	0.299
State FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
State · Week	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes

Robust standard errors in parentheses. Standard errors clustered by state.

Table of Contents

1 Research Question

2 Literature Review

3 Model

4 Data

5 IV Regression

6 Lasso

7 Result

8 Measurement Error Bias

9 Conclusion

10 Questions

Setting

Apply idea presented by Acemoglu et al. (2001).

- $\tilde{SD}_1$ (Descrates Lab): a random variable that measures SD with error, so that

$$\tilde{SD}_1 = SD + u_1$$

- $\tilde{SD}_2$ (Unacast): another random variable that measures x with error, so that

$$\tilde{SD}_2 = SD + u_2$$

- z: a good (relevant and exogenous) instrument for x

Bias Calculation

- The omitted variable bias if you observed $M_C SD$ and used it in the initial regression can be represented as the following:

$$\frac{Cov(\epsilon, M_C SD)}{Var(M_C SD)} = \beta^{IV}_{\tilde{x}_2} - \beta^{IV}_{z} = -1.19 \times 10^{-6} - 1.19 \times 10^{-5} \approx -1.31 \times 10^{-5}$$

- The bias due to the omitted variables in β^{OLS} can be represented as the following:

$$\begin{aligned}\frac{Cov(\epsilon, M_C SD)}{Var(M_C \tilde{SD}_1)} &= \frac{\beta^{OLS}}{\beta^{IV}_{\tilde{x}_2}} \left(\beta^{IV}_{\tilde{x}_2} - \beta^{IV}_{z}\right) \\ &= \frac{3.75 \times 10^{-6}}{-1.19 \times 10^{-6}}(-1.19 \times 10^{-6} - 1.19 \times 10^{-5}) \approx 4.13 \times 10^{-5}\end{aligned}$$

- The bias due to measurement error in β^{OLS} can be represented as the following:

$$\frac{Var(M_C SD)}{Var(M_C \tilde{SD}_1)} = \frac{\beta^{OLS}}{\beta^{IV}_{\tilde{x}_2}} = \frac{3.75 \times 10^{-6}}{-1.19 \times 10^{-6}} \approx -3.15$$

Table of Contents

1 Research Question

2 Literature Review

3 Model

4 Data

5 IV Regression

6 Lasso

7 Result

8 Measurement Error Bias

9 **Conclusion**

10 Questions

Conclusion

- We analyze direct causal relationship between social mobility and COVID-19 related mortality.
- We used lasso implementation to select lags of social mobility in our empirical model.
- We estimate a substantial positive effect of social mobility on COVID-19 related mortality.
- We measure the bias from measurement error and endogeneity. The bias from measurement error is much larger than bias from endogeneity.

Table of Contents

1 Research Question

2 Literature Review

3 Model

4 Data

5 IV Regression

6 Lasso

7 Result

8 Measurement Error Bias

9 Conclusion

10 Questions

Questions

- Measurement Bias: Can we have negative bias?
- Panel data identifying assumption and relation to IV regression?

References I

Acemoglu, Daron, Simon Johnson, and James A. Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation.," *American Economic Review*, 2001, *91* (5), 1369–1401.

Adda, Jérôme, "Economic Activity and the Spread of Viral Diseases: Evidence from High Frequency Data," *The Quarterly Journal of Economics*, 2016, *131* (2), 891–941.

Community Collaborative Rain, Hail, & Snow Network, "Daily Precipitation Reports." URL: https://www.cocorahs.org/ViewData/ListDailyPrecipReports.aspx.

Dave, Dhaval M, Andrew I Friedson, Kyutaro Matsuzawa, and Joseph J Sabia, "When Do Shelter-in-Place Orders Fight COVID-19 Best? Policy Heterogeneity Across States and Adoption Time," Technical Report, National Bureau of Economic Research 2020.

Fullman, Nancy, Bree Bang-Jensen, Grace Reinke, Kenya Amano, Christopher Adolph, and John Wilkerson, "State-Level Social Distancing Policies in Response to COVID-19 in the US," 2020.

Lowen, Anice and John Steel, "Roles of Humidity and Temperature in Shaping Influenza Seasonality," *Journal of Virology*, 2014.

References II

Pica, Natalie and Nicole Bouvier, "Environmental factors affecting the transmission of respiratory viruses," *Current Opinion in Virology*, 2012, *2*, 90–95.

Qiu, Yun, Xi Chen, and Wei Shi, "Impacts of Social and Economic Factors on the Transmission of Coronavirus Disease 2019 (COVID-19) in China," *Journal of Population Economics*, 2020, p. 1.

Systems Science and Engineering at Johns Hopkins University, "COVID-19 Data Repository," 2020. URL: https://github.com/CSSEGISandData/COVID-19.

Tosepu, Ramadhan, Joko Gunawan, Devi Effendy, La Ode Ali Imran Ahmad, Hariati Lestari, Hariati Bahar, and Pitrah Asfian, "Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia," *Science of The Total Environment*, 2020.

Unacast, "Unacast Social Distancing Dataset," 2020. URL: https://www.unacast.com/data-for-good.

Villas-Boas, Sofia B, James Sears, Miguel Villas-Boas, and Vasco Villas-Boas, "Are We #StayingHome to Flatten the Curve?," *UC Berkeley CUDARE Working Paper*, 2020.

References III

Warren, Michael and Samuel Skillman, "Mobility Changes in Response to COVID-19," 2020.

Xu, Paiheng, Mark Dredze, and David A Broniatowski, "The Twitter Social Mobility Index: Measuring Social Distancing Practices from Geolocated Tweets," *arXiv Preprint*, 2020.