Treatment Effects without a Control Group (Preliminary and Incomplete)

Botosaru, Giacomini, Weidner

February 11, 2022

Abstract

We propose a method for estimating the effect of a program or policy when all individuals in a population are treated. We show how individual pre-treatment information - even from very short panels - can be exploited to forecast individual counterfactuals, which can then be used to estimate the average treatment effect. We propose a simple estimator based on local polynomial regressions, which does not require correct specification of the individual forecast model or a long pre-treatment history. Our first contribution is to show that this estimator is unbiased and asymptotically normal for a broad class of data-generating processes (DGPs) that express the individual potential outcomes as the sum of (possibly) three unobserved components: a stationary process, a unit root process, and a polynomial time trend. Simulation results suggest that the choice of a larger polynomial order could mitigate the bias due to a "non-stationary" initial condition in short panels.

1 Introduction

The common approach to estimating average treatment effects (ATT) assumes the availability of a control group of untreated individuals. This paper proposes an alternative way to estimate ATT using panel data that does not require the existence of a control group. We propose an estimator of the ATT - the Forecasted Treatment Effects (FAT) - that is based on forecasting individual counterfactuals using pre-treatment data. The main insight of the paper is that a simple forecast that fits a polynomial time trend to the individual pre-treatment time series can deliver an unbiased estimate of the ATT with good finite-sample performance. The only requirement is that the chosen order of

the polynomial is larger than the true one. We show analytically that this result holds for a broad class of data-generating processes that express the individual outcomes as the sum of (possibly) three unobserved components: a stationary process, a unit root process and a polynomial time trend. We show in simulations that choosing a larger polynomial order could also mitigate the bias due to a "non-stationary" initial condition in a short time series.

The findings in this paper suggest that for unbiased estimation of the ATT it is sufficient to focus on forecasting the deterministic component of the individual time series, while the stochastic component could be left unspecified. This claim may seem counterintuitive from a forecasting perspective, as it suggests that even a misspecified forecasting model can deliver an unbiased estimate of the ATT. It also signals a shift in focus relative to the dynamic panel data literature: rather than trying to obtain unbiased estimates of the model's parameters, we want unbiased forecasts of counterfactuals. To this extent, our conclusion is that it seems important to avoid misspecifying the deterministic component, while correctly specifying the stochastic component (and using an unbiased estimator for its parameters) is not crucial and may actually worsen the finite-sample performance.

Our baseline estimator uses individual time series of pre-treatment outcomes to forecast the counterfactual outcome for each individual. The difference between the observed post-treatment outcome and the forecasted outcome estimates the individual treatment effect, while the sample average of individual effects estimates the ATT. Importantly, we do not require specifying a forecasting model but allow for a broad class of data-generating processes for the individual counterfactuals that potentially allow for a large degree of individual heterogeneity, including fixed effects, lagged outcomes with heterogeneous coefficients, and differential treatment times, even when the panel is short.

We then consider several extension of our method, including the presence of a control group and staggered adoption. Our method is particularly relevant for situations with a limited number of pre-treatment periods, or with periods in which all or most units are treated (for which two-way fixed effects estimation obtains negative weight on the treatment effects in later periods for early-adopters).

2 Related Literature

This paper is related to several strands of the literature on treatment effect estimation, forecasting and panel data econometrics. There are some examples in the literature of approaches to estimating treatment effects without assuming the existence of a control group. Some use Bayesian methods under strong parametric assumptions (e.g., Brodersen et al. (2015); Varian (2014)) and some (the so-called "interrupted time series" approach, e.g., Bernal et al. (2017), Baiker and Svoronos (2019), Miratrix (2020)) make strong assumptions such as a single treated unit and no variation in treatment timing. Unlike these approaches, our method is valid under more general assumptions on the data-generating process and allows for multiple treated units and staggered adoption.

Another approach to estimating treatment effects without a control group is the so-called "regression discontinuity in time" that is often applied in empirical environmental economics. This approach requires high-frequency data around the treatment time and it involves local estimation before and after the treatment. Our approach has a different theoretical justification and does not require high-frequency data. We also use local estimation but only on data before the treatment in order to forecast counterfactuals.

Our estimation of treatment effects is based on forecasting individual counterfactuals using panel data. There is a large literature on forecasting with panel data (for recent examples, see the Empirical Bayes approach of Liu, Moon and Schorfheide, 2020). This literature makes strong parametric assumptions and focuses on forecast accuracy. In contrast, we allow for a broad class of DGPs and we focus on forecast unbiasedness, rather than accuracy.

The focus on unbiasedness relates our work to the classical literature on unbiased forecasts in time series (e.g., Fuller and Hasza, 1980, Dufour, 1984) and to more recent contributions in panel data such as Mavroeidis et. al. (2015). Unlike our approach, this literature considers more restrictive DGPs, for example assuming stationarity and symmetric error distributions.

Finally, there has been a lively recent discussion in the literature on how the standard approach to estimating average treatment effects using Ordinary Least Squares or Two-way Fixed Effects regressions in panel data is generally inconsistent under treatment effect heterogeneity (e.g., Wooldridge (2005); Chernozhukov et al. (2013); Imai and Kim (2019); Sloczynski (2020); de Chaisemartin and D'Haultfoeuille (2020); Goodman-Bacon (2021). All proposed solutions assume the existence of a control group in every

period, see, e.g., Cengiz et al. (2019); Callaway and Sant'Anna (2020); Sun and Abraham (2020); Goodman-Bacon (2021); Baker et al. (2021); Borusyak et al. (2021); Liu et al. (2021). This means that there is presently no standard difference-in-difference consistent estimator for the case of staggered, heterogeneous treatment. In contrast, our solution allows for heterogeneous treatment effects, staggered adoption and does not require the existence of a control group.

3 Forecasted treatment effects: the baseline case

3.1 Setup and notation

Suppose we observe outcomes y_{it} for individuals $i = 1, \ldots, n$ and time periods $t = 1, \ldots, T$. All individuals in the population are untreated for time periods $t \leq \tau$, and treated for $t > \tau$. Each individual has two potential outcomes at time t: $y_{it}(0)$ if the individual is not treated and $y_{it}(1)$ if the individual is treated. Under SUTVA, the observed outcomes are then given by

$$y_{it} = (1 - d_{it})\, y_{it}(0) + d_{it}\, y_{it}(1), \qquad\qquad d_{it} = \mathbb{1}(t > \tau).$$

We consider more general situations later, in particular, for staggered adoption where $d_{it} = \mathbb{1}(t > \tau_i)$, with τ_i heterogeneous across i.

The parameter of interest is the average treatment effect on the treated $h > 0$ periods after the treatment:

$$\mathrm{ATT}_h := \mathbb{E}\left[y_{i\tau+h}(1) - y_{i\tau+h}(0)\right] = \mathbb{E}\left[y_{i\tau+h} - y_{i\tau+h}(0)\right], \tag{1}$$

where the expectation is taken with respect to the cross sectional distribution.

3.2 Proposed estimator of average treatment effects: the FAT

The challenge in identifying and estimating (1) is that the counterfactual $y_{i\tau+h}(0)$ is not observed for $h > 0$. If a control group is available, the usual strategy is to impose sufficient assumptions that allow to identify $\mathbb{E}[y_{i\tau+h}(0)]$ from the observed outcomes of the control group. In the absence of a control group, we exploit pre-treatment data to obtain a forecast for the counterfactual $y_{i\tau+h}(0)$. We denote this forecast by $\widehat{y}_{i\tau+h}(0)$.

Our proposed estimator of ATT_h, which we call the forecasted average treatment

effect estimator (FAT_h), is given by

$$\text{FAT}_h = \frac{1}{n}\sum_{i=1}^{n}\left[y_{i\tau+h} - \widehat{y}_{i\tau+h}\left(0\right)\right], \tag{2}$$

where $\widehat{y}_{i\tau+h}(0)$ is a measurable function of the information set $\{y_{it}, i = 1, ..., n; t = 1, ..., \tau\}$.

3.3 The unbiasedness condition

Assumption 1 (**Baseline model, cross-sectional sampling**). *The potential outcomes* $(y_{it}(0), y_{it}(1) : t = 1, \ldots, T)$ *are independent and identicially distributed across* $i = 1, \ldots, n$.

A key insight of the paper is that consistency of FAT_h for ATT_h is guaranteed by a cross-sectional weak law of large numbers, as long as the forecasts satisfy the following unbiasedness condition,

$$\mathbb{E}\left[\widehat{y}_{i\tau+h}(0) - y_{i\tau+h}\left(0\right)\right] = 0. \tag{3}$$

Lemma 1. *Let Assumption 1 hold. For each* $i = 1, \ldots, n$, *let the forecast* $\widehat{y}_{i\tau+h}(0)$ *be a function of* $(y_{i1}, \ldots, y_{i\tau})$ *such that* (3) *holds. Assume furthermore that* $V = \text{Var}(y_{i\tau+h}(1) - \widehat{y}_{i\tau+h}\left(0\right)$ *is finite. Then we have*

$$\sqrt{n}\left(\text{FAT}_h - \text{ATT}_h\right) \Rightarrow \mathcal{N}\left(0, V\right).$$

Proof. Let $u_i := y_{i\tau+h}(1) - \widehat{y}_{i\tau+h}\left(0\right)$. We have

$$\begin{aligned}
\text{FAT}_h - \text{ATT}_h &= \frac{1}{n}\sum_{i=1}^{n}\left[y_{i\tau+h}(1) - \widehat{y}_{i\tau+h}\left(0\right)\right] - \mathbb{E}\left[y_{i\tau+h}(1) - y_{i\tau+h}\left(0\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[y_{i\tau+h}(1) - \widehat{y}_{i\tau+h}\left(0\right)\right] - \mathbb{E}\left[y_{i\tau+h}(1) - \widehat{y}_{i\tau+h}\left(0\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(u_i - \mathbb{E}u_i\right).
\end{aligned}$$

Our assumptions guarantee that the $u_i - \mathbb{E}u_i$ have zero mean and finite variance and are i.i.d. across i. The Lindeberg-Levy CLT therefore gives the desired result. □

3.4 Local polynomial regression

We construct forecasts of counterfactuals via individual-specific local polynomial regressions using pre-treatment data. For individual i, let $q_i \in \{0, 1, 2, \ldots, \tau - 1\}$ be the maximal order of the polynomial time trend, and let $R_i \in \{q+1, \ldots, \tau\}$ be the number of pre-treatment time periods used for the regression. The h-period ahead forecast of the counterfactual is given by

$$\widehat{y}_{i,\tau+1}^{(q_i,R_i)} := \sum_{k=0}^{q_i} \widehat{\alpha}_k^{(i,q_i,R_i)} \, (\tau+1)^k, \qquad \widehat{\alpha}^{(i,q_i,R_i)} := \underset{\alpha \in \mathbb{R}^{q_i+1}}{\operatorname{argmin}} \sum_{t \in \mathcal{T}_i} \left(y_{it} - \sum_{k=0}^{q_i} \alpha_k \, t^k \right)^2, \tag{4}$$

where $\alpha = (\alpha_0, \ldots, \alpha_{q_i})$ is a $q_i + 1$ vector, $\mathcal{T}_i = \{\tau - R_i + 1, \ldots, \tau\}$ is the set of the R_i time periods directly preceding the treatment date.

Remark 1. *An interesting special case is $R_i = q_i + 1$ and $h = 1$. In that case, $\widehat{y}_{i\tau+1}^{(q_i)}(0) := \widehat{y}_{i\tau+1}^{(q_i,q_i+1)}(0)$ can equivalently be defined iteratively by*

$$\widehat{y}_{i\tau+1}^{(q_i)}(0) = \begin{cases} y_{i\tau} & \text{for } q_i = 0, \\ \widehat{y}_{i\tau+1}^{(q_i-1)}(0) - \left[\widehat{y}_{i\tau}^{(q_i-1)}(0) - y_{i\tau} \right] & \text{for } q_i > 0. \end{cases} \tag{5}$$

This iteration is quite intuitive: the order-q_i forecast is formed by subtracting the lagged forecast error $\widehat{y}_{i\tau}^{(q_i-1)}(0) - y_{i\tau}$ from the forecast $\widehat{y}_{i\tau+1}^{(q_i-1)}(0)$ of order $q_i - 1$. Explicit formulas for this case are given by

$$\begin{aligned} \widehat{y}_{i\tau+1}^{(0)}(0) &= y_{i\tau}, \\ \widehat{y}_{i\tau+1}^{(1)}(0) &= 2y_{i\tau} - y_{i\tau-1}, \\ \widehat{y}_{i\tau+1}^{(2)}(0) &= 3y_{i\tau} - 3y_{i\tau-1} + y_{i\tau-2}, \\ \widehat{y}_{i\tau+1}^{(q_i)}(0) &= \sum_{t=\tau-q_i}^{\tau} w_t^{(q_i,1)} \, y_{it}, \qquad\qquad w_t^{(q_i,1)} = (-1)^{(\tau-t)} \binom{q_i+1}{\tau-t+1}, \end{aligned}$$

where $\binom{a}{b} = \frac{a!}{b!(a-b)!}$ is the binomial coefficient.[1]

Remark 2. *The first alternative way to obtain the same estimate in that case is via the cross-sectional averages $\overline{y}_t = \frac{1}{n} \sum_{i=1}^n y_{it}$ of the observed outcomes in time period t.*

[1] Laderman and Laderman (1982) derive a similar expression in the context of forecasting a time series by polynomial regression using the entire available time series.

Due to linearity of the forecasting procedure we can rewrite FAT_h *as*

$$\mathrm{FAT}_h = \overline{y}_{\tau+h} - \sum_{k=0}^{q} \overline{\alpha}_k \, (\tau+h)^k, \qquad \overline{\alpha} := \underset{\alpha \in \mathbb{R}^{q+1}}{\operatorname{argmin}} \sum_{t \in \mathcal{T}} \left(\overline{y}_t - \sum_{k=0}^{q} \alpha_k \, t^k \right)^2, \tag{6}$$

where $\mathcal{T} = \{\tau - R + 1, \ldots, \tau\}$*, and we suppress the dependence on* τ*,* q*,* R*. Here, the cross-sectional averages for* $t \leq \tau$ *are used to obtain a forecast for* $t = \tau + h$*, which is then subtracted from the cross-sectional average observed at that time period.*

The second alternative way to obtain FAT in the case $R = R_i$ *and* $q = q_i$ *is as a simple pooled regression estimator, namely we have* $\mathrm{FAT}_h = \widehat{\beta}_h$*, where*

$$\left(\widehat{\beta}, \widehat{\alpha}\right) = \underset{\left\{\beta \in \mathbb{R}^h, \, \alpha \in \mathbb{R}^{n \times (q+1)}\right\}}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{t=\tau-R+1}^{\tau+h} \left(y_{it} - \sum_{k=1}^{h} \mathbb{1}\{t = \tau + k\} \, \beta_k - \sum_{k=0}^{q} \alpha_{ik} \, t^k \right)^2, \tag{7}$$

which is the OLS estimator obtained from regressing y_{it} *on a set of time dummies* $\mathbb{1}(t = \tau + k)$*, for* $k \in \{1, \ldots, h\}$*, and individual specific time trends.*[2]

The alternative estimation strategies in (6) *and* (7) *provide algebraically identical treatment effect estimates in the case of our baseline setting with* $R = R_i$ *and* $q = q_i$*. In a more general setting, however, those alternative estimation strategies do not give the same treatment effect estimator, and may indeed easily give inconsistent estimates for ATT if applied incorrectly.*

Assumption 2 (**Baseline model, time series process**)**.** *The potential outcome in the absence of the treatment follows the process:*

$$y_{it}(0) = y_{it}^{(1)}(0) + y_{it}^{(2)}(0) + y_{it}^{(3)}(0), \tag{8}$$

where $y_{it}^{(1)}(0)$ *is a mean stationary process,* $y_{it}^{(2)}(0) = y_{it-1}^{(2)}(0) + u_{it}(0)$ *is a unit-root process with innovations satisfying* $\mathbb{E}u_{it}(0) = 0$*, for all* $t \geq 2$*, and* $y_{it}^{(3)}(0) = \sum_{k=1}^{q_0} \alpha_{ik}^{(3)} \, t^k$ *is a polynomial trend of order* $q_0 \in \{0, 1, 2, \ldots\}$ *and coefficients* $\alpha_{ik}^{(3)} \in \mathbb{R}$*.*

Theorem 1. *Let Assumption 2 hold and let* $q_i \in \{q_0, \ldots, \tau - 1\}$*,* $R_i \in \{q_i + 1, \ldots, \tau\}$*, and* $h \in \{1, 2, \ldots\}$*. Then,*

$$\mathbb{E}\left[\widehat{y}_{i\tau+h}^{(q_i, R_i)}(0) - y_{i\tau+h}(0) \right] = 0.$$

[2]It actually does not matter for $\widehat{\beta}$ here whether we make the coefficients α on the time trend individual specific or not.

Proof. Since the forecast $\widehat{y}_{i\tau+h}^{(q_i,R_i)}(0)$ can be written as a linear combination of past outcomes, and given (8), it is sufficient to show that for each component $y_{it}^{(r)}(0)$, $r \in \{1,2,3\}$,

$$\mathbb{E}\left[\sum_{t\in\mathcal{T}_i} w_t^{(q_i,R_i,h)}\, y_{it}^{(r)} - y_{i\tau+h}^{(r)}(0)\right] = 0. \tag{9}$$

For both the mean stationary component ($r=1$) and the unit-root component ($r=2$) we have $\mathbb{E}\left(y_{it}^{(r)} - y_{i\tau+h}^{(r)}(0)\right) = 0$. Multiplying this equation by $w_t^{(q_i,R_i,h)}$, summing over $t \in \mathcal{T}_i$, and using that the non-random weights $w_t^{(q_i,R_i,h)}$ satisfy $\sum_{t\in\mathcal{T}_i} w_t^{(q_i,R_i,h)} = 1$, we obtain (9) for $r=1$ and $r=2$. To show (9) for the polynomial trend ($r=3$), note that by (4), $\sum_{t\in\mathcal{T}_i} w_t^{(q_i,R_i,h)}\, y_{it}^{(3)} = \sum_{k=0}^{q_i} \widetilde{\alpha}_k^{(i,\tau,q_i,R_i)}\,(\tau+h)^k$, where

$$\widetilde{\alpha}^{(i,\tau,q_i,R_i)} := \underset{\alpha\in\mathbb{R}^{q_i+1}}{\operatorname{argmin}} \sum_{t\in\mathcal{T}_i}\left(y_{it}^{(3)} - \sum_{k=0}^{q_i}\alpha_k\, t^k\right)^2.$$

Since $q_i \geq q_0$ for all i, the objective function in the last display is minimized (with value zero) at $\widetilde{\alpha}^{(i,\tau,q,R_i)} = \alpha_{ik}^{(3)}$, which implies $y_{i\tau+h}^{(3)}(0) = \sum_{t\in\mathcal{T}_i} w_t^{(q,R_i,h)}\, y_{it}^{(3)}$, that is, (9) holds for $r=3$ even without taking the expectations . □

Corollary 1. *Let Assumption 1 and the assumption of Theorem 1 hold. Let* $\mathrm{Var}(y_{it})$ *be finite for all* $t \in \{1,\ldots,T\}$*. Then we have*

$$\sqrt{n}\left(\mathrm{FAT}_h^{(q_i,R_i)} - \mathrm{ATT}_h\right) \Rightarrow \mathcal{N}(0,V),$$

with $V = \mathrm{Var}(\widehat{y}_{i\tau+h}^{(q_i,R_i)}(0) - \widehat{y}_{i\tau+h}(0))$.

4 Generalizations

In this section we discuss how to generalize our procedure to allow for modeling the stochastic component, the presence of a control group, covariates, limited anticipation, staggered adoption.

Our estimand, which we call Forecasted Average Treatment Effect, is then

$$FAT_h = \mathbb{E}\left(y_{i\tau+h} - \widehat{y}_{i\tau+h}(0)\right)$$

A sufficient condition for $FAT_h = ATT_h$ is that the forecast be unbiased, i.e.

$$\mathbb{E}\left(y_{i\tau+h}\left(0\right)\right) = \mathbb{E}\left(\widehat{y}_{i\tau+h}(0)\right)) \tag{10}$$

In this section we propose an estimator for the forecast that satisfies (10). Additionally, we describe the class of data generating processes for $y_{it}(0)$ for which this forecast is unbiased.

4.1 Modeling stochastic component

4.2 Control group

The presence of a control group makes it possible to consider a broader class of DGPs for the individual counterfactuals. In particular, it allows for common shocks.

Suppose that all individuals are untreated before the implementation of the treatment at time τ but that some individuals remain untreated after τ. Let $D_i = 1$ if individual i is untreated. The outcome variable is then given by:

$$y_{it} = D_i\left[1\left(t \leq \tau\right) y_{it}\left(0\right) + 1\left(t > \tau\right) y_{it}\left(1\right)\right] + \left(1 - D_i\right) y_{it}\left(0\right)$$

The parameter of interest is

$$\begin{aligned} ATT_h &= \mathbb{E}\left(y_{i\tau+h}\left(1\right) - y_{i\tau+h}\left(0\right)\middle|\, D_i = 1\right) \\ &= \mathbb{E}\left(y_{i\tau+h} - y_{i\tau+h}\left(0\right)\middle|\, D_i = 1\right) \end{aligned}$$

and our estimand is

$$FAT_h = \mathbb{E}\left(y_{i\tau+h} - y_{i\tau+h}^{(R_i)}\left(0\right)\middle|\, D_i = 1\right)$$

where $y_{i\tau+h}^{(R_i)}\left(0\right)$ is the forecast of $y_{i\tau+h}\left(0\right)$ given R_i outcomes $\{y_{it}\}_{t=\tau-R_i+1}^{\tau}$.

In Section 3, we showed that a sufficient condition for $FAT_h = ATT_h$ is that the forecast be unbiased, i.e.

$$\mathbb{E}\left(y_{i\tau+h}\left(0\right)\middle|\, D_i = 1\right) = \mathbb{E}\left(y_{i\tau+h}^{(R_i)}\left(0\right)\middle|\, D_i = 1\right). \tag{11}$$

The assumptions in Section 3 excluded processes subject to common shocks, such as shocks that affect all individuals in the same way before the treatment. The presence

of a control group allows us to weaken these assumptions and consider a larger class of processes. To see this, consider that the potential outcome $y_{it}(0)$ is subject to a shock at a *known* time $j \in (1, \tau)$, i.e.

$$y_{it}(0) = \gamma 1(t \geq j) + Y_{it}(0), Y_{it}(0) \in \mathcal{Y}_t.$$

The counterfactual outcome at $\tau + h$ satisfies:

$$\mathbb{E}(y_{i\tau+h}(0) | D_i = 1) = \mathbb{E}(\gamma | D_i = 1) + \mathbb{E}\left(y_{i\tau+h}^{(R_i)}(0) \middle| D_i = 1\right).$$

Should there be an untreated group subject to the same shock at time j, i.e.

$$\mathbb{E}(\gamma | D_i = 1) = \mathbb{E}(\gamma | D_i = 0),$$

the shock can be estimated from the control group as

$$\mathbb{E}(\gamma | D_i = 0) = \mathbb{E}(y_{it} 1(t \geq j) | D_i = 0) - \mathbb{E}(y_{it} 1(t < j) | D_i = 0).$$

Remark 3. *Notice that, in fact, one can recover* $\mathbb{E}(\gamma | D_i = 1)$ *from another time series for the treated group that is different from the outcome of interest, is unaffected by the treatment, and is subject to the same shock before the implementation of the treatment. Call this time series* z_{it}*. This assumption is not testable, however suggestive evidence for it can be found by plotting the aggregate time series* $\frac{1}{n}\sum_{i=1}^{n} z_{it}$ *against* $\frac{1}{n}\sum_{i=1}^{n} y_{it}$ *for* $t \in \{1, \ldots \tau\}$ *and checking that the series move together except at time* j*.*

Remark 4. *When the shock happens between* τ *and* $\tau + h$*,* $FAT_h^{treated} - FAT_h^{control}$ *estimates the effect of the treatment without the shock.*

4.3 Staggered adoption

Our approach naturally lends itself to the possible presence of treatment anticipation, as long as it is limited. One simply modifies the pre-treatment estimation window R_i to include past observations only up to the time $\tau - \delta$ at which it is still reasonable to assume that there was no treatment anticipation, so that $R_i \in \{q + 1, \ldots, \tau - \delta\}$.

5 Simulation Study

6 Empirical illustrations

References

Baker, A., D. F. Larcker, and C. C. Y. Wang (2021). How much should we trust staggered difference-in-differences estimates? Technical report. doi:10.2139/ssrn.3794018.

Bernal, J. L., S. Cummins, and A. Gasparrini (2017). Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International Journal of Epidemiology*, 348–355.

Borusyak, K., X. Jaravel, and J. Spiess (2021). Revisiting event study designs: Robust and efficient estimation. Technical report.

Brodersen, K., F. Gallusser, J. Koehler, N. Remy, and S. L. Scott (2015). Inferring causal impact using bayesian structural time-series models. *Annals of Applied Statistics 1*(9), 247–274.

Callaway, B. and P. H. C. Sant'Anna (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*. doi:10.1016/j.jeconom.2020.12.001.

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics 134*(3), 1405–1454.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013). Average and quantile effects in nonseparable panel models. *Econometrica 81*(2), 535–580.

de Chaisemartin, C. and X. D'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–2996.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*. doi:10.1016/j.jeconom.2021.03.014.

Imai, K. and I. S. Kim (2019). When should we use unit fixed effects regression models for causal inference with longitudinal data? *American Journal of Political Science 63*(2), 467–490.

Liu, L., Y. Wang, and Y. Xu (2021). A practical guide to counterfactual estimators for causal inferences with time-series cross-sectional data. Technical report.

Sloczynski, T. (2020). Interpreting OLS estimands when treatment effects are heterogeneous: Smaller groups get larger weights. *Review of Economics and Statistics*. doi: 10.1162/rest_a_00953.

Sun, L. and S. Abraham (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*. doi:10.1016/j.jeconom.2020.09.006.

Varian, H. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives 2*(28), 3–28.

Wooldridge, J. M. (2005). Fixed-effects and related estimators for correlated random coefficient and treatment-effect panel data models. *Review of Economics and Statistics 87*, 385–390.