Banco de México

Working Papers

N° 2022-07

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
Harvard University and NBER

August 2022

La serie de Documentos de Investigación del Banco de México divulga resultados preliminares de trabajos de investigación económica realizados en el Banco de México con la finalidad de propiciar el intercambio y debate de ideas. El contenido de los Documentos de Investigación, así como las conclusiones que de ellos se derivan, son responsabilidad exclusiva de los autores y no reflejan necesariamente las del Banco de México.

The Working Papers series of Banco de México disseminates preliminary results of economic research conducted at Banco de México in order to promote the exchange and debate of ideas. The views and conclusions presented in the Working Papers are exclusively the responsibility of the authors and do not necessarily reflect those of Banco de México.

Documento de Investigación
2022-07

Working Paper
2022-07

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**
Harvard University and NBER

Abstract: Linear panel models, and the "event-study plots" that often accompany them, are popular tools for learning about policy effects. We discuss the construction of event-study plots and suggest ways to make them more informative. We examine the economic content of different possible identifying assumptions. We explore the performance of the corresponding estimators in simulations, highlighting that a given estimator can perform well or poorly depending on the economic environment. An accompanying Stata package, xtevent, facilitates adoption of our suggestions.
Keywords: Linear Panel Data Models, Difference-in-Differences, Staggered Adoption, Pre-Trends, Event Study
JEL Classification: C23, C52

Resumen: Los modelos lineales de panel, y los "gráficos de estudio de evento" que usualmente los acompañan, son herramientas populares para aprender acerca de efectos de políticas. Discutimos la construcción de gráficos de estudio de evento y sugerimos maneras de hacerlos más informativos. Examinamos el contenido económico de diferentes posibles supuestos de identificación. Exploramos el desempeño de los estimadores correspondientes en simulaciones, enfatizando que un estimador particular puede tener un buen o mal desempeño dependiendo del contexto económico. El paquete acompañante de Stata -xtevent- facilita la adopción de nuestras sugerencias.
Palabras Clave: Modelos Lineales de Datos de Panel, Diferencias en Diferencias, Adopción Escalonada, Tendencias Previas, Estudio de Evento

---

*This is a draft of a chapter in progress for Advances in Economics and Econometrics: Twelfth World Congress, an Econometric Society monograph. Detailed acknowledgments and funding sources can be found at: https://jorgeperezperez.com/files/EventStudy.pdf. The views expressed herein are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia, the Federal Reserve System, Banco de México, or the funding sources. To install the Stata package associated with this document, please visit https://www.banxico.org.mx/DataSetsWeb/dataset-7.zip or type ssc install xtevent in Stata.

† Email: simon.freyaldenhoven@phil.frb.org.

‡ Booth School of Business. Email: chansen1@chicagobooth.edu.

§ Dirección General de Investigación Económica. Email: jorgepp@banxico.org.mx.

** Department of Economics. Email: jesse_shapiro@fas.harvard.edu.

1 Introduction

A common situation in empirical economics is one where we are interested in learning the dynamic effect of a scalar policy z_{it} on some outcome y_{it} in an observational panel of units $i \in \{1,...,N\}$ observed in a sequence of periods $t \in \{1,...,T\}$. Examples include

- the effect of participation in a training program (e.g., Ashenfelter 1978), where units i are individuals, the policy variable z_{it} is an indicator for program participation, and the outcome y_{it} is earnings;
- the effect of the minimum wage (e.g., Brown 1999), where units i are US states, the policy variable z_{it} is the level of the minimum wage, and the outcome y_{it} is the employment rate of some group; and
- the effect of Walmart (e.g., Basker 2005), where units i are cities, the policy variable z_{it} is an indicator for the presence of a Walmart, and the outcome y_{it} is a measure of retail prices.

A common model for representing this situation is a linear panel model with dynamic policy effects:

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

Here, α_i denotes a unit fixed effect, γ_t a time fixed effect, and q_{it} a vector of controls with conformable coefficients ψ. The scalar C_{it} denotes a (potentially unobserved) confound that may be correlated with the policy, and the scalar ε_{it} represents an unobserved shock that is not correlated with the policy. Estimates of models of the form in (1) are common in economic applications.[1]

The term $\sum_{m=-G}^{M} \beta_m z_{i,t-m}$ means that the policy can have dynamic effects. The outcome at time t can only be directly affected by the value of the policy at most $M \geq 0$ periods before t and at most $G \geq 0$ periods after t. The values of G and M are often assumed by the researcher. The parameters $\{\beta_m\}_{m=-G}^{M}$ summarize the magnitude of the dynamic effects and are often estimated. Estimates of $\{\beta_m\}_{m=-G}^{M}$ are often summarized in an "event-study plot" of the form

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

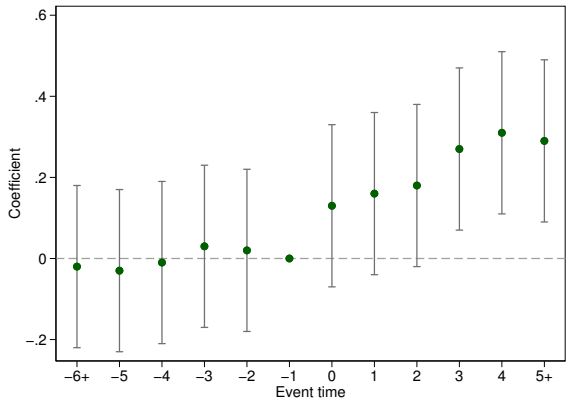


Figure 1: Exemplary event-study plot. The plot shows a hypothetical example of an event-study plot. See Section 2 for details on the construction of event-study plots.

in Figure 1, which we may loosely think of as depicting estimates of the cumulative effects $\sum_{m=-G}^{k} \beta_m$ at different horizons k.[2]

In economic settings, variation in the policy may be related to other determinants of the outcome. For example, a person's entry into a training program may reflect unobserved shocks to earnings, a state's decision to increase its minimum wage may be influenced by its economic performance, and Walmart's arrival in a city may signal trends in the local retail environment. Such factors make it challenging to identify $\{\beta_m\}_{m=-G}^{M}$, and are captured in (1) via the confound C_{it}. If the confound is unobserved, identification of $\{\beta_m\}_{m=-G}^{M}$ typically requires substantive restrictions on C_{it} that cannot be learned entirely from the data and must therefore be justified on economic grounds.

In this chapter, we have three main goals. The first, which we take up in Section 2, is to make suggestions on the construction of event-study plots of the form in Figure 1. These suggestions aim at improving the informativeness of these plots. They involve a mix of codifying what we consider common best practices, as well as suggesting some practices that are not currently in use. An accompanying Stata package, `xtevent`, makes it easier for readers to adopt our suggestions.

The second, which we take up in Section 3, is to consider possible approaches to identification of $\{\beta_m\}_{m=-G}^{M}$. Each approach requires some restrictions on the confound C_{it}. In accordance with our view that such restrictions should be motivated on economic grounds, our discussion emphasizes and contrasts the economic content of different possible restrictions.

The third, which we take up in Section 4 and an accompanying Appendix, is to illustrate the

[2]Roth (2021) identifies 70 papers published in three American Economic Association journals between 2014 and June 2018 that include such a plot.

performance of different estimators under some specific data-generating processes. We choose data-generating processes motivated by the economic settings of interest, and estimators corresponding to the approaches to identification that we discuss in Section 3. Our simulations highlight that there is not one "best" estimator—a given estimator may perform well or poorly depending on the underlying economic environment. The simulation results reinforce the importance of matching identifying assumptions (and the corresponding estimator) to the setting at hand.

The model in (1) is flexible in some respects. It allows the policy to have dynamic effects before and after its contemporaneous value. Additionally it does not restrict the form of z_{it}, which may in principle be continuous, discrete, or binary. One important case, referenced in the literature and throughout our discussion, is that of *staggered adoption*, in which 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 $t' \geq t$.[3]

The model in (1) is also restrictive in many respects. It assumes that all terms enter in a linear, separable way.[4] It includes neither lagged dependent variables nor predetermined variables. The model in (1) also assumes that the effect of the policy is homogeneous across units i. An active literature explores the implications of relaxing this homogeneity.[5] Although we maintain homogeneity throughout most of the exposition, we discuss in Section 3 how to adapt the approaches to identification we consider there to allow for some forms of heterogeneous policy effects, and we illustrate in Section 4 how to estimate policy effects in the presence of both heterogeneity and confounding. An important theme of our discussion is that accounting for heterogeneity in policy effects is not a substitute for accounting for confounding, and vice versa.

2 Plotting

For this section, we assume the researcher has adopted assumptions sufficient for the identification and estimation of $\{\beta_m\}_{m=-G}^{M}$, and consider how to plot the resulting estimates for the reader. We return to the question of how to obtain such estimates in later sections. Thus, we do not explicitly consider specific restrictions on the confound C_{it} in this section, but we emphasize that all of the approaches to restricting the confound that we discuss in Section 3,

[3] See, for example, Shaikh and Toulis (2021), Ben-Michael et al. (2021), and Athey and Imbens (forthcoming).

[4] Athey and Imbens (2006) and Roth and Sant'Anna (2021), among others, consider nonlinear models in the canonical difference-in-differences setting.

[5] See, for example, de Chaisemartin and D'Haultfoeuille (2020), Athey and Imbens (forthcoming), Goodman-Bacon (2021), Callaway and Sant'Anna (2021), and Sun and Abraham (2021).

and all of the corresponding estimators that we employ in Section 4, can be applied in tandem with the approaches to plotting that we develop in this section.

2.1 The Basics

There are two important ways in which a typical event-study plot departs from a plot of the estimates of $\{\beta_m\}_{m=-G}^{M}$ in (1). The first is that, for visual reasons, it is often appealing to depict cumulative estimated effects of the policy. To simplify estimating cumulative effects, we adopt a representation of (1) in terms of changes in the policy z_{it}. The second is that, to permit visualization of overidentifying information, it is often appealing to plot cumulative estimated effects of the policy at horizons outside of the range of horizons over which the policy is thought to affect the outcome. To make it possible to include such estimates on the plot, we generalize (1) to entertain that a change in the policy at time t leads to a change in the outcome $L_G \geq 0$ periods before $t-G$ and $L_M \geq 0$ periods after $t+M$.

Specifically, we estimate the following regression model:

$$\begin{aligned} y_{it} = &\sum_{k=-G-L_G}^{M+L_M-1} \delta_k \Delta z_{i,t-k} + \delta_{M+L_M} z_{i,t-M-L_M} + \delta_{-G-L_G-1}(1 - z_{i,t+G+L_G}) \\ &+ \alpha_i + \gamma_t + q_{it}'\psi + C_{it} + \varepsilon_{it}, \end{aligned} \tag{2}$$

where Δ denotes the first difference operator. To build intuition for the form of (2), note that in the special case of staggered adoption, $\Delta z_{i,t-k}$ is an indicator for whether unit i adopted the policy exactly k periods before period t, $z_{i,t-M-L_M}$ is an indicator for whether unit i adopted at least $M+L_M$ periods before period t, and $(1-z_{i,t+G+L_G})$ is an indicator for whether unit i will adopt more than $G+L_G$ periods after period t.

The parameters $\{\delta_k\}_{k=-G-L_G-1}^{k=M+L_M}$ can be interpreted as cumulative policy effects at different horizons. In particular, (1) implies that

$$\delta_k = \begin{cases} 0 \quad \text{for} \quad k < -G \\ \sum_{m=-G}^{k} \beta_m \quad \text{for} -G \leq k \leq M \\ \sum_{m=-G}^{M} \beta_m \quad \text{for} \quad k > M. \end{cases} \tag{3}$$

See, for example, Schmidheiny and Siegloch (2020) and Sun and Abraham (2021). The representation in (3) holds for general forms of the policy z_{it}, not only for the case of staggered adoption.

The core of an event-study plot, illustrated by Figure 1, is a plot of the points

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

where $\{\hat{\delta}_k\}_{k=-G-L_G-1}^{k=M+L_M}$ are the estimates of $\{\delta_k\}_{k=-G-L_G-1}^{k=M+L_M}$. We refer to k as *event time*, to the vector $\delta = (\delta_{-G-L_G-1}, ..., \delta_{M+L_M})'$ as the *event-time path* of the outcome variable, and to the corresponding estimate $\hat{\delta}$ as the estimated event-time path of the outcome variable. In the case of staggered adoption, the "event" is the adoption of the policy. Recall, however, that we allow for more general forms of the policy z_{it}, including that it exhibits multiple changes or "events," or that it is continuous. We include a + symbol in the x-axis label for event times $k = -G - L_G - 1$ and $k = M + L_M$ as a reminder of the interpretation of the corresponding terms in (2).

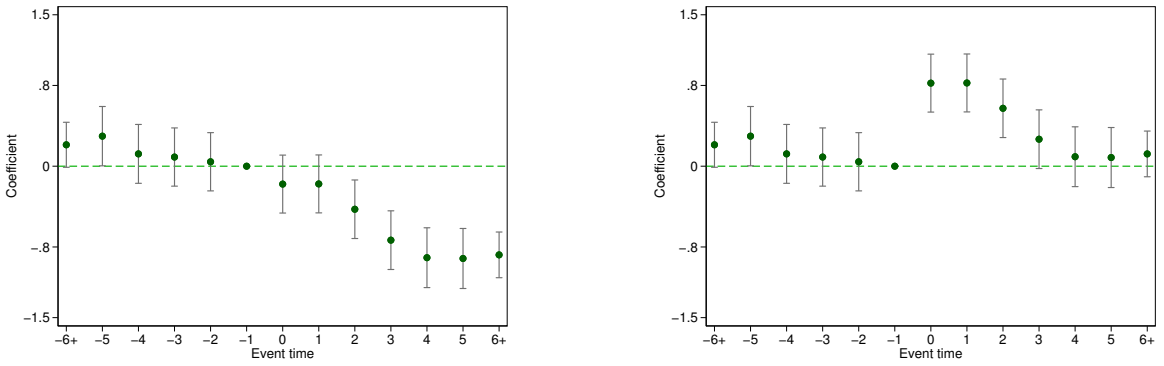
2.2 Normalization

The event-study plot is meant to illustrate the cumulative effect of the policy on the outcome. The effect of the policy must be measured with reference to some baseline. This manifests in (2) through the fact that the policy variables in the equation are collinear (they sum to zero), meaning that a normalization is required for identification of the event-time path. Indeed, one reason it is valuable to include the endpoint variables $z_{i,t+G+L_G}$ and $z_{i,t-M-L_M}$ in (2) is that, by saturating the model, inclusion of these variables forces a conscious choice of normalization.

A common choice of normalization, and one that we think serves as a good default, is $\delta_{-G-1} = 0$. In the leading case where $G = 0$, i.e., future values of the policy cannot affect the current value of the outcome, this default implies a normalization $\delta_{-1} = 0$. In the case of staggered adoption, the normalization $\delta_{-1} = 0$ means that the plotted coefficients can be interpreted as estimated effects relative to the period before adoption. More generally, the normalization $\delta_{-1} = 0$ means that the plotted coefficients can be interpreted as estimated effects relative to the effect of a one-unit change in the policy variable one period ahead.

Suggestion 1. *Normalize* $\delta_{-G-1} = 0$ *when estimating* (2).

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." The figure includes pointwise 95 percent confidence intervals for the elements of the event-time path. The normalization $\delta_{-1} = 0$ makes it easy to use the depicted confidence intervals to test the point hypothesis that the event-time path is identical at event times -1 and



(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. Each plot depicts the elements of the estimated event-time path of the outcome, $\hat{\delta}_k$, on the y-axis, against event time, k, on the x-axis. The intervals are pointwise 95 percent confidence intervals for the corresponding elements, δ_k, of the event-time path of the outcome. In the left panel the estimated event-time path of the outcome is "smooth." In the right panel the estimated event-time path of the outcome exhibits a "jump" at the time of the event. The value of δ_{-1} has been normalized to 0, in accordance with Suggestion 1.

k, i.e., $H_0 : \delta_{-1} = \delta_k$. For example, setting $k = 0$, a reader of Figure 2(a) can readily conclude that there is no statistically significant change in the outcome at event time 0. A reader of Figure 2(b) can readily conclude that there is a statistically significant change in the outcome at event time 0.

2.3 Magnitude

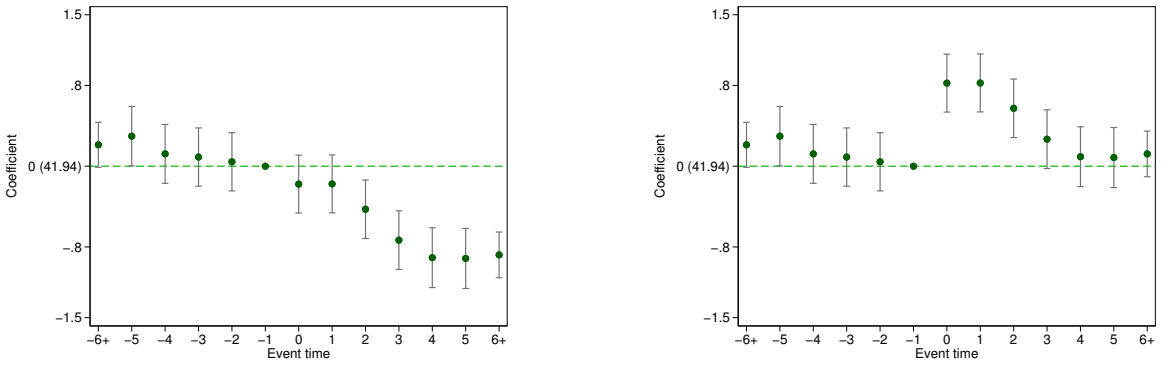
A given cumulative effect of the policy can have a different interpretation depending on the baseline level of the dependent variable. To make it easier for a reader to evaluate economic significance, we suggest to include a parenthetical label for the normalized coefficient δ_{k^*} that conveys a reference value for the outcome.

Suggestion 2. *Include a parenthetical label for the normalized coefficient δ_{k^*} that has value*

$$\frac{\sum_{(i,t):\Delta z_{i,t-k^*}\neq 0} y_{it}}{|(i,t) : \Delta z_{i,t-k^*} \neq 0|}. \tag{4}$$

The value in (4) corresponds to the mean of the dependent variable ($-k^*$) periods in advance of a policy change.[6] For example, if $G = 0$, then following Suggestion 1 we would normalize

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



(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 2.

$\delta_{-1} = 0$, and following Suggestion 2 we would include a parenthetical label with value $\frac{\sum_{(i,t):\Delta z_{i,t+1}\neq 0} y_{it}}{|(i,t):\Delta z_{i,t+1}\neq 0|}$. In the special case of staggered adoption, this expression corresponds to the sample mean of y_{it} one period before adoption. More generally, it corresponds to the sample mean of y_{it} one period before a policy change.

Figure 3 illustrates Suggestion 2 by adding the proposed parenthetical label next to the 0 point on the y-axis. The parenthetical label makes it easier to interpret the magnitude of the estimated cumulative policy effects depicted on the plot. For example, in Figure 3(a), the coefficient estimate $\hat{\delta}_3$ implies that the estimated cumulative effect of the policy on the outcome at event time 3, measured relative to event time -1, is roughly -0.73. This effect appears to be modest relative to 41.94, the sample mean of the dependent variable one period in advance of a policy change, shown in the parenthetical label.

2.4 Inference

It is common for researchers to include 95 percent pointwise confidence intervals for the elements of the event-time path, as we have done in Figures 1 through 3. As we discuss in Section 2.2, these confidence intervals make it easy for a reader to answer prespecified statistical questions about the value of an individual element of the event-time path, relative to that of the normalized element.

A reader may also be interested in answering statistical questions about the entire event-time path, rather than merely a single element. Because such questions involve multiple parameters and will typically not be prespecified, answering them reliably requires accounting

for the fact that the reader is implicitly testing multiple hypotheses at once. A convenient way to do this is to augment the plot by adding a 95 percent uniform confidence band. Just as a 95 percent pointwise confidence interval is designed to contain the true value of an individual parameter at least 95 percent of the time, a 95 percent uniform confidence band is designed to contain the true value of a set of parameters at least 95 percent of the time.

We specifically suggest plotting a sup-t confidence band. A sup-t band is straightforward to visualize in an event-study plot and can be computed fairly easily in the settings we consider by using draws from the estimated asymptotic distribution of $\hat{\delta}$ to compute a critical value. Interested readers may see, for example, Freyberger and Rai (2018) or Olea and Plagborg-Møller (2019) for details.

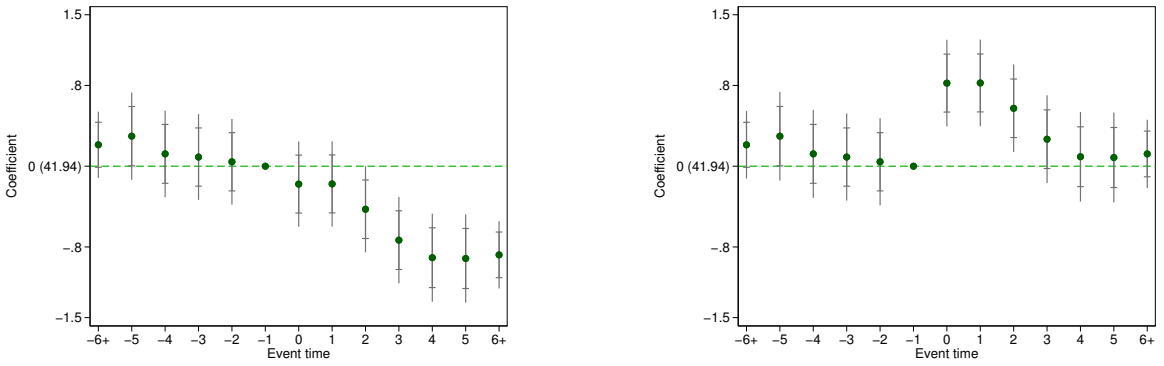
Suggestion 3. *Plot a uniform sup-t confidence band for the event-time path of the outcome δ, in addition to pointwise confidence intervals for the individual elements of the event-time path.*

Figure 4 illustrates Suggestion 3 by adding a set of outer lines comprising the sup-t band to the plots in Figure 3, retaining the pointwise confidence intervals as the inner bars. Just as any value of an element of the event-time path that lies outside the inner confidence interval can be regarded as statistically inconsistent with the corresponding estimate, any event-time path that does not pass entirely within the outer confidence band can be regarded as statistically inconsistent with the estimated path. For example, in both plots, we cannot reject the hypothesis that the event-time path is equal to zero in all pre-event periods using the uniform confidence bands. On the other hand, we would conclude that $\hat{\delta}_{-5}$ is statistically significant using the pointwise intervals. Unless we had prespecified that we were specifically interested in the null hypothesis that $\delta_{-5} = 0$, the first conclusion based on the uniform confidence bands appears more suitable.

2.5 Overidentification and Testing

The model in (1) asserts that changes in the policy variable more than G periods in the future do not change the outcome. Taking $L_G > 0$ and thus including event times earlier than $-G$ in the plot means that the plot contains information about whether this hypothesis is consistent with the data. A failure of this hypothesis might indicate anticipatory behavior, or, in what seems a more common interpretation in practice, it might indicate the presence of a confound. Because in many situations $G = 0$, researchers sometimes refer to a test of this hypothesis as a test for pre-event trends or "pre-trends."

The model in (1) also asserts that changes in the policy variable more than M periods in the past do not change the outcome. Taking $L_M > 0$ and thus including event times after M in



(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 3. The pointwise confidence intervals are illustrated by the inner bars, and the uniform, sup-t confidence bands are given by the outer lines.

the plot means that the plot contains information about whether this hypothesis is consistent with the data. A failure of this hypothesis might indicate richer dynamic effects of the policy than the researcher has contemplated.

Each of these hypotheses can be tested formally via a Wald test. The result of the Wald test cannot generally be inferred from the pointwise confidence intervals or uniform confidence band. We therefore propose to include the p-values corresponding to each of the two Wald tests.

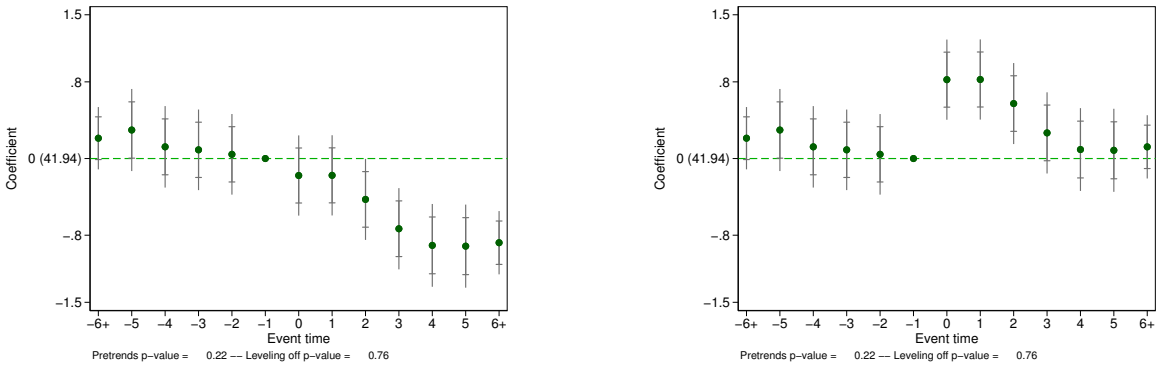
Suggestion 4. *Include in the plot legend p-values for Wald tests of each of the following hypotheses:*

$$H_0 : \delta_k = 0, \qquad -G - L_G - 1 \leq k < -G \qquad \text{(no pre-trends)}$$

$$H_0 : \delta_M = \delta_{M+k}, \qquad 0 < k \leq L_M \qquad \text{(dynamics level off)}$$

Figure 5 illustrates Suggestion 4 by adding the two suggested p-values to the plots from Figure 4, taking $G = 0$ and $L_M = 1$. In both cases, the tested hypotheses are not rejected at conventional levels.

The exact specification of the hypotheses listed in Suggestion 4 depends on the choice of L_G and L_M. Formally, these control the number of overidentifying restrictions that (1) imposes on (2). We do not offer a formal procedure for choosing L_G and L_M. Instead, we suggest a default rule of thumb that we expect will be reasonable in many cases.



(a) "Smooth" event-time trend

(b) "Jump" at the time of the event

Figure 5: Testing p-values. Exemplary event-study plot for two possible datasets. Relative to Figure 4, 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.

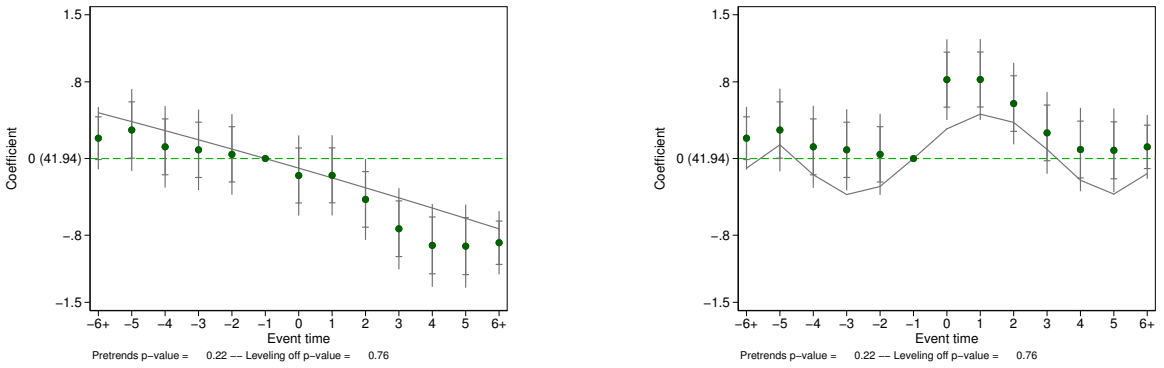
Suggestion 5. *Set* $L_M = 1$ *and set* $L_G = M + G$.

Setting $L_M = 1$ guarantees that it is possible to test the researcher's hypothesis that changes in the policy variable more than M periods in the past do not change the outcome. Setting $L_G = M + G$ makes the plot "symmetric," in the sense that it permits testing for pre-trends over as long of a horizon as the policy is thought to affect the outcome. If we take $M = 5$ and $G = 0$, then Figure 5 is consistent with Suggestion 5.

2.6 Confounds

A failure to reject the hypothesis of no pre-trends does not mean there is no confound, or even that there are no pre-trends (Roth 2021). Any estimated event-time path for the outcome can be reconciled with a devil's-advocate model in which the policy has no effect on the outcome, i.e., in which the true values of $\{\beta_m\}_{m=-G}^{M}$ are zero. In this devil's-advocate model, all of the dynamics of the estimated event-time path of the outcome variable are driven by an unmeasured confound.

In some cases, the sort of confound needed to fully explain the event-time path of the outcome may be economically implausible. To help the reader evaluate the plausibility of the confounds that are consistent with the data, we propose to add to the event-study plot a representation of the "most plausible" confound that is statistically consistent with the estimated event-time path of the outcome. There is no universal notion of what is plausible, but we expect that in many situations, less "wiggly" confounds will be more plausible than more "wiggly" ones.



(a) "Smooth" event-time trend

(b) "Jump" at the time of the event

Figure 6: Least "wiggly" path of confound. Exemplary event-study plot for two possible datasets. Relative to Figure 5, a curve has been added that illustrates the least "wiggly" confound that is consistent with the event-time path of the outcome, in accordance with Suggestion 6.

Among polynomial confounds that are consistent with the estimated event-time path of the outcome, we propose to define the least "wiggly" one as the one that has the lowest possible polynomial order and that, among polynomials with that order, has the coefficient with lowest possible magnitude on its highest-order term. Towards a formalization, 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 j^{th} 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.

Suggestion 6. *Plot the least "wiggly" confound whose path is contained in the Wald region $CR(\delta)$ for the event-time path of the outcome. Specifically, plot $\delta^*(v^*)$, where*

$$p^* = \min\{\dim(v) : \delta^*(v) \in CR(\delta)\} \text{ and} \tag{5}$$

$$v^* = \arg\min_v \{v_{p^*}^2 : \dim(v) = p^*, \delta^*(v) \in CR(\delta)\}. \tag{6}$$

This suggestion is closely related to the sensitivity analysis proposed in Rambachan and Roth (2021).

Figure 6 illustrates Suggestion 6 by adding the path $\delta^*(v^*)$ to the plot in Figure 5. In Figure 6(a), the estimated event-time path is consistent with a confound that follows a very "smooth" path, close to linear in event time, that begins pre-event and simply continues post-event. We suspect that in many economic settings such confound dynamics would be considered plausible, thus suggesting that a confound can plausibly explain the entire event-time path of

the outcome, and therefore that the policy might plausibly have no effect on the outcome.

In Figure 6(b), by contrast, the estimated event-time path demands a confound with a very "wiggly" path. We suspect that in many economic settings such confound dynamics would not be considered plausible, thus suggesting that a confound cannot plausibly explain the entire event-time path of the outcome, and therefore that the policy does affect the outcome.

2.7 Summarizing Policy Effects

Researchers sometimes present a scalar summary of the magnitude of the policy effect. One way to obtain such a summary is to compute an average or other scalar function of the post-event effects depicted in the plot (de Chaisemartin and D'Haultfoeuille 2021). Another is to estimate a static version of (1) including only the contemporaneous value of the policy variable (restricting $M = G = 0$ as in, e.g., Gentzkow et al. 2011). 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 (2), as follows.

Suggestion 7. *If using a more restrictive model than* (1) *to summarize magnitude, visualize the implied restrictions as follows:*

1. *Estimate* (1) *subject to the contemplated restrictions.*
2. *Use* (3) *to translate the estimates from* (1) *into cumulative effects.*
3. *Overlay the estimated cumulative effects on the event-study plot.*

Figure 7 illustrates this suggestion in a hypothetical example in which the more restrictive model assumes that the policy effect is static, i.e., that the current value of the policy affects only the current value of the outcome. Comparison of the two event-time paths provides a visualization of the fit of the more restrictive model. The uniform confidence band permits visual assessment of the more restrictive model. In the case shown in Figure 7, the more restrictive model is not included in the uniform confidence band, implying that we can reject the hypothesis that the effect of the policy is static. Figure 7 also displays the p-value from a Wald test of the more restrictive model, which also implies that the more restrictive model is rejected. The Wald test may be more powerful for testing this joint hypothesis than the test based on the sup-t band, but its implications are harder to visualize in this setting (Olea and Plagborg-Møller 2019).

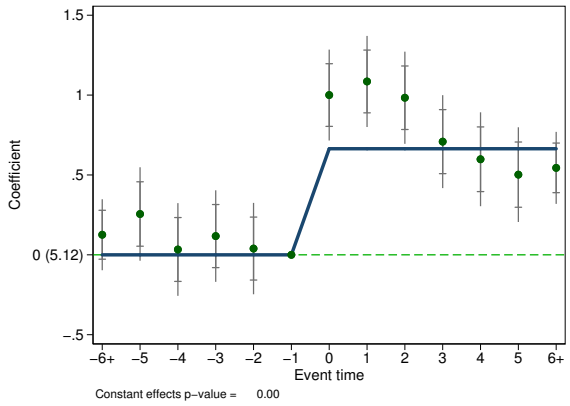


Figure 7: Evaluating the fit of a more restrictive model. Here, we overlay estimates from a model that imposes static treatment effects with estimates from a more flexible model that allows for dynamic treatment effects, in accordance with Suggestion 7.

3 Approaches to Identification

We now discuss approaches to identification of the parameters $\{\beta_m\}_{m=-G}^{M}$ in (1), which for ease of reference we collect in a vector β. Since, in general, the confound C_{it} will not be observed, identification of β will rely on restrictions on how observable and latent variables relate to C_{it} and z_{it}. To simplify statements, 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

$$C_{it} = \lambda_i' F_t + \xi \eta_{it}, \text{ and thus} \tag{7}$$

$$y_{it} = \alpha_i + \gamma_t + q_{it}' \psi + \sum_{m=-G}^{M} \beta_m z_{i,t-m} + \lambda_i' F_t + \xi \eta_{it} + \varepsilon_{it}. \tag{8}$$

In (7) and (8), F_t is a vector of (possibly unobserved) factors with (possibly unknown) unit-specific loadings λ_i, and η_{it} is an unobserved scalar with (possibly unknown) coefficient ξ.

We assume that G and M are known. In some settings it may be possible to estimate M, though inference on β would then need to account for data-driven choice of M. In some settings it may also be possible to estimate G, but because pre-trends are often used to diagnose confounding, we instead suggest choosing G on economic grounds. See Schmidheiny and Siegloch (2020) and Borusyak et al. (2021a) for related discussion.

To formalize the sense in which C_{it} is a confound, we maintain throughout that

$$\mathbb{E}[\varepsilon_{it}|F, \lambda_i, \eta_i, z_i, \alpha_i, \gamma, q_i] = 0, \tag{9}$$

with F, η_i, z_i, γ, and q_i collecting all T observations of F_t, η_{it}, z_{it}, γ_t, and q_{it}, respectively. Given this condition, if the factor F_t and the scalar η_{it} were observed, then with sufficiently rich data it would be possible to identify β simply by treating F_t and η_{it} as control variables. If F_t and η_{it} are not observed, then in general β is not identified.

Intuitively, then, identification of β will require restrictions on F_t and η_{it}. Typically the appropriate identifying assumptions cannot be learned from the data and so should ideally be justified on economic grounds. Our discussion therefore emphasizes and contrasts the economic content of different assumptions. We mostly focus on situations where only one of F_t or η_{it} is relevant, but in principle it seems possible to "mix and match" approaches when both components may be present.

3.1 Not Requiring Proxies or Instruments

The first class of approaches imposes sufficient structure on how latent variables enter the model so that the parameters β can be identified without the use of proxy or instrumental variables. We begin with a situation where the confound is low-dimensional ($\xi = 0$).

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.*

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.[7] 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}. For example, Neumark and Wascher (1992) adopt this approach in an

[7]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.

application where the policy z_{it} is the coverage-adjusted minimum wage in state i in year t, measured relative to the average hourly wage in the state, the outcome y_{it} is a measure of teenage or young-adult employment, and the controls q_{it} include economic and demographic variables.

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 confounding 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 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 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 $f(t) = t$, thus assuming that displacement is exogenous with respect to unobserved determinants of earnings, conditional on the worker fixed effect and the 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 using 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 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 (Bai 2009) or common correlated effects (Pesaran 2006) estimator. They may also be estimated via synthetic control methods (Abadie et al. 2010; Chernozhukov et al. 2020).

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.

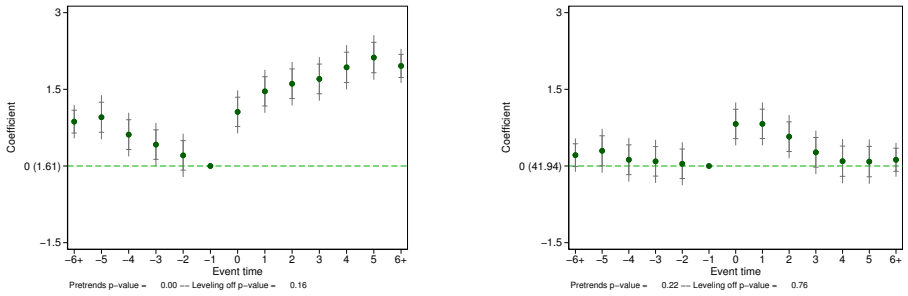
Assumption 2. *$F_t = 0 \; \forall \; t$ and the confound obeys*

$$\eta_{it} = \tilde{\alpha}_i + \tilde{\gamma}_t + q_{it}'\tilde{\psi} + \sum_m \phi' f(m) z_{i,t-m} \tag{10}$$

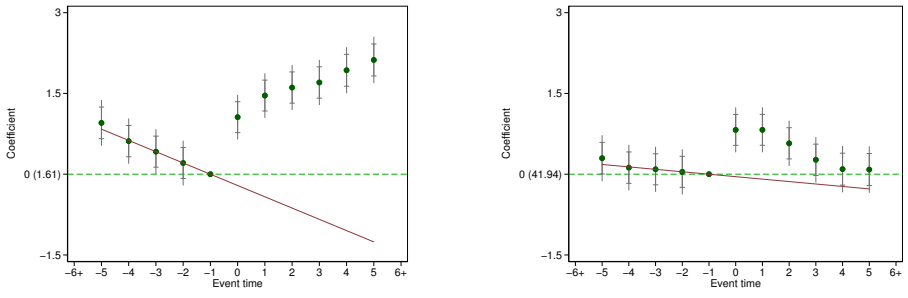
for $f(\cdot)$ a known low-dimensional set of basis functions, and $\tilde{\alpha}_i$, $\tilde{\gamma}_t$, $\tilde{\psi}$, and ϕ unknown parameters.

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. Then under Assumption 2 an event-study plot with $\xi\eta_{it}$ as the outcome would 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 trend in the outcome y_{it} can be used to learn the slope of the trend in $\xi\eta_{it}$. 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 $|m| \leq 3$.

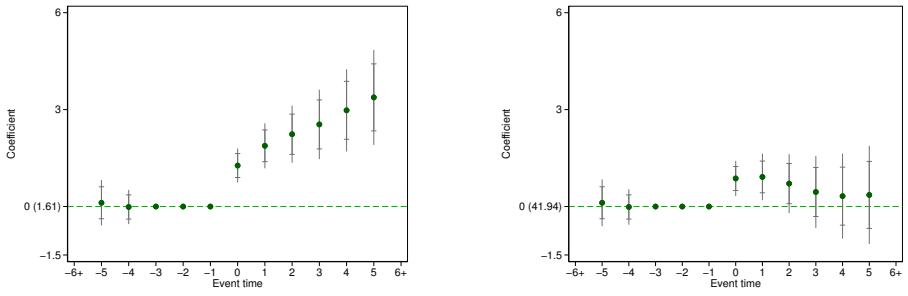
Figure 8 illustrates an estimator motivated by Assumption 2 in two possible datasets. Figure 8(a) depicts the event-time path of the outcome variable. The plot on the left exhibits a strongly declining pre-event trend, whereas the plot on the right exhibits a milder pre-event trend. Figure 8(b) extrapolates the confound from the pre-event event-time path of 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 periods. Each column of plots corresponds to a different possible dataset. Panel (a) shows the event-study plot. Panel (b) extrapolates a linear event-time trend from the three immediate pre-event periods, 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).

assuming a constant basis function in (10). 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.

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, and has no average trend in event time outside that horizon.

Assumption 2 is related to "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). Similar to an estimator motivated by Assumption 2, regression discontinuity in time can involve estimating local or global polynomial approximations to trends before and after an event such as adoption of a policy. 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.[8]

[8] 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).

3.2 Requiring Proxies or Instruments

The second class of approaches that we consider requires that we have additional variables available that can serve either as proxies for the unobserved confound C_{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}.

The following assumption relies on using additional observed variables to serve as proxies for η_{it}.

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

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

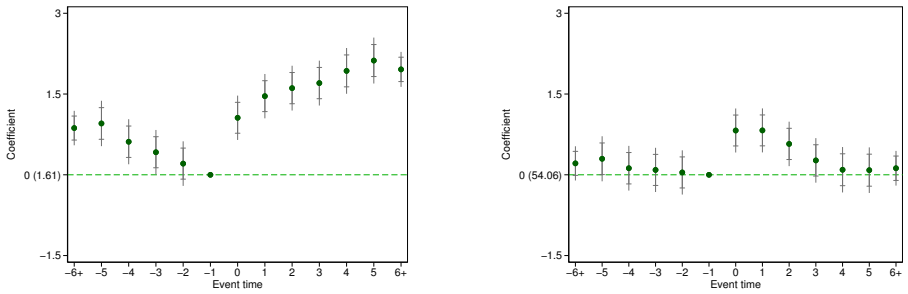
where α_i^x is an unobserved unit-specific vector, γ_t^x is an unobserved period-specific vector, ψ^x is an unknown matrix-valued parameter, Ξ^x is an unknown vector-valued parameter with all elements nonzero, and u_{it} is an unobserved vector. Moreover, one of the following holds:

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

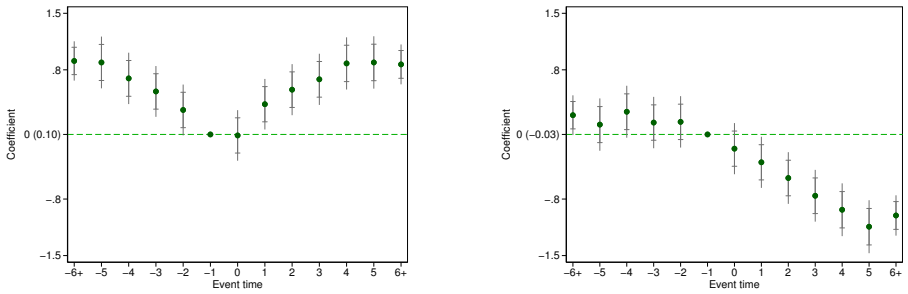
(b) The unobservable u_{it} satisfies $\mathbb{E}[u_{it}|z_i, \alpha_i^x, \gamma^x, q_i] = 0$, and the population projection of η_{it} on $\{z_{i,t-m}\}_{m=-G-L_G}^{M+L_M}$, q_{it}, and unit and time indicators, has at least one nonzero coefficient on $z_{i,t+m}$ for some $m > G$.

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 not suffice for recovery of 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.

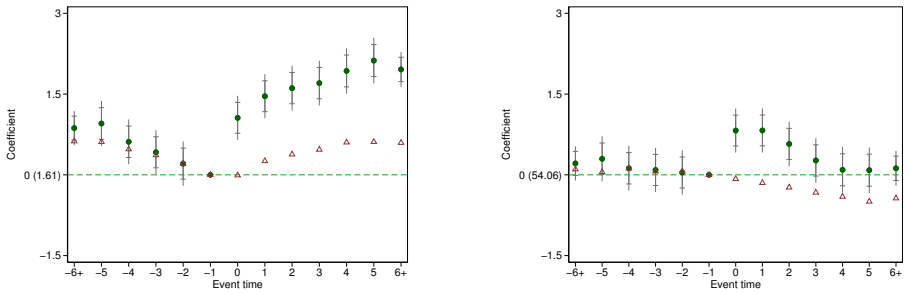
Under Assumption 3(b), we need only a single proxy for the confound η_{it}, and require that the noise in the proxy is conditionally mean-independent of the policy. Therefore, any relationship between the proxy and the policy reflects the relationship between the confound and the policy. Recall that 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



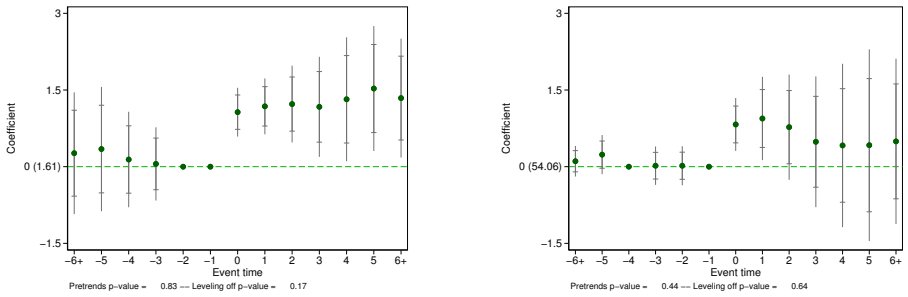
(a) Event-study plot for outcome



(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 two event times. Panel (d) adjusts for the confound using the 2SLS estimator proposed in Freyaldenhoven et al. (2019).

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. Freyaldenhoven et al. (2019) show in particular that the parameters β can be estimated via two-stage least squares, instrumenting for the proxy with leads of the policy.

Figure 9 illustrates an estimator motivated by Assumption 3(b) 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 two event times. 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).

Hastings et al. (2021) employ an approach motivated by Assumption 3(b) 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 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. (2021) 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*

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

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. In their setting, the 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).

Assumption	Economic interpretation	Estimator
1	The confound is low-dimensional and is...	
(a)	... the sum of time-invariant and cross-sectionally invariant components	Two-way fixed effects (TWFE)
(b)	... a known function $f(t)$ of calendar time	TWFE and unit indicators interacted with $f(t)$
(c)	... pervasive enough to be learned from the data	Interactive fixed effects or synthetic controls
2	Post-event dynamics of confound can be learned from pre-event data	TWFE controlling for extrapolated confound dynamics
3	We observe a noisy proxy for the confound and we have...	
(a)	... at least two proxies with uncorrelated measurement errors	IV estimation using one proxy as instrument for the other
(b)	... at least one proxy that (i) exhibits a pre-trend and (ii) whose measurement error is uncorrelated with the policy	IV estimation using leads of the policy as instruments for the proxy
4	We have an excluded instrument for the policy	IV estimation using the instrument for the policy

Table 1: Summary of identification approaches and corresponding estimators. This table summarizes the economic content of the various restrictions imposed on the confound in Sections 3.1 and 3.2 together with their corresponding estimators.

Table 1 summarizes the identification approaches discussed in Sections 3.1 and 3.2, and their corresponding estimators, focusing on the economic content of each restriction. We hope that Table 1 can serve as a useful reference for applied researchers by highlighting the economic content of, and connections among, the different approaches to identification.

3.3 Heterogeneous Policy Effects Under Staggered Adoption

Here we relax 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., Goodman-Bacon 2021; Athey and Imbens forthcoming; Callaway and Sant'Anna 2021; Sun and Abraham 2021), 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 of the policy on the outcome differs across cohorts t^*. Denoting the causal effects for cohort t^* by $\{\beta_{m,t^*}\}_{m=-G}^{M}$, collected in vector 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 Sections 3.1 and 3.2 can then be applied to recover the parameters β_{t^*} in (13). Estimation can proceed, for example, by interacting the policy variables in the model with indicators for cohort, and then adopting the controls, instruments, or other elements suggested by the given identifying assumption.

Next, consider the possibility that the causal effect 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, and in general it may not be possible to recover such an average at all. Exceptions include:

- Random assignment, i.e., $t^*(i)$ assigned independently of all variables in the model including $\{\beta_{m,i}\}_{m=-G}^{M}$. In such a setting, standard results on linear models imply that an estimate of β obtained from the usual two-way fixed effects estimator applied to (1) is a valid estimate of a proper weighted average of the unit-specific policy effects β_i for all units. (In the particular setting of event studies, see, for example, de Chaisemartin and D'Haultfoeuille 2021, Corollary 2.)

- No confound ($\xi = F_t = 0$), no control variables ($\psi = 0$), and static policy effects ($\beta_{m,i} = 0$ for all $m \neq 0$). In such a setting, results in de Chaisemartin and D'Haultfoeuille (2021, 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.

- No confound ($\xi = F_t = 0$), no control variables ($\psi = 0$), and some never-treated units. In such a setting, Sun and Abraham (2021) establish that an estimate of β_{t^*} obtained from the usual two-way fixed effects estimator applied to (13) is a valid estimate of a proper weighted average of the unit-specific policy effects β_i for units in the corresponding cohort.

Importantly, these settings rule out most of the forms of confounding that we considered in Sections 3.1 and 3.2. Developing approaches to estimate a proper weighted average of unit-specific policy effects in the presence of these forms of confounding therefore seems a useful direction for future work.

We expect that many economic settings will exhibit both confounding and heterogeneous policy effects. It is therefore useful to consider together the economic content of restrictions on both confounding and heterogeneity. For example, 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, 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 on observed variables and the unobserved confounds.

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. We then extend our analysis to allow for heterogeneous policy effects and consider the performance of different estimators in that setting.

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 A1.

Figure 10 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 (2) with $M + L_M = 5$ and $G + L_G = 5$ within each of 1,000 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 (2) with the normalization $\delta_{-1} = 0$ imposed, for a different DGP and dependent variable. All plots in Figure 10 are created from a two-way fixed effects estimator of (2). 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 10, labeled "Confound C_{it}." In each of these plots, we estimate (2) 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 pre-event, with the median estimated δ_k close to zero for $k < 0$.

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 seems 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).

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 10 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 10 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 11. The plots are defined similarly to those in Figure 10 with the lines in the plot corresponding to the median, 2.5th and 97.5th percentile of $\hat{\delta}_k$ across 1,000 realizations. These estimates of the event-time path are based on (2) with $M + L_M = 5$ 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 11 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.

The second row of Figure 11 corresponds to an interactive fixed effects estimator motivated by Assumption 1(c). We set the number of factors that interact with 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 11 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

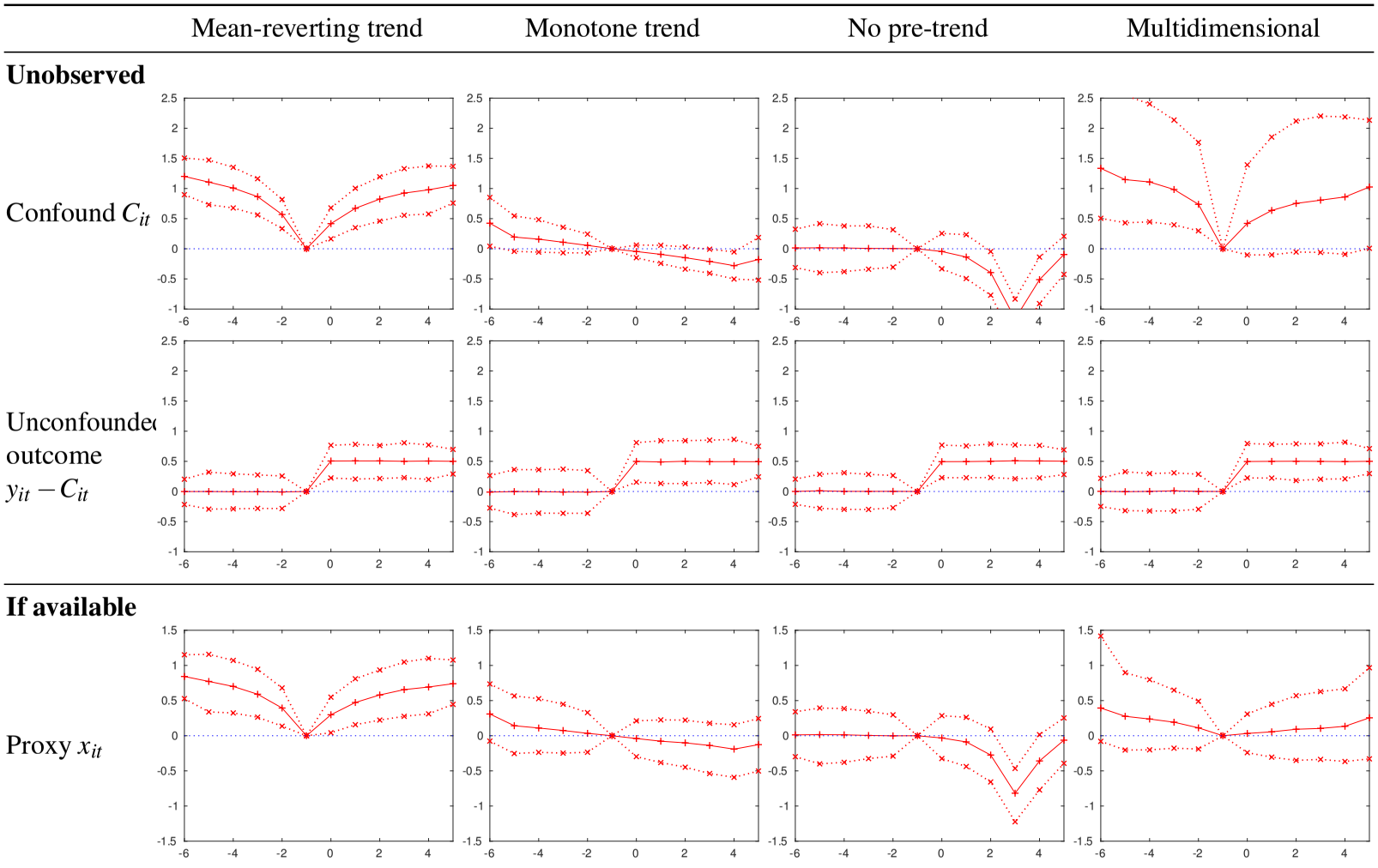


Figure 10: Graphical illustration of different outcomes across DGPs. Estimates are obtained via two-way fixed effects estimation of (2) with the dependent variable given in the row label. For each value of k indicated on the x-axis, the series correspond to the 2.5th (dotted, marked by x's), 50th (solid, marked by +'s), and 97.5th (dotted, marked by x's) percentiles across 1,000 simulations for each $\hat{\delta}_k$.

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 11 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 10 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 (2) 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 $z_{i,t+G+L_G}$, 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 permits free estimation of the remaining pre-event coefficients and thus visual inspection of the overidentifying restriction that the remaining pre-event effects are zero.[9]

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 10 that only the "Mean-reverting trend" DGP exhibits a strong pre-trend in the proxy. For this DGP, the final row of Figure 11 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 instruments in this design which results in a widely dispersed sampling distribution.

Figure 11 shows that, across the estimators and DGPs that we consider, no estimator performs uniformly well. The two-way fixed effects estimator exhibits substantial median bias in all DGPs we consider. Each other estimator exhibits low median bias for at least one

[9]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 replications due to sampling variation, so the second restriction is not visually transparent in Figure 11.

DGP and substantial median bias for at least one other DGP. No estimator exhibits low median bias for the "No pre-trends" DGP. Because a researcher will not know the true confound or policy effect, it will not in general be possible to choose estimators based solely on the data, further reinforcing the value of justifying identifying assumptions on economic grounds, and of performing sensitivity analysis in situations where multiple identifying assumptions are plausible.

In the Appendix, we extend our analysis to include five additional estimators. Specifically, Appendix Figure A1 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 A2 considers two versions of a synthetic control estimator motivated by Assumption 1(c). Finally, Appendix Figure A3 considers a measurement-error correction, assuming the availability of two proxies for the confound, motivated by Assumption 3(a). As with the estimators we focus on in the main text, none of these additional estimators performs uniformly well across all DGPs.

4.3 Heterogeneity-Robust Estimation and Heterogeneous Policy Effects

Up to this point we have considered estimators that estimate a single path of policy effects for all units i. Correspondingly, we have considered DGPs in which the effect of the policy is homogeneous across units i. In this subsection we consider the consequences of using estimators designed to be robust to heterogeneity in policy effects. We also consider the consequence for these and other estimators of drawing data from DGPs that feature such heterogeneity. We do so in three steps.

First, we consider some existing estimators from the literature that are designed to be robust to heterogeneity in policy effects, and apply those to data drawn from our baseline DGPs. Many recent papers have proposed approaches to estimating proper weighted averages of policy effects when policy effects differ across units or over time. Many of these estimators are grounded in models that rule out the forms of confounding exhibited by our DGPs. We might therefore expect these estimators to perform poorly when applied to data drawn from our DGPs, which feature homogeneous policy effects but substantial confounding.

Figure 12 illustrates the performance of the estimators proposed by de Chaisemartin and D'Haultfoeuille (2021), Sun and Abraham (2021), and Borusyak et al. (2021a) in data drawn from our DGPs. These estimators perform similarly to the two-way fixed effects estimator in terms of their median bias in these DGPs. Because there is substantial confounding present in these DGPs, the median bias of the estimators is often large. In many cases, the dispersion in

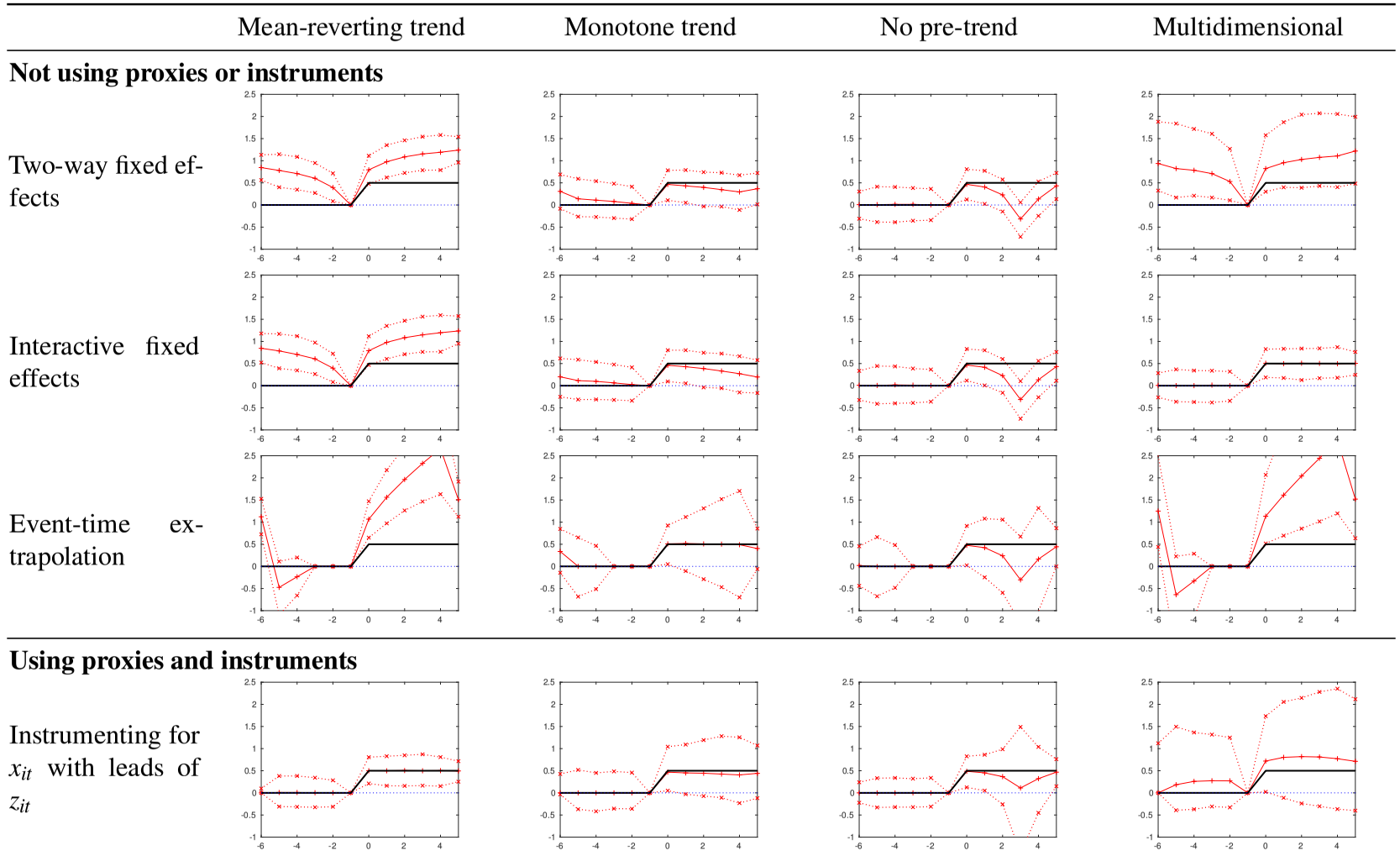


Figure 11: Graphical illustration of different estimators across DGPs. Estimates are obtained via estimation of (2) using the estimator given in the row label. True treatment effect depicted in solid black. For each value of k indicated on the x-axis, the series correspond to the 2.5th (dotted, marked by x's), 50th (solid, marked by +'s), and 97.5th (dotted, marked by x's) percentiles across 1,000 simulations for each $\hat{\delta}_k$.

the estimator is also similar to that of the two-way fixed effects estimator, suggesting little statistical cost or benefit from adopting the heterogeneity-robust estimator in place of the two-way fixed effects estimator in these DGPs.

Second, we apply some of the estimators discussed in Section 3 to data drawn from augmented versions of our baseline DGPs that feature heterogeneity in policy effects across units i. Under staggered adoption, a two-way fixed effects estimator of β_m in (1) can be represented as a weighted average of policy effects for different event times and cohorts (Sun and Abraham 2021). However, this estimator can put nonzero weight on effects at event times other than m, and can put negative weight on the effect at event time m for some cohorts. As a result, an estimator of β_m based on (1) need not estimate a proper weighted average of cohort-specific policy effects $\beta_{m,t^*(i)}$ in (13). Appendix Figure A4 depicts the weights across event times and cohorts for the "Mean-reverting trend" DGP. We indeed find that a two-way fixed effects estimator of β_m in (1) is not a properly weighted average of cohort-specific policy effects $\beta_{m,t^*(i)}$ in (13) in this design, suggesting that estimators based on models assuming homogeneous policy effects might perform poorly if the policy effects are heterogeneous.

Figure 13 directly illustrates the performance of some of the estimators discussed in Section 3 when the policy effect varies by cohort as in (13). Here, we extend the "Mean-reverting trend" DGP to allow for up to two possible dynamic profiles of policy effects. Each unit is assigned to one of the two profiles, with the assignment probability dependent on the unit's cohort $t^*(i)$.[10] Figure 13 illustrates the four resulting designs with the profiles of policy effects depicted in black.

The first column ("Static, homogeneous policy effect") corresponds to the "Mean-reverting trend" DGP, with a homogeneous, static effect of the policy on the outcome. The second column ("Static, heterogeneous policy effect") allows for two possible static policy effects, one positive and one null. The third column ("Dynamic, homogeneous policy effect") features a design with a homogeneous, dynamic policy effect that builds over time following adoption. The fourth column ("Dynamic, heterogeneous policy effect") features a design with two possible policy effects, one dynamic (building over time) and one static (no effect).

The first row of Figure 13 considers the infeasible two-way fixed effects estimator applied to the unconfounded outcome $y_{it} - C_{it}$. Under homogeneous policy effects we expect this estimator to be approximately median-unbiased, and this is what we find. However, as

[10] In particular, unit i's probability of being assigned to the larger of the two policy effects is given by logistic$(400 - 30t^*(i))$, so that earlier-adopting cohorts are more likely to be assigned to the larger policy effect. In tandem with our other design choices, this choice of assignment probability means that almost every unit in the first cohort experiences the larger policy effect, almost every unit in the last cohort experiences the smaller policy effect, and similar numbers of units experience each of the two profiles.

discussed above, this estimator may fail to recover a proper weighted average of the true policy effects under heterogeneous policy effects. This possibility manifests in a spurious pre-trend in the median series for both designs with heterogeneous policy effects.

The second row of Figure 13 considers the feasible two-way fixed effects estimator applied to the outcome y_{it}. The behavior of this estimator is dominated by the role of the confound. The third row of Figure 13 considers the estimator that instruments for the proxy x_{it} with the strongest lead of the policy z_{it}. This estimator accounts for confounding but not for heterogeneous policy effects. Accordingly, this estimator is approximately median-unbiased in the designs with homogeneous policy effects, but exhibits a spurious pre-trend in its median series in the designs with heterogeneous policy effects.

Our third and final step in this subsection is to extend the estimators in the second and third rows of Figure 13 to allow for different profiles of policy effects by cohort as in (13). We do this by interacting the policy variables in (2) with indicators for cohort, i.e. for the period in which the given unit adopts the policy. We then apply the resulting estimators to the same set of DGPs as in Figure 13, treating the average estimated policy profile across cohorts as our point estimate.

Figure 14 summarizes our findings. Consistent with Sun and Abraham (2021), the modification to allow for cohort-specific profiles of policy effects ensures that the infeasible estimator recovers a proper weighted average of the possible true profiles of policy effects. This improved performance extends to the interacted instrumental variables estimator, which no longer exhibits a spurious pre-trend. The interacted two-way fixed effects estimator performs poorly because of the role of confounding. In this setting, then, we find the best performance in terms of bias from an estimator that accounts both for heterogeneity in policy effects and for confounding.

Our findings in this section highlight that addressing the possibility of confounding and addressing the possibility of heterogeneous policy effects are not substitutes for one another. In settings with severe confounding, estimators designed to recover proper weighted averages of heterogeneous effects under no-confounding assumptions can perform poorly. Likewise, in settings with limited or no confounding but significant heterogeneity in policy effects, estimators motivated by models with homogeneous policy effects may exhibit undesirable behavior such as spurious pre-trends. Developing additional estimators robust to both confounding and heterogeneous policy effects seems like a useful direction for future research.

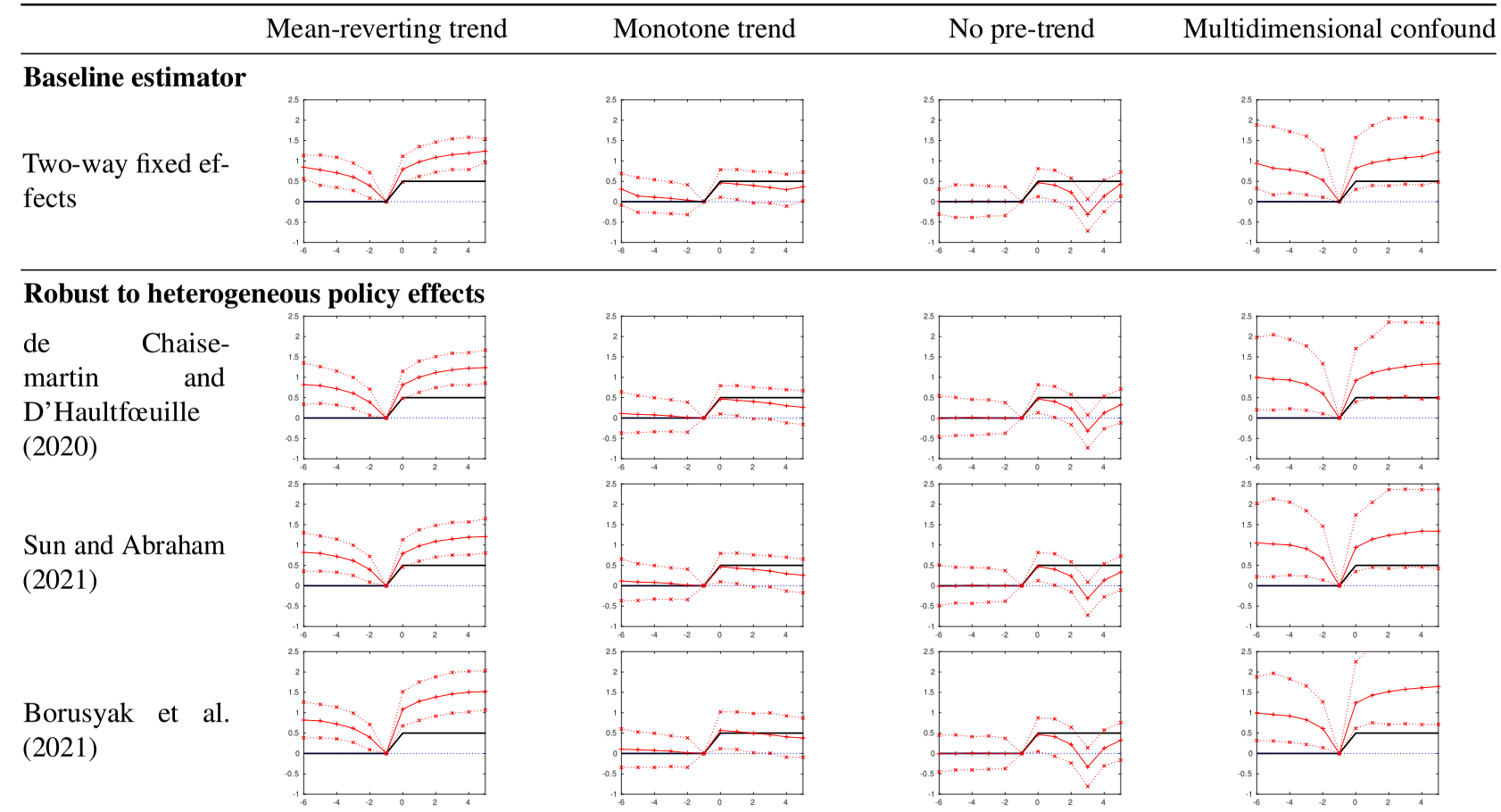


Figure 12: Estimators designed to allow for heterogeneous effects of the policy. The true treatment effect is depicted in solid black. For each value of k indicated on the x-axis, the series correspond to the 2.5th (dotted, marked by x's), 50th (solid, marked by +'s), and 97.5th (dotted, marked by x's) percentiles across 1,000 simulations for each $\hat{\delta}_k$. The first row ("Two-way fixed effects") implements a two-way fixed effects estimator. The second row ("de Chaisemartin and D'Haultfœuille (2020)") uses the `did_multiplegt` package (de Chaisemartin et al. 2019). The third row ("Sun and Abraham (2021)") uses the interaction-weighted estimator described in Sun and Abraham 2021 Section 4.1. The fourth row ("Borusyak et al. (2021)") uses the `did_imputation` package (Borusyak et al. 2021b).

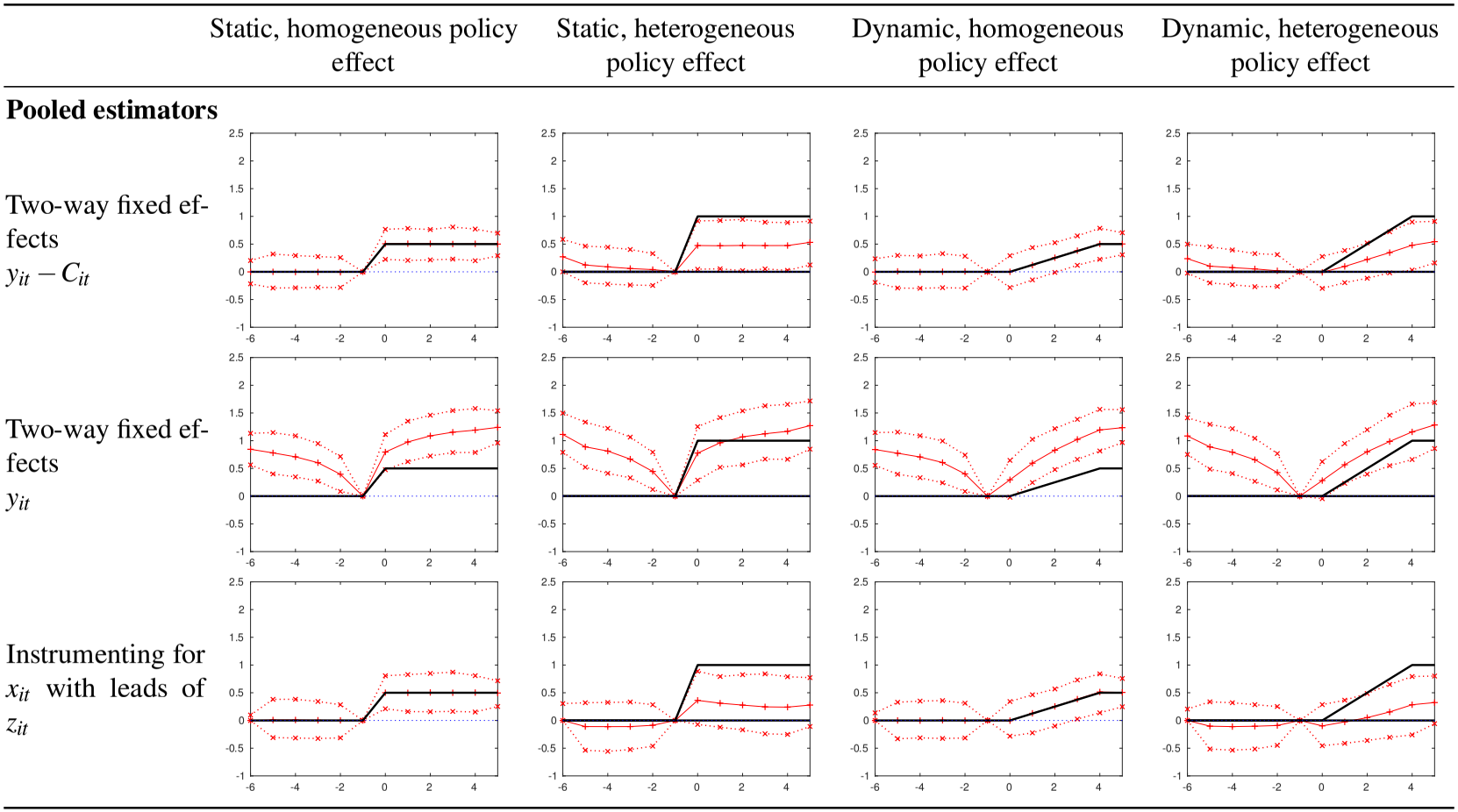


Figure 13: Graphical illustration of pooled estimators in DGPs with heterogeneous policy effects. All designs are based on the "Mean-reverting trend" DGP but allow for up to two profiles of possible policy effects, depicted in solid black. Unit i's probability of being assigned to the larger of the two policy effects is given by logistic$(400 - 30t^*(i))$ where $t^*(i)$ is the period in which unit i adopts the policy. Estimates are obtained via estimation of (2) using the estimator given in the row label. For the first row, the outcome variable is $y_{it} - C_{it}$. For each value of k indicated on the x-axis, the 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 $\hat{\delta}_k$.

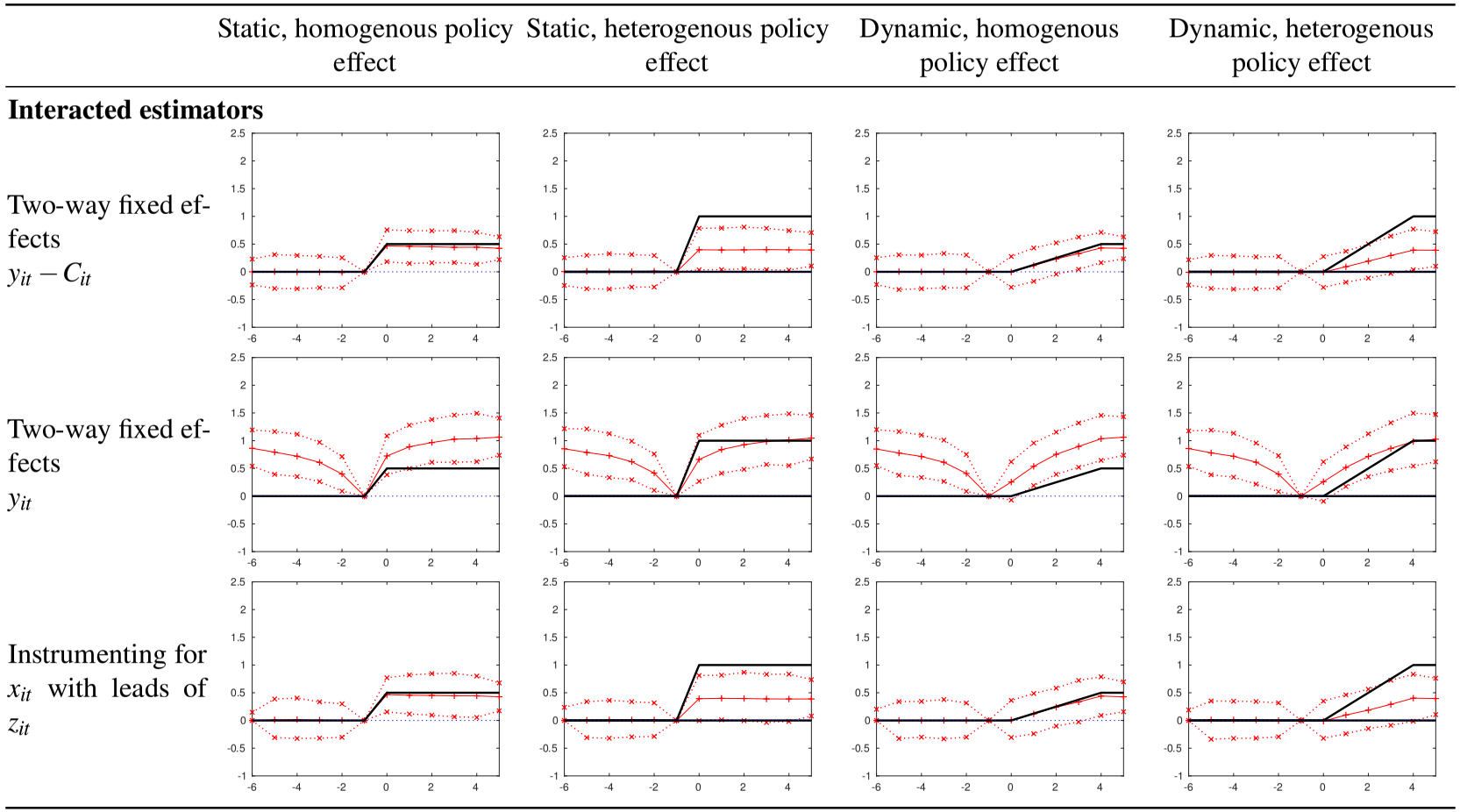


Figure 14: Graphical illustration of interacted estimators in DGPs with heterogeneous policy effects. All designs are based on the "Mean-reverting trend" DGP but allow for up to two possible profiles of policy effects, depicted in solid black. Unit i's probability of being assigned to the larger of the two policy effects is given by logistic$(400 - 30t^*(i))$ where $t^*(i)$ is the period in which unit i adopts the policy. Estimates are obtained via estimation of (13) using the estimator given in the row label. For the first row, the outcome variable is $y_{it} - C_{it}$. For each value of k indicated on the x-axis, the 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 $\hat{\delta}_k$.

5 Conclusion

Many important economic applications have been studied using a linear panel model and an accompanying event-study plot. We have tried to make three contributions in this chapter. The first is to suggest improvements to the event-study plot, both by codifying what we consider existing best practices and also by recommending some practices that are less frequently used. The second is to discuss some available approaches to identification in the presence of confounds, emphasizing and contrasting the economic content of the different identifying assumptions. The third is to examine the performance of estimators corresponding to these approaches to identification under different possible economically motivated data-generating processes.

We find that no estimator performs well uniformly under all reasonable DGPs, and that the performance of an estimator cannot typically be gauged from the data at hand. These findings highlight the importance of motivating modeling assumptions on economic grounds, and of developing new statistical tools grounded in economically motivated assumptions.

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.

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. SYNTH: Stata module to implement synthetic control methods for comparative case studies, 2020. URL `https://ideas.repec.org/c/boc/bocode/s457334.html`.

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 Zvi Griliches and Michael 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 W. Imbens, and Khashayar Khosravi. Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, pages 1–15, 2021. doi: 10.1080/01621459.2021.1891924.

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. AUGSYNTH: Augmented synthetic control method, 2020. URL https://github.com/ebenmichael/augsynth.

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, Xavier Jaravel, and Jann Spiess. Revisiting event study designs. Working paper, 2021a. URL https://6cd96a90-e111-4c99-bb2c-0fdc17ec7c2d.filesusr.com/ugd/bacd2d_c1d299c975204d57b2067545d51630e2.pdf.

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. DID_IMPUTATION: Stata module to perform treatment effect estimation and pre-trend testing in event studies, 2021b. URL https://ideas.repec.org/c/boc/bocode/s458957.html.

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*, 225(2):200–230, 2021.

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*, 2020. doi: 10.1080/01621459.2020.1751646.

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):2964–2996, 2020.

Clément de Chaisemartin and Xavier D'Haultfoeuille. Difference-in-differences estimators of intertemporal treatment effects. arXiv:2007.04267, 2021.

Clément de Chaisemartin, Xavier D'Haultfoeuille, and Yannick Guyonvarch. DID_MULTIPLEGT: Stata module to estimate sharp Difference-in-Difference designs with multiple groups and periods, 2019. URL `https://ideas.repec.org/c/boc/bocode/s458643.html`.

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): 308–352, 2018.

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. *Journal of Econometrics*, 225(2):254–277, 2021.

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*, 13(3), 2021.

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. Employment effects of minimum and submininum wages: Panel data on state minimum wage laws. *Industrial and Labor Relations Review*, 46 (1):55–81, 1992.

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 comparative case studies: Does the minimum wage reduce employment? *Journal of Business & Economic Statistics*, 2021. doi: 10.1080/07350015.2021.1927743.

Ashesh Rambachan and Jonathan Roth. An honest approach to parallel trends. Working paper, 2021. URL `https://jonathandroth.github.io/assets/files/HonestParallelTrends_Main.pdf`.

Jonathan Roth. Pre-test with caution: Event-study estimates after testing for parallel trends. Working paper, 2021. URL `https://jonathandroth.github.io/assets/files/roth_pretrends_testing.pdf`.

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. Working paper, 2021. URL `https://cowles.yale.edu/3a/fisherstaggered-randomization-tests-observational-studies-staggered-adoption-treatment.pdf`.

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. EVENTSTUDYWEIGHTS: Stata module to estimate the implied weights on the cohort-specific average treatment effects on the treated (CATTs) (event study specifications), 2021. URL `https://ideas.repec.org/c/boc/bocode/s458833.html`.

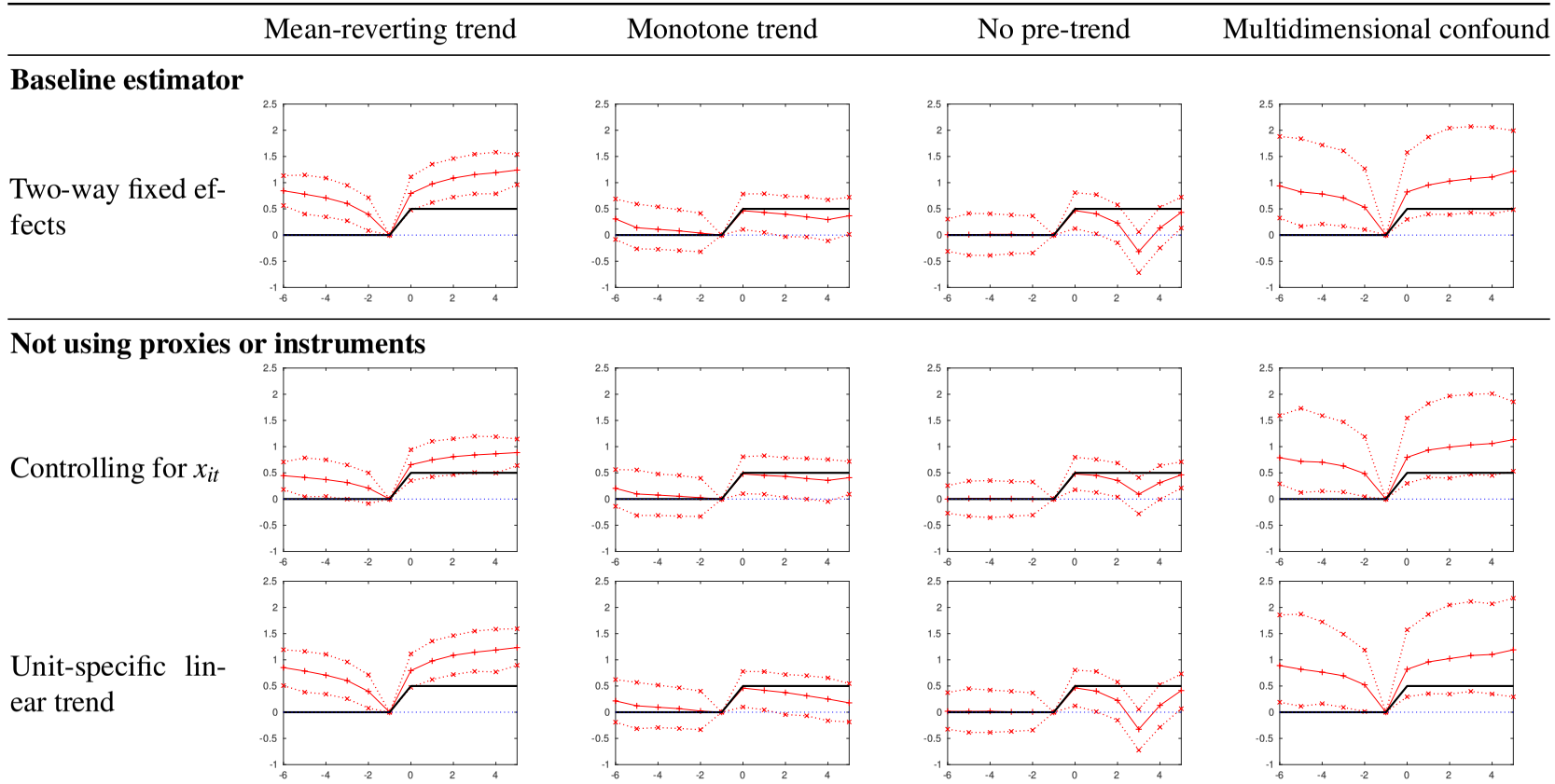
Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, 2021.

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.

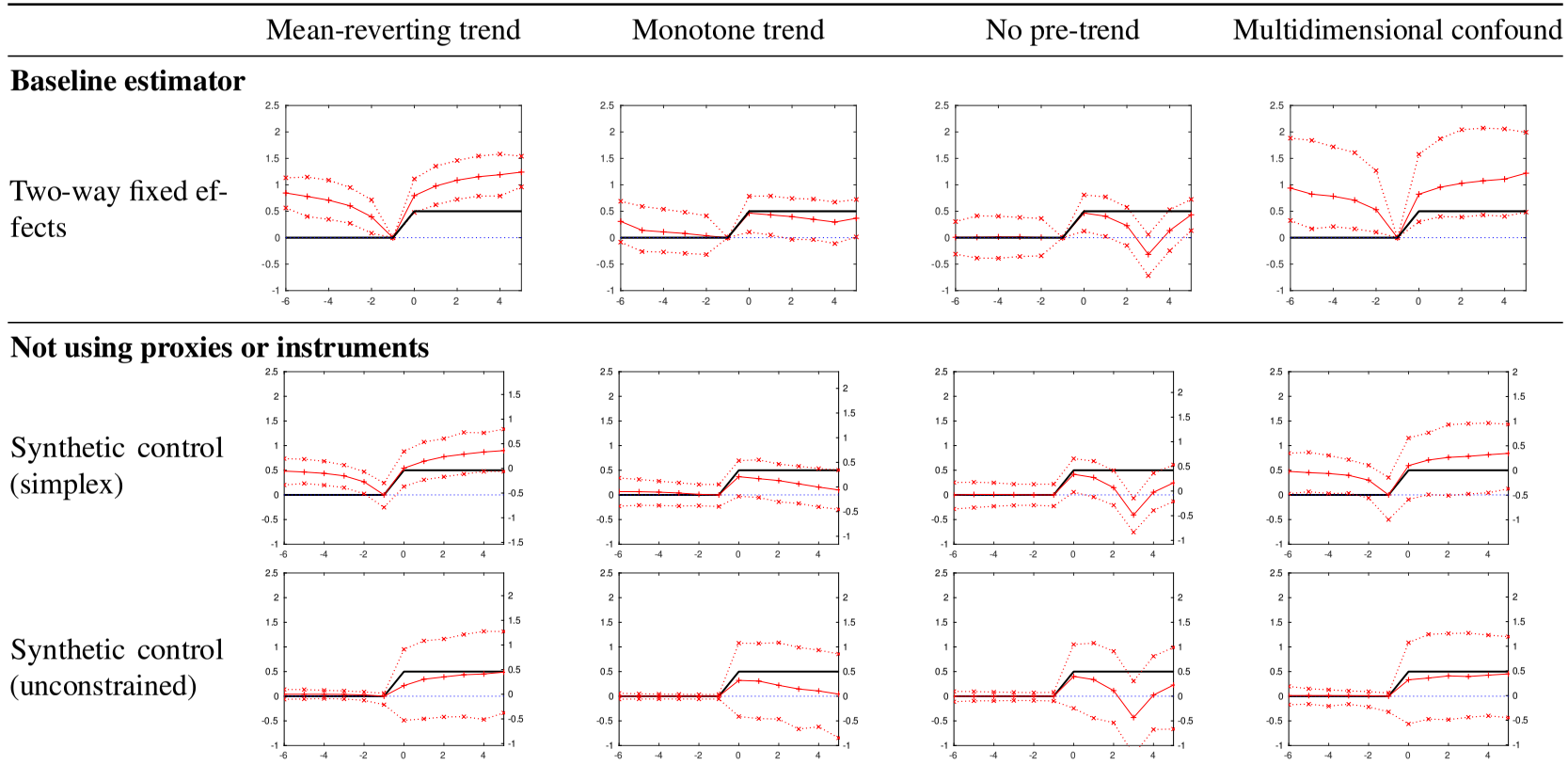
Appendix

	Mean-reverting trend	Monotone trend	No pre-trend	Multidimensional confound
Observed				
Outcome	$y_{it} = 0.5\left(\sqrt{0.5}\alpha_i + \sqrt{0.5}\gamma_t\right) + 0.5 z_{it} + \sqrt{0.5} C_{it} + 0.5\varepsilon_{it}$			
Policy	$z_{i1} = 0;\ \Delta z_{it} = \mathbf{1}\left(\left\{z_{i,t-1} = 0 \text{ and } C_{i,t+P} + 0.5\zeta^z_{i,t+P} < -2\right\}\right)$ for $t \geq 2$			
	$P = -1$	$P = 6$	$P = 3$	$P = -1$
Unobserved				
Fixed effects	$\gamma_t = 0.5\gamma_{t-1} + \sqrt{0.75}\zeta^\gamma_t$			
Confound C_{it}	$C_{it} = \sqrt{1-\sigma^2_\eta}\left[\sqrt{1-\sigma^2_a-\sigma^2_d}\lambda_i' F_t + \sqrt{\sigma^2_a} a_i + \sqrt{\sigma^2_d} d_t\right] + \sqrt{\sigma^2_\eta}\eta_{it}$, with $d_t = 0.5 d_{t-1} + \sqrt{0.75}\zeta^d_t$			
	$\sigma^2_a = \sigma^2_d = \sigma^2_\eta = 0.5,\ F_t = 0$			$\sigma^2_\eta = 0, \sigma^2_a = \sigma^2_d = 0.1, \dim(F_t) = 2$
	$\eta_{it} = \rho_\eta \eta_{i,t-1} + \sqrt{1-\rho^2_\eta}\zeta^\eta_{it}$			$\lambda_{ji} = \frac{1}{\sqrt{2}}\left(1 + \zeta^\lambda_{ji}\right),\ F_{jt} = \rho_{F_j} F_{j,t-1} + \sqrt{1-\rho^2_{F_j}}\zeta^F_{jt}$
	$\rho_\eta = 0.6$	$\rho_\eta = 0.95$	$\rho_\eta = 0.4$	$\rho_{F_1} = 0.9, \rho_{F_2} = 0.4$
If available				
Proxies	$x_{it} = 0.5\left(\sqrt{0.5}\alpha^x_i + \sqrt{0.5}\gamma^x_t\right) + \sqrt{0.5}\left[\sqrt{1-\theta}\left(0.5[a_i + d_t] + \sqrt{0.5}\eta_{it}\right) + \sqrt{\theta}\lambda_{1i} F_{1t}\right] + 0.5 u_{it}$, with $\gamma^x_t = 0.5\gamma^x_{t-1} + \sqrt{0.75}\zeta^{\gamma^x}_t;\ u_{it} = 0.5 u_{i,t-1} + \sqrt{0.75}\zeta^u_{it}$			
	$\theta = 0$			$\theta = 1$

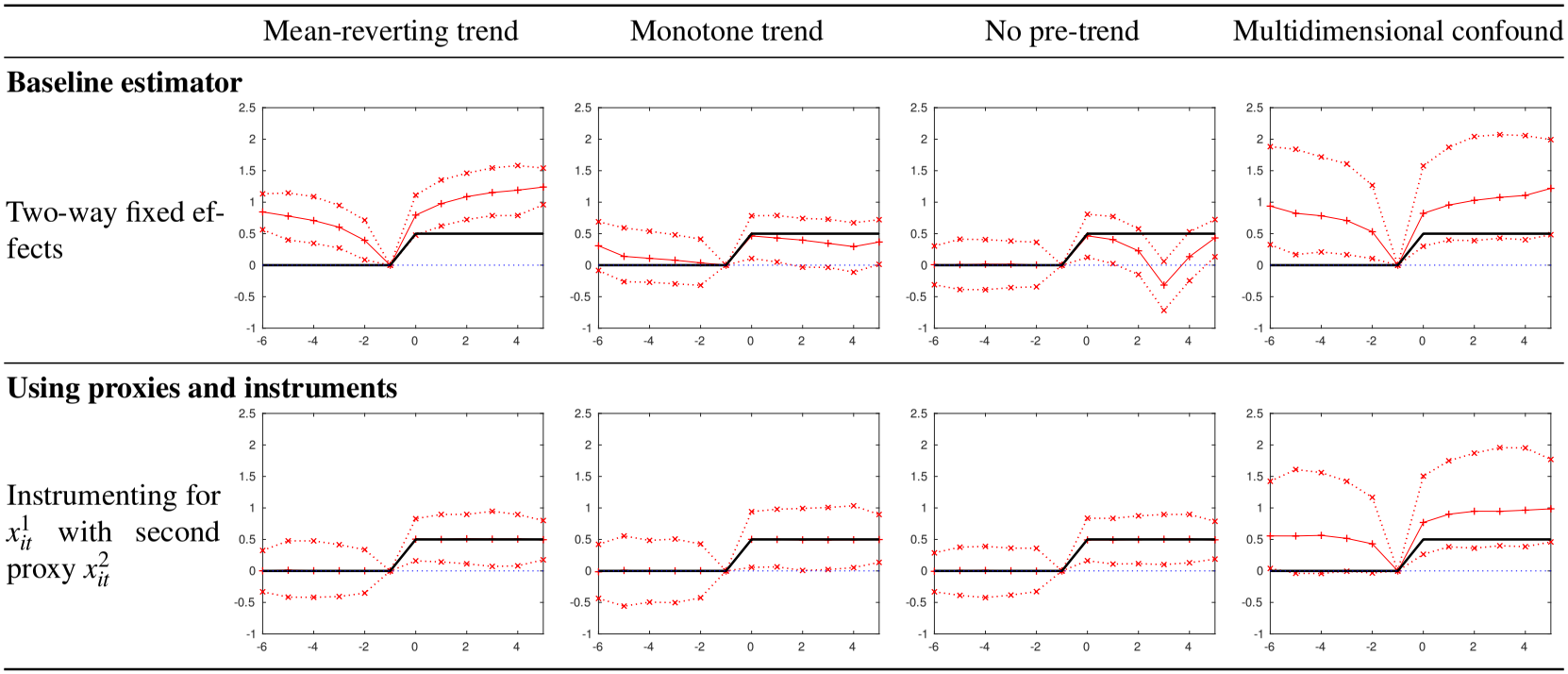
Appendix Table A1: Detailed description of the data-generating processes underlying the simulations presented in Section 4. We generate data for the time periods $t = -4, -3, \ldots, T+5$ to avoid missing observations for leads and lags of the policy and to allow for a policy adoption rule that is consistent with a forward/backward-looking decision-maker on either end of the sample. All random variables not otherwise defined, including initial conditions for autoregressive processes, are distributed as i.i.d. standard normal. We set $z_{it} = z_{i,t-1}$ if $t + P > T + 5$.



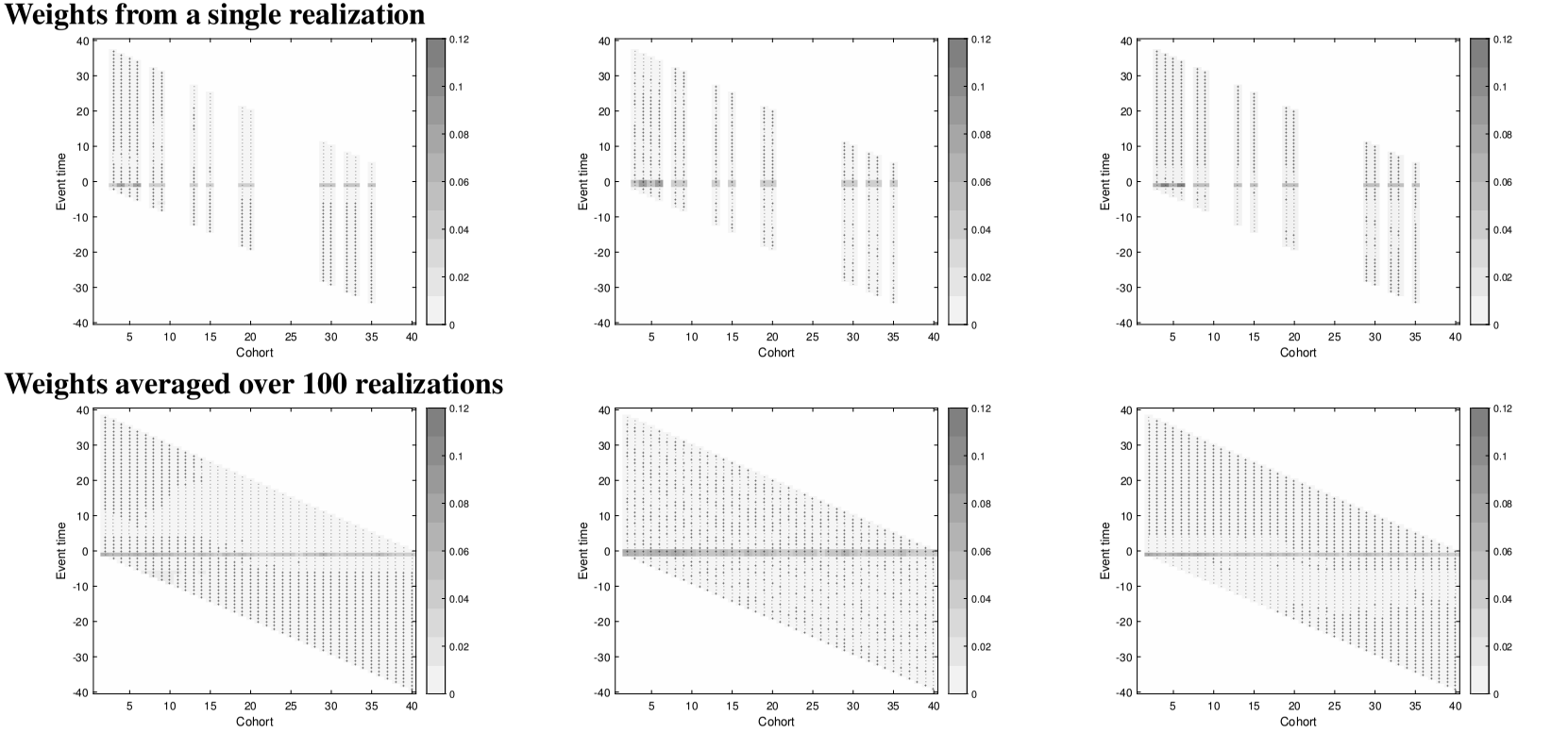
Appendix Figure A1: Additional estimators not using proxies or instruments across DGPs. The true treatment effect is depicted in solid black. For each value of k indicated on the x-axis, the series correspond to the 2.5th (dotted, marked by x's), 50th (solid, marked by +'s), and 97.5th (dotted, marked by x's) percentiles across 1000 simulations for each $\hat{\delta}_k$. The first row ("Two-way fixed effects") implements a two-way fixed effects estimator. The second row ("Controlling for x_{it}") includes the proxy x_{it} directly in the controls q_{it}. The third row ("Unit-specific linear trend") allows for unit-specific linear time trends.



Appendix Figure A2: Synthetic control estimators across DGPs. The true treatment effect is depicted in solid black. For each value of k indicated on the x-axis, the 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 right axis depicts the estimated treatment effect relative to the synthetic control. On the left axis, estimates are normalized such that the median treatment effect is zero the period before the event. The donor pool for unit i in cohort $t^*(i)$ consists of all units j with $t^*(j) > t^*(i)+5$ and all units j that never adopt the policy, and weights are chosen to fit the path of the outcome variable in all periods prior to $t^*(i)$. The first row ("Two-way fixed effects") implements a two-way fixed effects estimator. The second row ("Synthetic control (Simplex)") estimates $\hat{\delta}_k$ using the `synth` package (Abadie et al. 2020), restricting the weights on the donor units to the simplex. The third row ("Synthetic control (Unconstrained)") estimates $\hat{\delta}_k$ using the `augsynth` package (Ben-Michael et al. 2020), allowing donor weights outside the simplex but imposing an ℓ_2-penalty of 0.001.



Appendix Figure A3: Additional estimator using proxies and instruments across DGPs. The true treatment effect is depicted in solid black. For each value of k indicated on the x-axis, the 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 first row ("Two-way fixed effects") implements a two-way fixed effects estimator. The second row ("Instrumenting for x_{it}^1 with second proxy x_{it}^2") considers a measurement-error correction, assuming the availability of two proxies for the confound, by instrumenting for one proxy, x_{it}^1, with the other, x_{it}^2.



Appendix Figure A4: Graphical illustration of weights underlying coefficients. The figure shows the estimated weights underlying the coefficients δ_{-6} ("Furthest lead"), δ_0 ("Event-time coefficient"), and δ_5 ("Furthest lag") from the model in (2) with $M + L_M = 5$ and $G + L_G = 5$. In each plot, the shading of each cell denotes the absolute estimated weight that corresponds to the given cohort and event time, and the symbol ($+$ or $-$) denotes the sign of the estimated weight. The first row ("Weights from a single realization") shows the estimated weights from a single realization from the "Mean-reverting trend" DGP. The second row ("Weights averaged over 100 realizations") shows the average estimated weights across 100 realizations of the "Mean-reverting trend" DGP. The weights are defined following Proposition 1 of Sun and Abraham (2021) and estimated following equation (13) of Sun and Abraham (2021) using the package `eventstudyweights` (Sun 2021).