Visualization, Identification, and Estimation in the Linear Panel Event-Study Design

Simon Freyaldenhoven
Federal Reserve Bank of Philadelphia

Christian Hansen
University of Chicago

Jorge Pérez Pérez
Banco de México

Jesse M. Shapiro
Brown University and NBER[*]

This version: May 7, 2021

Abstract

We discuss the role of "event-study plots" commonly presented in applied research, and make some suggestions on their construction. We review approaches to identification in the linear panel event study design, discuss their economic content, and point out some of their connections. We illustrate strengths and weaknesses of the corresponding estimators using simulations.

---

1 Introduction

We have an observational panel of units $i \in \{1, ..., N\}$ in a sequence of periods $t \in \{1, ..., T\}$. We are interested in learning the causal effect of an observed scalar policy variable z_{it} on an observed scalar outcome y_{it}. Economic situations that can be cast into this setting include:

- **Participation in a public program.** Here i indexes individuals or households, z_{it} is an indicator for participation in a public program (e.g., a training program, as in Ashenfelter 1978), and y_{it} is an outcome variable (e.g., earnings).

- **Firm entry.** Here i indexes markets (e.g., cities), z_{it} is an indicator for periods following the entry of a firm (e.g., Walmart, as in Basker 2005), and y_{it} is an outcome variable (e.g., average retail prices).

- **State law.** Here i indexes US states, z_{it} is the level of a policy variable (e.g., the minimum wage, as in Brown 1999) or an indicator for periods following the adoption of a policy (e.g., unilateral divorce, as in Wolfers 2006), and y_{it} is an outcome variable (e.g., employment rates, divorce rates).

We will focus attention throughout on the linear model:

$$y_{it} = \alpha_i + \gamma_t + q'_{it}\psi + \sum_{m=-G}^{M} \beta_m z_{i,t-m} + C_{it} + \varepsilon_{it}, \tag{1}$$

where α_i denotes a unit fixed effect, γ_t a time fixed effect, and q_{it} a vector of controls. The scalar C_{it} denotes a (potentially unobserved) confound that may be correlated with the policy. For simplicity, we assume throughout that ε_{it} is an unobserved scalar that obeys $\mathbb{E}[\varepsilon_{it}|C_i, z_i, \alpha_i, \gamma_t, q_i] = 0$, with C_i, z_i, and q_i collecting all T observations for unit i of C_{it}, z_{it}, and q_{it} respectively, and further maintain that neither lagged dependent variables nor other predetermined variables enter the model. Depending on the specific estimator and treatment of the fixed effects, it may be possible to relax these conditions in practice.

If the confound C_{it} were observed, it would thus be straightforward, with sufficient data and policy variation, to learn the parameters of interest, $\{\beta_m\}_{m=-G}^{M}$ for $G \geq 0$ and $M \geq 0$, which summarize the dynamic effects of the policy. In general, the confound C_{it} will not be observed, and identification of the parameters of interest will rely on restrictions on how observable and latent variables relate to C_i and z_i.

A special case of interest is where the policy is binary, $z_{it} \in \{0, 1\}$, all units begin without the policy, $z_{i1} = 0$ for all i, and once the policy is adopted by a given unit it is never reversed, $z_{it'} \geq z_{it}$ for all i and $t' \geq t$. Following Athey and Imbens (forthcoming) and others (e.g., Shaikh and Toulis

2019 and Ben-Michael et al. 2021), we will call this situation one of *staggered adoption*. We adopt the convention of referring to the index m as "event time," with event time 0 corresponding to the contemporaneous value of the policy variable, event time 1 to the first lag of the policy variable, event time -1 to the first lead of the policy, and so forth. In cases of staggered adoption, the "event" is the adoption of the policy. In more general designs, including those with non-binary policy z_{it}, there need not be a single "event," and the meaning of event time comes instead from the specification of (1).

The model allows that the policy for unit i at time t affects the outcomes for unit i from periods $t - G$ through $t + M$, with $G = 0$ corresponding to the assumption of no anticipatory effects. We will assume that G and M are known. It is not generally possible to learn G and M from the data unless there are never-treated units in the sample (Borusyak and Jaravel 2018; Schmidheiny and Siegloch 2020).

Estimates of models of the form in (1) are very common in economic applications. Freyaldenhoven et al. (2019) identify 16 articles using a linear panel model in the 2016 issues of the *American Economic Review*. During the July 2020 meeting of the NBER Labor Studies Program, estimates of a model of the form in (1) played an important role in at least one presentation on three of the four meeting dates.[1]

Though commonly estimated, the model in (1) is restrictive in many ways. The restriction that all sources of confounding enter the model in an additively separable way is substantial. See, for example, Athey and Imbens (2006) and Roth and Sant'Anna (2021) who consider nonlinear models in the canonical difference-in-differences setting.

The model in (1) also imposes that the effects of the policy at different time horizons are linear and separable. Formally, the model in (1) is a special case of

$$y_{it} = \alpha_i + \gamma_t + q'_{it}\psi + g(\{z_{it-m}\}_{m=-G}^{M}) + C_{it} + \varepsilon_{it}, \tag{2}$$

where $g(\cdot)$ is some (possibly unknown) function. Under staggered adoption, the restriction from (2) to (1) is without loss of generality. Under more general forms of the policy variable, the restriction from (2) to (1) is substantial.

Estimates of models like (1) are often summarized in empirical research with an "event-study plot," similar to Figure 1(a), that depicts model-based estimates of policy effects in event time and their confidence intervals. Roth (2020) identifies 70 papers published in three AEA journals between 2014 and June 2018 that include such a plot. Section 2 reviews the construction and interpretation of these plots and discusses some suggestions to make them more informative.

Interest in plots of the sort in Figure 1(a) often centers on trends in the outcome variable prior

---

[1] See Aghion et al. (2020), Ayromloo et al. (2020), and Derenoncourt et al. (2020).

(a) Outcome variable

(b) Effect due to confound
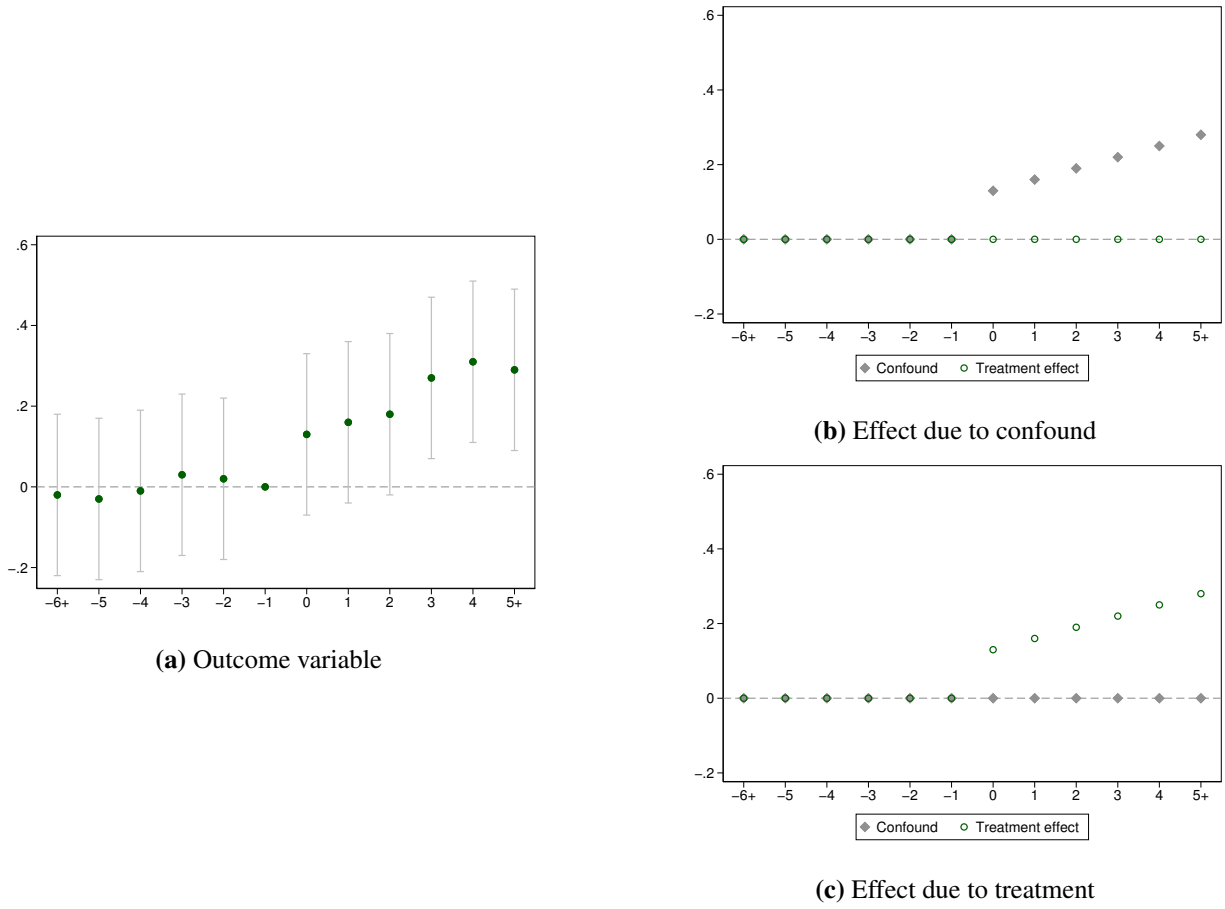
(c) Effect due to treatment

Figure 1: Exemplary event-study plot with two possible interpretations. Panel (a) shows the event-study plot of the outcome variable as it might be observed in practice. Panels (b) and (c) provide two potential scenarios that are equally consistent with Panel (a). In Panel (b), there is no effect of the policy and the post-event dynamics of the outcome variable are entirely driven by the confound. In Panel (c), there is no confound and the post-event dynamics of the outcome variable reflect the effect of the policy.

to the change in policy. If the economic mechanism is such that future values of a policy variable (or values of a policy variable more than some known number of periods in the future) cannot affect the current value of the outcome, such pre-event trends or "pre-trends" provide a mechanism to visually assess these restrictions and could be used to formally test overidentifying restrictions and potentially falsify the researcher's model. However, Figures 1(b) and 1(c) also illustrate the well-known fact that it is not possible in general to determine the post-event confound entirely from pre-event trends. The figures show that the post-event path of the outcome can, in general, be attributed either to the effect of the policy or to the effect of a confound that is correlated with the policy.

That it will not generally be possible to infer confound behavior from pre-event trends means that substantive economic restrictions on C_{it} are crucial for identification of $\{\beta_m\}_{m=-G}^{M}$ in (1). As

these restrictions will generally involve some untestable components, it is important that they be motivated on economic grounds based on the setting and data at hand. Section 3 reviews some existing approaches to identification and discusses connections among them.

Different restrictions on C_{it} will tend to suggest different estimation strategies. A given estimator's performance will vary depending on whether the restrictions that motivate the estimator are well suited to the underlying economic process. Section 4, and an accompanying Appendix, present simulation evidence on the performance of a range of estimators under a number of economically motivated data-generating processes. We highlight settings in which different estimators perform relatively well, and those in which they perform relatively poorly.

An important restriction in (1) is that the effects of the policy are the same for different units i and in different time periods t. An active literature explores the implications of relaxing this homogeneity. See, for example, Athey and Imbens (forthcoming), Callaway and Sant'Anna (forthcoming), Sun and Abraham (forthcoming), de Chaisemartin and D'Haultfoeuille (2020a), and Goodman-Bacon (2020). Section 5 discusses how to adapt the approaches in Section 3 to deal with some forms of heterogeneity in policy effects.

We hope that this article provides a framework for data visualization, identification, and estimation in linear panel event-study settings that is useful for applied researchers, and especially that the article helps to highlight the economic content of, and connections among, different approaches to identification.

2   Plotting

In this section, we review event-study plots. Although on their own these plots are insufficient for inference about policy effects, in tandem with beliefs about the economic setting they are a useful tool for assessing the plausibility of different possible confounds and dynamic policy effects. As part of our review, we offer some suggestions that we think could make these plots even more useful. An accompanying Stata package, `xtevent`, makes these suggestions easier to adopt.

2.1   Definitions

There are many ways to construct an event-study plot. We consider plotting the points

$$\{k, \hat{\delta}_k\}_{k=-G-L_G-1}^{k=M+L_M}$$

and their confidence intervals, obtained from estimating

$$y_{it} = \sum_{k=-G-L_G}^{M+L_M-1} \delta_k \Delta z_{i,t-k} + \delta_{M+L_M} z_{i,t-M-L_M} + \delta_{-G-L_G-1}(-z_{i,t+G+L_G+1})$$
$$+ \alpha_i + \gamma_t + q'_{it}\psi + C_{it} + \varepsilon_{it}, \tag{3}$$

where Δ denotes the first difference operator, $L_G \geq 0$ denotes the number of pre-event periods to be plotted in which anticipatory effects are ruled out, and $L_M \geq 0$ denotes the number of periods to be plotted after dynamics are believed to have died off. Let $\delta = (\delta_{-G-L_G-1}, ..., \delta_{M+L_M})'$ be the vector that collects the $\{\delta_k\}_{k=-G-L_G-1}^{k=M+L_M}$ and define $\hat{\delta}$ similarly. For brevity, we refer to δ as the *event-time path* of the outcome variable.

Estimating δ requires assumptions about C_{it}, with different assumptions suggesting different approaches. We defer this topic to Section 3, and for the purposes of this section assume the researcher has adopted some approach to account for the confound C_{it}.

2.2 Interpretation

The interpretation of the key terms in (3) is particularly clear in the case of staggered adoption. In this case, $\Delta z_{i,t+k}$ is an indicator for whether unit i adopted the policy exactly k periods after period t, $z_{i,t+G+L_G+1}$ is an indicator for whether unit i will have adopted as of $G + L_G + 1$ periods after period t, and $z_{i,t-M-L_M}$ is an indicator for whether unit i adopted at least $M + L_M$ periods before period t.

The parameters in (3) have a useful interpretation even outside the staggered adoption setting. In particular, if $L_G = L_M = 0$, then (3) is equivalent to (1), and if $L_G > 0$ or $L_M > 0$, then (3) is equivalent to (1) with L_G additional leads or L_M additional lags included. Whenever (1) holds, we have that

$$\delta_k = \begin{cases} 0 & \text{for} \quad k < -G \\ \sum_{m=-G}^{k} \beta_m & \text{for} -G \leq k \leq M \\ \sum_{m=-G}^{M} \beta_m & \text{for} \quad k > M \end{cases} ; \tag{4}$$

i.e., that the event-study plot depicts the cumulative dynamic policy effects from (1). See, for example, Sun and Abraham (forthcoming) and Schmidheiny and Siegloch (2020).

2.3 Normalization

The policy variables in (3) are collinear, so some normalization is typically required for identification of the coefficients δ. Indeed, one reason it is valuable to include the endpoint variables
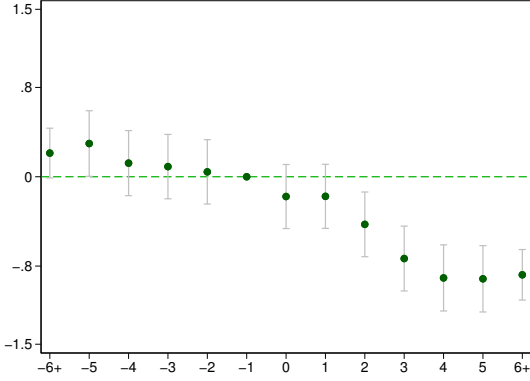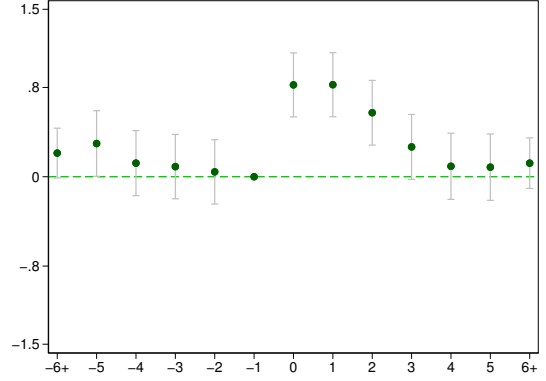
(a) "Smooth" event-time trend **(b)** "Jump" at the time of the event

Figure 2: Baseline event-study plot. Exemplary event-study plot for two possible datasets. The left panel illustrates an event-study plot with potentially "smooth" event-time dynamics, and the right panel illustrates an event-study plot with a "jump" at the time of the event.

$z_{i,t+G+L_G+1}$ and $z_{i,t-M-L_M}$ in (3) is that, by saturating the model, inclusion of these variables forces a conscious choice of normalization. A common choice, and one that we think serves as a good default, is to normalize $\delta_{-G-1} = 0$. In the leading case where $G = 0$, i.e. there are no anticipatory effects, this will mean normalizing $\delta_{-1} = 0$. In the case of staggered adoption, the normalization $\delta_{-1} = 0$ means that the plotted coefficients can be interpreted as effects relative to the period before adoption. More generally, the normalization $\delta_{-1} = 0$ means that the plotted coefficients can be interpreted as effects relative to the effect of a one-unit change in the policy variable one period ahead.

Figure 2 illustrates an event-study plot with the normalization $\delta_{-1} = 0$ for two possible datasets. Both figures exhibit identical pre-event dynamics. After event time -1, Figure 2(a) exhibits a "smooth" continuation of pre-event trends, whereas Figure 2(b) exhibits a "jump."

Because a given change relative to baseline can have a different interpretation depending on the level of the dependent variable at baseline, our first suggestion for improving the information content of event-study plots is to include a parenthetical label for the normalized coefficient δ_{k^*}.

Suggestion 1. *Include a parenthetical label for the normalized coefficient δ_{k^*} that has value*

$$\frac{\sum_{(i,t):\Delta z_{i,t+k^*}\neq 0} y_{it}}{|(i,t): \Delta z_{i,t+k^*} \neq 0|}. \tag{5}$$

The value in (5) corresponds to the mean of the dependent variable k^* periods in advance of a policy change.[2] For example, under our default recommendation that δ_{-1} be normalized to zero, we would

---

[2]When multiple coefficients in δ are normalized, we propose to label the value of the one whose index is closest to zero.

6
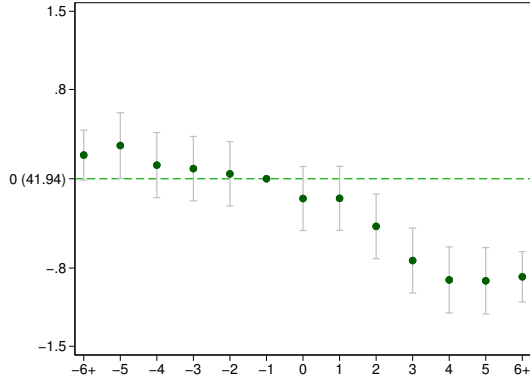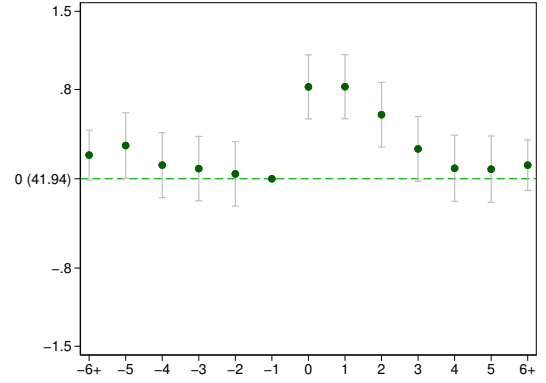
(a) "Smooth" event-time trend   **(b)** "Jump" at the time of the event

Figure 3: Label for normalized coefficient. Exemplary event-study plot for two possible datasets. Relative to Figure 2, a parenthetical label for the average value of the outcome corresponding to the normalized coefficient has been added, in accordance with Suggestion 1.

include a parenthetical label with value $\frac{\sum_{(i,t):\Delta z_{i,t-1}\neq 0} y_{it}}{|(i,t):\Delta z_{i,t-1}\neq 0|}$, i.e. the average of y_{it} at event time -1, to provide a sensible benchmark value for comparison when interpreting the other estimated effects.

Figure 3 illustrates by adding the proposed label to the plots in Figure 2. The additional value shows up as a parenthetical label next to the 0 point on the y-axis. The estimates of the un-normalized coefficients provide estimated effects relative to the normalized coefficient and are read off the y-axis. For example, in Figure 3(a), the coefficient estimate corresponding to the third lag implies that the estimated effect of being at event time 3 relative to event time -1 is roughly -0.73. This appears to be a modest effect relative to 41.94, the mean of the dependent variable in the reference period, shown in the parenthetical label.

2.4 Inference

We suspect that the popularity of event-study plots arises in part from the fact that some paths of the plot are more suggestive of the presence of confounds than others. It is therefore important for the plot's reader to be able to judge which paths are consistent with the data.

Suggestion 2. *Plot a uniform confidence band for the path of $\{k, \delta_k\}_{k=-G-L_G-1}^{k=M+L_M}$, in addition to pointwise confidence intervals for the elements of δ.*

Uniform confidence bands are suitable for testing prespecified hypotheses about the entire path of the coefficients (e.g., do they all lie on a line?), whereas pointwise confidence intervals are suitable only for testing prespecified hypotheses about individual coefficients (e.g., is there a change one period after the policy goes into effect?). Providing uniform confidence bands thus allows visual assessment of entire paths of the coefficients that are not rejected by the data which is not
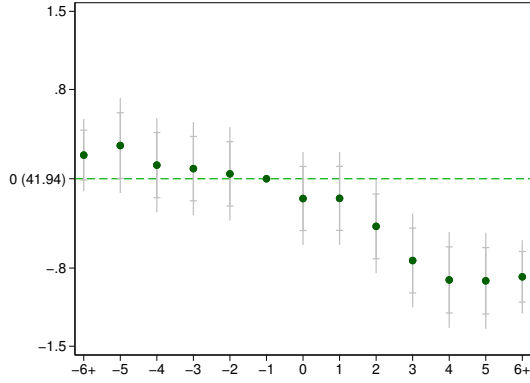
(a) "Smooth" event-time trend
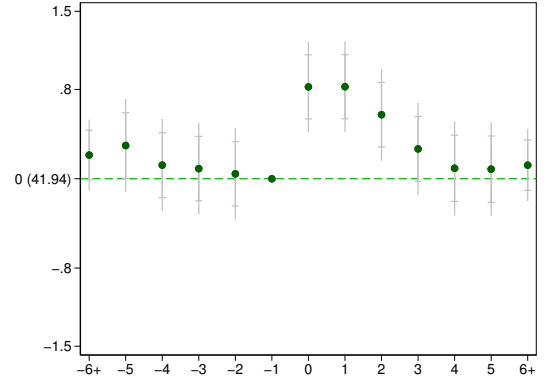
(b) "Jump" at the time of the event

Figure 4: Uniform confidence band. Exemplary event-study plot for two possible datasets. Relative to Figure 3, uniform confidence bands have been added, in accordance with Suggestion 2. The original pointwise confidence bands are illustrated by the inner bars and the uniform, sup-t confidence bands are given by the outer lines.

possible if only pointwise confidence intervals are provided. We propose to use sup-t confidence bands, which are straightforward to visualize and also convenient to compute in the settings we consider (Freyberger and Rai 2018; Olea and Plagborg-Møller 2019). Figure 4 illustrates by adding sup-t confidence bands as outer lines to the plots in Figure 3, retaining the pointwise confidence intervals as the inner bars.

2.5 Confounds

Any path $\{k, \hat{\delta}_k\}_{k=-G-L_G-1}^{k=M+L_M}$ can be reconciled with a devil's-advocate model in which the policy has no effect on the outcome. In this devil's-advocate model, all of the dynamics of the event-time path of the outcome variable are driven by an unmeasured confound. The devil's-advocate model is most plausible when it implies economically reasonable dynamics for the confound. Though there is no universal notion of what is economically reasonable, in many settings researchers may expect that the dynamics of a confound are "smooth," in the sense of not changing radically at the time of policy adoption. Under such a prior, the smoother the event-time paths that are not rejected, the more concerned we may be about potential confounds.

Suggestion 3. *Plot the confound with the "smoothest" event-time dynamics that are consistent with the data under the hypothesis of no treatment effect.*

This suggestion is closely related to the sensitivity analysis proposed in Rambachan and Roth (2020).

We adopt an unconventional notion of smoothness that we think maps well to informal discussions around pre-event trends which tend to focus on event-time dynamics that are well-approximated
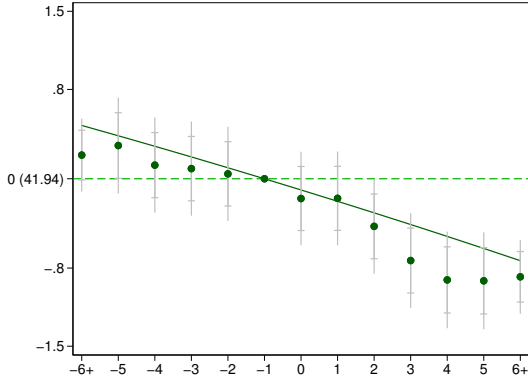
(a) "Smooth" event-time trend
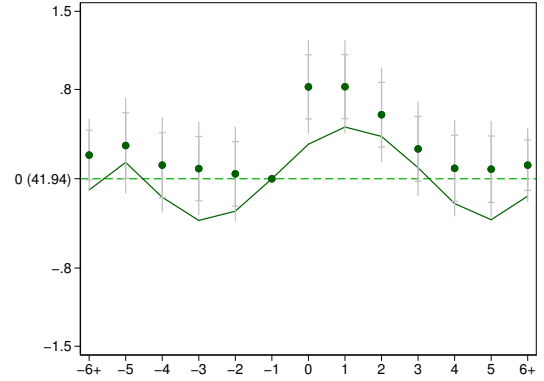
(b) "Jump" at the time of the event

Figure 5: "Smoothest" event-time path. Exemplary event-study plot for two possible datasets. Relative to Figure 4, a "smoothest" line that lies within the Wald confidence region has been added, in accordance with Suggestion 3.

by low-order polynomial deterministic trends (see, e.g., Dobkin et al. 2018). Specifically, we adopt a lexicographic notion in which lower-order polynomials are preferred to higher-order polynomials and then, within a given order of polynomial, those with lower curvature as measured by the magnitude of the coefficient on the highest order term are preferred to those with higher curvature.

To implement this notion, we use a two-step procedure. Formally, for v a finite-dimensional coefficient vector and k an integer, define the polynomial term $\delta_k^*(v) = \sum_{j=1}^{\dim(v)} v_j k^{j-1}$, where v_j denotes the jth element of coefficient vector v and $\dim(v)$ denotes the dimension of this vector. Let $\delta^*(v)$ collect the elements $\delta_k^*(v)$ for $-G - L_G - 1 \leq k \leq M + L_M$, so that $\delta^*(v)$ reflects a polynomial path in event time with coefficients v.

First, we find the lowest order p^* that the polynomial path $\delta^*(v)$ can have and still produce coefficients δ that are not rejected by a Wald test. Formally we define

$$p^* = \min\{\dim(v) : \delta^*(v) \in CR(\delta)\}, \tag{6}$$

where $CR(\delta)$ is the Wald confidence region for some desired significance level for δ.

Second, among all polynomial paths $\delta^*(v)$ with order p^* we choose the one $\delta^*(v^*)$ with the smallest coefficient on the highest-order term, again requiring that the resulting coefficients δ are not rejected by a Wald test. Formally this is given by

$$v^* = \arg\min_v\{v_{p^*}^2 : \dim(v) = p^*, \delta^*(v) \in CR(\delta)\}. \tag{7}$$

The resulting path $\delta^*(v^*)$ thus provides a visualization of the "most reasonable" path that cannot be rejected by the data according to a Wald test, where we deem lower-order polynomial trends

9

more reasonable and, for a fixed polynomial order, those with smaller coefficients on higher-order terms.

Figure 5 illustrates by adding the path $\delta^*(v^*)$ to the plot in Figure 4. In Figure 5(a), the estimated event-time path is consistent with a confound that follows a very "smooth" path that begins pre-event and simply continues post-event. In Figure 5(b), by contrast, the estimated event-time path demands a confound with a very "wiggly" path. Whether such wiggliness is economically plausible will depend on the setting, but we expect that in many settings, a confound of the form depicted in Figure 5(a) will be more plausible than one of the form depicted in Figure 5(b).

2.6   Overidentification and Testing

Some of the interest in event-study plots stems from the visual test for pre-trends, that is, of the overidentifying restriction that $\delta_k = 0$ for $-(G + L_G) \leq k < -G$. We suggest to include the p-value for this test in the plot legend, as well as one for a test that the dynamic effects have "leveled off" at the end of the plot window.

Suggestion 4. *Include in the plot legend p-values for Wald tests of the following hypotheses:*
   *(no pre-trends)* $H_0 : \delta_k = 0$ *for* $-(G + L_G) \leq k < -G$
   *(dynamics level off)* $H_0 : \delta_M = \delta_{M+k}$ *for* $0 < k \leq L_M$.

Figure 6 illustrates these suggestions by adding these two p-values to the plots from Figure 5, taking $G = 0$ and $L_M = 1$. As Roth (2020) notes, tests for pre-event trends often have low power in practice. The p-values are therefore not a good substitute for the presentation of the plot itself. We suggest $L_M = 1$ as a default setting to permit the proposed test of whether dynamics have leveled off within the plot window.

Summarizing Magnitudes with a More Restrictive Model

Researchers sometimes present a scalar summary of the magnitude of the policy effect (see, e.g., Gentzkow et al. 2011). One way to obtain such a summary is to average the post-event effects depicted in the plot. Another is to estimate a static version of (1) including only the contemporaneous value of the policy variable (restricting $M = G = 0$). In the case where the researcher imposes this (or some other) restriction, we suggest to evaluate the fit of the more restrictive model to the dynamics of a more flexible model that allows for dynamic effects, such as (3), as follows. First, estimate (1) subject to the contemplated restrictions (e.g. a static version with only the contemporaneous value of the policy variable). Second, use (4) to translate the restricted coefficients to the implied estimates $\hat{\delta}_k$ of the event-time path of the outcome. Third, overlay the resulting estimates $\hat{\delta}_k$ on those from the less restrictive model.

10
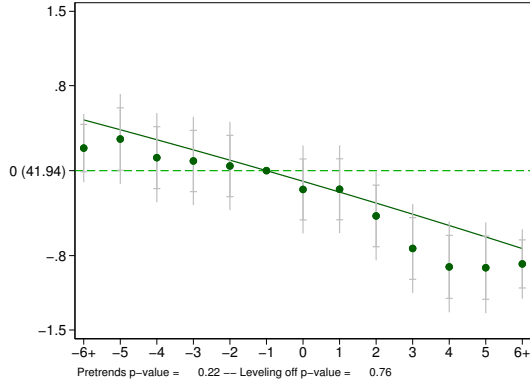
(a) "Smooth" event-time trend
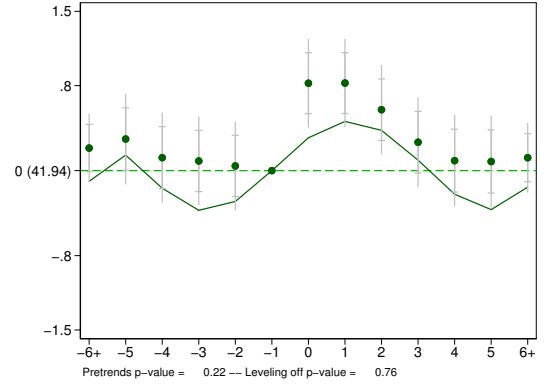
(b) "Jump" at the time of the event

Figure 6: Testing p-values. Exemplary event-study plot for two possible datasets. Relative to Figure 5, p-values for testing for the absence of pre-event effects and p-values for testing the null that dynamics have leveled off have been added, in accordance with Suggestion 4.

Comparison of the two event-time paths provides a visualization of the fit of the more restrictive model. The uniform confidence bands also allow ready visual assessment of more restrictive models. The null of constant treatment effects will be rejected, using the sup-t procedure, whenever the uniform confidence band does not include 0 everywhere before event time 0 and include a horizontal line after event time 0. Figure 7 illustrates a case where the more restrictive model is not included in the uniform confidence band suggesting that the treatment effects are indeed dynamic.

Though less visually convenient, a Wald test can also be used for more efficient discrimination among hypotheses. Figure 7 includes the p-value from a Wald test of the more restrictive model, showing that the more restrictive model is rejected.

3 Approaches to Identification

We now discuss approaches to identification of the parameters $\{\beta_m\}_{m=-G}^{M}$ in (1). In what follows, we decompose the confound C_{it} into a component that depends on a low-dimensional set of time-varying factors with individual-specific coefficients, and an idiosyncratic component such that

$$C_{it} = \lambda_i' F_t + \xi \eta_{it}, \text{ and thus} \tag{8}$$

$$y_{it} = \alpha_i + \gamma_t + q_{it}' \psi + \sum_{m=-G}^{M} \beta_m z_{i,t-m} + \lambda_i' F_t + \xi \eta_{it} + \varepsilon_{it}. \tag{9}$$

In (8) and (9), F_t is a vector of (possibly unobserved) factors with (possibly unknown) unit-specific loadings λ_i, η_{it} is an unobserved scalar with (possibly unknown) coefficient ξ, and we maintain throughout that $\mathbb{E}[\varepsilon_{it}|F, \lambda_i, \eta_i, z_i, \alpha_i, \gamma_t, q_i] = 0$, with F and η_i collecting all T observations of F_t
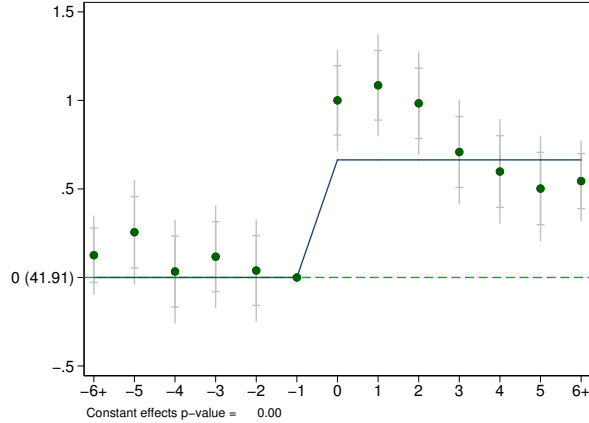
11

Figure 7: Evaluating the fit of a more restrictive model. Here, we overlay estimates from a model that imposes static treatment effects with estimates from a more flexible model that allows for dynamic treatment effects.

and η_{it}, respectively.

To identify $\{\beta_m\}_{m=-G}^M$ in (9) a researcher must, in principle, account for both the low-dimensional component $\lambda_i' F_t$ and the idiosyncratic component η_{it}. We next review a number of existing approaches that impose sufficient structure on each of these components to yield identification of $\{\beta_m\}_{m=-G}^M$, and we point out some connections between them. Which approach is the "correct approach" cannot generally be learned from the data, and therefore needs to be motivated on economic grounds based on the setting.

For clarity, we only discuss situations in which the researcher's economic model dictates that the confound is entirely low-dimensional or entirely idiosyncratic. In situations where a researcher is concerned with the presence of both types of confound, it may be possible to achieve identification by "mixing and matching" across approaches.

3.1 Conditional on Observables

The first class of approaches imposes sufficient structure on how latent variables enter the model so that the parameters $\{\beta_m\}_{m=-G}^M$ can be identified without additional observable variables.

Assumption 1. $\xi = 0$ *and one of the following holds:*

(a) $F_t = 0$ *for all* t.

(b) $F_t = f(t)$ *for* $f(\cdot)$ *a known low-dimensional set of basis functions.*

(c) *The dimension of* F_t *is small.*

12

Note that although we state Assumption 1 as a set of possible restrictions on the latent factor F_t, sufficient conditions for identification may also be available as restrictions on the loadings λ_i.[3] Note also that, while we focus on the conditions in Assumption 1, formal conditions for identification will typically include side conditions such as rank conditions and conditions guaranteeing a long enough panel.

In case (a), the parameters β can be estimated via the usual two-way fixed effects estimator with controls q_{it}. The assumption that all sources of latent confounding are either time-invariant, and thus captured by the α_i, or cross-sectionally invariant, and thus captured by the γ_t, is parsimonious but restrictive. If, for example, the time effects γ_t represent the effects of macroeconomic shocks or changes in federal policy, then Assumption 1(a) implies that these factors influence all units (e.g., states) in the same way, ruling out many sensible models in which the influence of aggregate factors may depend on unit-level features such as the sectoral composition of the local economy or the nature of local policies. In applications, these restrictions should ideally be motivated on economic grounds.

In case (b), the parameters β can be estimated by further controlling for a unit-specific trend with shape $f(t)$. Common choices include polynomials, such as $f(t) = t$ (unit-specific linear trend, e.g., Jacobson et al. 1993) or $f(t) = (t, t^2)$ (unit-specific quadratic trends, e.g., Friedberg 1998). For example, Jacobson et al. (1993) adopt this approach in an application where the policy z_{it} is an indicator for quarters t following the displacement of worker i from a job. The outcome y_{it} is earnings. The worker fixed effect α_i accounts for time-invariant worker-specific differences in earnings. Jacobson et al. (1993) further allow for worker-specific differences in earnings trends by taking $f(t) = t$, and thus assuming that displacement is exogenous with respect to unobserved determinants of earnings, conditional on the worker fixed effect and worker-specific time trend.

Case (b) generalizes case (a) by parameterizing time-varying confounds as deterministic trends and allowing for unit-specific coefficients. Researchers employing this approach should ideally be able to argue on economic grounds why the chosen trend structure plausibly approximates latent sources of confounding. For example, the assumption that confounds are well-approximated by a linear trend may be more plausible in a short panel than in a very long one, where changes in trend are more likely and where a constant linear trend may imply implausible long-run behavior.

Assumption 1(c) differs from Assumption 1(b) by treating the time-varying confound F_t as unknown. Since F_t now has to be learned from the data, approaches utilizing Assumption 1(c) will generally require larger N and T and that $(1/N) \sum_i^N \lambda_i \lambda_i'$ is positive definite (see, e.g., Bai 2009). An implication of the latter condition is that factors need to be pervasive in the sense of having non-negligible association with many of the individual units. Assumption 1(c) therefore seems especially appealing when the latent confound is thought to arise from aggregate factors (such as

---

[3]For example, in case (a) we may say that $\lambda_i = 0$ for all i. In case (c) we may say that the dimension of λ_i is small.

macroeconomic shocks) that affect all units but to a different extent. It seems less appealing when, say, the policy of interest is adopted based on idiosyncratic local factors that are not related to aggregate shocks.

In case (c), the parameters β can be estimated via an interactive fixed effects or common correlated effects estimator (Bai 2009; Pesaran 2006). They may also be estimated via synthetic control methods (Abadie et al. 2010; Chernozhukov et al. 2020), though in practice this approach is most often used when the number of treated units is small and the policy follows staggered adoption (but see Dube and Zipperer 2015; Powell 2021).

Powell (2021) adopts a structure similar to that implied in case (c) to examine the effect of the log of the prevailing minimum wage in state i in quarter t, z_{it} on the youth employment rate, y_{it}. A concern is that states tend to increase the minimum wage when the local economy is strong (Neumark and Wascher 2007). Assumption 1(c) captures this concern by allowing that each state i has a unique, time-invariant response λ_i to the state F_t of the aggregate economy. Powell (2021) proposes a generalization of synthetic controls to account for this confound and recover the causal effect of interest.

Assumption 1 holds that the confound is low-dimensional ($\xi = 0$). We next turn to a scenario where the confound is idiosyncratic.

Assumption 2. $F_t = 0 \; \forall \; t$ and the confound obeys

$$\mathbb{E}[\eta_{it}|z_i, \alpha_i, \gamma_t, q_i] = \tilde{\alpha}_i + \tilde{\gamma}_t + q_{it}'\tilde{\psi} + \sum_m \phi' f(m) z_{i,t-m} \tag{10}$$

for $f(\cdot)$ a known low-dimensional set of basis functions, and $\tilde{\alpha}_i$, $\tilde{\gamma}_t$, $\tilde{\psi}$, and ϕ unknown parameters.

Under Assumption 2, it is possible to control for the effect of the latent confound η_{it} by extrapolation. For intuition, consider the case of staggered adoption without anticipatory effects ($G = 0$). Suppose that $f(m)$ is equal to 1 when $m \in [-3, 3]$ and is equal to 0 otherwise. The model therefore predicts that a projection of η_{it} into event time will have a linear trend that begins three periods before adoption and continues for three periods after. Because the policy has no causal effect on the outcome before adoption, the pre-adoption periods can be used to learn the slope of the trend. Extrapolating this slope into the post-adoption periods then permits accounting for the confound as in, for example, Dobkin et al. (2018). It is easy to extend this approach to extrapolation of a richer function, for example by supposing instead that $f(m) = (1, m)$ when $|m| \leq 3$.

Figure 8 illustrates in two hypothetical scenarios. Figure 8(a) depicts the event-time path of the outcome variable. Figure 8(b) extrapolates the confound from the pre-event event-time path of the outcome. Figure 8(c) adjusts the path of the outcome for the estimated path of the confound, thus (under the maintained assumptions) revealing the effect of the policy on the outcome.
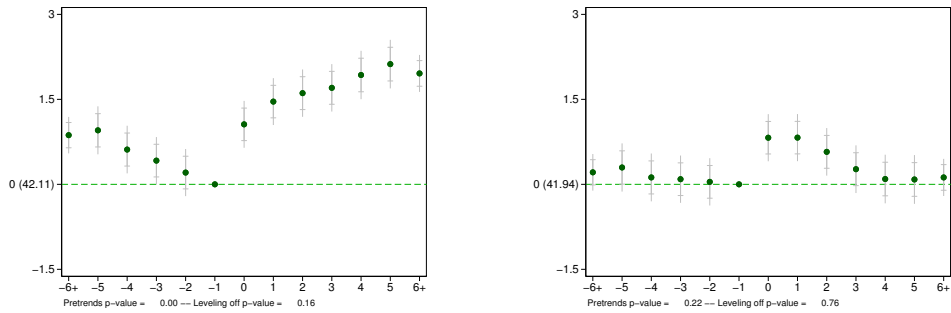
(a) Event-study plot
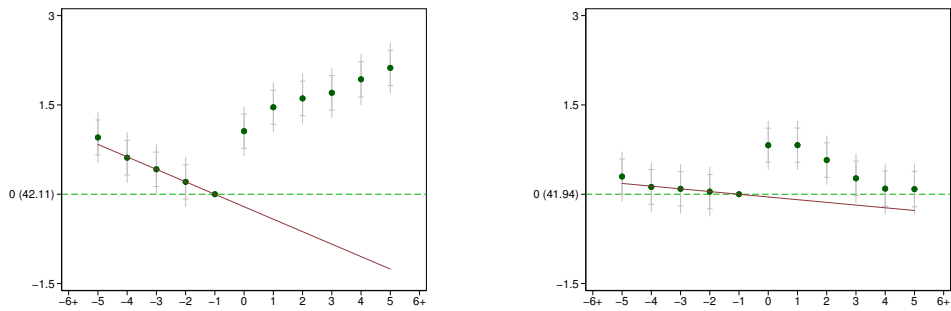
(b) Overlay extrapolation line
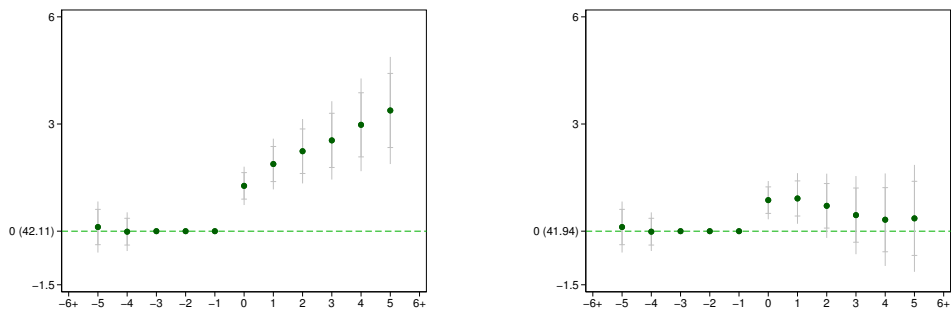
(c) Subtract extrapolated trend

Figure 8: Illustration of approach based on extrapolating confound from pre-event period. Each column of plots corresponds to a different possible dataset. Panel (a) shows the event-study plot. Panel (b) extrapolates a linear trend between time periods using the three periods before the event, as in Dobkin et al. (2018). It then overlays the event-time coefficients for the trajectory of the dependent variable and the extrapolated linear trend. In Panel (c), the estimated effect of the policy is the deviation of the event-time coefficients of the dependent variable from the extrapolated linear trend. Note that the y-axis in Panel (c) differs from that in (a) and (b).

It is useful to compare the economic content of Assumption 2 with that of Assumption 1(b) under a similar basis $f(\cdot)$ (e.g., a linear trend). In Dobkin et al. (2018), the outcome y_{it} is a measure of an individual's financial well-being such as a credit score. The policy z_{it} is an indicator for periods after the individual was hospitalized. The confound C_{it} might reflect the individual's underlying health state, which affects both hospitalization and credit scores. Assumption 1(b) would hold if the confound is equal to $\lambda_i' t$, which imposes that every individual's health follows a deterministic linear trend in calendar time with individual-specific slope λ_i. Assumption 2, by contrast, would hold if the confound is equal to $\phi(t - t_i^*)$, where t_i^* is the date of hospitalization for individual i, so that every individual's health evolves with the same slope ϕ in time relative to hospitalization. Assumption 2 would also hold if, say, the confound is equal to 3ϕ when $t - t_i^* > 3$, -3ϕ when $t - t_i^* < -3$, and $\phi(t - t_i^*)$ otherwise, such that every individual's health evolves with the same slope ϕ in time relative to hospitalization within 3 periods of the hospitalization.

Assumption 2 is related to a literature on "regression discontinuity in time," an approach to learning the effect of the policy by "narrowing in" on a short window around a policy change (see, e.g., Hausman and Rapson 2018). To us, such an approach seems most appealing when there is economic interest in the very short-run effect of the policy. For example, estimates of the effect of a policy announcement on equity prices based on comparing prices just after and just before the announcement may be very informative about the market's view of the effect of the policy. By contrast, estimates of the effect of the minimum wage on employment based on comparing employment the day (or hour) after a change in the minimum wage to employment the day (or hour) before the change may not reveal much about the economically important consequences of the minimum wage.[4]

3.2 Proxies and Instruments

The second class of approaches that we consider envisions that we have additional variables available that can serve either as proxies for the unobserved confound η_{it} or as instruments for the policy z_{it}. For simplicity, we assume throughout this subsection that $F_t = 0$, so that the only unobserved confound is η_{it}.

Our first set of assumptions relies on using additional observed variables to serve as proxies for the latent individual and time-varying source of confounding η_{it}.

Assumption 3. *There is an observed vector x_{it} that obeys*

$$x_{it} = \alpha_i^x + \gamma_t^x + \psi^x q_{it} + \Xi^x \eta_{it} + u_{it}, \tag{11}$$

---

[4]Our observation here relates to work that studies inference on causal effects away from the cutoff in regression-discontinuity designs (see, e.g., Cattaneo et al. 2020).

where α_i^x is an unobserved unit-specific vector, γ_t^x is an unobserved period-specific vector, ψ^x is an unknown matrix-valued parameter, Ξ^x is an unknown vector-valued parameter with all elements nonzero, and u_{it} is an unobserved vector. Moreover, one of the following holds:

(a) *The vector x_{it} has $\dim(x_{it}) \geq 2$ and the unobservable u_{it} satisfies*
   $E[u_{it}|z_i, \alpha_i^x, \gamma_t^x, q_i, \varepsilon_{it}] = 0$ with $E[u_{it}'u_{it}|z_i, \alpha_i^x, \gamma_t^x, q_i, \varepsilon_{it}]$ diagonal.

(b) *The unobservable u_{it} satisfies $\mathbb{E}[u_{it}|z_i, \alpha_i^x, \gamma_t^x, q_i] = 0$, and the population projection of η_{it}*
   on $\{z_{i,t-m}\}_{m=-G-L_G}^{M+L_M}$, q_{it}, and unit and time indicators, has at least one nonzero coefficient
   on $z_{i,t+m}$ for some $m > G$.

With only a single proxy available and no further restrictions, controlling for the noisy proxy x_{it} instead of the true confound η_{it} will generally lead to invalid inference for β. Under Assumption 3(a), we have at least two proxies available for the confound η_{it}. The proxies are possibly noisy, but because the noise is uncorrelated between the proxies, the parameters β can be estimated via two-stage least squares, instrumenting for one proxy, say x_{it}^1, with the other, say x_{it}^2, as in a standard measurement error model (e.g., Aigner et al. 1984). See Griliches and Hausman (1986) and Heckman and Scheinkman (1987) for related discussion.

Under Assumption 3(b), we need only a single proxy for the confound η_{it}, and require that the noise in the proxy is conditionally mean-independent of the policy. Therefore, any relationship between the proxy and the policy reflects the relationship between the confound and the policy. Recall that we also assume that the policy does not affect the outcome more than G periods in advance. Therefore, any relationship between the outcome and leads of the policy more than G periods in the future must reflect the relationship between the confound and the outcome. Intuitively, then, the relationship between the pre-trend in the proxy and the pre-trend in the outcome reveals the magnitude of confounding. Knowing this, it is possible to identify the effect of the policy on the outcome. Specifically, Freyaldenhoven et al. (2019) show that the parameters β can be estimated via two-stage least squares, instrumenting for the proxy with leads of the policy.

Figure 9 illustrates in the two hypothetical scenarios from Figure 8. Figure 9(a) repeats the event-study plots for the outcome. Figure 9(b) shows hypothetical event-study plots for the proxy. Figure 9(c) aligns the scale of the proxy so that its event-study coefficients agree with those of the outcome at event times -2 and -1. With this alignment, under Assumption 3(b) the event-study plot of the proxy mirrors that of the latent confound. By subtracting the rescaled event-study coefficients for the proxy from those for the outcome, we therefore arrive at an unconfounded event-study plot for the outcome, illustrated in Figure 9(d).

As a concrete example of the use of Assumption 3(b), Hastings et al. (forthcoming) employ this approach in a setting where the policy z_{it} is an indicator for household i's participation in the Supplemental Nutrition Assistance Program (SNAP) in quarter t and the outcome y_{it} is a measure

17
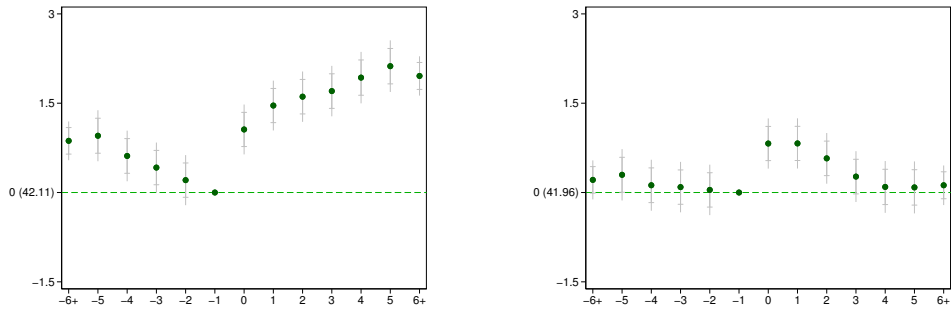
(a) Event-study plot for outcome
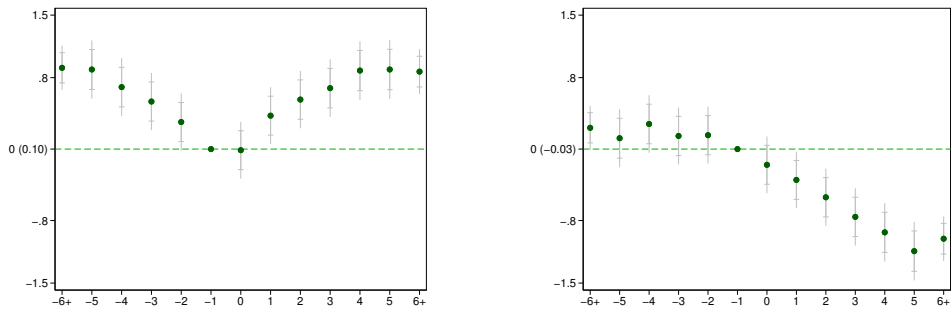
(b) Event-study plot for proxy
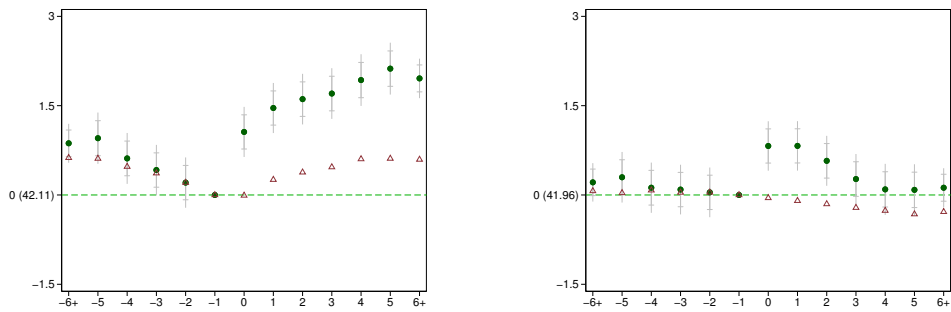
(c) Align proxy to outcome
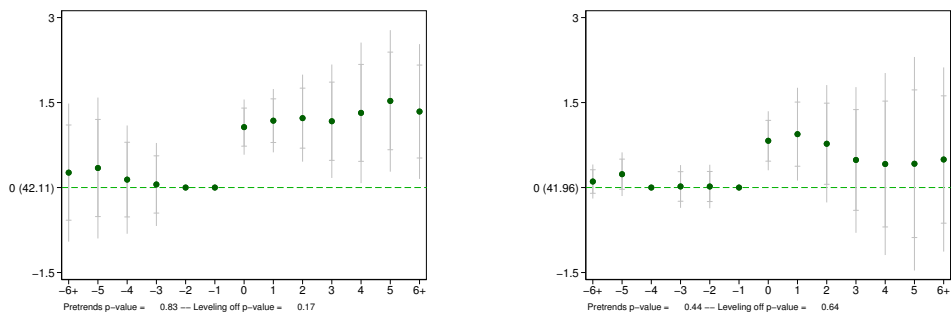
(d) Subtract rescaled confound from outcome

Figure 9: Illustration of use of a proxy to adjust for confound. Each column of plots corresponds to a different possible dataset. Panel (a) shows the event-study plot for the outcome. Panel (b) shows the event-study plot for the proxy. Panel (c) overlays the point estimates from Panel (b) on the plot from Panel (a), aligning the coefficients at event times -1 and 0. Panel (d) adjusts for the confound using the 2SLS estimator proposed in Freyaldenhoven et al. (2019).

of the healthfulness of the foods the household purchases at a retailer. The confound η_{it} might be income, which can influence both the policy (because SNAP is a means-tested program) and the outcome (because healthy foods may be a normal good). Hastings et al. (forthcoming) estimate the effect of SNAP participation on food healthfulness via two-stage least squares, instrumenting for an income proxy x_{it} with future participation in SNAP.

In some settings it may be possible to form a proxy by measuring the outcome for a group unaffected by the policy. Freyaldenhoven et al. (2019) discuss an example similar to the minimum wage application in Powell (2021), where the goal is to estimate the effect of the state minimum wage on youth employment. If adult employment x_{it} is unaffected by the minimum wage but is affected by the state η_{it} of the local economy, adult unemployment can be used as a proxy variable following Assumption 3(b).

Of course, one may also have access to a more conventional instrumental variable for the endogenous policy itself.

Assumption 4. *There is an observed vector w_{it} whose sequence w_i obeys*

$$E[\eta_{it}|\alpha_i, \gamma_t, q_i, w_i] = 0. \tag{12}$$

Under Assumption 4 and a suitable relevance condition, the parameters β can be estimated via two-stage least squares, instrumenting for the policy z_{it} with the instruments w_{it}.

Besley and Case (2000) employ this type of instrumental variables approach to study the impact of workers' compensation on labor market outcomes. Specifically, their policy variable, z_{it}, is a measure of the generosity of workers' compensation benefits in state i in year t and the outcome y_{it} is a measure of employment or wages. The confound η_{it} in this example might include the strength of the economy in the state, which could influence both the generosity of benefits and the levels of employment or wages. Besley and Case (2000) propose to instrument for z_{it} with a measure w_{it} of the fraction of state legislators who are women, a variable that has been found to influence public policy but that may plausibly be otherwise unrelated to the strength of a state's economy.

Relative to the conditions outlined in Section 3.1, Assumption 3 and Assumption 4 have the desirable feature of allowing for more general forms of confounding captured by η_{it}. Their most obvious drawback is the requirement of finding and justifying the validity of the proxy or instrument. The use of proxies à la Assumption 3 should ideally be justified by economic arguments about the likely source of confounding. The requirement of instrument validity to motivate Assumption 4 is also substantive. As with any approach based on instrumental variables, the ones justified by Assumption 3 and Assumption 4 are subject to issues of instrument weakness, which our experience suggests can be especially important in practice for the approach suggested by Assumption 3(b).
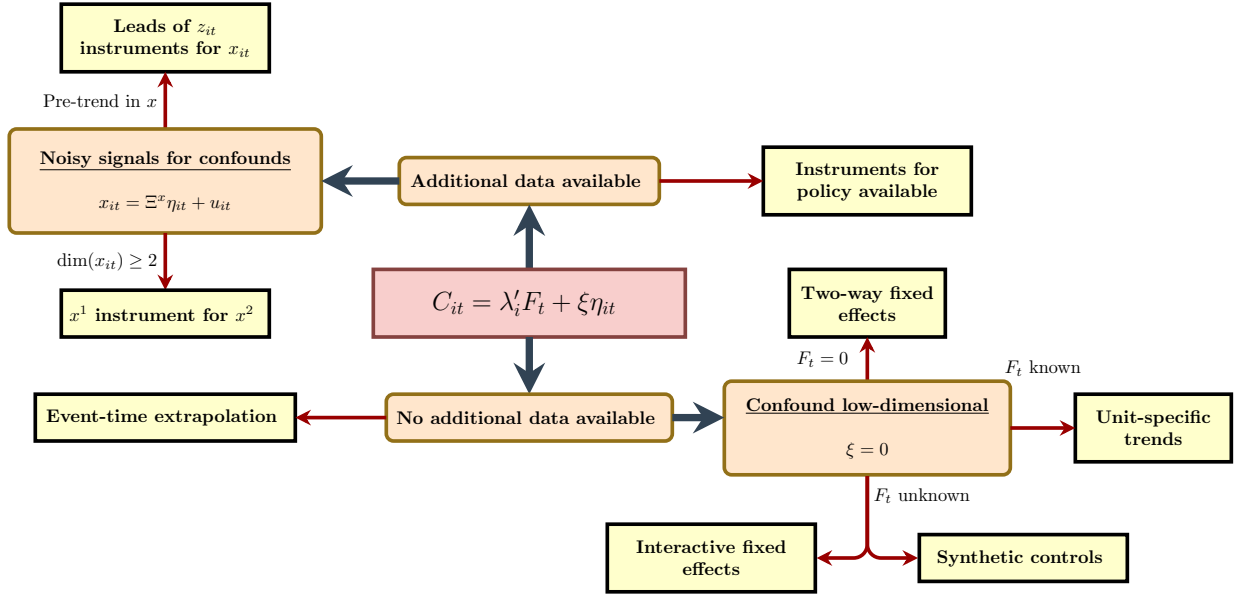
Figure 10: Schematic representation of approaches to identification. The diagram illustrates connections among the approaches to identification described in Section 3. The parameters of interest are $\{\beta_m\}_{m=-G}^{M}$ in (1).

Figure 10 summarizes Section 3 and visualizes the connections among the various approaches that we have described in this section.

4   Simulations

We now apply estimation strategies based on the approaches to identification discussed in Section 3 to simulated data. We present results for four different data-generating processes (DGPs) in a staggered adoption setting. In the following, we briefly summarize some key features of the DGPs and then present our findings on the behavior of the estimators.

4.1   Designs

Our simulation is stylized after United States state-level panel data where we set $N = 50$ and $T = 40$. Detailed descriptions of the DGPs are provided in Appendix Table 1.

Figure 11 provides a graphical summary of key features of each DGP. Each plot in the figure summarizes estimated coefficients obtained from estimating an event-study model of the form in (3) with $M + L_M = 4$ and $G + L_G = 5$ within each of 1000 simulation replications. Each column corresponds to a different DGP. Each row corresponds to a different dependent variable. That is, each figure provides estimates of δ_k for $k = -6, ..., 5$, defined in (3) with the normalization $\delta_{-1} = 0$ imposed, for a different DGP and dependent variable. All plots in Figure 11 are created from a

two-way fixed effects estimator of (3). To summarize estimation results, we report the simulation median as well as the 2.5th and 97.5th percentile of each estimated δ_k.

One of the important features distinguishing the four reported DGPs is the event-time path of the confound illustrated in the first row of Figure 11, labeled "Confound C_{it}." In each of these plots, we estimate (3) using the confound C_{it} as the dependent variable. Producing this plot is infeasible in practice but allows us to highlight the dynamics of the confound under the different DGPs. The confound in the "Mean-reverting trend" and "Multidimensional" DGPs (columns 1 and 4) shows mean-reverting behavior reminiscent of Ashenfelter's dip (Ashenfelter 1978). The confound in the "Monotone trend" DGP has an event-time path that tends to decline over the window considered. Finally, the confound in the "No pre-trend" DGP has a flat event-time path until after event time.

A key ingredient to the confound dynamics is the rule for generating the policy variable. In each DGP, the policy variable starts at zero and turns to one the first period after $C_{i,t+P}$ plus an independent noise term exceeds a pre-specified threshold. This policy process provides a stylized model for a decision-maker (say, the state legislature) who makes a P-period-ahead forecast of the confound (say, the state of the economy) and then adopts the policy if the forecast is sufficiently favorable. Our simulation DGPs make use of $P = -1$ which can be thought of as a backward-looking decision-maker ("Mean-reverting trend" and "Multidimensional"), $P = 3$ ("No pre-trend"), and $P = 6$ ("Monotone trend"). The choice of P then interacts with other design parameters to produce the specific event-time dynamics in the confound.

Following the discussion in Section 3, the other important design feature that we vary is whether the confound is low-dimensional or idiosyncratic. The confound is idiosyncratic ($F_t = 0$) in the "Mean-reverting trend," "Monotone trend," and "No pre-trend" DGPs, and low-dimensional ($\xi = 0$) with two common factors in the "Multidimensional" DGP.

When the confound is idiosyncratic, we expect estimators motivated by Assumption 1, such as interactive fixed effects or synthetic controls, to perform poorly. Following our simplified discussion in Section 3, we generate η_{it} as a scalar in our simulation DGPs. In economic settings where a scalar confound is plausible, it is more likely that additional data in the form of proxies is available and that having a single proxy may be sufficient for identification via the proxy-based instrumental variables approach motivated by Assumption 3(b).

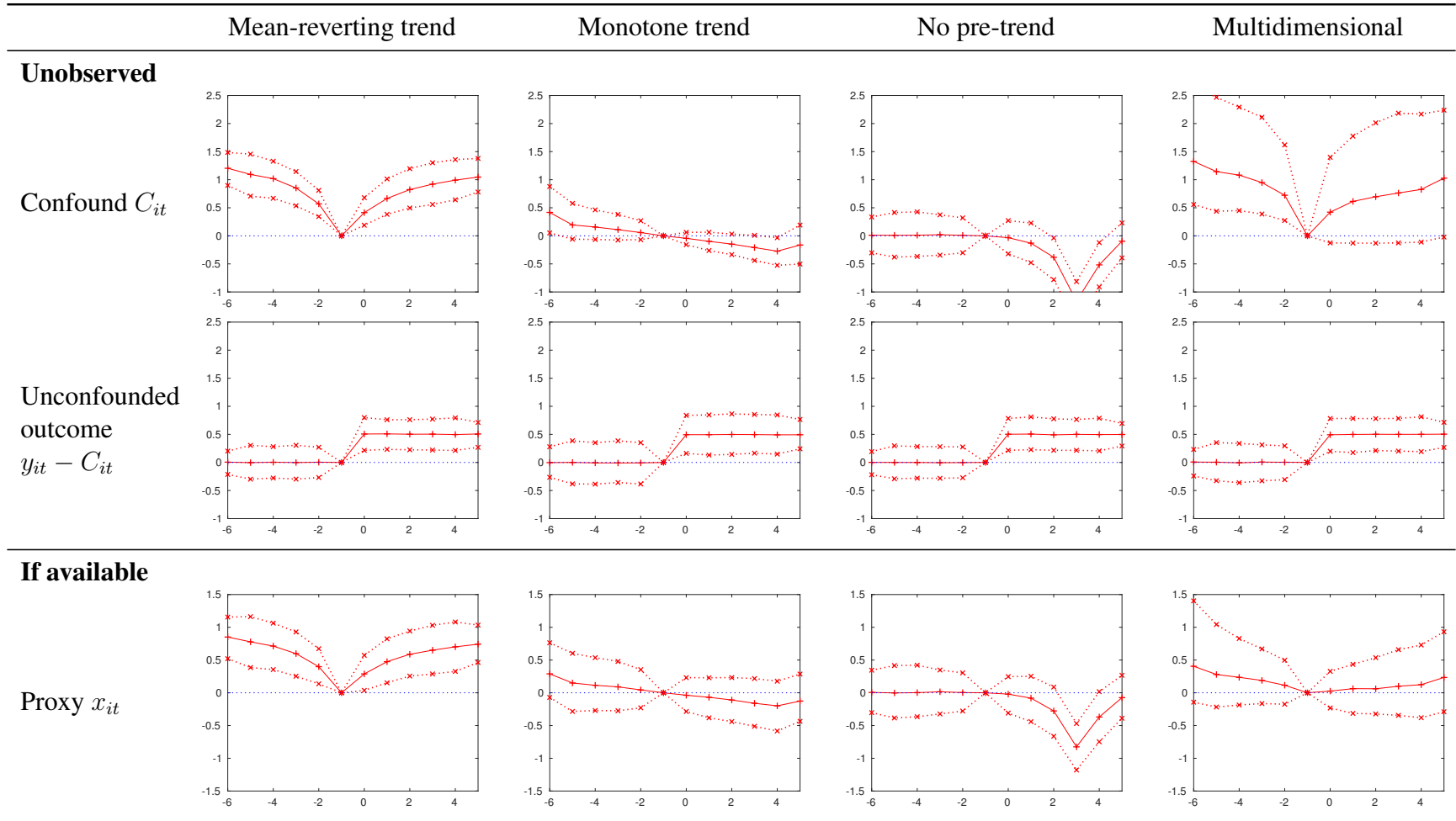**Figure 11: Graphical illustration of different outcomes across DGPs.** Estimates are obtained via two-way fixed effects estimation of (3) with the dependent variable given in the row label. For each value of k indicated on the x-axis, red series correspond to the 2.5th (dotted, marked by x's), 50th (solid, marked by +'s), and 97.5th (dotted, marked by x's) percentiles across 1000 simulations for each $\hat{\delta}_k$.

On the other hand, when the confound is low-dimensional, the approaches based on Assumption 1 will tend to be more appealing. With two common factors, the presence of multiple sources of confounding also complicates the use of proxy-based instrumental variables strategies as now one needs adequate proxies for both sources of confounding. We use the "Multidimensional" DGP to illustrate this complication by generating a proxy variable x_{it} that is only informative about one of the two factors.

The second row in Figure 11 takes the unconfounded outcome $y_{it} - C_{it}$ as the dependent variable. Producing this plot is infeasible in practice but allows us to highlight the true effect of the policy under the different DGPs. Under all four DGPs, adoption permanently increases the outcome by 0.5 units. We also see that the simulation results in this case are similar across designs suggesting that differences in the simulation results based on feasible estimators discussed below are driven by the differing behavior of the confounding variation.

The third row in Figure 11 takes the proxy x_{it} as the dependent variable. Producing this plot is feasible in practice. In the "Mean-reverting trend," "Monotone trend," and "No pre-trend" DGPs, the proxy is a noisy, linear function of the confound and therefore has event-time dynamics similar to, but noisier than, those of the confound. In the "Multidimensional" DGP, the proxy is a noisy, linear function of the first dimension of the two-dimensional confound, and therefore has different event-time dynamics from those of the confound. Since the confound is not observed in practice, a researcher will not generally know how well the event-time dynamics of the proxy mirror those of the confound, and the choice of proxy should therefore ideally be motivated on economic grounds.

4.2 Estimates

We next examine the performance of feasible estimators for the dynamic treatment effect of the policy z_{it} on the outcome y_{it} based on the approaches to identification outlined in Section 3 in each of the four DGPs. We summarize results in Figure 12. The plots are defined similarly to those in Figure 11 with the lines in the plot corresponding to the median, 2.5th and 97.5th percentile of $\hat{\delta}_k$ across 1000 realizations. These estimates of the event-time path are based on (3) with $M + L_M = 4$ and $G + L_G = 5$. Each column corresponds to a different DGP. Each row now corresponds to a different estimator. The black line in each figure gives the true values of the treatment effect for ease of reference.

The first row of Figure 12 corresponds to the two-way fixed effects estimator motivated by Assumption 1(a). Across all DGPs the estimated effect of the policy is severely median-biased, with dynamics heavily influenced by the confound.

	Mean-reverting trend	Monotone trend	No pre-trend	Multidimensional

Conditional on observables

Two-way fixed effects

Interactive fixed effects

Event-time extrapolation

Proxies and instruments

Instrumenting for x_{it} with leads of z_{it}
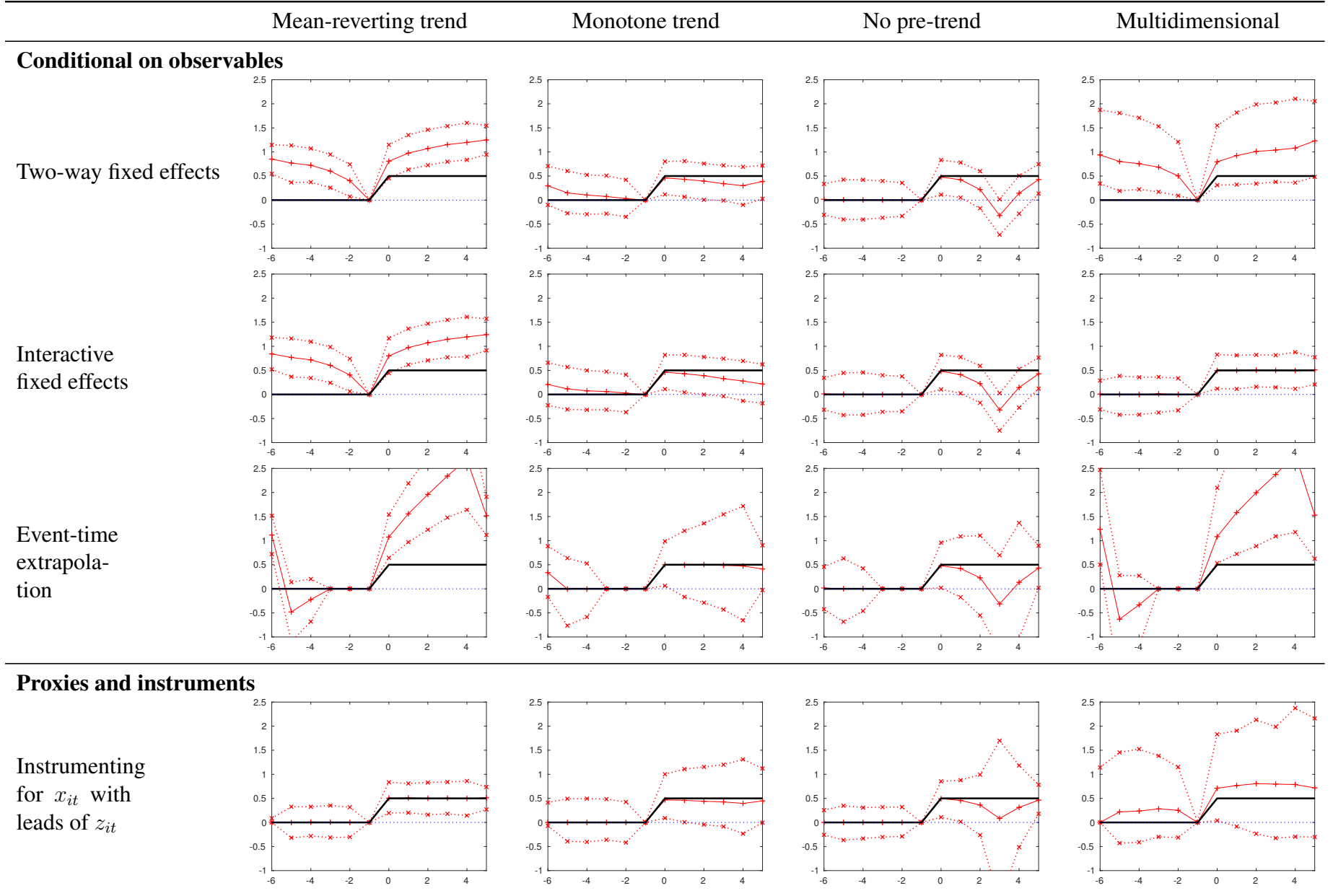
Figure 12: Graphical illustration of different estimators across DGPs. Estimates are obtained via estimation of (3) using the estimator given in the row label. True treatment effect depicted in solid black. For each value of k indicated on the x-axis, red series correspond to the 2.5th (dotted, marked by x's), 50th (solid, marked by +'s), and 97.5th (dotted, marked by x's) percentiles across 1000 simulations for each $\hat{\delta}_k$.

The second row of Figure 12 corresponds to an interactive fixed effects estimator motivated by Assumption 1(c). We set the number of estimated interactive fixed effects equal to the number of sources of confounding: one for the "Mean-reverting trend," "Monotone trend," and "No pre-trend" DGPs, and two for the "Multidimensional" DGP. Since the DGPs in the first three columns include confounding variation that is not captured by a model with shared factors, we see in these three scenarios that interactive fixed effects performs similarly to the baseline two-way fixed effects estimator. However, the interactive fixed effects estimator performs very well in the final DGP where confounding is generated by two shared factors. In this case, the interactive fixed effects estimator is essentially median-unbiased and overall appears to perform similarly to the infeasible regression that makes use of the unconfounded outcome.

The third row of Figure 12 corresponds to an estimator motivated by Assumption 2, taking $f(m) = \mathbf{1}_{|m| \leq 3}$ following Dobkin et al. (2018). This estimator exhibits substantial median bias in all DGPs except for "Monotone trend," where the estimator is approximately median-unbiased but exhibits substantial variability. Given that it is based on extrapolating pre-event dynamics into the post-event period, it seems unsurprising that this estimator exhibits median bias in DGPs in which the confound's post-event path is not well-approximated by linear extrapolation of its fitted pre-event path. Because the post-event path of the confound cannot generally be learned from the data, this finding highlights the importance of motivating assumptions such as Assumption 2 on economic grounds.

The final row of Figure 12 corresponds to an instrumental variables (IV) estimator motivated by Assumption 3(b), following Freyaldenhoven et al. (2019). To obtain the estimates in this case, we make use of the proxy depicted in Figure 11 to stand in for the unobserved confound and then instrument for this proxy using a lead of the (change in) policy variable. Under the assumption of no anticipatory effects, all leads of the policy variable are potential instruments. For instrument choice, we estimate (3) via two-way fixed effects, with the proxy as the outcome variable. We then choose, from among the leads $\{\Delta z_{i,t+k}\}_{k=G+1}^{G+L_G}$ and $z_{i,t+G+L_G+1}$, the one with the largest absolute t-statistic to serve as the excluded instrument for estimation of the structural equation. Graphically, this corresponds to imposing the additional restriction that the coefficient on this selected lead is equal to zero. Using only a single lead is appealing because it allows free estimation of the remaining pre-event coefficients and thereby a visual inspection of the identifying assumption that the remaining pre-event effects are zero.[5]

Having strong identification based on Assumption 3(b) requires that there is a strong association between leads and the proxy variable. We can see in the final row of Figure 11 that only

---

[5]Note that two zero restrictions — that δ_{-1} and the coefficient on the strongest lead from the first stage are zero — are imposed in each simulation replication. The strongest lead varies across simulation replication due to sampling variation, so the second restriction is not visually transparent in Figure 12.

the "Mean-reverting trend" DGP exhibits a strong pre-trend in the proxy. For this DGP, the final row of Figure 12 shows that the IV estimator appears to perform well, delivering essentially median-unbiased results with sampling variability that is similar to the infeasible regression using the unconfounded outcome. A smaller pre-trend in the proxy in the "Monotone trend" DGP means identification is weaker in this DGP. While the median bias remains relatively low, the weaker identification results in a wider sampling distribution. The leads of the policy variable are roughly unrelated to the policy variable in the "No pre-trend" DGP, leading to a loss of identification and a very wide sampling distribution. The IV estimator also performs poorly in the "Multidimensional confound" DGP. As noted above, the proxy is only related to one of the two sources of confounding, leaving an unaccounted-for source of confounding resulting in substantial median bias. The leads are also relatively weak in this design which leads to a widely dispersed sampling distribution.

In the Appendix, we consider five additional estimators. Specifically, Appendix Figure 1 considers an estimator that includes the proxy x_{it} directly in the controls q_{it}, and an estimator motivated by Assumption 1(b), with $f(t) = t$ and so including unit-specific linear time trends as a control. Appendix Figure 2 considers two versions of a synthetic control estimator motivated by Assumption 1(c). Finally, Appendix Figure 3 considers a measurement-error correction, assuming the availability of two proxies for the confound, motivated by Assumption 3(a).

These simulations illustrate the more general point that no estimator works uniformly across all plausible types of confounding. We also reiterate it will generally not be possible to learn the appropriate modeling approach using data-dependent model selection techniques and that it is therefore important that researchers carefully justify their identifying assumptions on economic and institutional grounds, rather than statistical convenience. The inability to uniformly determine the appropriate estimation strategy based solely on the data also suggests there is scope for sensitivity analysis in situations where multiple plausible economic stories exist.

5 Heterogeneous Effects of the Policy in a Staggered Adoption Setting

In this section we explore the implications of relaxing the assumption in (1) that the effect of the policy is identical across units. In economic applications, heterogeneity in policy effects across units i may arise for many reasons, including economic differences between the units. For example, minimum wage laws may have different effects on employment in different geographic regions (e.g., Wang et al. 2019). In line with some recent literature (e.g., Athey and Imbens forthcoming; Callaway and Sant'Anna forthcoming; Sun and Abraham forthcoming; Goodman-Bacon 2020), we focus on the staggered adoption setting. Accordingly, let $t^*(i)$ denote the period in which unit i adopts the policy, which we will refer to as the unit's *cohort*.

First, consider the possibility that the causal effect β of the policy on the outcome differs across cohorts t^*. Denoting the causal effect for cohort t^* by β_{t^*}, we can modify equation (1) as follows:

$$y_{it} = \alpha_i + \gamma_t + q'_{it}\psi + \sum_{m=-G}^{M} \beta_{m,t^*(i)}z_{i,t-m} + C_{it} + \varepsilon_{it}. \tag{13}$$

The approaches to identification discussed in Section 3 can then be applied to recover the parameters β_{t^*} in (13).

Next, consider the possibility that the causal effect β of the policy on the outcome differs across units. In this case, estimates of β_{m,t^*} in (13) need not be valid estimates of a proper weighted average of the unit-specific policy effects $\beta_{m,i}$ for units in the corresponding cohort. An exception is a setting with some never-treated units, no included control variables ($\psi = 0$), and a parallel trends assumption that rules out all the forms of confounding considered in Section 3 outside of Assumption 1(a). In such a setting, Sun and Abraham (forthcoming) establish that estimates of β_{m,t^*} obtained from the usual two-way fixed effects estimator applied to (13) are valid estimates of a proper weighted average of the unit-specific policy effects $\beta_{m,i}$ for units in the corresponding cohort.[6] Developing approaches to estimate a proper weighted average of unit-specific policy effects in the presence of the various forms of confounding considered in Section 3 seems a useful direction for future work.

Because we expect parallel trends to fail under most of the forms of confounding considered in Section 3, we expect estimators that are designed to be robust to heterogeneity in treatment effects under the assumption of parallel trends to exhibit median bias in our DGPs. Appendix Figure 4 illustrates this in our simulated DGPs for the estimators proposed by de Chaisemartin and D'Haultfoeuille (2020b), Sun and Abraham (forthcoming), and Borusyak and Jaravel (2018).

If the policy has different effects on different units, it may be natural to assume that latent sources of confounding likewise have different effects on different units. Bonhomme and Manresa (2015) and Su et al. (2016) develop approaches to recovering heterogeneous policy effects in models where coefficients are homogeneous within subgroups of observations but heterogeneous across groups, and where group membership is unknown and must be inferred from the data. Pesaran (2006) provides an approach to recover heterogeneous policy effects that differ across units but are constant over time, while allowing the unobserved confounds to have an interactive fixed effects structure, as in Assumption 1(c).

In some situations, we may expect that policy effects differ across periods t as well as units i,

---

[6]Another setting of interest is one with no included control variables, parallel trends, and only contemporaneous effects of the policy, i.e. $\beta_{m,i} = 0$ for all $m \neq 0$. In such a setting, results in de Chaisemartin and D'Haultfoeuille (2020b, Online Appendix Section 3.1) imply that an estimate of β from a static two-way fixed effects estimator applied to (1) is a valid estimate of a proper weighted average of unit-specific policy effects.

say because the effect of the policy depends on aggregate macroeconomic conditions whose influence further depends on specific features of the unit-level environment. Within this setting, Athey et al. (2021) show that matrix completion methods can be used to recover heterogeneous effects of the policy under the assumption of no dynamic effects and with no included control variables. Feng et al. (2017) provide estimators of heterogeneous effects that are unit-specific and allowed to depend on observed time-varying categorical variables. Su and Wang (2017) also provide estimators of time-varying unit-specific heterogeneous effects under the restriction that effects vary smoothly over time. Finally, Chernozhukov et al. (2019) provide estimators of heterogeneous policy effects that vary across units and over time under a factor structure for both slopes of observed variables and the unobserved confounds.

6   Conclusion

The event-study plot is a widely used statistical tool that conveys helpful information about the data when estimating policy effects. In Section 2, we suggest some ways to make it more informative. Making the plot and the underlying data useful for causal or structural inference requires economic assumptions about the setting at hand. In Section 3, we review some approaches to identification, discuss some connections among them, and discuss their economic interpretation with examples. The discussion makes clear that no approach applies uniformly to all situations. In Section 4, we demonstrate this by applying several different estimators to a set of simulated DGPs designed to evoke a state-level panel with a policy change. We show that the performance of a given estimator can vary greatly depending on features of the underlying DGP that cannot typically be inferred from the data. Although most of our discussion assumes homogeneous policy effects, in Section 5 we connect with an important recent literature by discussing how to extend the approaches in Section 3 to allow for some forms of heterogeneity in policy effects. There, we also discuss the role of confounding in settings with such heterogeneity.

References

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505, 2010.

Philippe Aghion, Céline Antonin, Simon P. Bunel, and Xavier Jaravel. What are the labor and product market effects of automation? New evidence from France. Presentation to the National Bureau of Economic Research (NBER) Summer Institute Labor Studies, 2020.

Dennis J. Aigner, Cheng Hsiao, Arie Kapteyn, and Tom Wansbeek. Latent variable models in econometrics. In Z. Griliches and M. D. Intriligator, editors, *Handbook of Econometrics*, volume 2, chapter 23, pages 1321–1393. Elsevier, 1984.

Orley Ashenfelter. Estimating the effect of training programs on earnings. *Review of Economics and Statistics*, 60(1):47–57, 1978.

Susan Athey and Guido W. Imbens. Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431 – 497, 2006.

Susan Athey and Guido W Imbens. Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, forthcoming.

Susan Athey, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, pages 1–41, 2021.

Shalise Ayromloo, Benjamin Feigenberg, and Darren Lubotsky. Employment eligibility verification requirements and local labor market outcomes. Presentation to the National Bureau of Economic Research (NBER) Summer Institute Labor Studies, 2020.

Jushan Bai. Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279, 2009.

Emek Basker. Selling a cheaper mousetrap: Wal-mart's effect on retail prices. *Journal of Urban Economics*, 58(2):203–229, 2005.

Eli Ben-Michael, Avi Feller, and Jesse Rothstein. Synthetic controls and weighted event studies with staggered adoption. arXiv:1912.03290, 2021.

Timothy Besley and Anne Case. Unnatural experiments? Estimating the incidence of endogenous policies. *Economic Journal*, 110(467):672–694, 2000.

Stéphane Bonhomme and Elena Manresa. Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184, 2015.

Kirill Borusyak and Xavier Jaravel. Revisiting event study designs. Working paper, 2018.

Charles Brown. Minimum wages, employment, and the distribution of income. In Orley C. Ashenfelter and David Card, editors, *Handbook of Labor Economics*, volume 3, chapter 32, pages 2101 – 2163. Elsevier, 1999.

Brantly Callaway and Pedro H. C. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, forthcoming.

Matias D. Cattaneo, Luke Keele, Rocío Titiunik, and Gonzalo Vazquez-Bare. Extrapolating treatment effects in multi-cutoff regression discontinuity designs. *Journal of the American Statistical Association*, 0(0):1–12, 2020.

Victor Chernozhukov, Christian Hansen, Yuan Liao, and Yinchu Zhu. Inference for heterogeneous effects using low-rank estimation of factor slopes. arXiv:1812.08089, 2019.

Victor Chernozhukov, Kaspar Wüthrich, and Yinchu Zhu. An exact and robust conformal inference method for counterfactual and synthetic controls. arXiv:1712.09089, 2020.

Clément de Chaisemartin and Xavier D'Haultfoeuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9), 2020a.

Clément de Chaisemartin and Xavier D'Haultfoeuille. Difference-in-differences estimators of intertemporal treatment effects. arXiv:2007.04267, 2020b.

Ellora Derenoncourt, Clemens Noelke, and David Weil. Spillover effects from voluntary employer minimum wages. Presentation to the National Bureau of Economic Research (NBER) Summer Institute Labor Studies, 2020.

Carlos Dobkin, Amy Finkelstein, Raymond Kluender, and Matthew J. Notowidigdo. The economic consequences of hospital admissions. *American Economic Review*, 108(2):308–52, 2018.

Arindrajit Dube and Ben Zipperer. Pooling multiple case studies using synthetic controls: An application to minimum wage policies. IZA Discussion Paper No. 8944, 2015.

Guohua Feng, Jiti Gao, Bin Peng, and Xiaohui Zhang. A varying-coefficient panel data model with fixed effects: Theory and an application to US commercial banks. *Journal of Econometrics*, 196 (1):68–82, 2017.

Simon Freyaldenhoven, Christian Hansen, and Jesse M. Shapiro. Pre-event trends in the panel event-study design. *American Economic Review*, 109(9):3307–3338, 2019.

Joachim Freyberger and Yoshiyasu Rai. Uniform confidence bands: Characterization and optimality. *Journal of Econometrics*, 204(1):119–130, 2018.

Leora Friedberg. Did unilateral divorce raise divorce rates? Evidence from panel data. *American Economic Review*, 88(3):608–627, 1998.

Matthew Gentzkow, Jesse M. Shapiro, and Michael Sinkinson. The effect of newspaper entry and exit on electoral politics. *American Economic Review*, 101(7):2980–3018, 2011.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. Working paper, 2020.

Zvi Griliches and Jerry A Hausman. Errors in variables in panel data. *Journal of Econometrics*, 31(1):93–118, 1986.

Justine S. Hastings, Ryan Kessler, and Jesse M. Shapiro. The effect of snap on the composition of purchased foods: Evidence and implications. *American Economic Journal: Economic Policy*, forthcoming.

Catherine Hausman and David S. Rapson. Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics*, 10(1):533–552, 2018.

James Heckman and Jose Scheinkman. The importance of bundling in a gorman-lancaster model of earnings. *Review of Economic Studies*, 54(2):243–255, 1987.

Louis S. Jacobson, Robert J. LaLonde, and Daniel G Sullivan. Earnings losses of displaced workers. *American Economic Review*, 83(4):685–709, 1993.

David Neumark and William L. Wascher. Minimum wages and employment. *Foundations and Trends in Microeconomics*, 3(1–2):1–182, 2007.

José Luis Montiel Olea and Mikkel Plagborg-Møller. Simultaneous confidence bands: Theory, implementation, and an application to SVARs. *Journal of Applied Econometrics*, 34(1), 2019.

M. Hashem Pesaran. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012, 2006.

David Powell. Synthetic control estimation beyond case studies: Does the minimum wage reduce employment? Working paper, 2021.

Ashesh Rambachan and Jonathan Roth. An honest approach to parallel trends. Working paper, 2020.

Jonathan Roth. Pre-test with caution: Event-study estimates after testing for parallel trends. Working paper, 2020.

Jonathan Roth and Pedro H. C. Sant'Anna. When is parallel trends sensitive to functional form? arXiv:2010.04814, 2021.

Kurt Schmidheiny and Sebastian Siegloch. On event study designs and distributed-lag models: Equivalence, generalization and practical implications. CEPR Discussion Paper 13477, 2020.

Azeem Shaikh and Panos Toulis. Randomization tests in observational studies with staggered adoption of treatment. arXiv:1912.10610, 2019.

Liangjun Su and Xia Wang. On time-varying factor models: Estimation and testing. *Journal of Econometrics*, 198(1):84–101, 2017.

Liangjun Su, Zhentao Shi, and Peter C.B. Phillips. Identifying latent structures in panel data. *Econometrica*, 84(6):2215–2264, 2016.

Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, forthcoming.

Wuyi Wang, Peter C.B. Phillips, and Liangjun Su. The heterogeneous effects of the minimum wage on employment across states. *Economics Letters*, 174:179 – 185, 2019.

Justin Wolfers. Did unilateral divorce laws raise divorce rates? A reconciliation and new results. *American Economic Review*, 96(5):1802–1820, 2006.