# Supplementary Appendix: Policy Evaluation during a Pandemic

Brantly Callaway* Tong Li†

October 21, 2022

This supplementary appendix contains (i) additional theoretical results related to estimation and inference of the parameters of interest in the main text, (ii) an explanation for how to extend the arguments in the main text to the case where the timing of the policy can vary across locations, (iii) additional discussion regarding the compatibility of the unconfoundedness approach with several extensions to the Stochastic SIRD Model for Untreated Potential Outcomes, and (iv) additional details about the application on shelter-in-place orders.

## SA Additional Theoretical Results

This section contains several additional results related to verifying double robustness and inference. The results in this section are, for the most part, not new, but rather slightly adapt existing arguments to the particular case considered in the paper. We provide these results primarily to complete the arguments presented in the paper.

We consider the case where a researcher implements parametric working models for the propensity score and outcome regression. We denote the propensity score working model by $p(\mathcal{F}_{t^*-1};\pi)$ where $\pi$ is a finite dimensional parameter, and we denote the pseudo true value of the parameter by $\pi^*$ and the estimated value of the parameter by $\hat{\pi}$. Likewise, we denote the outcome regression working model by $m^C_{0,t}(\mathcal{F}_{t^*-1};\mu_t)$ where $\mu_t$ is a finite dimensional parameter, and we denote the pseudo true value of the parameter by $\mu_t^*$ and the estimated value of the parameter by $\hat{\mu}_t$.

We make the following assumptions

**Assumption SA.1** (Random Sample)**.** *The data consists of* $\{Y_{l1}, Y_{l2}, \ldots, Y_{l\mathcal{T}}, \mathcal{F}_{l1}, \mathcal{F}_{l2}, \ldots, \mathcal{F}_{l\mathcal{T}}, D_l\}_{l=1}^n$ *which are iid across locations.*

**Assumption SA.2** (Assumptions for propensity score)**.** *(i)* $p(\mathcal{F}_{t^*-1};\pi) = \Lambda(h_{ps}(\mathcal{F}_{t^*-1})'\pi)$ *is a parametric working model for* $p(\mathcal{F}_{t^*-1})$ *where* $\Lambda(z) = 1/(1+\exp(z))$ *and* $h_{ps}$ *allows for transformations of*

*Department of Economics. University of Georgia. brantly.callaway@uga.edu
†Department of Economics. Vanderbilt University. tong.li@vanderbilt.edu

$\mathcal{F}_{t^*-1}$, *(ii) the pseudo true value* $\pi^*$ *is in the interior of the parameter space* $\Pi$ *which is a compact subset of* $\mathbb{R}^k$ *(*$k$ *being the dimension of* $h_{ps}(\mathcal{F}_{t^*-1})$*), (iii)* $Q_{ps} := \mathbb{E}[h_{ps}(\mathcal{F}_{t^*-1})h_{ps}(\mathcal{F}_{t^*-1})'\Lambda(h_{ps}(\mathcal{F}_{t^*-1})'\pi^*)(1-\Lambda(h_{ps}(\mathcal{F}_{t^*-1})'\pi^*)]$ *is positive definite, and (iv)* $\mathbb{E}\left[||h_{ps}(\mathcal{F}_{t^*-1})||^4\right] < \infty$.

**Assumption SA.3** (Assumptions for outcome regression)**.** *For all* $t = t^*, \ldots, \mathcal{T}$*, (i)* $m_{0,t}(\mathcal{F}_{t^*-1}; \mu_t) = h_{or}(\mathcal{F}_{t^*-1})'\mu_t$ *is a parametric working model for* $m^C_{0,t}(\mathcal{F}_{t^*-1})$ *and* $h_{or}$ *allows for transformations of* $\mathcal{F}_{t^*-1}$*, (ii)* $Q_{0,or} := \mathbb{E}[h_{or}(\mathcal{F}_{t^*-1})h_{or}(\mathcal{F}_{t^*-1})'|D=0]$ *is positive definite, (iii)* $\mathbb{E}[C_t^4|D=0] < \infty$*, (iv)* $\mathbb{E}\left[||h_{or}(\mathcal{F}_{t^*-1})||^4|D=0\right] < \infty$

**Assumption SA.4.** *For all* $t = t^*, \ldots, \mathcal{T}$*, at least one (but not necessarily both) of the following conditions hold: (i)* $p(\mathcal{F}_{t^*-1}) = p(\mathcal{F}_{t^*-1}; \pi^*)$*, (ii)* $m^C_{0,t}(\mathcal{F}_{t^*-1}) = m^C_{0,t}(\mathcal{F}_{t^*-1}; \mu_t^*)$

Assumptions SA.1 to SA.4 are all standard assumptions. Assumption SA.2 says that we use a logit model for the propensity score working model and invokes standard conditions for logit models. Assumption SA.3 says that we use a linear model for the outcome regression working model and invokes standard assumptions for linear models. Assumption SA.4 says that at least one of the propensity score working model or outcome regression working model is correctly specified.

It is helpful to define some additional notation regarding the weights in Theorem 2. First,

$$\omega_1(D) := \frac{D}{\mathbb{E}[D]} \qquad \text{and} \qquad \hat{\omega}_1(D) := \frac{D}{\bar{D}}$$

where $\bar{D} = n^{-1}\sum_{l=1}^n D_l$. In addition, define

$$\tilde{\omega}_0(D, \mathcal{F}_{t^*-1}, \pi) := \frac{p(\mathcal{F}_{t^*-1}; \pi)}{(1 - p(\mathcal{F}_{t^*-1}; \pi))}(1-D)$$

$$\omega_0(D, \mathcal{F}_{t^*-1}, \pi) := \frac{\tilde{\omega}_0(D, \mathcal{F}_{t^*-1}, \pi)}{\mathbb{E}[\tilde{\omega}_0(D, \mathcal{F}_{t^*-1}, \pi)]}$$

$$\hat{\omega}_0(D, \mathcal{F}_{t^*-1}; \pi) := \frac{\tilde{\omega}_0(D, \mathcal{F}_{t^*-1}; \pi)}{\frac{1}{n}\sum_{h=1}^n \tilde{\omega}_0(D_h, \mathcal{F}_{ht^*-1}; \pi)}$$

We also denote the estimated weights by $\hat{\omega}(D, \mathcal{F}_{t^*-1}, \hat{\pi}) = \hat{\omega}_1(D) - \hat{\omega}_0(D, \mathcal{F}_{t^*-1}, \hat{\pi})$.

The next result shows that the estimand in Equation (9) in Theorem 2 is indeed doubly robust.

**Proposition SA.1.** *Under the Stochastic SIRD Model for Untreated Potential Outcomes and Assumptions 1 and SA.1 to SA.4,*

$$\widehat{ATT}^C_t = \frac{1}{n}\sum_{l=1}^n \hat{\omega}(D_l, \mathcal{F}_{lt^*-1}, \hat{\pi})(C_{lt} - m^C_{0,t}(\mathcal{F}_{lt^*-1}; \hat{\mu}_t) \xrightarrow{p} ATT^C_t$$

*Proof.* Under Assumptions SA.2 and SA.3, it immediately holds by the weak law of large numbers

and continuous mapping theorem that

$$\widehat{ATT}_t^C \xrightarrow{p} \mathbb{E}\left[\omega(D, \mathcal{F}_{t^*-1}, \pi^*)(C_t - m_{0,t}^C(\mathcal{F}_{t^*-1}; \mu_t^*))\right]$$

It remains to show that this expression is equal to $ATT_t^C$.

*Case 1: Propensity Score is Correctly Specified*

First, consider the case where the propensity score is correctly specified so that $p(\mathcal{F}_{t^*-1}) = p(\mathcal{F}_{t^*-1}; \pi^*)$, but where the outcome regression may be misspecified so that it can be the case that $m_{0,t}^C(\mathcal{F}_{t^*-1}) \neq m_{0,t}^C(\mathcal{F}_{t^*-1}; \mu_t^*)$. In this case, $\omega(D, \mathcal{F}_{t^*-1}) = \omega(D, \mathcal{F}_{t^*-1}; \pi^*)$. From the arguments in Equation (B.10) in the proof of Theorem 2, it holds that

$$ATT_t^C = \mathbb{E}[\omega(D, \mathcal{F}_{t^*-1})C_t] \tag{SA.1}$$

Then, notice that

$$\begin{aligned}\mathbb{E}[\omega(D, \mathcal{F}_{t^*-1})m_{0,t}^C(\mathcal{F}_{t^*-1}; \mu_t^*)] &= \mathbb{E}\left[m_{0,t}^C(\mathcal{F}_{t^*-1}; \mu_t^*)\mathbb{E}[\omega(D, \mathcal{F}_{t^*-1})|\mathcal{F}_{t^*-1}]\right] \\ &= 0\end{aligned} \tag{SA.2}$$

where the last equality holds by Equation (B.11).

Combining, Equations (SA.1) and (SA.2) implies the result for this part when the propensity score is correctly specified.

*Case 2: Outcome Regression is Correctly Specified*

Next, we consider the case where the outcome regression is correctly specified so that $m_{0,t}^C(\mathcal{F}_{t^*-1}) = m_{0,t}^C(\mathcal{F}_{t^*-1}; \mu_t^*)$, but the propensity score may not be correctly specified so that it may be the case that $p(\mathcal{F}_{t^*-1}) \neq p(\mathcal{F}_{t^*-1}; \pi^*)$. In this case,

$$\begin{aligned}&\mathbb{E}\left[\omega(D, \mathcal{F}_{t^*-1}; \pi^*)(C_t - m_{0,t}^C(\mathcal{F}_{t^*-1}; \mu_t^*))\right] \\ &\qquad = \mathbb{E}\left[\omega(D, \mathcal{F}_{t^*-1}; \pi^*)(C_t - m_{0,t}^C(\mathcal{F}_{t^*-1}))|D=1\right]\mathbb{E}[D] \\ &\qquad + \mathbb{E}\left[\omega(D, \mathcal{F}_{t^*-1}; \pi^*)(C_t - m_{0,t}^C(\mathcal{F}_{t^*-1}))|D=0\right](1 - \mathbb{E}[D]) \\ &\qquad := A + B\end{aligned}$$

where the first equality holds by the law of iterated expectations. We consider Term A and Term B in turn next. Next, notice that

$$\begin{aligned}A &= \mathbb{E}[(C_t - m_{0,t}^C(\mathcal{F}_{t^*-1}))|D=1] \\ &= \mathbb{E}[C_t|D=1] - \mathbb{E}\left[\mathbb{E}[C_t(0)|\mathcal{F}_{t^*-1}, D=1]|D=1\right] \\ &= ATT_t^C\end{aligned}$$

where the first equality holds by the definition of the weights (and by the expectation being conditional on $D = 1$), the second equality holds by unconfoundedness (which holds here by Theorem 2), and the last equality holds by the law of iterated expectations and the definition of $ATT_t^C$.

Finally, consider Term B

$$\begin{aligned} B &= \mathbb{E}\left[\omega(D, \mathcal{F}_{t^*-1}; \pi^*)(\mathbb{E}[C_t|\mathcal{F}_{t^*-1}, D = 0] - m_{0,t}^C(\mathcal{F}_{t^*-1}))|D = 0\right](1 - \mathbb{E}[D]) \\ &= 0 \end{aligned}$$

where the first equality holds by the law of iterated expectations and the last equality holds by the definition of $m_{0,t}^C(\mathcal{F}_{t^*-1})$.

Combining the results for Term A and Term B implies the result for this part when the outcome regression is correctly specified. □

Next, we provide additional results related to conducting inference. For this part, we primarily follow Sant'Anna and Zhao (2020) who consider similar doubly robust estimands to the ones considered in the current paper. Interestingly, that paper considers estimation and inference in a (conditional) difference in differences setup; however, most of their arguments apply to the current paper with only minor modifications.

Notice that, under Assumptions SA.1 to SA.3,

$$\sqrt{n}\left(\hat{\pi} - \pi^*\right) = Q_{ps}^{-1}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\iota^{ps}(D_l, \mathcal{F}_{lt^*-1}) + o_p(1) \tag{SA.3}$$

$$\sqrt{n}\left(p(\mathcal{F}_{t^*-1}; \hat{\pi}) - p(\mathcal{F}_{t^*-1}; \pi^*)\right) = \kappa_{ps}(\mathcal{F}_{t^*-1})'Q_{ps}^{-1}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\iota^{ps}(D_l, \mathcal{F}_{lt^*-1}) + o_p(1) \tag{SA.4}$$

where $\iota^{ps}(D, \mathcal{F}_{t^*-1}) = h_{ps}(\mathcal{F}_{t^*-1})(D - p(\mathcal{F}_{t^*-1}; \pi^*))$ and $\kappa_{ps}(\mathcal{F}_{t^*-1}) = \lambda(h_{ps}(\mathcal{F}_{t^*-1})'\pi^*)h_{ps}(\mathcal{F}_{t^*-1})$ with $\lambda$ the derivative of $\Lambda$. Similarly,

$$\sqrt{n}\left(\hat{\mu}_t - \mu_t^*\right) = Q_{0,or}^{-1}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\iota^{or}(C_{lt}, D_l, \mathcal{F}_{lt^*-1}) + o_p(1) \tag{SA.5}$$

$$\sqrt{n}\left(m_{0,t}^C(\mathcal{F}_{t^*-1}; \hat{\mu}_t) - m_{0,t}^C(\mathcal{F}_{t^*-1}; \mu_t^*)\right) = h_{or}(\mathcal{F}_{t^*-1})'Q_{0,or}^{-1}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\iota^{or}(C_{lt}, D_l, \mathcal{F}_{lt^*-1}) + o_p(1) \tag{SA.6}$$

where $\iota^{or}(C_t, D, \mathcal{F}_{t^*-1}) = \frac{(1-D)}{(1-\mathbb{E}[D])^{1/2}}h_{or}(\mathcal{F}_{t^*-1})(C_t - h_{or}(\mathcal{F}_{t^*-1})'\mu_t^*)$.

Next, we provide an asymptotically linear representation for estimating $ATT_t^C$ as well as its limiting distribution. Before doing that, we introduce some additional notation. First, let $W_t := (C_t, D, \mathcal{F}_{t^*-1}')'$, $W_{lt} := (C_{lt}, D_l, \mathcal{F}_{lt^*-1}')'$, and $W := (W_{t^*}', \ldots, W_{\mathcal{T}}')'$. Further, define

$$\psi_{1,t}^{A}(W_{lt}) := -\frac{\mathbb{E}\left[D(C_t - m_{0,t}^{C}(\mathcal{F}_{t^*-1};\mu_t^*)\right]}{\mathbb{E}[D]^2}(D_l - \mathbb{E}[D])$$

$$\psi_{0,t}^{B_1}(W_{lt}) := \frac{1}{\mathbb{E}\left[\tilde{\omega}_0(D,\mathcal{F}_{t^*-1};\pi^*)\right]}\mathbb{E}\left[\frac{(1-D)(C_t - m_{0,t}^{C}(\mathcal{F}_{t^*-1};\mu_t^*))}{(1-p(\mathcal{F}_{t^*-1};\pi^*))^2}\kappa_{ps}(\mathcal{F}_{t^*-1})'\right] Q_{ps}^{-1}\iota^{ps}(D_l,\mathcal{F}_{lt^*-1})$$

$$\psi_{0,t}^{B_{21}}(W_{lt}) := \zeta_t\mathbb{E}\left[\left(\frac{(1-D)}{(1-p(\mathcal{F}_{t^*-1};\pi^*))^2}\right)\kappa_{ps}(\mathcal{F}_{t^*-1})'\right] Q_{ps}^{-1}\iota^{ps}(D_l,\mathcal{F}_{lt^*-1})$$

$$\psi_{0,t}^{B_{22}}(W_{lt}) := \zeta_t\left(\tilde{\omega}_0(D_l,\mathcal{F}_{lt^*-1};\pi^*) - \mathbb{E}\left[\tilde{\omega}_0(D,\mathcal{F}_{t^*-1};\pi^*)\right]\right)$$

$$\psi_{t}^{C}(W_{lt}) := \mathbb{E}\left[\omega(D_l,\mathcal{F}_{lt^*-1};\pi^*)h_{or}(\mathcal{F}_{lt^*-1})'\right] Q_{0,or}^{-1}\iota^{or}(C_{lt},\mathcal{F}_{lt^*-1})$$

$$\psi_{t}^{D}(W_{lt}) := \omega(D_l,\mathcal{F}_{lt^*-1};\pi^*)(C_{lt} - m_{0,t}^{C}(\mathcal{F}_{lt^*-1};\mu_t^*)) - ATT_t^C$$

and where

$$\zeta_t := \frac{\mathbb{E}\left[\omega_0(D,\mathcal{F}_{t^*-1};\pi^*)(C_t - m_{0,t}^{C}(\mathcal{F}_{t^*-1};\mu_t^*))\right]}{\mathbb{E}\left[\tilde{\omega}_0(D,\mathcal{F}_{t^*-1};\pi^*)\right]}$$

Next, define

$$\psi_t(W_{lt}) := \psi_{1,t}^{A}(W_{lt}) - \psi_{0,t}^{B_1}(W_{lt}) + \psi_{0,t}^{B_{21}}(W_{lt}) + \psi_{0,t}^{B_{22}}(W_{lt}) - \psi_{t}^{C}(W_{lt}) + \psi_{t}^{D}(W_{lt})$$

and $\Psi(W) := (\psi_{t^*}(W_{t^*}),\ldots,\psi_{\mathcal{T}}(W_{\mathcal{T}}))'$. Moreover, let $ATT^C = (ATT_{t^*}^C,\ldots,ATT_{\mathcal{T}}^C)'$, and, likewise, $\widehat{ATT}^C = (\widehat{ATT}_{t^*}^C,\ldots,\widehat{ATT}_{\mathcal{T}}^C)'$.

**Proposition SA.2.** *Under the Stochastic SIRD Model for Untreated Potential Outcomes and Assumptions 1 and SA.1 to SA.4, for any $t^* \leq t \leq \mathcal{T}$,*

$$\sqrt{n}(\widehat{ATT}_t^C - ATT_t^C) = \frac{1}{\sqrt{n}}\sum_{l=1}^{n}\psi_t(W_{lt}) + o_p(1)$$

*In addition,*

$$\sqrt{n}(\widehat{ATT}^C - ATT^C) \xrightarrow{d} N(0,V)$$

*where $V = \mathbb{E}[\Psi(W)\Psi(W)']$.*

Before proving Proposition SA.2, we provide an additional helpful result that is used in the proof.

**Lemma SA.1.** *Under the Stochastic SIRD Model for Untreated Potential Outcomes and Assumptions 1 and SA.1 to SA.4,*

$$\sqrt{n}\left(\tilde{\omega}_0(D,\mathcal{F}_{t^*-1},\hat{\pi})-\tilde{\omega}_0(D,\mathcal{F}_{t^*-1},\pi^*)\right)=\left(\frac{(1-D)}{(1-p(\mathcal{F}_{t^*-1};\pi^*))^2}\right)\sqrt{n}\left(p(\mathcal{F}_{t^*-1};\hat{\pi})-p(\mathcal{F}_{t^*-1};\pi^*)\right)+o_p(1)$$

*Proof.* First, notice that

$$\begin{aligned}
&\sqrt{n}\left(\tilde{\omega}_0(D,\mathcal{F}_{t^*-1},\hat{\pi})-\tilde{\omega}_0(D,\mathcal{F}_{t^*-1},\pi^*)\right)\\
&\qquad=\sqrt{n}\left(\frac{p(\mathcal{F}_{t^*-1};\hat{\pi})}{(1-p(\mathcal{F}_{t^*-1};\hat{\pi}))}-\frac{p(\mathcal{F}_{t^*-1};\pi^*)}{(1-p(\mathcal{F}_{t^*-1};\pi^*))}\right)(1-D)\\
&\qquad=\left(\frac{(1-D)}{(1-p(\mathcal{F}_{t^*-1};\hat{\pi}))(1-p(\mathcal{F}_{t^*-1};\pi^*))}\right)\sqrt{n}\left(p(\mathcal{F}_{t^*-1};\hat{\pi})-p(\mathcal{F}_{t^*-1};\pi^*)\right)\\
&\qquad=\left(\frac{(1-D)}{(1-p(\mathcal{F}_{t^*-1};\pi^*))^2}\right)\sqrt{n}\left(p(\mathcal{F}_{t^*-1};\hat{\pi})-p(\mathcal{F}_{t^*-1};\pi^*)\right)+o_p(1)
\end{aligned}$$

where the first equality holds by the definition of $\tilde{\omega}_0$, the second equality by cross-multiplying and rearranging, and the last equality by the weak law of large numbers and the continuous mapping theorem. □

*Proof of Proposition SA.2.* By adding and subtracting terms (and using the result from Proposition SA.1), we can write

$$\begin{aligned}
\sqrt{n}\left(\widehat{ATT}_t^C-ATT_t^C\right)&=\frac{1}{\sqrt{n}}\sum_{l=1}^{n}(\hat{\omega}_1(D_l)-\omega_1(D_l))(C_{lt}-m_{0,t}^C(\mathcal{F}_{lt^*-1};\hat{\mu}_t))\\
&\quad-\frac{1}{\sqrt{n}}\sum_{l=1}^{n}(\hat{\omega}_0(D_l,\mathcal{F}_{lt^*-1};\hat{\pi}))-\omega_0(D_l,\mathcal{F}_{lt^*-1};\pi^*))(C_{lt}-m_{0,t}^C(\mathcal{F}_{lt^*-1};\hat{\mu}_t))\\
&\quad-\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\omega(D_l,\mathcal{F}_{lt^*-1};\pi^*)(m_{0,t}^C(\mathcal{F}_{lt^*-1};\hat{\mu}_t)-m_{0,t}^C(\mathcal{F}_{lt^*-1};\mu_t^*))\\
&\quad+\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\omega(D_l,\mathcal{F}_{lt^*-1};\pi^*)(C_{lt}-m_{0,t}^C(\mathcal{F}_{lt^*-1};\mu_t^*))-ATT_t^C\\
&:=A-B-C+D
\end{aligned}$$

Term A involves the estimation effect of the first component of the weights, Term B involves the estimation effect of the second component of the weights, Term C involves the estimation effect of the outcome regression, and Term D is the estimation effect if the weights and outcome regression were known. We consider each of these terms in turn next.

First, we consider Term A. Notice that we can re-write it as

$$\begin{aligned}
A &= \frac{1}{\sqrt{n}}\sum_{l=1}^{n}(\hat{\omega}_1(D_l) - \omega_1(D_l))(C_{lt} - m_{0,t}^C(\mathcal{F}_{lt^*-1}; \mu_t^*)) \\
&- \frac{1}{\sqrt{n}}\sum_{l=1}^{n}(\hat{\omega}_1(D_l) - \omega_1(D_l))(m_{0,t}^C(\mathcal{F}_{lt^*-1}; \hat{\mu}_t) - m_{0,t}^C(\mathcal{F}_{lt^*-1}; \mu_t^*)) \\
&= A_1 - A_2
\end{aligned}$$

Now consider Term $A_1$.

$$\begin{aligned}
A_1 &= \frac{1}{\sqrt{n}}\sum_{l=1}^{n}\left(\frac{D_l}{\bar{D}} - \frac{D_l}{\mathbb{E}[D]}\right)(C_{lt} - m_{0,t}^C(\mathcal{F}_{lt^*-1}; \mu_t^*)) \\
&= -\frac{1}{\sqrt{n}}\sum_{l=1}^{n} D_l \frac{(\bar{D} - \mathbb{E}[D])}{\bar{D}\mathbb{E}[D]}(C_{lt} - m_{0,t}^C(\mathcal{F}_{lt^*-1}; \mu_t^*)) \\
&= -\frac{1}{\sqrt{n}}\sum_{l=1}^{n} \frac{\mathbb{E}\left[D(C_t - m_{0,t}^C(\mathcal{F}_{t^*-1}; \mu_t^*)\right]}{\mathbb{E}[D]^2}(D_l - \mathbb{E}[D]) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{l=1}^{n}\psi_{1,t}^A(W_{lt}) + o_p(1)
\end{aligned}$$

where the first equality holds by the definition of the weights, the second equality holds by combining terms, the third equality holds by the weak law of large numbers and continuous mapping theorem, and the last equality by the definition of $\psi_{1,t}^A$.

Next, consider Term $A_2$,

$$\begin{aligned}
A_2 &= \frac{1}{n}\sum_{l=1}^{n}\left(\frac{D_l}{\bar{D}} - \frac{D_l}{\mathbb{E}[D]}\right) h_{or}(\mathcal{F}_{lt^*-1})'\sqrt{n}(\hat{\mu}_t - \mu_t^*) + o_p(1) \\
&= -\frac{1}{n}\sum_{l=1}^{n}\left(\frac{D_l(\bar{D} - \mathbb{E}[D])}{\bar{D}\mathbb{E}[D]}\right) h_{or}(\mathcal{F}_{lt^*-1})'\sqrt{n}(\hat{\mu}_t - \mu_t^*) + o_p(1) \\
&= -\frac{\mathbb{E}[Dh_{or}(\mathcal{F}_{t^*-1})']}{\mathbb{E}[D]^2}(\bar{D} - \mathbb{E}[D])\sqrt{n}(\hat{\mu}_t - \mu_t^*) + o_p(1) \\
&= o_p(1)
\end{aligned}$$

where the first equality holds from the definitions of the weights and by Equation (SA.6), the second equality holds by combining terms, the third equality by the weak law of large numbers and continuous mapping theorem, and the last equality holds because $(\bar{D} - \mathbb{E}[D]) = o_p(1)$ and $\sqrt{n}(\hat{\mu}_t - \mu_t^*) = O_p(1)$.

Next, consider Term B,

$$
\begin{aligned}
B &= \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \left( \frac{\tilde{\omega}_0(D_l, \mathcal{F}_{lt^*-1}; \hat{\pi}) - \tilde{\omega}_0(D_l, \mathcal{F}_{lt^*-1}; \pi^*)}{\frac{1}{n}\sum_{h=1}^{n} \tilde{\omega}_0(D_h, \mathcal{F}_{ht^*-1}; \hat{\pi})} \right) (C_{lt} - m_{0,t}^{C}(\mathcal{F}_{lt^*-1}; \hat{\mu}_t)) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \left( \frac{\tilde{\omega}_0(D_l, \mathcal{F}_{lt^*-1}; \pi^*)}{\frac{1}{n}\sum_{h=1}^{n} \tilde{\omega}_0(D_h, \mathcal{F}_{ht^*-1}; \hat{\pi})} - \frac{\tilde{\omega}_0(D_l, \mathcal{F}_{lt^*-1}; \pi^*)}{\mathbb{E}\left[\tilde{\omega}_0(D, \mathcal{F}_{t^*-1}; \pi^*)\right]} \right) (C_{lt} - m_{0,t}^{C}(\mathcal{F}_{lt^*-1}; \hat{\mu}_t)) \\
&:= B_1 + B_2
\end{aligned}
$$

For Term $B_1$,

$$
\begin{aligned}
B_1 &= \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \left( \frac{\tilde{\omega}_0(D_l, \mathcal{F}_{lt^*-1}; \hat{\pi}) - \tilde{\omega}_0(D_l, \mathcal{F}_{lt^*-1}; \pi^*)}{\mathbb{E}\left[\tilde{\omega}_0(D, \mathcal{F}_{t^*-1}; \pi^*)\right]} \right) (C_{lt} - m_{0,t}^{C}(\mathcal{F}_{lt^*-1}; \mu_t^*)) + o_p(1) \\
&= \frac{1}{n} \sum_{l=1}^{n} \left( \frac{(1 - D_l)(C_{lt} - m_{0,t}^{C}(\mathcal{F}_{lt^*-1}; \mu_t^*))}{(1 - p(\mathcal{F}_{lt^*-1}; \pi^*))^2 \mathbb{E}\left[\tilde{\omega}_0(D, \mathcal{F}_{t^*-1}; \pi^*)\right]} \right) \sqrt{n} \left( p(\mathcal{F}_{lt^*-1}; \hat{\pi}) - p(\mathcal{F}_{lt^*-1}; \pi^*) \right) + o_p(1) \\
&= \frac{1}{\mathbb{E}\left[\tilde{\omega}_0(D, \mathcal{F}_{t^*-1}; \pi^*)\right]} \mathbb{E}\left[ \frac{(1 - D)(C_t - m_{0,t}^{C}(\mathcal{F}_{t^*-1}; \mu_t^*))}{(1 - p(\mathcal{F}_{t^*-1}; \pi^*))^2} \kappa_{ps}(\mathcal{F}_{t^*-1})' \right] Q_{ps}^{-1} \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \iota^{ps}(D_l, \mathcal{F}_{lt^*-1}) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \psi_{0,t}^{B_1}(W_{lt}) + o_p(1)
\end{aligned}
$$

where the first equality follows from similar arguments as above, the second equality uses Lemma SA.1, and the last equality holds from Equation (SA.4) and by the weak law of large numbers and continuous mapping theorem.

Next, for term $B_2$,

$$
\begin{aligned}
B_2 &= -\frac{1}{n} \sum_{l=1}^{n} \frac{\omega_0(D_l, \mathcal{F}_{lt^*-1}; \pi^*)(C_{lt} - m_{0,t}^{C}(\mathcal{F}_{lt^*-1}; \hat{\mu}_t))}{\frac{1}{n}\sum_{h=1}^{n} \tilde{\omega}_0(D_h, \mathcal{F}_{ht^*-1}; \hat{\pi})} \left( \frac{1}{\sqrt{n}} \sum_{h=1}^{n} \tilde{\omega}_0(D_h, \mathcal{F}_{ht^*-1}; \hat{\pi}) - \mathbb{E}\left[\tilde{\omega}_0(D, \mathcal{F}_{t^*-1}; \pi^*)\right] \right) \\
&= -\frac{\mathbb{E}\left[\omega_0(D, \mathcal{F}_{t^*-1}; \pi^*)(C_t - m_{0,t}^{C}(\mathcal{F}_{t^*-1}; \mu_t^*))\right]}{\mathbb{E}\left[\tilde{\omega}_0(D, \mathcal{F}_{t^*-1}; \pi^*)\right]} \left( \frac{1}{\sqrt{n}} \sum_{h=1}^{n} \tilde{\omega}_0(D_h, \mathcal{F}_{ht^*-1}; \hat{\pi}) - \mathbb{E}\left[\tilde{\omega}_0(D, \mathcal{F}_{t^*-1}; \pi^*)\right] \right) + o_p(1) \\
&= -\frac{\mathbb{E}\left[\omega_0(D, \mathcal{F}_{t^*-1}; \pi^*)(C_t - m_{0,t}^{C}(\mathcal{F}_{t^*-1}; \mu_t^*))\right]}{\mathbb{E}\left[\tilde{\omega}_0(D, \mathcal{F}_{t^*-1}; \pi^*)\right]} \left( \frac{1}{\sqrt{n}} \sum_{h=1}^{n} \tilde{\omega}_0(D_h, \mathcal{F}_{ht^*-1}; \hat{\pi}) - \tilde{\omega}_0(D_h, \mathcal{F}_{ht^*-1}; \pi^*) \right) \\
&\quad - \frac{\mathbb{E}\left[\omega_0(D, \mathcal{F}_{t^*-1}; \pi^*)(C_t - m_{0,t}^{C}(\mathcal{F}_{t^*-1}; \mu_t^*))\right]}{\mathbb{E}\left[\tilde{\omega}_0(D, \mathcal{F}_{t^*-1}; \pi^*)\right]} \left( \frac{1}{\sqrt{n}} \sum_{h=1}^{n} \tilde{\omega}_0(D_h, \mathcal{F}_{ht^*-1}; \pi^*) - \mathbb{E}\left[\tilde{\omega}_0(D, \mathcal{F}_{t^*-1}; \pi^*)\right] \right) + o_p(1) \\
&:= -B_{21} - B_{22}
\end{aligned}
$$

where the first equality holds by cross-multiplying and the definition of $\omega_0$, the second equality holds by the weak law of large numbers and the continuous mapping theorem, and the third equality by adding and subtracting terms.

For $B_{21}$, notice that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{h=1}^{n}\tilde{\omega}_0(D_h,\mathcal{F}_{ht^*-1};\hat{\pi})-\tilde{\omega}_0(D_h,\mathcal{F}_{ht^*-1};\pi^*)\\
&\qquad=\frac{1}{n}\sum_{h=1}^{n}\left(\frac{(1-D_h)}{(1-p(\mathcal{F}_{ht^*-1};\pi^*))^2}\right)\sqrt{n}\left(p(\mathcal{F}_{ht^*-1};\hat{\pi})-p(\mathcal{F}_{ht^*-1};\pi^*)\right)+o_p(1)\\
&\qquad=\frac{1}{n}\sum_{h=1}^{n}\left(\frac{(1-D_h)}{(1-p(\mathcal{F}_{ht^*-1};\pi^*))^2}\right)\kappa_{ps}(\mathcal{F}_{ht^*-1})'Q_{ps}^{-1}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\iota^{ps}(D_l,\mathcal{F}_{lt^*-1})+o_p(1)\\
&\qquad=\mathbb{E}\left[\left(\frac{(1-D)}{(1-p(\mathcal{F}_{t^*-1};\pi^*))^2}\right)\kappa_{ps}(\mathcal{F}_{t^*-1})'\right]Q_{ps}^{-1}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\iota^{ps}(D_l,\mathcal{F}_{lt^*-1})+o_p(1)
\end{aligned}$$

where the first equality holds by Lemma SA.1, the second equality holds by Equation (SA.4), and the last equality holds by the weak law of large numbers and continuous mapping theorem. This implies that

$$B_{21}=\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\psi_{0,t}^{B_{21}}(W_{lt})+o_p(1)$$

For $B_{22}$, notice that it is immediately given by

$$B_{22}=\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\psi_{0,t}^{B_{22}}(W_{lt})$$

Now, we turn to Term C.

$$\begin{aligned}
C&=\frac{1}{n}\sum_{l=1}^{n}\omega(D_l,\mathcal{F}_{lt^*-1};\pi^*)h_{or}(\mathcal{F}_{lt^*-1})'Q_{0,or}^{-1}\frac{1}{\sqrt{n}}\sum_{h=1}^{n}\iota^{or}(C_{ht},D_h,\mathcal{F}_{ht^*-1})+o_p(1)\\
&=\mathbb{E}\left[\omega(D,\mathcal{F}_{t^*-1};\pi^*)h_{or}(\mathcal{F}_{t^*-1})'\right]Q_{0,or}^{-1}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\iota^{or}(C_{lt},D_l,\mathcal{F}_{lt^*-1})+o_p(1)\\
&=\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\psi_t^{C}(W_{lt})+o_p(1)
\end{aligned}$$

where the first equality holds by Equation (SA.6) and the second equality by the weak law of large numbers and continuous mapping theorem.

Finally, for Term D, notice that it is immediately given by

$$D=\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\psi_t^{D}(W_{lt})$$

Combining the results for Terms A-D establishes the first part of the result. Asymptotic normality holds by applying the central limit theorem jointly for $t=t^*,\ldots,\mathcal{T}$.

□

In order to actually conduct inference, we use the multiplier bootstrap. To start with, we describe the multiplier bootstrap procedure that we use. First, define $\hat{\Psi}$ as the sample analogue of $\Psi$. Next, let $\xi$ denote an $n$-dimensional vector of iid random variables with mean zero, variance one, finite third moment, and that is independent of the original data (e.g., two common choices for $\xi$ are either iid draws from $N(0,1)$ or to draw $\xi$ equal to $-1$ or 1 each with probability $1/2$). Then, we consider a bootstrapped version of $\widehat{ATT}^C$ given by

$$\widehat{ATT}^{C,*} = \widehat{ATT}^C + \frac{1}{n}\sum_{l=1}^{n} \xi_l \hat{\Psi}(W_l)$$

Relative to the more common nonparametric bootstrap, there are two main advantages of the multiplier bootstrap. First, it is very fast to compute as it essentially only involves making random draws from a simple distribution rather than re-estimating $\widehat{ATT}^C$ at every bootstrap iteration. Second, since the multiplier bootstrap perturbs the influence function rather than re-drawing data, this approach does not run into the practical problem of particular bootstrap iterations not being able to estimate the parameters of interest (e.g., this can occur when there are discrete covariates where some combinations occur infrequently).

The next result shows that the proposed multiplier bootstrap procedure follows the same limiting distribution as the original estimator of $ATT^C$.

**Proposition SA.3.** *Under the Stochastic SIRD Model for Untreated Potential Outcomes and Assumptions 1 and SA.1 to SA.4,*

$$\sqrt{n}(\widehat{ATT}^{C,*} - \widehat{ATT}^C) \xrightarrow{d^*} N(0, V)$$

*where $V$ is the same as in Proposition SA.2 and $\xrightarrow{d^*}$ denotes convergence in bootstrap distribution.*

*Proof.* Given the results in Proposition SA.2, the proof of Proposition SA.3 follows from the same arguments as in the proof of Theorem 3 in Callaway and Sant'Anna (2021). □

Recall that $ATT^C = (ATT_{t^*}, \ldots, ATT^C_{\mathcal{T}})'$ so that the results in Propositions SA.2 and SA.3 hold jointly across post-treatment periods. It is therefore straightforward to construct uniform confidence bands that asymptotically cover $ATT^C$ simultaneously with fixed probability $1-\alpha$. In particular, one can construct a uniform confidence band as follows.

**Algorithm SA.1** (Multiplier Bootstrap for Uniform Confidence Band)**.**

*Step 1: Draw $\xi_l$, $l = 1, \ldots, n$ which are iid across $l$, have mean 0, variance 1, and finite third moment*

*Step 2: Set* $\widehat{ATT}^{C,*} = \widehat{ATT}^{C} + \frac{1}{n}\sum_{l=1}^{n} \xi_l \hat{\Psi}(W_l)$.

*Step 3: For* $t = t^*, \ldots, \mathcal{T}$, *compute* $\hat{R}_t^* = \sqrt{n}(\widehat{ATT}_t^{C,*} - \widehat{ATT}_t^{C})$ *where* $\widehat{ATT}_t^{C,*}$ *is a particular element of* $\widehat{ATT}^{C,*}$ *from Step 2.*

*Repeat Steps 1-3* $B$ *times where* $B$ *is the (large) number of bootstrap iterations.*

*Step 4: Compute* $\hat{V}_t^{1/2} = (q_{0.75,t} - q_{0.25,t})/(z_{0.75} - z_{0.25})$ *where* $q_{p,t}$ *is the* $p$*th quantile of* $\hat{R}_t^*$ *across the* $B$ *bootstrap iterations and* $z_p$ *is the* $p$*th quantile of the standard normal distribution.*

*Step 5: For each bootstrap draw, compute sup-t* $= \max_{t \in \{t^*, \ldots, \mathcal{T}\}} |\hat{R}_t^*| \hat{V}_t^{-1/2}$

*Step 6: Construct the critical value* $\hat{c}_{1-\alpha}$ *as the* $(1-\alpha)$ *quantile of the* $B$ *bootstrap draws of sup-t.*

*Step 7: Construct the uniform confidence band* $\hat{C}_t = [\widehat{ATT}_t^{C} \pm \hat{c}_{1-\alpha} \hat{V}_t^{-1/2}/\sqrt{n}]$

The next result shows that the uniform confidence band from Algorithm SA.1 has the asymptotically correct coverage.

**Proposition SA.4.** *Under the Stochastic SIRD Model for Untreated Potential Outcomes and Assumptions 1 and SA.1 to SA.4,*

$$P(ATT_t^C \in \hat{C}_t \ \text{ for all } t \in \{t^*, \ldots, \mathcal{T}\}) \to 1 - \alpha$$

*as* $n \to \infty$.

*Proof.* Given the result in Proposition SA.3, the result holds by the same argument as for Theorem 3 (and Corollary 1) in Callaway and Sant'Anna (2021). □

To conclude, we provide the limiting distribution and briefly discuss an inference procedure for the adjusted regression DID estimator of $ATT^Y$ discussed in the main text. The arguments only require a minor extension of the results for $ATT^C$ given above; therefore, we only briefly sketch the additional arguments for adjusted regression DID here.

For the arguments below, suppose that the Stochastic SIRD Model for Untreated Potential Outcomes, Assumption 1, Assumption SA.1, assumptions analogous to Assumptions SA.2 to SA.4 but for current Covid-19 cases rather than cumulative cases all hold. In addition, for all $t = t^*, \ldots, \mathcal{T}$, we assume that $Q_t^{0,I} := \mathbb{E}[(1, \Delta^{(t^*-1,t)} I_t)'(1, \Delta^{(t^*-1,t)} I_t)|D = 0]$ is positive definite and that $\mathbb{E}[\Delta^{(t^*-1,t)} Y_t^4 | D = d] < \infty$ for $d \in \{0, 1\}$.

We also define

$$\psi^A_{Y,t}(W_{lt}) := -\left(\frac{\mathbb{E}[D\Delta^{(t^*-1,t)}Y_t]}{\mathbb{E}[D]^2}(D_l - \mathbb{E}[D])\right) + \left(\frac{D_l}{\mathbb{E}[D]}\Delta^{(t^*-1,t)}Y_{lt} - \mathbb{E}\left[\frac{D}{\mathbb{E}[D]}\Delta^{(t^*-1,t)}Y_t\right]\right)$$

$$\psi^B_{Y,t}(W_{lt}) := \eta_{\tau,\alpha}(D_l, \Delta^{(t^*-1,t)}I_{lt}, \Delta^{(t^*-1,t)}Y_{lt})'(1, \mathbb{E}[\Delta^{(t^*-1,t)}I_t(0)|D=1])'$$

$$\psi^C_{Y,t}(W_l t) := \alpha\left\{-\left(\frac{\mathbb{E}[D\Delta^{(t^*-1,t)}I_t]}{\mathbb{E}[D]^2}(D_l - \mathbb{E}[D])\right) + \left(\frac{D_l}{\mathbb{E}[D]}\Delta^{(t^*-1,t)}I_{lt} - \mathbb{E}\left[\frac{D}{\mathbb{E}[D]}\Delta^{(t^*-1,t)}I_t\right]\right) - \psi_{I,t}(W_{lt})\right\}$$

where $\eta_{\tau,\alpha}(D, \Delta^{(t^*-1,t)}I_t, \Delta^{(t^*-1,t)}Y_t)$ is defined below and where $\psi^I_t$ is the same as $\psi_t$ in Proposition SA.2 except with $I$ (the current number of Covid-19 cases) replacing $C$ (the cumulative number of Covid-19 cases) everywhere. Next, define

$$\psi_{Y,t}(W_{lt}) := \psi^A_{Y,t}(W_{lt}) - \psi^B_{Y,t}(W_{lt}) - \psi^C_{Y,t}(W_{lt})$$

and $\Psi_Y(W) = (\psi_{Y,t^*}(W_{t^*}), \ldots, \psi_{Y,\mathcal{T}}(W_{\mathcal{T}}))'$. Define $ATT^Y = (ATT^Y_{t^*}, \ldots, ATT^Y_{\mathcal{T}})'$ and $\widehat{ATT}^Y = (\widehat{ATT}^Y_{t^*}, \ldots, \widehat{ATT}^Y_{\mathcal{T}})'$. Then, the following results all hold using essentially the same arguments as above:

$$\sqrt{n}(\widehat{ATT}^Y_t - \widehat{ATT}^Y_t) = \frac{1}{\sqrt{n}}\sum_{l=1}^n \psi_{Y,t}(W_{lt}) + o_p(1)$$

Moreover,

$$\sqrt{n}(\widehat{ATT}^Y - ATT^Y) \xrightarrow{d} N(0, V_Y)$$

where $V_Y = \mathbb{E}[\Psi_Y(W)\Psi_Y(W)']$. In addition, the multiplier bootstrap can be used to conduct inference and uniform confidence bands can be constructed analogously as above. We show the result for the influence function below. The remaining results hold immediately from the same arguments as above given the distinct expression for the influence function in this case.

Start by noticing that

$$\begin{aligned}
\sqrt{n}(\widehat{ATT}^Y_t - ATT^Y_t) &= \frac{1}{\sqrt{n}}\sum_{l=1}^n \left(\hat{\omega}_1(D_l)\Delta^{(t^*-1,t)}Y_{lt} - \mathbb{E}\left[\frac{D}{\mathbb{E}[D]}\Delta^{(t^*-1,t)}Y_t\right]\right) \\
&\quad - \sqrt{n}\left((\hat{\tilde{\tau}}_t - \tilde{\tau}_t) + (\hat{\alpha} - \alpha)\hat{\mathbb{E}}[\Delta^{(t^*-1,t)}I_t(0)|D=1]\right) \\
&\quad - \alpha\sqrt{n}\left(\hat{\mathbb{E}}[\Delta^{(t^*-1,t)}I_t(0)|D=1] - \mathbb{E}[\Delta^{(t^*-1,t)}I_t(0)|D=1]\right) \\
&:= A - B - C
\end{aligned}$$

where

$$\hat{\mathbb{E}}[\Delta^{(t^*-1,t)}I_t(0)|D=1] = \frac{1}{n}\sum_{l=1}^{n}\left(\frac{D_l}{\bar{D}}\Delta^{(t^*-1,t)}I_{lt} - \hat{\omega}(D_l,\mathcal{F}_{lt^*-1})(I_{lt} - \hat{m}^C_{0,t}(\mathcal{F}_{lt^*-1}))\right)$$

Following similar arguments as above, it follows that

$$\begin{aligned}A &= -\frac{\mathbb{E}[D\Delta^{(t^*-1,t)}Y_t]}{\mathbb{E}[D]^2}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}(D_l - \mathbb{E}[D]) + \frac{1}{\sqrt{n}}\sum_{l=1}^{n}\left(\frac{D_l}{\mathbb{E}[D]}\Delta^{(t^*-1,t)}Y_{lt} - \mathbb{E}\left[\frac{D}{\mathbb{E}[D]}\Delta^{(t^*-1,t)}Y_t\right]\right) + o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{l=1}^{n}\psi^A_{Y,t}(W_{lt}) + o_p(1)\end{aligned}$$

For term B, first notice that

$$\begin{aligned}\sqrt{n}\begin{pmatrix}\hat{\tilde{\tau}}_t - \tilde{\tau}_t\\ \hat{\alpha} - \alpha\end{pmatrix} &= Q_t^{0,I^{-1}}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\frac{(1-D_l)}{(1-\mathbb{E}[D])^{1/2}}(1,\Delta^{(t^*-1,t)}I_{lt})'\Delta^{(t^*-1,t)}Y_{lt} + o_p(1)\\ &:= \frac{1}{\sqrt{n}}\sum_{l=1}^{n}\eta_{\tau,\alpha}(D_l,\Delta^{(t^*-1,t)}I_{lt},\Delta^{(t^*-1,t)}Y_{lt}) + o_p(1)\end{aligned}$$

Thus,

$$\begin{aligned}B &= \frac{1}{\sqrt{n}}\sum_{l=1}^{n}\eta_{\tau,\alpha}(D_l,\Delta^{(t^*-1,t)}I_{lt},\Delta^{(t^*-1,t)}Y_{lt})'(1,\mathbb{E}[\Delta^{(t^*-1,t)}I_t(0)|D=1])' + o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{l=1}^{n}\psi^B_{Y,t}(W_{lt}) + o_p(1)\end{aligned}$$

For Term C, first notice that

$$\begin{aligned}&\sqrt{n}\left(\hat{\mathbb{E}}[\Delta^{(t^*-1,t)}I_t(0)|D=1] - \mathbb{E}[\Delta^{(t^*-1,t)}I_t(0)|D=1]\right)\\ &\quad= \frac{1}{\sqrt{n}}\sum_{l=1}^{n}\left(\hat{\omega}_1(D_l)\Delta^{(t^*-1,t)}I_{lt} - \mathbb{E}\left[\frac{D}{\mathbb{E}[D]}\Delta^{(t^*-1,t)}I_t\right]\right)\\ &\quad- \frac{1}{\sqrt{n}}\sum_{l=1}^{n}\left(\hat{\omega}(D_l,\mathcal{F}_{lt^*-1})(I_{lt} - \hat{m}^I_{0,t}(\mathcal{F}_{lt^*-1})) - \mathbb{E}\left[\omega(D,\mathcal{F}_{t^*-1})(I_t - m^I_{0,t}(\mathcal{F}_{t^*-1}))\right]\right)\\ &\quad:= C_1 - C_2\end{aligned}$$

Using the same arguments as above, it immediately follows that

$$C_1 = -\frac{\mathbb{E}[D\Delta^{(t^*-1,t)}I_t]}{\mathbb{E}[D]^2}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}(D_l - \mathbb{E}[D]) + \frac{1}{\sqrt{n}}\sum_{l=1}^{n}\left(\frac{D_l}{\mathbb{E}[D]}\Delta^{(t^*-1,t)}I_{lt} - \mathbb{E}\left[\frac{D}{\mathbb{E}[D]}\Delta^{(t^*-1,t)}I_t\right]\right)$$

For $C_2$, notice that it is exactly the same (up to $I$ replacing $C$) as in Proposition SA.2, which implies

that

$$C_2 = \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \psi_t^I(W_{lt}) + o_p(1)$$

Multiplying the expressions for $C_1$ and $C_2$ by $\alpha$ implies that

$$C = \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \psi_{Y,t}^C(W_{lt}) + o_p(1)$$

Plugging back in the expressions for Terms A, B, and C provides the influence function.

# SB Variation in Treatment Timing

In the application in the paper, an extra complication is that there is variation in the timing of implementing Covid-related policies across different states. In this section, we provide additional details along these lines that expand and formalize the discussion in Remark 2 in the main text. For simplicity, we focus on the case where the researcher is interested in the effect of the policy on cumulative Covid-19 cases rather than economic outcomes, but we note that analogous arguments to the ones presented here apply to that case as well. We make the following assumption

**Assumption SB.1** (Staggered Treatment Adoption)**.** *For all* $t = 2, \ldots, \mathcal{T}$, $D_{lt-1} = 1 \implies D_{lt} = 1$

Staggered treatment adoption says that once a location becomes treated, they remain treated in subsequent periods. This assumption applies for shelter-in-place orders at least over the relatively short time horizons that we consider in the application; even in cases where a location removes an early pandemic policy, these policies are arguably "scarring" in the sense that locations do not go back to an "untreated state" in periods after they removed the policy (see Sun and Abraham (2021) for a good discussion of how scarring policies fit into the framework of staggered treatment adoption).

To deal with variation in treatment timing across locations, for this section, we slightly modify the notation relative to the main text. First, we define a location's "group" by the time period when it becomes treated; that is, we set $G_l$ to be the time period when location $l$ implements the policy. For locations that do not implement the policy in any time period, we set $G_l = 0$; to conserve on notation below, it is helpful to additional define $U_l = \mathbf{1}\{G_l = 0\}$. Under Assumption SB.1, knowing a location's group implies that one knows that location's entire path of participating in the treatment. Because there is variation in treatment timing, in this section, we index potential outcomes by group; that is, let $S_{lt}(g), I_{lt}(g), R_{lt}(g), \delta_{lt}(g)$, and $C_{lt}(g)$ denote the values of the pandemic-related outcomes for location $l$ in time period $t$ if the policy had been implemented in time period $g$ for that location. In this notation, as in the main text, untreated potential outcomes can be written as $S_{lt}(0), I_{lt}(0), R_{lt}(0), \delta_{lt}(0)$, and $C_{lt}(0)$. In this setup, the researcher observes $S_{lt} = S_{lt}(G_l), I_{lt} = I_{lt}(G_l), R_{lt} = R_{lt}(G_l), \delta_{lt} = \delta_{lt}(G_l)$, and $C_{lt} = C_{lt}(G_l)$.

**Assumption SB.2** (No Anticipation)**.** $S_{lt}(G_l) = S_{lt}(0)$, $I_{lt}(G_l) = I_{lt}(0)$, $R_{lt}(G_l) = R_{lt}(0)$, $\delta_{lt}(G_l) = \delta_{lt}(0)$, *and* $C_{lt}(G_l) = C_{lt}(0)$ *for all* $t < G_l$.

Assumption SB.2 says that, in pre-treatment periods, outcomes are not affected by participating in the treatment in the future. Next, let $\mathcal{G} := \text{support}(G) \setminus \{0\}$ denote the set of groups that ever participate in the treatment. The remainder of this section focuses on identifying treatment effects for this set of groups. To proceed along these lines, following Callaway and Sant'Anna (2021), we target identifying group-time average treatment effects which are defined as

$$ATT^C(g,t) := \mathbb{E}[C_{lt}(g) - C_{lt}(0)|G = g]$$

This is the mean difference between observed Covid-19 cases for locations in group $g$ in time period $t$ relative to the number of Covid-19 cases that would have occurred if the policy had not been implemented. Given that $ATT^C(g,t)$ can be identified, the arguments in Callaway and Sant'Anna (2021) imply that a number of other, more aggregated treatment effect parameters can recovered as weighted averages of $ATT^C(g,t)$. One main example, is an event study parameter $ATT^C_{ES}(e)$ which is defined as the average treatment effect across groups in the time period (if it is available) when that group has been treated for exactly $e$ periods; this parameter is equal to a weighted average of all available $ATT^C(g,t)$ parameters such that $t = g + e$ where weights are given by each group's relative size. Another parameter is an overall treatment effect parameter $ATT^C_O$ which is the average treatment effect experienced by all locations that participate in the treatment in any time period which can be recovered from $ATT(g,t)$ by, for each group, averaging all of their available post-treatment $ATT(g,t)$'s and then combining these group-specific parameters by averaging them with weights given by relative group size. See Callaway and Sant'Anna (2021) for additional details and other possible parameters of interest. Below we focus on identifying $ATT^C(g,t)$.

The next proposition shows that $ATT^C(g,t)$ is identified for $t \geq g$ (i.e., post-treatment periods) under the Stochastic SIRD Model for Untreated Potential Outcomes from the main text. Towards this end, we define some additional notation. Define $p_g := \text{P}(G = g)$, $p_{g|\{g,u\}} := \text{P}(G = g|\mathbf{1}\{G = g\} + U = 1)$, $p_g(\mathcal{F}_{g-1}) := \text{P}(G = g|\mathcal{F}_{g-1}, \mathbf{1}\{G = g\} + U = 1)$. We also need to slightly modify the Stochastic SIRD Model for Untreated Potential Outcomes to account for different groups and variation in treatment timing. In particular, here we replace Equation (A.6) with

$$\mathbb{E}[u_t|\mathcal{F}_{t-1}(0), \ldots, \mathcal{F}_1(0), G] = 0 \text{ for all } t = 2, \ldots, \mathcal{T} \tag{SB.1}$$

and we replace Equation (A.7) with

$$u_t \perp\!\!\!\perp (G, \mathcal{F}_{t-1}(0), \ldots, \mathcal{F}_1(0))|\mathcal{F}_{t-1}(0) \text{ for all } t = 2, \ldots, \mathcal{T} \tag{SB.2}$$

These are analogous to the corresponding expressions in the main text but are imposed at the group level rather than for just for treated and untreated locations (in the case where treatment timing

is constant, these conditions are the same as the ones in the main text). Using the same sorts of arguments as for Proposition 1 in the main text, one can show that, for all $t \geq g$ (i.e., post-treatment periods)

$$\mathbb{E}[C_t(0)|\mathcal{F}_{g-1}, G = g] = \mathbb{E}[C_t(0)|\mathcal{F}_{g-1}, U = 1] \tag{SB.3}$$

which implies that, conditional on the pre-treatment state of the pandemic, untreated locations can be used to recover untreated cumulative cases that group $g$ would have experienced if it had not been treated. Notice that, in this expression, the "base period" $g-1$ (which is the period right before treatment for group $g$) plays the same role as $t^* - 1$ in the main text (the period right before the policy was implemented in the case where there is no variation in treatment timing).

**Proposition SB.1.** *In the Stochastic SIRD Model for Untreated Potential Outcomes with the additional conditions in Equations (SB.1) and (SB.2), and under the additional overlap condition that there exists some $\epsilon > 0$ such that $p_g > \epsilon$ and $p_g(\mathcal{F}_{g-1}) < 1 - \epsilon$, then for any $t \geq g$, $ATT^C(g,t)$ is identified and can be expressed as*

$$ATT^C(g,t) = \mathbb{E}\Big[\omega_g(G, \mathcal{F}_{g-1})(C_t - m^C_{U,g,t}(\mathcal{F}_{g-1}))\Big|\mathbf{1}\{G = g\} + U = 1\Big]$$

*where*

$$\omega_g(G, \mathcal{F}_{g-1}) := \frac{\mathbf{1}\{G = g\}}{p_{g|\{g,u\}}} - \frac{\frac{p_g(\mathcal{F}_{g-1})U}{p_{g|\{g,u\}}(1-p_g(\mathcal{F}_{g-1}))}}{\mathbb{E}\left[\frac{p_g(\mathcal{F}_{g-1})U}{p_{g|\{g,u\}}(1-p_g(\mathcal{F}_{g-1}))}\Big|\mathbf{1}\{G = g\} + U = 1\right]}$$

*and*

$$m^C_{U,g,t}(\mathcal{F}_{g-1}) := \mathbb{E}[C_t|\mathcal{F}_{g-1}, U = 1]$$

*Proof.* To start with, define the following notation,

$$\omega^1_g(G) := \frac{\mathbf{1}\{G = g\}}{p_{g|\{g,u\}}} \quad \text{and} \quad \omega^0_g(G, \mathcal{F}_{g-1}) := \frac{\tilde{\omega}^0_g(G, \mathcal{F}_{g-1})}{\mathbb{E}[\tilde{\omega}^0_g(G, \mathcal{F}_{g-1})|\mathbf{1}\{G = g\} + U = 1]}$$

and where

$$\tilde{\omega}^0_g(G, \mathcal{F}_{g-1}) := \frac{p_g(\mathcal{F}_{g-1})U}{p_{g|\{g,u\}}(1 - p_g(\mathcal{F}_{g-1}))}$$

so that we can re-express the weights in the proposition as

$$\omega_g(G, \mathcal{F}_{g-1}) = \omega^1_g(G) - \omega^0_g(G, \mathcal{F}_{g-1})$$

Now, notice that

$$\begin{aligned}
\mathbb{E}[C_t(g)|G=g] &= \mathbb{E}\left[\frac{\mathbf{1}\{G=g\}}{p_g}C_t\right] \\
&= \mathbb{E}\left[\frac{\mathbf{1}\{G=g\}}{p_g}C_t\middle|\mathbf{1}\{G=g\}+U=1\right](p_g+p_u) \\
&= \mathbb{E}\left[\frac{\mathbf{1}\{G=g\}}{p_{g|\{g,u\}}}C_t\middle|\mathbf{1}\{G=g\}+U=1\right] \\
&= \mathbb{E}\left[\omega_g^1(G)\,C_t\middle|\mathbf{1}\{G=g\}+U=1\right] \qquad \text{(SB.4)}
\end{aligned}$$

where the first equality holds because $C_t(g)$ are observed outcomes for group $g$, the second equality holds by the law of iterated expectations, the third equality holds because $p_{g|\{g,u\}} = p_g/(p_g+p_u)$, and the last equality holds by the definition of $\omega_g^1$. Next, notice that

$$\begin{aligned}
\mathbb{E}[C_t(0)|G=g] &= \mathbb{E}\Big[\mathbb{E}[C_t(0)|\mathcal{F}_{g-1},G=g]\Big|G=g\Big] \\
&= \mathbb{E}\Big[\mathbb{E}[C_t(0)|\mathcal{F}_{g-1},U=1]\Big|G=g\Big] \\
&= \mathbb{E}\Big[m^C_{U,g,t}(\mathcal{F}_{g-1})\Big|G=g\Big] \\
&= \mathbb{E}\left[\frac{\mathbf{1}\{G=g\}}{p_{g|\{g,u\}}}m^C_{U,g,t}(\mathcal{F}_{g-1})\middle|\mathbf{1}\{G=g\}+U=1\right] \\
&= \mathbb{E}\left[\omega_g^1(G)\,m^C_{U,g,t}(\mathcal{F}_{g-1})\middle|\mathbf{1}\{G=g\}+U=1\right] \qquad \text{(SB.5)}
\end{aligned}$$

where the first equality holds by the law of iterated expectations, the second equality holds by unconfoundedness of cumulative cases conditional on the pre-treatment state of the pandemic (i.e., in period $g-1$) as in Equation (SB.3), the third equality holds by the definition of $m^C_{U,g,t}$, the fourth equality holds from similar arguments as were used for Equation (SB.4), and the last equality holds by the definition of $\omega_g^1$. Subtracting Equation (SB.5) from Equation (SB.4), we have that

$$\begin{aligned}
ATT^C(g,t) &= \mathbb{E}[C_t(g)-C_t(0)|G=g] \\
&= \mathbb{E}\left[\omega_g^1(G)(C_t - m^C_{U,g,t}(\mathcal{F}_{g-1}))\middle|\mathbf{1}\{G=g\}+U=1\right] \qquad \text{(SB.6)}
\end{aligned}$$

Now consider

$$\begin{aligned}
&\mathbb{E}\left[\tilde{\omega}_g^0(G,\mathcal{F}_{g-1})(C_t - m_{U,g,t}^C(\mathcal{F}_{g-1}))\middle|\mathbf{1}\{G=g\}+U=1\right] \\
&\qquad = \mathbb{E}\left[\frac{p_g(\mathcal{F}_{g-1})U}{p_{g|\{g,u\}}(1-p_g(\mathcal{F}_{g-1}))}(C_t - m_{U,g,t}^C(\mathcal{F}_{g-1}))\middle|\mathbf{1}\{G=g\}+U=1\right] \\
&\qquad = \mathbb{E}\left[\frac{(1-p_{g|\{g,u\}})p_g(\mathcal{F}_{g-1})}{p_{g|\{g,u\}}(1-p_g(\mathcal{F}_{g-1}))}(C_t - m_{U,g,t}^C(\mathcal{F}_{g-1}))\middle|U=1\right] \\
&\qquad = \mathbb{E}\left[\frac{(1-p_{g|\{g,u\}})p_g(\mathcal{F}_{g-1})}{p_{g|\{g,u\}}(1-p_g(\mathcal{F}_{g-1}))}\underbrace{\left(\mathbb{E}[C_t|\mathcal{F}_{g-1},U=1]-m_{U,g,t}^C(\mathcal{F}_{g-1})\right)}_{=0}\middle|U=1\right] \\
&\qquad = 0
\end{aligned} \tag{SB.7}$$

where the first equality holds by the definition of $\tilde{\omega}_g^0$, the second equality holds by the law of iterated expectations, the third equality holds by the law of iterated expectations, and the last equality holds immediately from the previous one. Finally, consider

$$\begin{aligned}
&\mathbb{E}\left[\tilde{\omega}_g^0(G,\mathcal{F}_{g-1})\middle|\mathbf{1}\{G=g\}+U=1\right] \\
&\qquad = \mathbb{E}\left[\frac{p_g(\mathcal{F}_{g-1})U}{p_{g|\{g,u\}}(1-p_g(\mathcal{F}_{g-1}))}\middle|\mathbf{1}\{G=g\}+U=1\right] \\
&\qquad = \mathbb{E}\left[\frac{p_g(\mathcal{F}_{g-1})}{p_{g|\{g,u\}}(1-p_g(\mathcal{F}_{g-1}))}\underbrace{\mathbb{E}[U|\mathcal{F}_{g-1},\mathbf{1}\{G=g\}+U=1]}_{=1-p_g(\mathcal{F}_{g-1})}\middle|\mathbf{1}\{G=g\}+U=1\right] \\
&\qquad = \mathbb{E}\left[\frac{\mathbb{E}[\mathbf{1}\{G=g\}|\mathcal{F}_{g-1},\mathbf{1}\{G=g\}+U=1]}{p_{g|\{g,u\}}}\middle|\mathbf{1}\{G=g\}+U=1\right] \\
&\qquad = \mathbb{E}\left[\frac{\mathbf{1}\{G=g\}}{p_{g|\{g,u\}}}\middle|\mathbf{1}\{G=g\}+U=1\right] \\
&\qquad = 1
\end{aligned} \tag{SB.8}$$

where the first equality holds by the definition of $\tilde{\omega}_g^0$, the second equality holds by the law of iterated expectations, the third equality holds by canceling terms and by the definition of $p_g(\mathcal{F}_{g-1})$, the fourth equality holds by the law of iterated expectations, and the last equality holds by the definition of $p_{g|\{g,u\}}$. Combining Equation (SB.7) and Equation (SB.8), we have that

$$\mathbb{E}\left[\omega_g^0(G,\mathcal{F}_{g-1})(C_t - m_{U,g,t}^C(\mathcal{F}_{g-1}))\middle|\mathbf{1}\{G=g\}+U=1\right] = 0 \tag{SB.9}$$

Finally, subtracting Equation (SB.9) from Equation (SB.6), we have that

$$\begin{aligned}
ATT^C(g,t) &= \mathbb{E}\left[\left(\omega_g^1(G) - \omega_g^0(G,\mathcal{F}_{g-1})\right)(C_t - m_{U,g,t}^C(\mathcal{F}_{g-1}))\middle|\mathbf{1}\{G=g\}+U=1\right] \\
&= \mathbb{E}\left[\omega_g(G,\mathcal{F}_{g-1})(C_t - m_{U,g,t}^C(\mathcal{F}_{g-1}))\middle|\mathbf{1}\{G=g\}+U=1\right]
\end{aligned}$$

where the second equality holds by the definition of $\omega_g$. This completes the proof. □

# SC Extensions to Stochastic SIRD Model

In this section, we provide more details related to Remark 3 in the main text that concerns extensions to the Stochastic SIRD Model for Untreated Potential Outcomes. For simplicity, we focus on the case where the infection rate may vary across locations/time but suppose that the recovery rate $\lambda$ and the death rate $\gamma$ are constant across locations and time as in the main text. We also focus on whether or not Proposition 1 holds; given that this condition holds the same sorts of arguments in Theorem 2 will hold and can be used to recover $ATT_t^C$. We consider a slight variation of the Stochastic SIRD Model for Untreated Potential Outcomes:

$$\Delta C_{lt}(0) = \underbrace{\exp(\theta_{lt})}_{=\beta_{lt}} \frac{I_{lt-1}(0)}{N_l} S_{lt-1}(0)$$

where $\Delta C_{lt}(0)$ is the number new cases if the policy were not implemented. This generalizes the main case that we consider in the paper by letting $\beta_{lt}$ vary arbitrarily across locations and time (for some of the arguments below, it is useful to enforce that infection rates must be positive which is the reason that we introduce the exp term and $\theta_{lt}$ above). Without further restrictions, this is too general of a model to get any traction on because there are essentially location-time specific pandemics. The main case that we consider in the paper is closely related to the case when $\theta_{lt} = \theta + v_{lt}$ where, similarly to the case in the main text, we impose the following assumptions regarding $v_{lt}$:

$$\mathbb{E}[v_t | \mathcal{F}_{t-1}(0), \ldots, \mathcal{F}_1(0), D = d] = 1 \text{ for all } t = 2, \ldots, \mathcal{T}$$

and that

$$v_t \perp\!\!\!\perp (D, \mathcal{F}_{t-2}(0), \ldots, \mathcal{F}_1(0) | \mathcal{F}_{t-1}(0) \text{ for all } t = 2, \ldots, \mathcal{T}$$

The assumptions are analogous to Equations (A.6) and (A.7) in the main text with the only exception being that, in the main text, we consider the case where the unobservables are additive rather than multiplicative. That said, the same arguments for unconfoundedness as in the main text go through for either case up to minor adjustments related to multiplicative unobservables.

For the remainder of this section, we consider other leading choices for structure to put on $\theta_{lt}$.

*Case 1: $\theta_{lt} = \theta_t + v_{lt}$:* This setup allows for the infection rate to change over time even in the absence of the policy. Early in the pandemic, infection rates varying over time could occur due to general changes in the availability of masks or common sources of information about how Covid-19 is transmitted across people. Unconfoundedness holds in this setup using very similar arguments to those in the main text. In particular, Equation (B.2) can be replaced by $\mathcal{F}_{lt} = r_t(\mathcal{F}_{lt-1}, v_{lt})$ where

the difference here is that $r$ can vary over time. In this case, the results in Lemmas 1 and 2 go through essentially immediately (just adjusting to have $r_t$ instead of $r$ in the proofs). Given those modifications, the remainder of the proof goes through. This implies that the unconfoundedness strategy in the paper continues to be applicable in a setting where the infection rate varies over time with a component that is common across locations.

*Case 2:* $\theta_{lt} = \theta_t + \theta_l + v_{lt}$ This case allows for the infection rate to vary over time and to be systematically different across different locations. For the arguments below, we follow the large-$n$, small-$\mathcal{T}$ paradigm of treating $\theta_t$ as fixed parameters and $\theta_l$ as random. This is an empirically relevant case and would allow for the transmission rate to depend on unobserved location-specific characteristics (these could include things like a location's population density, a location's distribution of age and/or pre-existing health conditions, among other possibilities). This is a case, however, where the unconfoundedness strategy in the paper breaks down. In particular, notice that, we can write

$$\log(\Delta C_{lt}(0)) = \theta_t + \theta_l + v_{lt} + \log(I_{lt-1}(0))$$

where, for simplicity, we omit the term involving $S_{lt-1}(0)/N_l$ which is likely to be very close to 1 in early pandemic applications (the same issues discussed below still arise if we keep this term). Then, if we write down the same equation for $\log(\Delta C_{lt-1}(0))$ and subtract it, we get

$$\Delta \log(\Delta C_{lt}(0)) = \Delta\theta_t + \log\left(\frac{I_{lt-1}(0)}{I_{lt-2}(0)}\right) + \Delta v_{lt}$$

which, similar to many fixed effects-type arguments, has differenced out the location-specific fixed effect $\theta_l$. The expression in the previous equation involves the "state" of the pandemic for the two periods before treatment. However, even if $v_{lt}$'s are serially uncorrelated, like a dynamic panel model, there is endogeneity due to differencing out $\theta_l$, and this implies that the strategy of just conditioning on even two periods of the pre-treatment pandemic state would not work for unconfoundedness. That said, it seems like a viable alternative strategy would be to use ideas from the dynamic panel literature (e.g., using further lags of pandemic variables as IVs), to estimate the parameters of the pandemic model and then one could use that as the basis for estimating the $ATT$. As an interesting side-comment, motivated by relatively similar arguments, several papers have used this same transformation, $\Delta \log(\Delta C_{lt})$, as the outcome in two-way fixed effects event study regressions without accounting for potential endogeneity. In contrast to the discussion above, this line of argument holds in the case of a deterministic (rather than stochastic) SIRD model and then adding error terms after arriving at the above transformation (roughly, this leads to $v_{lt}$ in the previous equation rather than $\Delta v_{lt}$). This approach side-steps the endogeneity discussed above, but, arguably, starting from the stochastic SIRD model is more natural than adding error terms at the end. We do not provide a proof, but if one accepts the sort of arguments that avoid endogeneity here, then it appears that a version of unconfoundedness that conditions on two pre-treatment periods (rather than just one)

would go through in this case.

*Case 3:* $\theta_{lt} = \theta_t + h(X_l) + v_{lt}$*:* The motivation for this case is similar to previous one where there may be trends in the infection rate over time and where different locations may have systematically different infection rates. The main additional restriction here is that the systematic component of location-specific differences in the infection rate is driven by observed location-specific characteristics $X_l$ (see Chernozhukov, Kasahara, and Schrimpf (2021) for a related discussion of parameterizing the infection rate in this sort of way). For brevity, we do not provide the full proof here, but, heuristically, the arguments from Case 1 above apply here after conditioning on location-specific characteristics throughout the argument, and suggest that the unconfoundedness strategy considered in the paper is compatible with this setup as long as the conditioning variables in unconfoundedness include both $\mathcal{F}_{t^*-1}$ (the pre-treatment state of the pandemic) and $X_l$ (the characteristics of location $l$ that can affect the infection rate). This is also an empirically relevant case as, at least arguably, many of the most important characteristics of particular locations related to their infection rates (e.g., population density, etc.) are observable. Our arguments in the application are, at least implicitly, related to this case as we additionally condition on a location's population and region of the country (or for the county-level results compare counties in bordering states).

# SD Additional Results from Application

### Possible Issues using County-Level Data

In the main text, we briefly mentioned some possible issues that could arise with using county-level data relative to using state-level data; here, we expand on those comments. First, in the main text, we emphasized that, oftentimes, there existed more similar treated and untreated counties than treated and untreated states, both in terms of pre-treatment pandemic characteristics and other characteristics like population density. And, while good county-level data exists, one drawback of using county-level data is that the number of Covid-19 cases at the county level is generally small in the early part of the pandemic. For example, on April 1, 20% of counties in Arkansas and 40% of counties in Iowa had been recorded as having had exactly 0 Covid-19 cases. The unconfoundedness approach discussed in the paper could be violated if, among counties reporting 0 cases right before a SIPO was implemented, the first Covid-19 cases would have (absent the policy) tended to occur earlier or later for treated counties relative to untreated counties.

Another issue is that spatial correlations may be more important with county-level data because Covid-19 cases could spill over from one county to another; see Chandrasekhar, Goldsmith-Pinkham, Jackson, and Thau (2021), Oka, Wei, and Zhu (2021), and Bisin and Moro (2022) for more discussion along these lines. It would be an interesting extension to our approach to allow for spatial correlations and spillovers.

There are also potentially some issues related to conducting inference. One issue is that, because

we observe all counties in each state, it is somewhat unnatural to consider the counties as being draws from a larger population of counties. One alternative approach is to think about conditional inference where instead of considering the sampling procedure where outcomes, treatments, and covariates are draws from a large population, instead one considers the treatment status and covariates fixed but where the outcomes are sampled conditional on the treatment and covariates. Given existing work on conditional inference (see Abadie, Imbens, and Zheng (2014) in general and Borusyak, Jaravel, and Spiess (2021) as an example of conditional inference in the context of panel data and treatment effects), these suggest that, at least in some cases, our inference procedures discussed in Appendix SA may be conservative. Next, as discussed in the main text, our procedure that trims out some observations from the treated group can change the interpretation of the estimated parameter from an overall $ATT$ parameter to an $ATT$ that is local to locations where there are "similar" locations in the untreated state. Regarding inference, we view this trimming as a pre-processing step along the lines of Ho, Imai, King, and Stuart (2007) and Rubin (2008) because it does not involve concern the outcome. Finally, since the policies that we consider are at the state-level, a common choice in empirical research would be to cluster data at the state-level. In our setting, this is infeasible, especially for our results that only use two states. To get around this sort of issue, our approach is to carefully check for other common policies/shocks (and their timing) which are likely to be the main source of a "common shock" across counties in the same state. We generally find quite similar timing of other policies across bordering states (see Table 4 in the main text for more details), but differences in the mix/timing of policies across a pair of states would cast doubt on interpreting differences in outcomes across counties with similar pre-treatment characteristics as being (fully) due to a SIPO. See Section 5 of Roth, Sant'Anna, Bilinski, and Poe (2022) for a an interesting discussion of viewing "common shocks" as an identification issue rather than just an inference issue in a similar setting to the current one where the policy occurs at a more aggregate level than the unit of observation and there are very few aggregate units.

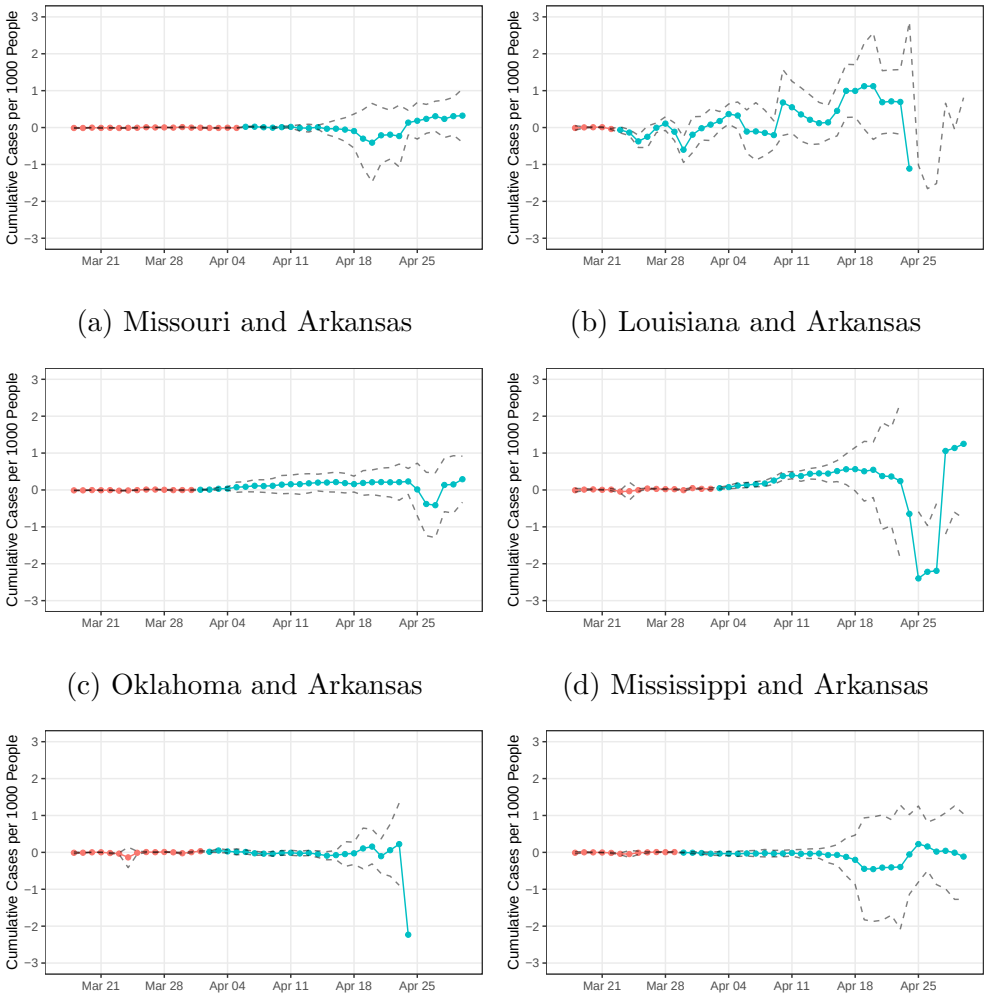
### Additional Details on Covid-Related Policies

Table 4, in the main text, provides the timing of five main Covid-related policies for the group of thirteen states considered in the part of the paper that uses county-level data. This section briefly provides some clarifications on particular policies, primarily regarding untreated states that did not implement a shelter-in-place order. See Fullman et al. (2021) and Weill, Stigler, Deschenes, and Springborn (2021) for substantially more details about the timing of policies and more local (e.g., city- or county-level policies) during the early part of the pandemic. Fullman et al. (2021) do not record any type of SIPO for Arkansas. On May 1, Iowa "strongly encourage[d] all vulnerable Iowans (i.e. those with pre-existing medical conditions and those older than 65) to limit activities outside of their home". On April 6, Nebraska's governor urged Nebraskans to stay at home. On March 17 and March 24, Oklahoma recommended that older and vulnerable people stay at home. South Dakota had a stay at home mandate for two counties that was primarily aimed at older and vulnerable

individuals and issued a statewide stay at home recommendation for vulnerable individuals on April 6. Finally, Fullman et al. (2021) classified both Tennessee and South Dakota's school closure policies as not being mandates (we listed both states as having school closures in Table 4 in the main text), but their notes indicate that it is debatable whether these policies should be considered as mandates, and it appears that schools in both states were closed for the remainder of the school year.

## Additional Figures

Figure 1: County-Level Estimates under Unconfoundedness for States Bordering Arkansas




(a) Missouri and Arkansas

(b) Louisiana and Arkansas

(c) Oklahoma and Arkansas

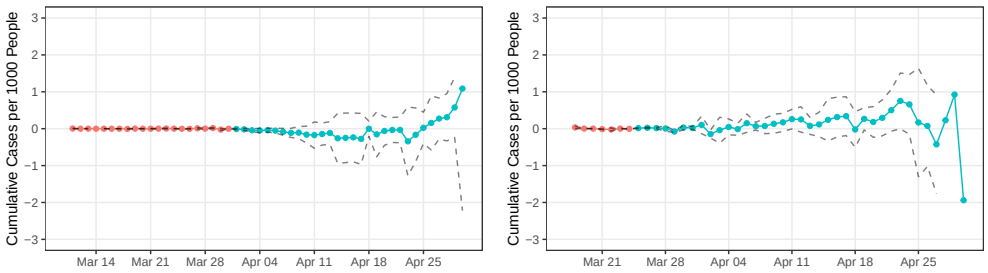
(d) Mississippi and Arkansas

(e) Tennessee and Arkansas

(f) Kansas and Arkansas

*Notes:* The figure provides estimates of SIPOs effects on Covid-19 cases for all states bordering Arkansas using the unconfoundedness approach discussed in the main text and using Arkansas as the comparison state. Of these states, Oklahoma did not actually implement a SIPO, and we use a placebo policy date for Oklahoma of April 1. Pre-treatment estimates are reported in red while post-treatment estimates are reported in blue. The pre-treatment estimates use the immediately preceding day as the base period while all the post-treatment periods use the period immediately before the treatment as the base period. The dashed lines provide 90% pointwise confidence intervals. Finally, the vertical axis is the same as for the corresponding figures in the main text and constant across panels in the figure; in some cases, this results in the confidence intervals falling outside of the displayed values.

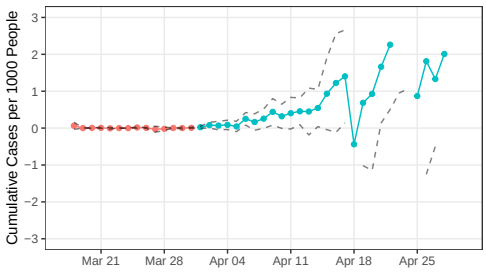
Figure 2: County-Level Estimates under Unconfoundedness for States Bordering Iowa



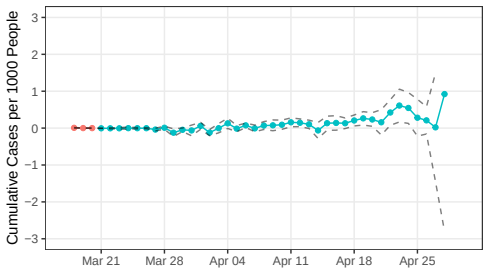

(a) South Dakota and Iowa



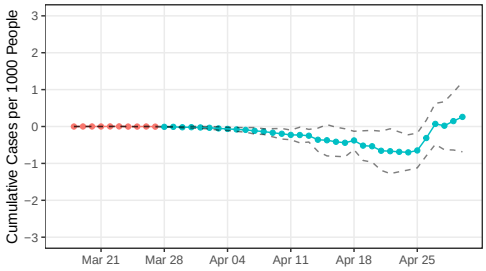

(b) Wisconsin and Iowa



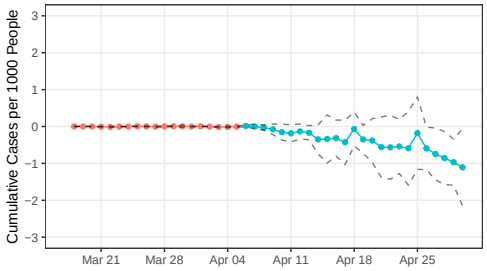

(c) Nebraska and Iowa



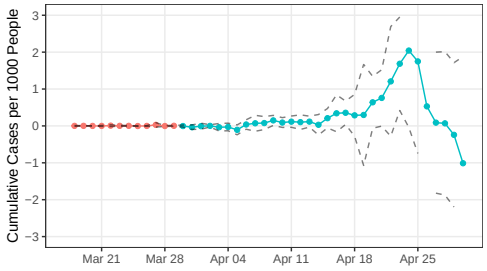

(d) Illinois and Iowa




(e) Minnesota and Iowa



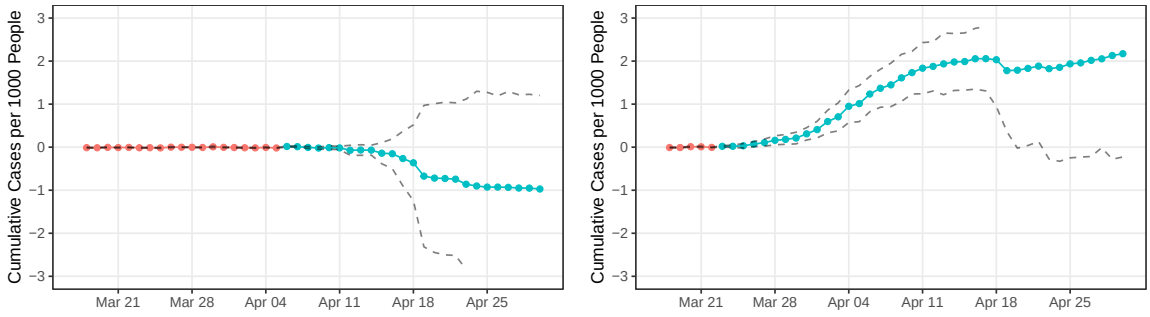

(f) Missouri and Iowa




(g) Kansas and Iowa

*Notes:* The figure provides estimates of SIPOs effects on Covid-19 cases for all states bordering Iowa using the unconfoundedness approach discussed in the main text and using Iowa as the comparison state. Of these states, Nebraska and South Dakota did not actually implement a SIPO, and we use a placebo policy date for these states of April 1. Pre-treatment estimates are reported in red while post-treatment estimates are reported in blue. The pre-treatment estimates use the immediately preceding day as the base period while all the post-treatment periods use the period immediately before the treatment as the base period. The dashed lines provide 90% pointwise confidence intervals. Finally, the vertical axis is the same as for the corresponding figures in the main text and constant across panels in the figure; in some cases, this results in the confidence intervals falling outside of the displayed values.

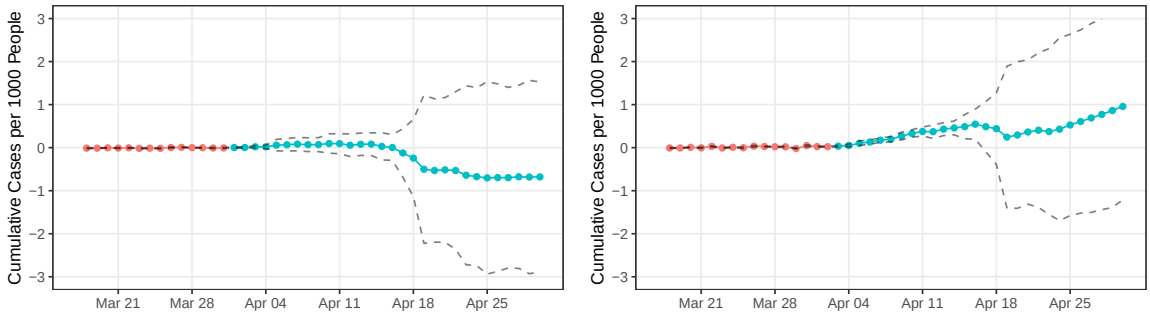
Figure 3: County-Level Estimates under Difference-in-Differences for States Bordering Arkansas




(a) Missouri and Arkansas



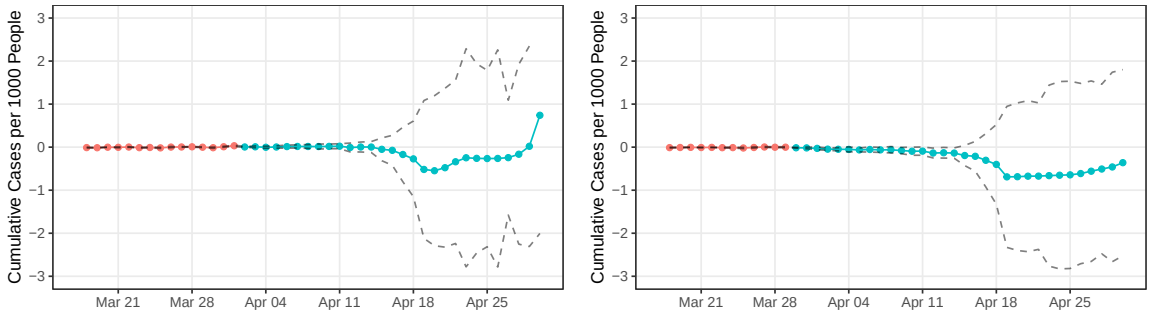

(b) Louisiana and Arkansas




(c) Oklahoma and Arkansas




(d) Mississippi and Arkansas



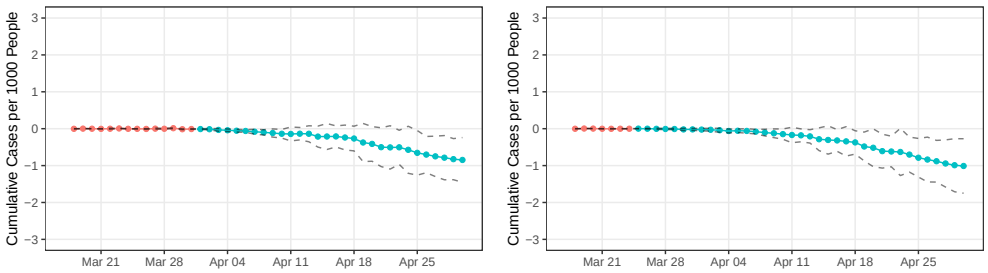

(e) Tennessee and Arkansas




(f) Kansas and Arkansas

*Notes:* The figure provides estimates of SIPOs effects on Covid-19 cases for all states bordering Arkansas using difference-in-differences and using Arkansas as the comparison state. Of these states, Oklahoma did not actually implement a SIPO, and we use a placebo policy date for Oklahoma of April 1. Pre-treatment estimates are reported in red while post-treatment estimates are reported in blue. The pre-treatment estimates use the immediately preceding day as the base period while all the post-treatment periods use the period immediately before the treatment as the base period. The dashed lines provide 90% pointwise confidence intervals. Finally, the vertical axis is the same as for the corresponding figures in the main text and constant across panels in the figure; in some cases, this results in the confidence intervals falling outside of the displayed values.

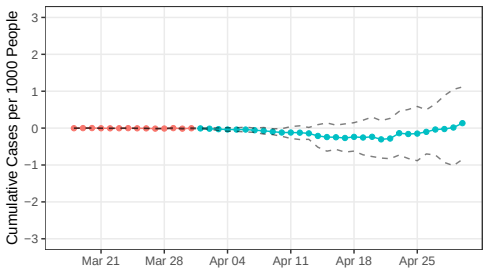
Figure 4: County-Level Estimates under Difference-in-Differences for States Bordering Iowa



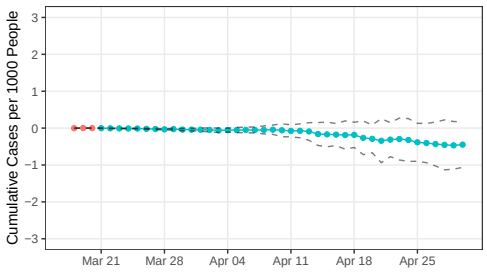

(a) South Dakota and Iowa



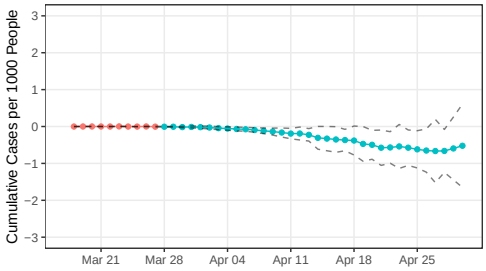

(b) Wisconsin and Iowa



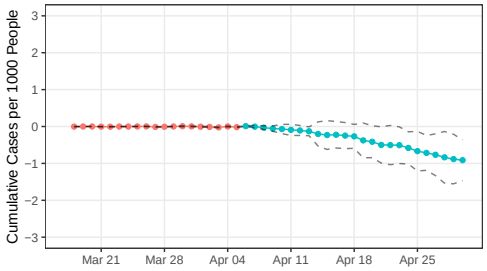

(c) Nebraska and Iowa



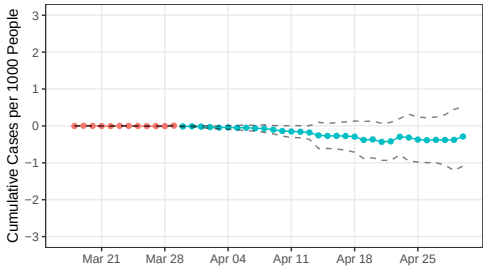

(d) Illinois and Iowa




(e) Minnesota and Iowa



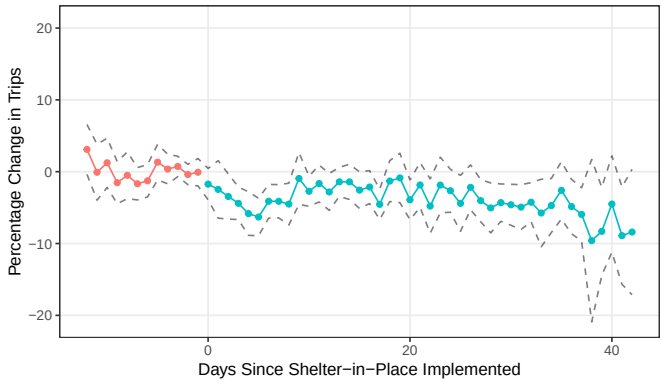

(f) Missouri and Iowa



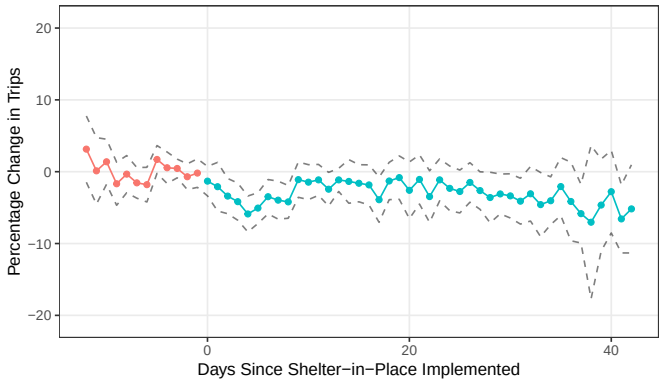

(g) Kansas and Iowa

*Notes:* The figure provides estimates of SIPOs effects on Covid-19 cases for all states bordering Iowa using the unconfoundedness approach discussed in the main text and using Iowa as the comparison state. Of these states, Nebraska and South Dakota did not actually implement a SIPO, and we use a placebo policy date for these states of April 1. Pre-treatment estimates are reported in red while post-treatment estimates are reported in blue. The pre-treatment estimates use the immediately preceding day as the base period while all the post-treatment periods use the period immediately before the treatment as the base period. The dashed lines provide 90% pointwise confidence intervals. Finally, the vertical axis is the same as for the corresponding figures in the main text and constant across panels in the figure; in some cases, this results in the confidence intervals falling outside of the displayed values.

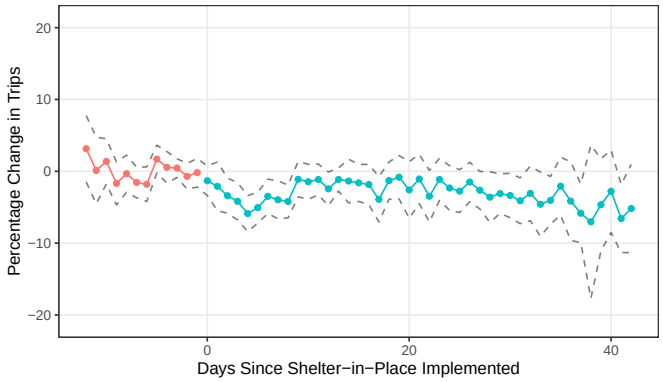
Figure 5: Estimates of SIPO Orders on Travel using State-Level Data




(a) Standard DID




(b) Regression DID




(c) Adjusted Regression DID

*Notes:* The figure contains event study type estimates of the effect of SIPOs on the percentage change in retail and recreation travel using the state-level data discussed in the main text. $e = 0$ corresponds to the time period when the policy was implemented. Negative values of $e$ correspond to pre-treatment estimates of the effect of the policy and can be thought of as pre-tests, and positive values of $e$ correspond to estimates of the effect of the policy at different lengths of exposure to the treatment. Panel (a) provides estimates using standard DID (without accounting for cases), Panel (b) provides regression DID estimates (accounting for cases but not that the policy may have a direct effect on cases), and Panel (c) provides adjusted regression DID estimates (accounting for cases and allowing for the policy to have had an effect on cases as is proposed in the text). The dashed line provides 90% pointwise confidence intervals.

# References


[1] Abadie, Alberto, Guido W Imbens, and Fanyin Zheng. "Inference for misspecified models with fixed regressors". *Journal of the American Statistical Association* 109.508 (2014), pp. 1601–1614.

[2] Bisin, Alberto and Andrea Moro. "Learning epidemiology by doing: The empirical implications of a Spatial-SIR model with behavioral responses". *Journal of Urban Economics* 127 (2022), p. 103368.

[3] Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. "Revisiting event study designs: Robust and efficient estimation". Working Paper. 2021.

[4] Callaway, Brantly and Pedro HC Sant'Anna. "Difference-in-differences with multiple time periods". *Journal of Econometrics* 225.2 (2021), pp. 200–230.

[5] Chandrasekhar, Arun G, Paul Goldsmith-Pinkham, Matthew O Jackson, and Samuel Thau. "Interacting regional policies in containing a disease". *Proceedings of the National Academy of Sciences* 118.19 (2021), e2021520118.

[6] Chernozhukov, Victor, Hiroyuki Kasahara, and Paul Schrimpf. "Causal impact of masks, policies, behavior on early Covid-19 pandemic in the US". *Journal of Econometrics* 220.1 (2021), pp. 23–62.

[7] Fullman, Nancy et al. *State-level social distancing policies in response to COVID-19 in the US*. Version 1.145, http://www.covid19statepolicy.org/. Sept. 2021.

[8] Ho, Daniel E, Kosuke Imai, Gary King, and Elizabeth A Stuart. "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference". *Political analysis* 15.3 (2007), pp. 199–236.

[9] Oka, Tatsushi, Wei Wei, and Dan Zhu. "The effect of human mobility restrictions on the COVID-19 transmission network in China". *PloS one* 16.7 (2021), e0254403.

[10] Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe. "What's trending in difference-in-differences? A synthesis of the recent econometrics literature". Working Paper. 2022.

[11] Rubin, Donald B. "For objective causal inference, design trumps analysis". *The Annals of Applied Statistics* 2.3 (2008), pp. 808–840.

[12] Sant'Anna, Pedro HC and Jun Zhao. "Doubly robust difference-in-differences estimators". *Journal of Econometrics* 219.1 (2020), pp. 101–122.

[13] Sun, Liyang and Sarah Abraham. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects". *Journal of Econometrics* 225.2 (2021), pp. 175–199.

[14] Weill, Joakim A, Matthieu Stigler, Olivier Deschenes, and Michael R Springborn. "Researchers' degrees-of-flexibility and the credibility of difference-in-differences estimates: Evidence from the pandemic policy evaluations". Working Paper. 2021.