NBER WORKING PAPER SERIES

PRE-EVENT TRENDS IN THE PANEL EVENT-STUDY DESIGN

Simon Freyaldenhoven
Christian Hansen
Jesse M. Shapiro

Working Paper 24565
http://www.nber.org/papers/w24565

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2018

We thank Isaiah Andrews, Jeff Clemens, Josh Gottlieb, David Neumark, Emily Oster, Bryce Steinberg, and seminar participants at Brown University, Stanford University, Tel Aviv University, and Hebrew University for helpful comments. We thank Justine Hastings for sharing regression output. We acknowledge financial support from the National Science Foundation under Grant No. 1558636 and Grant No. 1658037, and from the Brown University Population Studies and Training Center. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w24565.ack

Pre-event Trends in the Panel Event-study Design
Simon Freyaldenhoven, Christian Hansen, and Jesse M. Shapiro
NBER Working Paper No. 24565
April 2018
JEL No. C23,C26

ABSTRACT

We consider a linear panel event-study design in which unobserved confounds may be related both to the outcome and to the policy variable of interest. We provide sufficient conditions to identify the causal effect of the policy by exploiting covariates related to the policy only through the confounds. Our model implies a set of moment equations that are linear in parameters. The effect of the policy can be estimated by 2SLS, and causal inference is valid even when endogeneity leads to pre-event trends ("pre-trends") in the outcome. Alternative approaches, such as estimation following a test for pre-trends, perform poorly.

Simon Freyaldenhoven
Economics Department
Box B
Brown University
Providence, RI 02912
simon_freyaldenhoven@brown.edu

Christian Hansen
University of Chicago
Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637
chansen1@chicagobooth.edu

Jesse M. Shapiro
Economics Department
Box B
Brown University
Providence, RI 02912
and NBER
jesse_shapiro_1@brown.edu

An online appendix is available at http://www.nber.org/data-appendix/w24565

1 Introduction

We are interested in estimating the causal effect β of a policy variable z_{it} on an outcome y_{it} in a linear panel data model, where i indexes units and t indexes time. We are concerned that the strict exogeneity of z_{it} may fail due to the presence of a time-varying unobservable η_{it} that is correlated with both z_{it} and y_{it}. In the literature on the effects of the minimum wage, y_{it} is youth employment, i indexes states, t indexes calendar years, and z_{it} is an indicator for years after passage of a minimum-wage increase. The unobserved confound η_{it} is labor demand. The concern is that states tend to pass minimum-wage increases during good economic times (Card and Krueger 1995; Neumark and Wascher 2006).

A common diagnostic approach in such settings is to look at whether the policy change appears to have an effect on the outcome before it actually occurs.[1] The presence of such pre-event trends or "pre-trends" is taken as evidence against the strict exogeneity of the policy change.

This approach is incomplete. If pre-trends are not detected, it may be that there are no pre-trends, or that pre-trends are present but undetected due to limited statistical power. In the latter case, estimation under the assumption of strict exogeneity is typically inappropriate. If pre-trends are detected, it is understood that strict exogeneity is likely to fail, but it is not clear what to do.

In both cases, what is needed is a notion of magnitude: given some pre-trend in the outcome, how much of the apparent effect of the policy is due to confounds, and how much to the causal effect of the policy? Armed with such a notion, a researcher can conduct valid inference on β whether or not pre-trends are detected.

In this paper, we propose to obtain such a notion from the behavior of a covariate x_{it} that is affected by the confound η_{it} but not by the policy z_{it}. In the minimum wage context, adult employment responds to labor demand η_{it} but not to the minimum wage (Brown 1999). Instead of using adult employment as a control variable, as is commonly done in the literature,[2] we propose to look at its dynamics around minimum wage increases, and use these to infer the dynamics of η_{it}.

To fix ideas, suppose we observe the outcome y_{it} in periods $t = 1, ..., T$ and the policy z_{it} in

[1] Of the 16 papers in the 2016 *American Economic Review* that use a linear panel data model, 11 are concerned with the existence of pre-trends as a sign of endogeneity. Of these 11, 9 include a plot of pre-trends, of which 2 provide a formal test of whether pre-trends are zero. In the minimum wage context, Allegretto et al. (2011) provide a plot of pre-trends.

[2] Brown (1999, table 3) describes 13 models of the effect of the minimum wage on teenage or young adult unemployment that have been estimated using state-level panel data. In 5 of these there is a control for the contemporaneous or past employment-to-population ratio or the prime-age unemployment rate.

periods $t = 1 - L, ..., T + L$ for some $L \geq 1$. Say that

$$y_{it} = \beta z_{it} + \gamma \eta_{it} + \varepsilon_{it} \tag{1}$$

$$\mathbb{E}(\varepsilon_{it} | \eta_{it}, \{z_{it}\}_{t=1-L}^{T+L}) = 0 \tag{2}$$

$$\mathbb{E}(x_{it} | \eta_{it}, \{z_{it}\}_{t=1-L}^{T+L}) = \lambda \eta_{it}, \tag{3}$$

where (1) defines the causal model, (2) is strict exogeneity of the policy with respect to the unobserved error ε_{it}, and (3) defines the relationship of the covariate x_{it} to the confound η_{it} up to the nonzero parameter λ. When the parameter γ is equal to zero, the confound does not affect the outcome, and identification of β is immediate.

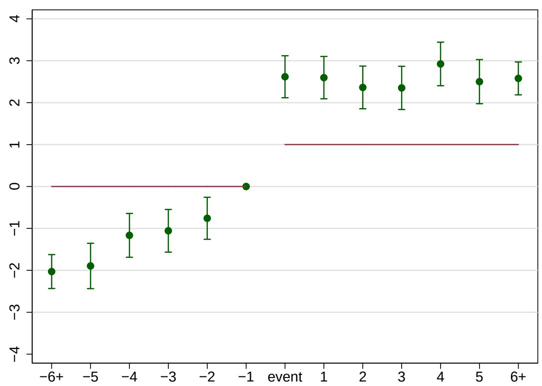
Figure 1a plots coefficients from a regression of y_{it} on $\left\{\Delta z_{i,t+l}\right\}_{l=-L}^{L}$ in data simulated from an example of (1). Here and throughout, Δ denotes the first difference operator. Because the figure resembles event-study plots in finance (Ball and Brown 1968; MacKinlay 1997), the corresponding estimates are sometimes called "event-study estimates" (Hoynes and Schanzenbach 2009; Duggan et al. 2016).

Figure 1a shows a clear pre-trend in the outcome, indicating that $\gamma \neq 0$. Figure 1b shows that the covariate x_{it} exhibits a pre-trend similar to that of the outcome, and a relatively smaller increase at the event time. We would like to use the covariate x_{it} to correct for the role of the confound η_{it}. Including the covariate x_{it} as a control variable will suffice only if x_{it} is a perfect proxy for η_{it} (i.e., $x_{it} = \lambda \eta_{it}$). Subtracting the covariate from the outcome (yielding dependent variable $y_{it} - x_{it}$) will suffice only if the effects of the confound are exactly parallel between the outcome and covariate (i.e., $\gamma = \lambda$).

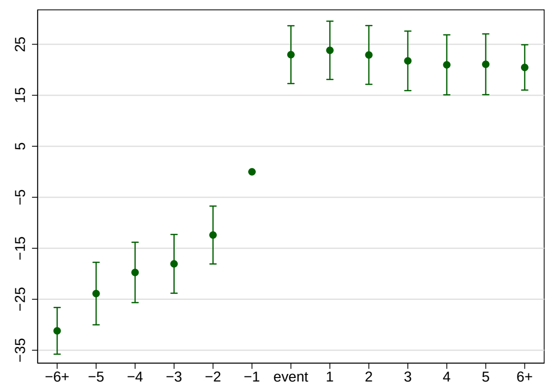
The alternative that we propose can be understood with reference to Figure 1c. Here, we rescale the series in Figure 1b so that it exactly matches that in Figure 1a in the two periods immediately before the event. Under our maintained assumptions, comparing the two series in Figure 1c allows us to decompose the change in the outcome at the event time into a component due to the causal effect of the policy and a component due to the confound η_{it}. The adjusted plot in Figure 1d removes the estimated effect of the pre-trend from Figure 1a, revealing the dynamics of the outcome net of the confound, and hence β, the causal effect of interest.

The geometry of these plots suggests an instrumental variables setup, in which Figure 1a plots the reduced form for the outcome and Figure 1b plots the first stage. Indeed, we show that β can be estimated by two-stage least squares (2SLS) regression of the outcome y_{it} on the policy z_{it} and covariate x_{it}, with leads (e.g., $z_{i,t+1}$) of the policy serving as excluded instruments. An essential assumption is that the dynamic relationship of x_{it} to z_{it} mirrors the dynamic relationship of η_{it} to z_{it}. This means, in particular, that x_{it} is affected by η_{it} but not by z_{it}.

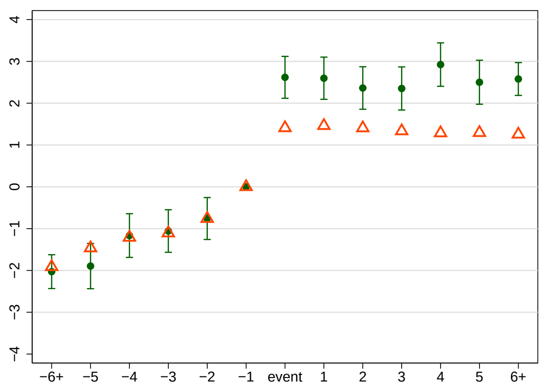
We also require that there be a pre-trend in the covariate x_{it}. We argue that a pre-trend in η_{it}



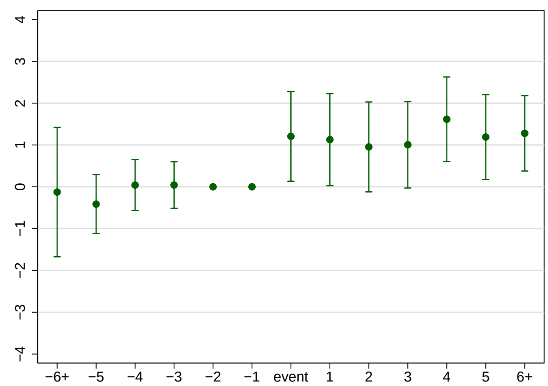
(a) Outcome of interest y_{it} around event time. Solid line depicts true causal effect.



(b) Unaffected covariate x_{it} around event time.



(c) Overlaying outcome of interest (with confidence intervals) y_{it} and rescaled unaffected covariate x_{it} (triangles) around event time.



(d) Outcome of interest y_{it} around event time, using the behavior of the covariate to net out the effect of the confound.

Figure 1: Hypothetical event plots. An unobserved factor potentially causes endogeneity, manifested as a pre-trend in the outcome y_{it}. A covariate x_{it} affected by the confound, but not by the policy, permits us to learn the dynamics of the confound and adjust for them.

is natural in the many economic settings in which the policy z_{it} changes when some unobserved state variable η_{it} crosses a threshold. Indeed, the common approach of using pre-trends to diagnose failures of exogeneity ($\gamma = 0$) is presumably motivated, in part, by the belief that the confound η_{it} is likely to exhibit a pre-trend. Our assumptions imply that a pre-trend in η_{it} manifests as a pre-trend in the covariate x_{it}, and may or may not manifest as a pre-trend in the outcome y_{it}.

Section 2 generalizes the setup in (1)-(3) to allow for multiple confounds, additive unit-specific fixed effects, and exogenous controls. We show that the model admits a GMM representation, from which standard results on estimation and inference (with large N and fixed T) are available.

Section 3 presents Monte Carlo evidence on the finite-sample performance of our proposed estimator under a range of alternative data-generating processes, varying both the quality of the proxy x_{it} and the strength of identification. In these simulations, when strongly identified our estimator outperforms the approach of controlling directly for x_{it} except when x_{it} is a nearly perfect proxy for η_{it}. Our estimator also outperforms the approach of conducting a test for pre-trends before proceeding with estimation. We show that this "pre-test" approach is unreliable because it is vulnerable to undetected (but important) confounds.

The main requirement that our approach imposes on a practitioner is to find a covariate x_{it} that is related to the confound η_{it} but unaffected by the policy z_{it}. This is similar in difficulty to finding a suitable control variable, but without the additional burden of ensuring that x_{it} proxies perfectly for η_{it}. (Of course, as our simulations reinforce, x_{it} must still provide a reasonable signal of η_{it} in order to permit strong identification.) We believe suitable covariates are present in many, though by no means all, applied settings of interest.

Section 4 presents applications of our proposed approach to the effect of SNAP on household spending (Hastings and Shapiro 2018), the effect of newspaper entry on voter turnout (Gentzkow et al. 2011), and the effect of the minimum wage on youth employment (Neumark et al. 2014; Allegretto et al. 2017). These applications illustrate a range of possibilities, including cases with clear pre-trends in the outcome, a case without meaningful pre-trends, and a case in which it is hard to tell. In some cases our proposed adjustment makes a small difference to point estimates, in some cases a larger difference, and in some cases it simply implies greater statistical uncertainty.

We are not aware of an existing formal proposal to use an unaffected covariate to adjust causal inference for pre-trends in a panel data model. In their appendix, Gentzkow et al. (2011) implement an estimator that is similar in spirit to the one that we propose, but that is not formally justified by our setup.[3] Borusyak and Jaravel (2017) study the identification and estimation of pre-trends in a dynamic panel data model, but do not consider the use of a covariate to address endogeneity, as we do here.

Our framework is closely related to classical work on models with measurement error and on panel data models with strict exogeneity. Replacement of η_{it} with x_{it} produces a factor model or a measurement error model (Aigner et al. 1984). A large literature, partially reviewed in Abbring and Heckman (2007), Heckman and Vytlacil (2007), and Matzkin (2007), shows how to establish identification in such models, typically by imposing covariance restrictions across equations governing multiple imperfect measurements of the latent factor. Instead, we impose strict exogeneity of the policy variable z_{it} with respect to the measurement error in x_{it} to achieve identification using only a single covariate.[4]

[3]Specification (6) of Table B1 in Gentzkow et al. (2011) uses a dynamic first stage analogous to Figure 1b and a static second stage analogous to (1). They provide a heuristic justification of their estimator in their footnote 5 but do not justify it formally.

[4]In the applications we have in mind, the number of covariates is small. If there are instead many covariates that contain independent information about the unobserved confounds, one may alternatively adapt methods from the

There are other ways to address policy endogeneity in linear panel data models like (1). One is to find an instrument for policy changes (Besley and Case 2000). This is an appealing approach when feasible, but in many settings such instruments are not readily available. Our approach replaces the requirement of an instrument that impacts the policy but not the outcome with the requirement of a covariate that is related to the confound but unaffected by the policy. Another approach is to impose dynamic restrictions on the relationship between x_{it} and η_{it}. In a panel data setting with mismeasured regressors, Griliches and Hausman (1986) propose to use lags of x_{it} to construct valid instruments for x_{it}. This approach requires either that the measurement errors are serially uncorrelated or that the correlation structure of the measurement error is known (Wansbeek 2001; Xiao et al. 2010). Our approach allows for arbitrary correlation in the measurement errors, but requires that the policy be strictly exogenous with respect to these errors.

2 Setup and Proposed Estimator

2.1 Model and Identifying Assumptions

We consider a static linear panel data model:

$$y_{it} = \beta z_{it} + q_{it}'\theta + \eta_{it}'\gamma + \alpha_i + \varepsilon_{it} \tag{4}$$

$$x_{it} = q_{it}'\psi + \Lambda\eta_{it} + \nu_i + u_{it}, \tag{5}$$

where y_{it} and z_{it} are observed scalars, q_{it} is an observed $Q \times 1$ vector, x_{it} is an observed $K \times 1$ vector, the $R \times 1$ vector η_{it}, the $K \times 1$ vector u_{it} and scalar ε_{it} are time-varying unobservables, α_i is a time-invariant unobserved scalar, ν_i is a time-invariant unobserved $K \times 1$ vector, and the remaining objects are conformably defined parameters. The target of interest is β. We require that $K \geq R$ and suppose for simplicity that $K = R$. We observe data $\{y_{it}, q_{it}, x_{it}\}_{i=1,t=1}^{N,T}$ and $\{z_{it}\}_{i=1,t=1-m}^{N,T+\ell}$ for $m \geq 0$ and $\ell \geq R$. We do not require that z_{it} is binary.

Vector q_{it} collects all observed exogenous variables (e.g., time period indicators) in the sense that we impose $\mathbb{E}[\eta_{it}|\{q_{it}\}_{t=1}^{T}] = \mathbb{E}[\varepsilon_{it}|\{q_{it}\}_{t=1}^{T}] = \mathbb{E}[u_{it}|\{q_{it}\}_{t=1}^{T}] = 0$ for all i and t. The vector is low-dimensional in the sense that $Q \ll N$. We do not impose any restrictions on the α_i and ν_i and thus treat them as fixed effects.

We take two steps to simplify the presentation of the results. First, we set $\theta = \psi = 0$. Statements carry over to the more general case by interpreting all data matrices as residuals from the projection of the remaining variables onto the exogenous variables. Second, we remove the fixed

literature on factor models in high dimensions as in, e.g., Stock and Watson (2002), Bai (2003), Bai and Ng (2010), and Hansen and Liao (2016).

effects. Let $\tilde{k}_{it} = k_{it} - \frac{1}{T}\sum_{s=1}^{T} k_{is}$ denote the within transformation for any variable k_{it}.[5] Then, we can simplify (4) and (5) to obtain

$$\tilde{y}_{it} = \beta\tilde{z}_{it} + \tilde{\eta}_{it}'\gamma + \tilde{\varepsilon}_{it} \tag{6}$$

$$\tilde{x}_{it} = \Lambda\tilde{\eta}_{it} + \tilde{u}_{it}. \tag{7}$$

We now state two assumptions that suffice to identify β.

Assumption 1 (Orthogonality conditions)**.** *There exists a set of non-negative integers $l = 0, 1, \ldots, L$ such that*

(a) $\mathbb{E}\left[\tilde{z}_{i,t+l}\tilde{\varepsilon}_{it}\right] = 0 \;\; \forall \;\; l$

(b) $\mathbb{E}\left[\tilde{z}_{i,t+l}\tilde{u}_{it}\right] = 0 \;\; \forall \;\; l.$

Assumption 2 (Rank conditions)**.** *Let $w_{it} = (\tilde{z}_{it}, \tilde{z}_{i,t+1}, \cdots, \tilde{z}_{i,t+L})'$ and define a matrix H as $H = \mathbb{E}(w_{it}[\tilde{z}_{it}, \tilde{x}_{it}'])$. Then:*

(a) $\mathrm{rank}(\Lambda) = R.$

(b) $\mathrm{rank}(H) = (R+1).$

Assumptions 1 and 2 are analogous, respectively, to the exclusion and relevance conditions in a linear instrumental variables setup. Strict exogeneity of z_{it} in (4), as is commonly assumed in panel event studies, implies Assumption 1(a). Strict exogeneity of z_{it} in the first stage relationship (5) implies Assumption 1(b), which allows $\tilde{z}_{it}$ and its leads to be correlated with $\tilde{x}_{it}$ only through $\tilde{\eta}_{it}$.

Remark 1. We do not require the orthogonality of $\tilde{\varepsilon}_{it}$ and $\tilde{u}_{it}$. The covariates x_{it} may be correlated with the outcome y_{it} through channels other than the confound η_{it}.

Remark 2. We may think of (7) as structural, or as a projection of $\tilde{x}_{it}$ on $\tilde{\eta}_{it}$. In principle, the latter interpretation permits the structural relationship between x_{it} and η_{it} to be nonlinear, provided the projection residuals respect Assumption 1(b).

Assumption 2(a) imposes that the covariates x_{it} contain information about all of the latent factors η_{it}. Assumption 2(b) is the equivalent of the usual instrumental variables relevance assumption and can in principle be checked in the data. It requires a nonzero correlation between the noisy proxy x_{it} and leads of z_{it}, i.e., a pre-trend in x_{it}.

Remark 3. Although Assumption 2(a) rules out that $\Lambda = 0$, we allow for $\gamma = 0$. This means that our assumptions do not imply a pre-trend in y_{it}.

[5] We also use this convention for leads and lags of a variable, so, for example, $\tilde{k}_{i,t-m} = k_{i,t-m} - \frac{1}{T}\sum_{s=1}^{T} k_{i,s-m}$.

Remark 4. If $\operatorname{rank}(\mathbb{E}(w_{it}[\tilde{z}_{it}, \tilde{\eta}'_{it}]) = (R+1)$, then Assumption 2(b) follows from Assumption 1, Assumption 2(a), and (7). That is, a pre-trend in the confound implies a pre-trend in the covariate.

Remark 5. Suppose that $z_{it} = \mathbf{1}\left(\exists t^* \leq t : \eta_{it^*} > \eta^*\right)$ for some threshold η^*. Then, Assumption 2(b) will hold for a wide range of processes. Intuitively, if η_{it} is autocorrelated, a threshold crossing at time $t+1$ provides a signal that the latent η_{it} was already large (close to the threshold) in the previous period. Economic settings covered by this case include:

- **Means-tested program.** We are interested in the effect of a household's participation z_{it} in a means-tested program on some outcome y_{it} as in Hastings and Shapiro (2018). Each household i becomes eligible for the program when the gap η_{it} between the household's income and a poverty line exceeds a threshold η^*. This setting is closely related to that in Ashenfelter (1978), who found that an individual's earnings tend to decline prior to the individual's entry into a job training program.

- **Firm entry.** We are interested in the effect of firm entry into a market on some outcome y_{it} as in Gentzkow et al. (2011). At any given time t, a single potential entrant can pay a one-time cost to enter market i and earn a stream of cash flows whose expected present discounted value is η_{it}. Under appropriate assumptions on η_{it} (for example, that it evolves as a random walk with i.i.d. innovations), the firm enters the first time that η_{it} exceeds a threshold η^* (McDonald and Siegel 1986; Dixit and Pindyck 1994). The policy z_{it} is an indicator for the presence of a firm in the market.

- **State law change.** We are interested in the effect of the passage of a law on some outcome y_{it}. A given state i passes the law when the underlying strength η_{it} of its economy exceeds some threshold η^*. The policy z_{it} is an indicator for periods after passage of the law.

Remark 6. It is straightforward to allow the dimension of z_{it} to be greater than one, and to allow for dynamic treatment effects such that $y_{it} = \sum_{j=0}^{m} \beta_j z_{i,t-j} + q'_{it}\theta + \eta'_{it}\gamma + \alpha_i + \varepsilon_{it}$. We do not pursue these in order to keep notation and statements simple.

Remark 7. Our model rules out anticipatory effects of the policy. Specifically, (6) and Assumption 1(a) exclude any direct effect of leads of z_{it} on y_{it}.

Remark 8. Although we have written (6) and (7) in terms of within-transformed variables, our analysis would apply to first-differenced variables, with corresponding changes in the interpretation of the assumptions.

2.2 GMM Representation and 2SLS Estimator

To move towards a GMM representation, use Assumption 2(a) to define the $R \times 1$ matrix $\tilde{\Gamma} = \Lambda(\Lambda'\Lambda)^{-1}\gamma$. Now define

$$\begin{aligned} \tilde{v}_{it} &\equiv \tilde{\varepsilon}_{it} - \tilde{u}_{it}'\tilde{\Gamma} & (8) \\ &= \tilde{y}_{it} - \beta\tilde{z}_{it} - \tilde{x}_{it}'\tilde{\Gamma} & (9) \end{aligned}$$

where (9) follows from (6) and (7) given the definition of $\tilde{\Gamma}$. Now from Assumption 1:

$$\mathbb{E}\left[w_{it}\tilde{v}_{it}\right] = 0. \tag{10}$$

Assumption 2(b) guarantees that the moment conditions in (10) are sufficient to identify β (and, incidentally, $\tilde{\Gamma}$).

Estimation may proceed by GMM using the sample analogues of (10) as moment conditions. For the case where T is fixed and N grows large, estimation and inference results are available under standard regularity conditions (Newey and McFadden 1994).

One convenient estimator justified by (10) is a 2SLS regression of $\tilde{y}_{it}$ on $\tilde{z}_{it}$ and $\tilde{x}_{it}$, treating the covariates $\tilde{x}_{it}$ as mismeasured regressors and the leads of $\tilde{z}_{it}$ as the excluded instruments. We will use this 2SLS estimator in our simulations and applications.

Remark 9. In principle, any functions of the leads of $\tilde{z}_{it}$ are valid instruments. In practice, we expect that T will often be moderately sized, and that the closest leads will be most informative. As a default we therefore suggest choosing the R closest leads of $\tilde{z}_{it}$ as instruments. Because the number of potential instruments will usually be small and the instruments are ordered (with closer leads more likely to be informative), BIC will often be a natural choice among formal methods for instrument selection.

Remark 10. Extending our framework to the case where $K > R$ provides many additional moment restrictions in principle, though one suspects the usual complications from using many moment conditions (Han and Phillips 2006; Newey and Windmeijer 2009) would arise in this setting.

Remark 11. Suppose that we observe $\tilde{y}_{it}$ and $\tilde{z}_{it}$ in one sample and $\tilde{x}_{it}$ and $\tilde{z}_{it}$ in another. Then we may proceed with two-sample instrumental variables estimation (Angrist and Krueger 1992; Inoue and Solon 2010) using the leads of $\tilde{z}_{it}$ as instruments for $\tilde{x}_{it}$.

3 Simulations

This section presents results from Monte Carlo simulations. These allow us to compare the performance of alternative estimators and to assess the adequacy of standard asymptotic approximations of the finite-sample distributions of our estimator.

3.1 Data-generating Processes and Estimators

Definition 1 (Data-generating processes)**.** *Throughout this section we consider the following data-generating processes (DGPs):*

- $\eta_{it} = \rho\eta_{it-1} + \zeta_{it}$, *where* $\zeta_{it} \sim N(0, \sigma_\zeta^2)$ *are i.i.d across* i *and* t.
- $z_{it} = \mathbf{1}(\{\exists t^* \leq t : \eta_{it^*} > \eta^*\})$, *where* η^* *is chosen so that the average number of events is approximately constant across different values of the simulation parameters.*[6]
- $K = 1$ *and*
$$x_{it} = \lambda\eta_{it} + u_{it} \tag{11}$$
where $u_{it} \sim N(0, \sigma_u^2)$ *are i.i.d. across* i *and* t.
- *The outcome is generated by:*
$$y_{it} = \beta z_{it} + 0.25\eta_{it} + 0.2t + \alpha_i + \varepsilon_{it} \tag{12}$$
where $\beta = 1$, $\alpha_i \sim N(0, 1)$ *are i.i.d. across* i, $\varepsilon_{it} \sim N(0, 1)$ *are i.i.d. across* i *and* t, *and* α_i *and* ε_{it} *are independent for all* i *and* t.

All of the simulations are based on the DGPs specified in Definition 1. Section 3.2 presents benchmark results for a design with $\lambda = \rho = 1$, $\sigma_\zeta^2 = 1$ and $\sigma_u^2 = 4$. We initialize η_{it} with $\eta_{i1} = 0$ and generate 20 time series observations for each i. We then use the ten time periods $t \in \{6, 15\}$ as the sample for estimation.

Section 3.3 presents more extensive results for a variety of designs with $\rho \in [0, 1)$. For these we choose σ_ζ^2 and σ_u^2 such that $\text{Var}(\tilde{\eta}_{it}) = 1$ and $\text{Var}(\tilde{x}_{it}) = 2$. To simulate these designs, we generate 20 time-series observations for each of 1000 cross-sectional units i. We initialize $\eta_{i,-19}$ as i.i.d. draws from a standard normal distribution and use the initial 20 observations $t = -19, -18, \ldots, 0$ as burn-in. We then keep an estimation sample of 10 time-series observations consisting of the periods $t = 6, 7, \ldots, 15$, retaining the full history of z_{it} so that we can construct leads and lags. Applying this procedure leaves us with $T = 10$ time-series observations on $N = 1000$ units, of which approximately 200 experience an event.[7]

[6] Specifically $\eta^* = \mathbf{1}(\rho < 0.9)(1.96 + 0.2\rho) + \mathbf{1}(\rho = 0.9)1.85 + \mathbf{1}(\rho = 1)4$. Online Appendix Figure 4 shows how the performance of our estimator changes as we change the importance of η_{it} in determining z_{it}.

[7] Online Appendix Figure 1 shows the mean number of cross-sectional observations in which an event occurs across the design space considered in the stationary case. Within each set of simulation parameters, the number of units with an event generally lies between 160 and 240. We include in our analysis all cross-sectional units, including those in which an event does not occur (Borusyak and Jaravel 2017).

To vary the strength of identification we will consider varying values of ρ in $[0, 0.9]$. As ρ increases, our instruments, the leads of z_{it}, will become stronger, resulting in better identification. On the other hand, as the autocorrelation in η_{it} approaches zero, we lose identification. Within this design, stronger persistence in η_{it} will tend to exacerbate bias that arises from failing to account for η_{it}.

To vary the quality of x_{it} as a proxy for η_{it}, we vary λ to control the population R^2 from the infeasible regression of x_{it} on η_{it} in (11). When this R^2 equals one, x_{it} is a perfect proxy, and the best possible control for η_{it} is x_{it}. As this R^2 approaches zero, the proxy x_{it} provides no signal about the latent variable η_{it}, and identification fails.

We consider four different feasible estimators for the policy effect β and its dynamic counterparts, and include individual and time fixed effects in all specifications. The first estimator we consider ignores η_{it} entirely and simply regresses the outcome y_{it} on the event indicator z_{it} ("Failing to control for η_{it}"). The second estimator uses x_{it} as a proxy for η_{it} and corresponds to the regression of the outcome y_{it} on the event indicator z_{it} and the covariate x_{it} ("Using x_{it} as proxy for η_{it}"). The third estimator is from our proposed 2SLS regression of the outcome y_{it} on the event indicator z_{it} and the covariate x_{it}, using $z_{i,t+1}$ as an excluded instrument for x_{it} ("2SLS - one lead"). Online Appendix Figure 2 presents corresponding results using the BIC to choose the number of first-stage leads.

The last estimator that we consider formalizes the idea of testing for pre-trends that is common in applied work ("Pre-testing for pre-trend"). To compute this estimator, we first compute the typical event study estimates, normalized so that the coefficient on $z_{i,t+1}$ is equal to zero. We then perform a conventional test that the coefficient on $z_{i,t+2}$ is equal to 0 at the 5% level. If we fail to reject the hypothesis, we conclude that there is no pre-trend and proceed with the analysis as in "No control."[8] If we reject the null, we conclude that there is a pre-trend and "give up." We formalize the notion of "giving up" by returning a confidence interval of $(-\infty, \infty)$ and no point estimate. When evaluating point estimates for this procedure, we consider only those cases where we do not give up. Online Appendix Figure 3 summarizes the rejection frequency of the pre-test.

3.2 Results for a Benchmark Data-generating Process

Figure 2 presents event-study estimates for a single realization from the DGP with $\rho = 1$. Specifically, each panel of Figure 2 depicts estimates of the coefficients δ_k from a different method of

[8]We designed this implementation of the pre-test procedure to match practice in empirical research based on our survey of the 2016 *American Economic Review*. For example, Bustos et al. (2016) estimate the effect of their policy variable one period in advance (Equation 13, Table A6) and report that, depending on the outcome variable, pre-trends are either not statistically different from zero or are opposite to the causal effect they estimate (section V.B). Pierce and Schott (2016) (Equation 3, Figure 4, and p. 1644) report that the estimated effect of their policy variable is statistically indistinguishable from zero in all periods prior to the policy change.

estimating the parameters of the following model:

$$y_{it} = \delta_{-6+}(1 - z_{i,t+5}) + \delta_{6+} z_{i,t-6} + \sum_{k=-5}^{5} \delta_{-k} \Delta z_{i,t+k} + \omega_t + \alpha_i + \eta_{it} + \varepsilon_{it}, \tag{13}$$

where ω_t are time effects, $(1 - z_{i,t+5})$ indicates that the event is more than five time periods in the future, and $z_{i,t-6}$ indicates the event took place more than five periods in the past. We use the normalization that $\delta_{-1} = 0$.

Figure 2 shows both pointwise 95% confidence intervals and uniform 95% sup-t confidence bands (Olea and Plagborg-Møller 2017). Applied papers commonly include pointwise confidence intervals in event plots. These permit testing only of preselected pointwise hypotheses. Uniform bands such as those we show here are designed to contain the true path of the coefficients 95% of the time, and are therefore arguably more useful for the goal of giving readers a sense of what kinds of pre-trends are consistent with the data.

Figure 2a reports results from estimating (13) including η_{it} as an additional regressor. Because η_{it} is unobserved, this approach is infeasible, but it provides a useful benchmark of best-case performance. Point estimates of pre-event trends are reasonably small and well-estimated. Estimates of the policy effects (δ_k for $k > 0$) are reasonably close to one, the true value.

Figure 2b reports estimates without any control for η_{it} and shows both strong pre-trends and substantial bias in the estimated effects of the policy. Figure 2c reports estimates based on including the observable x_{it} in place of the latent variable η_{it}. As x_{it} is a noisy measure of η_{it}, controlling for x_{it} only partially mitigates the pre-trends and the bias in the estimated policy effects relative to Figure 2b.

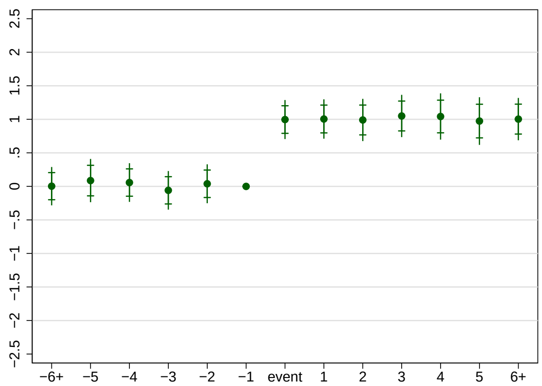
Figure 2d shows the event plot using our proposed 2SLS estimator to account for the unobserved factor η_{it}. Specifically, we proxy for η_{it} with x_{it} and instrument for x_{it} with $z_{i,t+1}$.[9] As expected, the proposed estimator delivers sensible estimates of pre-trends and policy effects, though there is a loss of precision relative to the infeasible benchmark in Figure 2a.

Figure 2e reports estimates after pre-testing. As no pre-trend is detected in this particular realization, this plot is identical to Figure 2b.

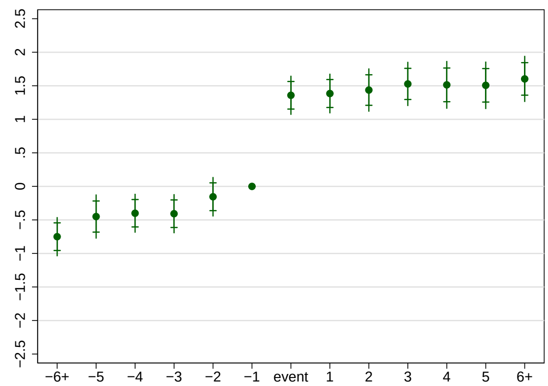
Figure 3 shows the median and uniform confidence band for the estimates in Figure 2 across repeated simulations from the same benchmark DGP. Figure 3 reinforces the conclusion from Figure 2 that, among the feasible estimators, only the 2SLS approach is centered at the true value.[10] Online Appendix Figure 6 shows that including unit-specific deterministic linear trends as a control—another common strategy for modeling confounds in a panel setting (Jacobson et al. 1993)—does

[9]Using $z_{i,t+1}$ as an instrument means that we need to normalize δ_k for an additional k. In Figure 2d we set $\delta_{-2} = 0$.

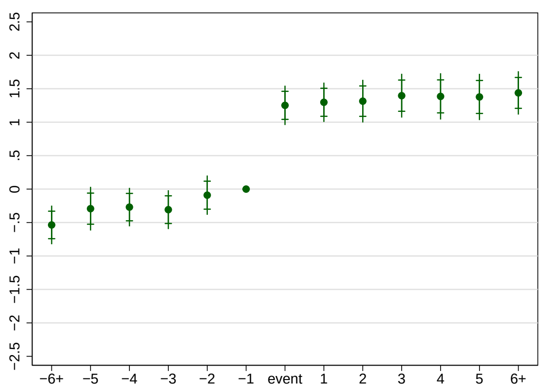
[10]Online Appendix Figure 5 shows how the appearance of plots based on our proposed 2SLS estimator depends on the choice of normalization.



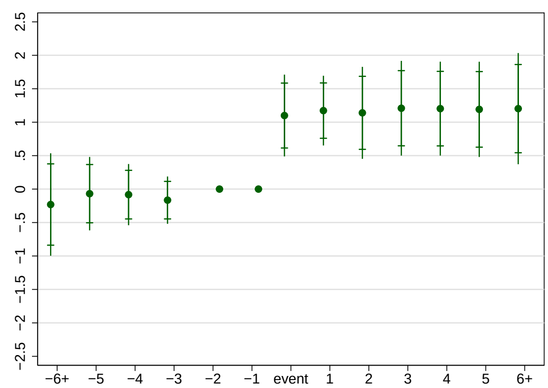
(a) Controlling directly for η_{it} (infeasible)



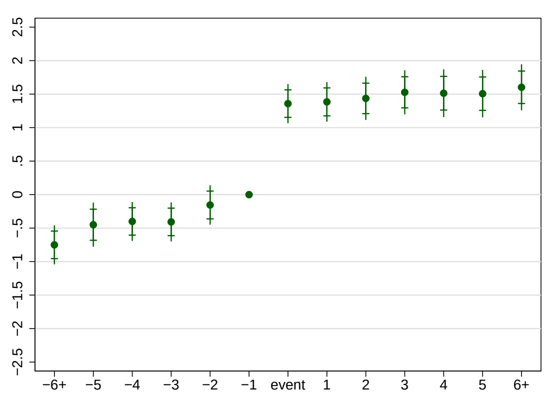
(b) Failing to control for η_{it}



(c) Using x_{it} as proxy for η_{it}

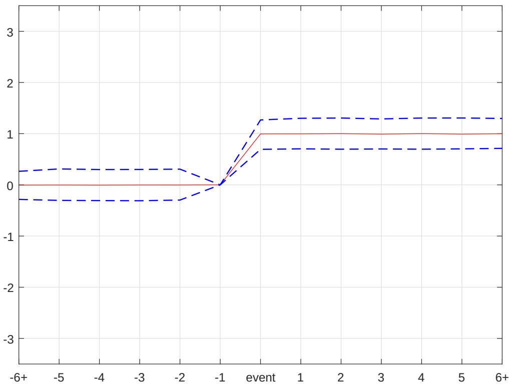


(d) Proposed 2SLS estimator, with closest lead of z_{it} as excluded instrument

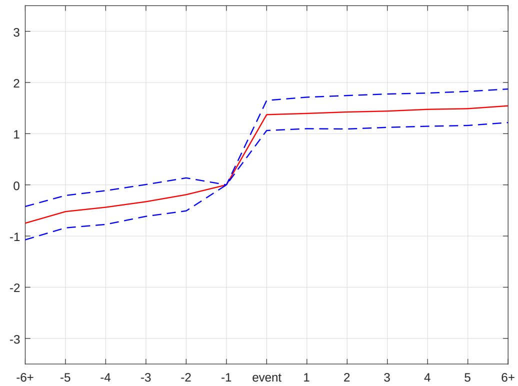


(e) Pre-testing for pre-trend

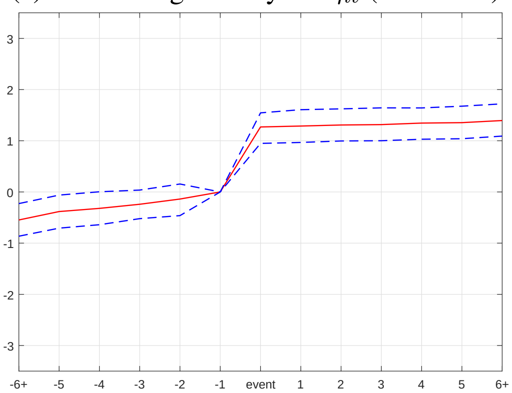
Figure 2: Exemplary event plots in the presence of a confounding factor using simulated data. Each plot shows estimates of the coefficients δ_k from (13) using either the infeasible estimator or one of the four feasible estimators defined in Section 3.1. Inner confidence sets as indicated by the dashes correspond to 95% pointwise confidence intervals, while outer confidence sets are the uniform 95% sup-t bands (with critical values obtained via simulation). Standard errors are clustered at the individual level. Data are a single draw from the benchmark DGP defined in Section 3.1 with a true causal effect of $\beta = 1$.



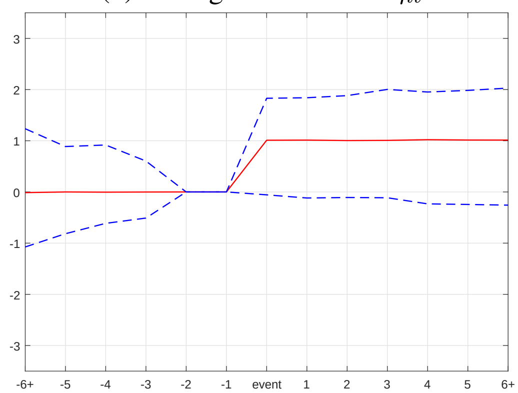
(a) Controlling directly for η_{it} (infeasible)



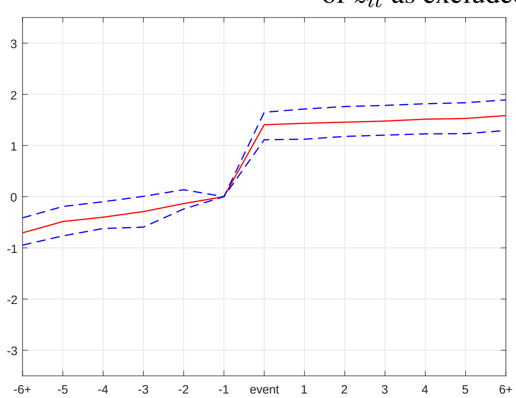
(b) Failing to control for η_{it}



(c) Using x_{it} as proxy for η_{it}



(d) Proposed 2SLS estimator, with closest lead of z_{it} as excluded instrument



(e) Pre-testing for pre-trend

Figure 3: Distribution of event plots under the presence of a confounding factor using simulated data from the benchmark DGP defined in Section 3.1 with a true causal effect of $\beta = 1$. Each plot shows estimates of the coefficients δ_k from (13) using either the infeasible estimator or one of the four feasible estimators defined in Section 3.1. The red solid line in the center represents the median estimate across 10,000 realizations, while the blue dotted lines depict the uniform 95% confidence band: 95% of the estimated sets of coefficients lie within this band. In the plot labeled "Pre-testing for pre-trend", we depict estimates that fail to control for η_{it} from the 584 realizations in which we do not detect a pre-trend.

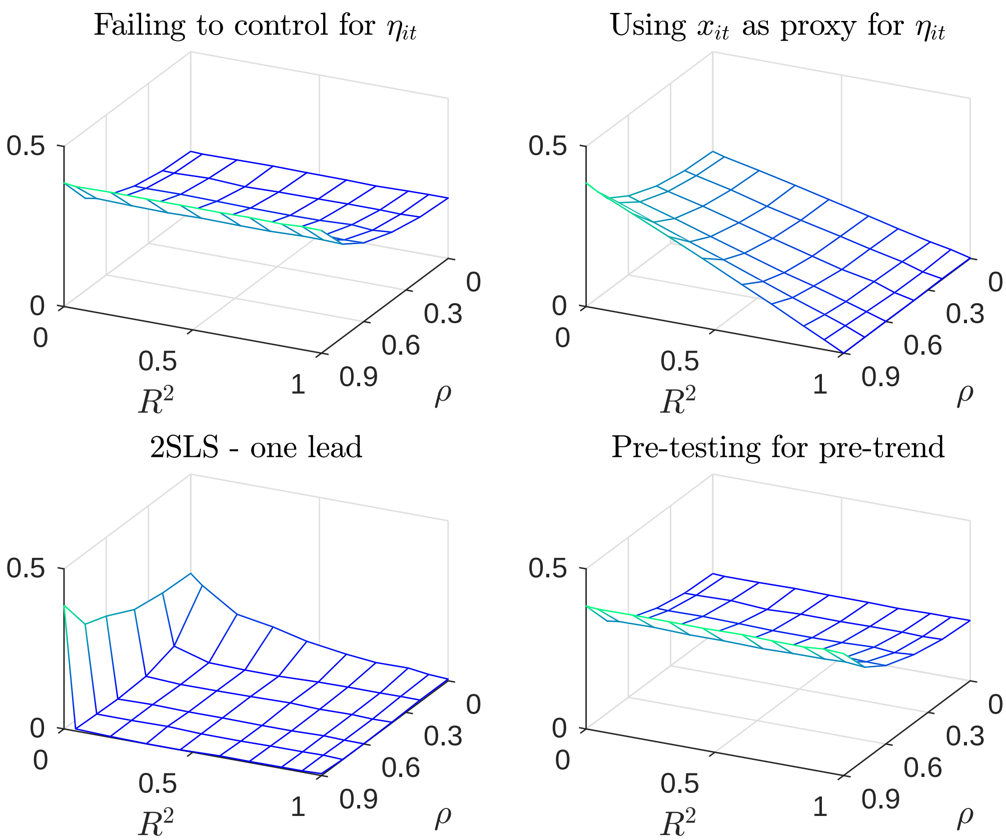


Figure 4: Median bias for each estimator defined in Section 3.1. Each point represents the median bias across 2000 simulation replications from the DGPs in Definition 1 with $\rho \in [0, 0.9]$. The horizontal axes in each panel correspond to the different values of ρ and of the population R^2 from the infeasible regression of x_{it} onto η_{it} in (11).

not lead to a correctly centered estimator.

3.3 Results for a Set of Data-generating Processes

We turn next to an exploration of the full space covered by the stationary variant of the DGPs. We consider estimates $\hat{\beta}$ from

$$y_{it} = \beta z_{it} + \omega_t + \alpha_i + \eta_{it}\gamma + \varepsilon_{it}, \tag{14}$$

where ω_t are time effects. We consider the four feasible estimators defined in Section 3.1.

Figure 4 depicts the absolute median bias of each estimator. As expected, the presence of the unobserved confound severely biases the estimator that completely fails to control for η_{it} (top left panel). Using x_{it} directly to control for η_{it} also results in severe bias except when the R^2 from the infeasible regression of x_{it} on η_{it} is very large, in which case x_{it} is a nearly perfect proxy for η_{it} (top right panel). Also in line with our expectations, the median of our proposed 2SLS estimator

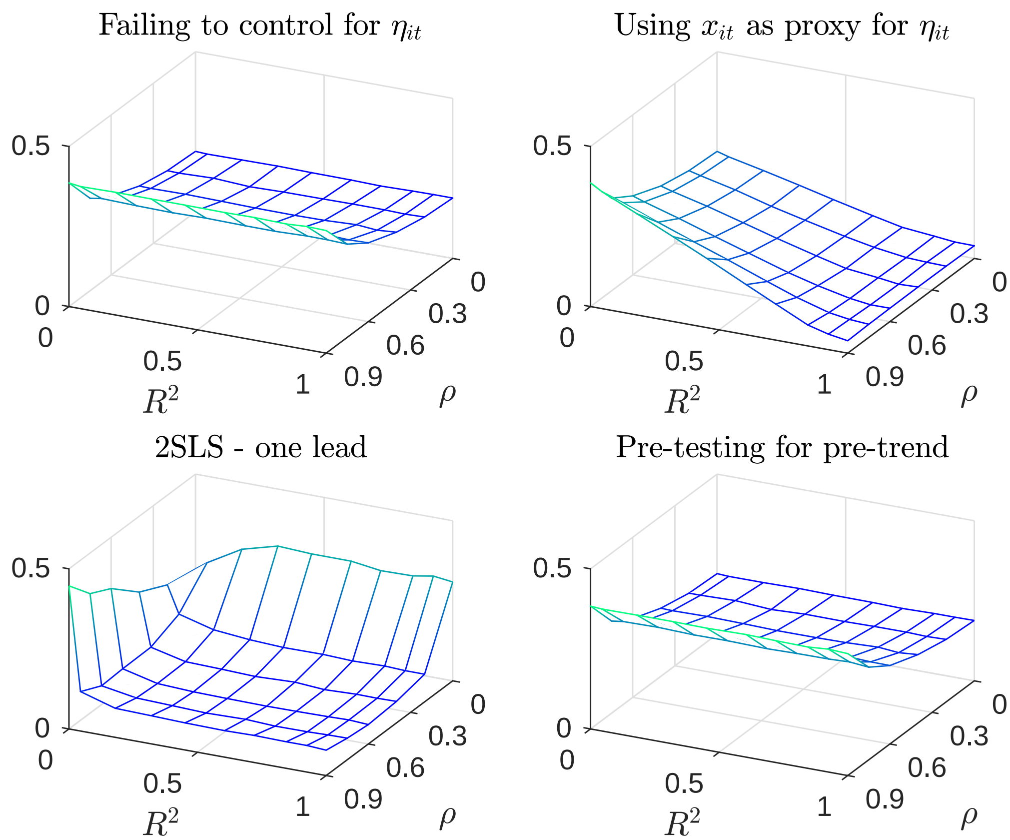


Figure 5: Median absolute deviation from the true parameter value for each estimator defined in Section 3.1. Each point represents the median absolute deviation across 2000 simulation replications from the DGPs in Definition 1 with $\rho \in [0, 0.9]$. The horizontal axes in each panel correspond to the different values of ρ and of the population R^2 from the infeasible regression of x_{it} onto η_{it} in (11).

is close to the true value across most of the parameter space (bottom left panel). The exceptions occur in the regions of weak identification, where there is either little correlation between x_{it} and η_{it} or little autocorrelation in η_{it}. Finally, the test for pre-trends leads to little improvement relative to no controls at all (bottom right panel). The reason is that the test often fails to detect pre-trends, even though they are present.

Figure 5 depicts the median absolute deviation of each estimator from the true parameter value. The sampling distribution of estimators other than our proposed 2SLS estimator are dominated by bias. Therefore, for these estimators, the plots in Figure 5 closely resemble those in Figure 4. In contrast, our proposed estimator performs well except in regions of the parameter space in which identification is weak.

Figure 6 depicts the coverage of the 95% confidence interval for each estimator constructed from the usual asymptotic approximation assuming the underlying sampling distribution is approximately normal and correctly centered. Failing to do anything to account for η_{it} results in

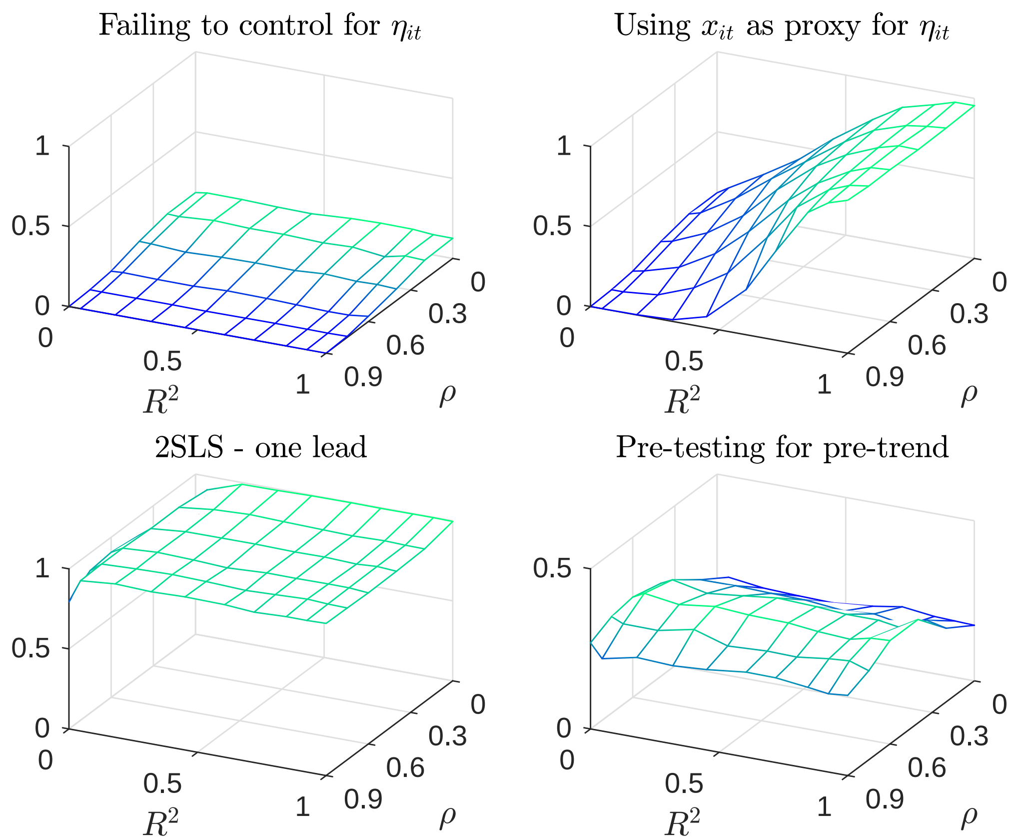


Figure 6: Coverage of the 95% confidence interval for each estimator defined in Section 3.1. Each point represents the coverage of the 95% confidence interval across 2000 simulations replications from the DGPs defined in Definition 1 with $\rho \in [0, 0.9]$. The confidence interval is constructed from the usual asymptotic approximation, with standard errors clustered at the individual level. The horizontal axes in each panel correspond to the different values of ρ and of the population R^2 from the infeasible regression of x_{it} onto η_{it} in (11).

severe size distortions across the entire parameter space (top left panel). Coverage is likewise poor when x_{it} is used directly as a proxy for η_{it}, except when x_{it} proxies η_{it} very well (top right panel). In contrast, empirical coverage for the 2SLS estimator is close to 95% throughout the parameter space, except where identification is weak (bottom left panel).[11]

The pre-test estimator exhibits uniformly poor coverage in this simulation design (bottom right panel). The observed coverage is a consequence of two offsetting patterns. When we reject the null of no pre-trend, coverage is necessarily 1 as we conclude we cannot use the data to learn about β. When we fail to detect a pre-trend and proceed as if no confound is present, coverage is close to 0 as the estimator is severely biased.

Although we present quantitative results for one particular form of pre-testing, we expect sim-

[11]Poor coverage in regions of weak identification could be corrected by applying appropriate weak-identification robust procedures (Stock et al. 2002; Andrews and Mikusheva 2016).

ilar issues would arise for any analogous attempt to weed out bad cases in advance of estimation. The reason is that such approaches are vulnerable to the presence of confounds that are small enough that they cannot be reliably detected in finite samples, but still large enough to significantly bias results obtained assuming such confounds are absent. By contrast, under the assumptions in Section 2, our proposed estimator delivers valid inferential statements when there are large confounds (in scenarios where pre-trends are detected), when there are modest confounds (where pre-trends are present but too small to be detected), and when there are no confounds (where omitting η_{it} would also be fine). Our recommendation to use the 2SLS estimator is in line with the large statistics and econometrics literature regarding the use of pre-tests for model/specification choice. Guggenberger (2010), in particular, makes a very similar argument in the context of choosing between OLS or IV estimation.

4 Applications

In this section, we apply our proposed estimator to empirical settings corresponding to the three examples discussed in Remark 5. Together these capture many of the scenarios a practitioner might encounter:

- A clear pre-trend in the outcome variable and a clear pre-trend in the covariate (Section 4.1).
- No pre-trend in the outcome variable and a clear pre-trend in the covariate (Section 4.1).
- An unclear pre-trend in the outcome variable and a clear pre-trend in the covariate (Section 4.2).
- An unclear pre-trend in the outcome variable and an unclear pre-trend in the covariate (Section 4.3).

4.1 The Effects of SNAP Participation on Household Spending Patterns

Hastings and Shapiro (2018) study the effect of participation in the Supplemental Nutrition Assistance Program (SNAP) on household spending in a panel event-study design. Here i indexes households and t indexes calendar quarters. The outcome y_{it} is either at-home food expenditures or the share of food spending going to store-brand items. The policy z_{it} is an indicator for time periods following entry into the program. SNAP is means-tested, so households become eligible when income η_{it} is sufficiently low. Past research shows that lower household income is associated with lower at-home food expenditures (Castner and Mabli 2010) and greater store-brand share (Bronnenberg et al. 2015), so income is a potential confound. Hastings and Shapiro (2018) have access to Rhode Island administrative data which includes SNAP participation z_{it} and a measure x_{it} of household income, and separate data from a grocery retailer which includes SNAP participation z_{it} and the outcomes y_{it}.[12]

Figure 7 reproduces from Hastings and Shapiro (2018) a plot of the time path of household income around the adoption of SNAP. Specifically, denoting average monthly household income during the quarter as x_{it}, we depict estimates $\hat{\delta}$ from

$$x_{it} = \delta_{-5+}(1 - z_{i,t+4}) + \delta_{5+} z_{i,t-5} + \sum_{k=-4}^{4} \delta_{-k} \Delta z_{i,t+k} + \phi_t + \nu_i + u_{it}, \qquad (15)$$

where ϕ_t are time effects.

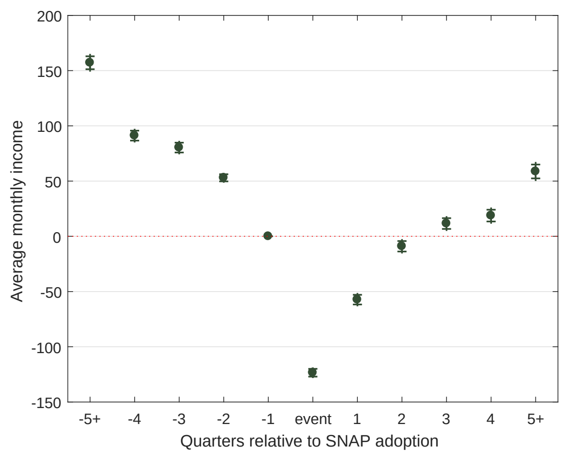


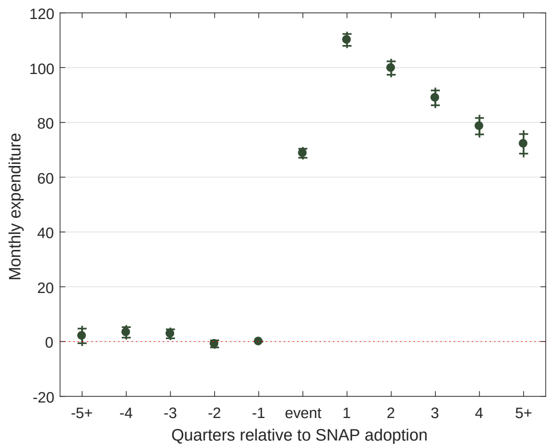
Figure 7: Estimated changes in household income at quarters around SNAP adoption. Figure plots estimates of coefficients δ from (15), with the time period one quarter prior to SNAP adoption ("-1") as the omitted category. Inner confidence sets as indicated by the dashes correspond to 95% pointwise confidence intervals, while outer confidence sets are the uniform 95% sup-t bands (with critical values obtained via simulation). Standard errors are clustered at the household level.

The patterns in Figure 7 are consistent with a model in which household income is a determinant of SNAP eligibility as in Remark 5. We see a clear decline in income in the time periods leading up to a household's adoption of SNAP. Following the adoption, we observe an increase in household income.

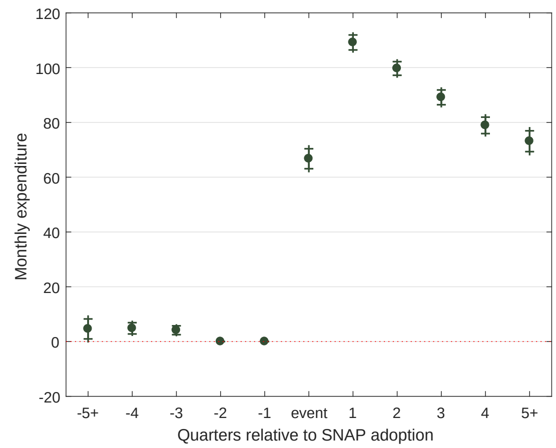
Figure 8 depicts estimates $\hat{\delta}$ from two specifications of

$$y_{it} = \delta_{-5+}(1 - z_{i,t+4}) + \delta_{5+} z_{i,t-5} + \sum_{k=-4}^{4} \delta_{-k} \Delta z_{i,t-k} + \gamma \eta_{it} + \alpha_i + \omega_t + \varepsilon_{it}, \qquad (16)$$

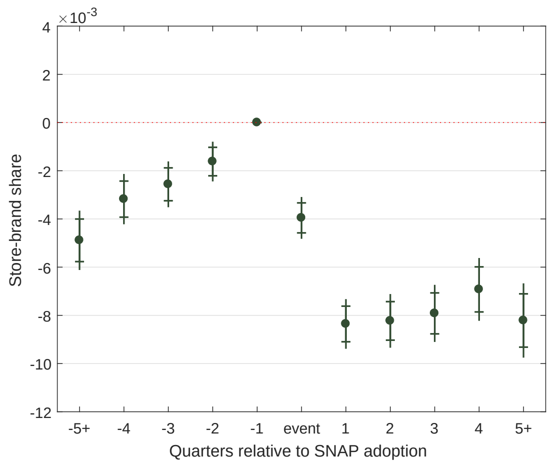
[12]The results in this section are based on regression output obtained from the authors at `http://www.brown.edu/Research/Shapiro/data/government.zip` and `http://www.brown.edu/Research/Shapiro/data/retailer.zip` on January 11, 2018.



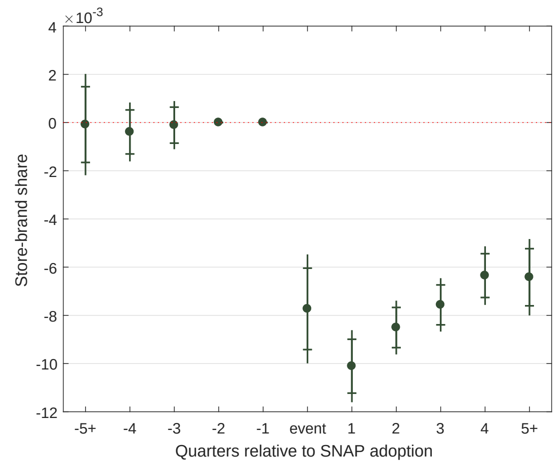
(a) At-home food expenditure around SNAP adoption, not controlling for household income.



(b) At-home food expenditure around SNAP adoption. Proposed 2SLS estimator, with z_{it+1} as excluded instrument.



(c) Store-brand share of food expenditures around SNAP adoption, not controlling for household income.



(d) Store-brand share of food expenditures around SNAP adoption. Proposed 2SLS estimator, with $z_{i,t+1}$ as excluded instrument.

Figure 8: Estimated changes in outcomes at quarters around SNAP adoption. Each figure plots estimates of coefficients δ from (16), with the time period one quarter prior to SNAP adoption ("-1") as the omitted category. Inner confidence sets as indicated by the dashes correspond to 95% pointwise confidence intervals, while outer confidence sets are the uniform 95% sup-t bands (with critical values obtained via simulation). Standard errors are clustered at the household level.

where ω_t are time effects and the outcome y_{it} represents either monthly at-home food expenditure (Figures 8a and 8b) or store-brand share of food expenditures (Figures 8c and 8d).

In Figures 8a and 8c, the term in (16) involving η_{it} is ignored and so no attempt is made to control for confounds. Figure 8a shows that there is no economically meaningful pre-trend in monthly at-home food expenditure. This is consistent with the argument in Hastings and Shapiro (2018) that the effect of cash income on food spending is small. By contrast, Figure 8c shows a

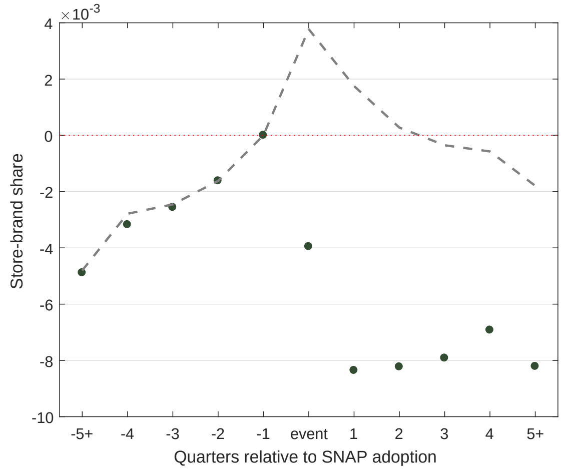


Figure 9: Geometric illustration of proposed estimator. Dashed line depicts household income (Figure 7, rescaled). Round markers depict store-brand share of food expenditures (Figure 8c) around SNAP adoption respectively. Difference depicted in Figure 8d.

clear pre-trend in store-brand share that is small in absolute terms but large relative to the change on adoption. We note that, since SNAP adoption can occur at any time in the quarter, period 0 is "partially treated".

Figures 8b and 8d use our proposed estimator, with the closest lead of z_{it} serving as an excluded instrument for x_{it}. Because y_{it} and x_{it} are not observed jointly, we use a two-sample instrumental variables estimator (Angrist and Krueger 1992; Inoue and Solon 2010). In the case of at-home food expenditures, Figure 8b shows that, as expected, taking the income confound into account does not alter the conclusions from the uncorrected plot in Figure 8a.

By contrast, Figure 8d differs markedly from Figure 8c. This is because the relatively large pre-trend in store-brand share implies a significant response to changes in income. The instrumental variable approach we take accounts for this pre-trend through the presence of the confound η_{it}, and eliminates the pre-trend from the plot. The dynamics of store-brand share that we observe following adoption likely reflects households' gradual exit from the program following adoption.

Figure 9 provides a geometric intuition for our proposed procedure. It combines a rescaled version of Figure 7 with Figure 8c. Our proposed estimator uses the dynamics in both the household income and store-brand share in the two quarters prior to the event to infer the size of the effect of the confound. Geometrically this amounts to aligning the two plots in the two-period window prior to the event. We interpret the remaining difference, depicted in Figure 8d, as an approximation of the causal effect of SNAP adoption on the store-brand share.

Figure 8 illustrates two possible scenarios for applying our approach in the presence of a clear potential confound. In the first scenario, confidence sets exclude a meaningful pre-trend in the outcome, and our proposed method formalizes the intuitive notion that the confound does not cause significant bias in the estimation of the policy effect. In the second scenario, there is a clear pre-trend in the outcome, and our method adjusts causal inference for the presence of the confound.

Table 1 presents estimates $\hat{\beta}$ from the static analogue of (16). Although a static model does not capture the post-treatment dynamics of the outcomes, it is a common way to summarize the effect size (Borusyak and Jaravel 2017). The first row shows that, with no control for household income, the estimated effect of adopting SNAP on monthly expenditure is 86 dollars, while SNAP adoption leads to a decrease in the store-brand share of 0.4%. The second row notes that controlling for household income directly is infeasible, as household income and the outcomes of interest are not observed in the same data. The third row shows that, using our proposed 2SLS estimator, the estimated effect of SNAP adoption on monthly food expenditure is 84 dollars, similar to the first row. On the other hand, the estimated effect on the store-brand share is a decrease of 0.7%, an increase in magnitude of almost 60% compared to the first row. As expected from Figure 7, the first stage is highly significant. The last row shows the estimate from pre-testing for a pre-trend. This estimate coincides with that in the first row in the case of monthly food expenditure, where there is no detectable pre-trend, and is undefined in the case of the store-brand share, where there is a clear pre-trend.

Estimator	Effect of SNAP adoption on		Coefficient on lead
	monthly expenditure	store-brand share	in first-stage
No control	85.97	-0.0044	
	(1.23)	(0.0004)	
Controlling for x_{it}	infeasible	infeasible	
Proposed 2SLS estimator	84.35	-0.0070	-151.81
(one lead)	(1.11)	(0.0004)	(2.55)
Pre-testing for pre-trend	85.97	—	
	(1.23)	(—)	

Table 1: Estimates of the effect of SNAP adoption. In the first two columns, each row corresponds to a different estimate $\hat{\beta}$ from $y_{it} = \beta z_{it} + \omega_t + \gamma \eta_{it} + \alpha_i + \varepsilon_{it}$. The first row uses no control for household income. The second row reports that controlling directly for household income is infeasible. The third row uses our proposed 2SLS estimator, treating the closest lead of SNAP adoption as an excluded instrument for household income. The last row depicts our pre-test estimator. For monthly expenditure, we cannot reject $\delta_{-2} = 0$ at the 5% level in (16) (see Figure 8a), and therefore proceed with estimation as in the first row. In contrast, for store-brand share, we reject $\delta_{-2} = 0$ at the 5% level in (16) (see Figure 8c), and therefore do not proceed with estimation. The last column shows the coefficient on the excluded instrument in the first stage of the 2SLS estimator. Standard errors in parentheses are clustered at the household level.

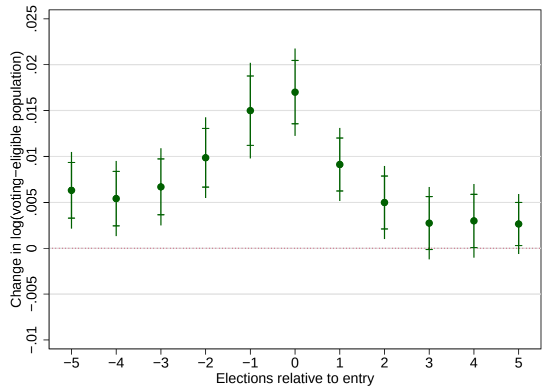


Figure 10: Estimated changes in population at election years around newspaper entries/exits. The plot shows estimates of coefficients δ_k from (17). Inner confidence sets as indicated by the dashes correspond to 95% pointwise confidence intervals, while outer confidence sets are the uniform 95% sup-t bands (with critical values obtained via simulation). Standard errors are clustered at the county level.

4.2 The Effect of Newspaper Entry and Exit on Electoral Politics

Gentzkow et al. (2011) study the effect of newspapers on voter turnout, exploiting variation generated by the entries and exits of daily newspapers to local markets in the US. Here, i indexes local markets (counties) and t indexes presidential election years. The outcome y_{it} is voter turnout. The policy z_{it} is the number of English-language daily newspapers in the market. Following Remark 5, it is reasonable to expect the entry of a newspaper to coincide with an improvement in market profitaiblity η_{it}. Because the state of the local economy could also affect voter turnout, market profitability is a potential confound. Gentzkow et al. (2011) have proxies for profitability, including a measure x_{it} of the log of the voting-eligible population.[13]

Figure 10 depicts estimates of the coefficients δ_k from

$$\Delta x_{it} = \sum_{k=-5}^{5} \delta_{-k} \Delta z_{i,t+k} + \Delta \phi_{st} + \Delta u_{it}, \tag{17}$$

where ϕ_{st} is a state-year fixed effect. The patterns in the figure are consistent with a model in which the voting-eligible population approximates newspaper profitability: We see a clear increase in population growth in the time periods leading up to a market entry, and then population growth flattens out again after an entry has occurred.

[13]We use the authors' original data, available at `https://www.aeaweb.org/articles?id=10.1257/aer.101.7.2980`, in our analysis.

Figure 11 depicts estimates of the coefficients δ_k from three specifications of the equation:

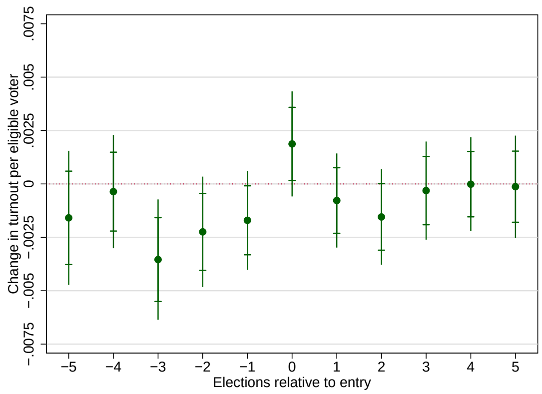
$$\Delta y_{it} = \sum_{k=-5}^{5} \delta_{-k} \Delta z_{i,t+k} + \Delta\omega_{st} + \gamma\Delta\eta_{it} + \Delta\varepsilon_{it}, \tag{18}$$

where ω_{st} is a state-year fixed effect. We omit additional control variables but show in the Online Appendix how their inclusion affects our results. In Figure 11a the term involving η_{it} is omitted from (18). This specification therefore does not control for newspaper profitability. Figure 11b controls for market profitability by directly substituting the observed x_{it} for η_{it} in (18). Figure 11c uses our proposed 2SLS estimator, with the closest lead of Δz_{it} serving as an excluded instrument for Δx_{it}. Based on the uniform bands, we reject the hypothesis of no pre-trends when no control is used, but fail to reject in the other two specifications.

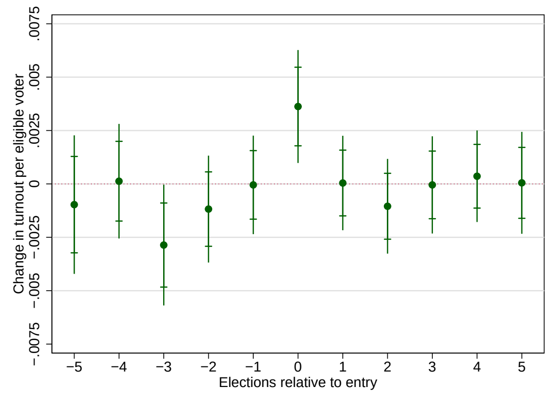
Table 2 presents estimates $\hat{\beta}$ from the static analogue of (18), which represents the causal effect of an additional newspaper on voter turnout. The first row shows that with no controls the estimated effect is 0.26 percentage points per newspaper. The second row shows that controlling for the log of voting-eligible population leads the estimate to increase to 0.37 percentage points per newspaper. The third row shows that our proposed 2SLS estimator gives an estimate of 0.34 percentage points per newspaper, which is statistically and economically similar to the estimate in

Estimator	Effect of newspaper entry	Coefficient on lead in first-stage
No control	0.0026 (0.0009)	
Controlling for x_{it}	0.0037 (0.0010)	
Proposed 2SLS estimator (one lead)	0.0034 (0.0013)	0.0128 (0.0017)
Pre-testing for pre-trend	— (—)	

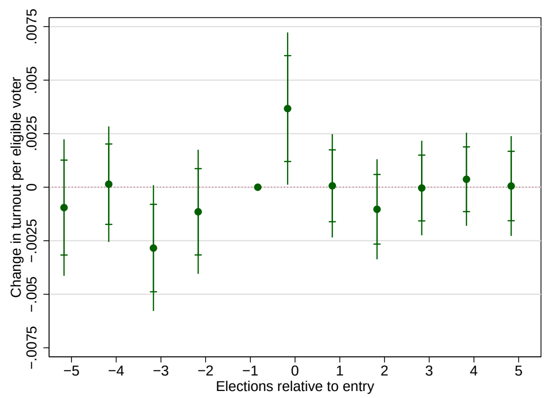
Table 2: Estimates of the effect of newspapers on voter turnout. In the first column, each row corresponds to a different estimate $\hat{\beta}$ from $\Delta y_{it} = \beta\Delta z_{it} + \Delta\omega_{st} + \gamma\Delta\eta_{it} + \Delta\varepsilon_{it}$. The first row uses no control for market profitability. The second row uses the log of the voting-eligible population as a proxy. The third row uses our proposed 2SLS estimator, treating the closest lead of the number of newspapers as an excluded instrument for the log of the voting-eligible population. The last row depicts our pre-test estimator. Because we reject $\delta_{-1} = 0$ at the 5% level in (18) (see Figure 11a), we do not proceed with estimation. The second column shows the coefficient on the excluded instrument in the first stage of the 2SLS estimator. Standard errors in parentheses are clustered at the county level.



(a) Not controlling for market profitability



(b) Using the log of voting-eligible population as a proxy for market profitability



(c) Proposed 2SLS estimator, with $\Delta z_{c,t+1}$ as excluded instrument

Figure 11: Estimated effects on presidential turnout at election years around newspaper entries/exits. The plot shows estimates of coefficients δ_k from (18). Inner confidence sets as indicated by the dashes correspond to 95% pointwise confidence intervals, while outer confidence sets are the uniform 95% sup-t bands (with critical values obtained via simulation). Standard errors are clustered at the county level.

the second row.[14]

The fourth row shows the results from the pre-test procedure defined in Section 3.1. A test of the hypothesis that the coefficient on $\Delta z_{i,t+1}$ in Figure 11a is significantly different from zero yields a $t-$statistic of -2.07 with a $p-$value of 0.039. The pre-test procedure therefore suggests not to proceed with estimation.

[14]The $p-$values for equality of estimates relative to controlling for x_{it} directly are 0.000 for the estimator with no control and 0.714 for our proposed 2SLS estimator. These $p-$values are based on 100 cluster-bootstrap replications.

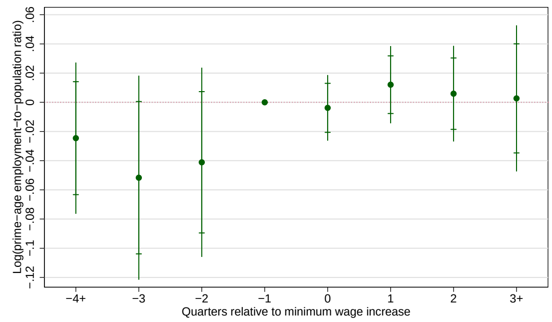


Figure 12: Prime-age employment at quarters around minimum wage increases. The plot shows estimates of coefficients δ from (19). Inner confidence sets as indicated by the dashes correspond to 95% pointwise confidence intervals, while outer confidence sets are the uniform 95% sup-t bands (with critical values obtained via simulation). Standard errors are clustered at the state level.

4.3 The Effect of the Minimum Wage on Youth Employment

There is an ongoing debate about the effect of the minimum wage on youth employment (Neumark et al. 2014; Allegretto et al. 2017). Let i index states and t index quarters. The outcome y_{it} is the log of the teen (16-19) employment-to-population ratio. The policy z_{it} is the log of the state minimum wage. The control q_{it} is the share of teenagers in the population. We may be concerned that states implement minimum-wage increases when demand η_{it} for labor is strong (Card and Krueger 1995; Neumark and Wascher 2006). We proxy for labor market conditions using a measure x_{it} of the log of the prime-age (25-55) employment-to-population ratio. For prime-age workers the effect of minimum wages is minor compared to other sources of variation (Brown 1999). However, directly controlling for x_{it}, as is commonly done, fails to allow for mismeasurement of the true demand for youth labor.

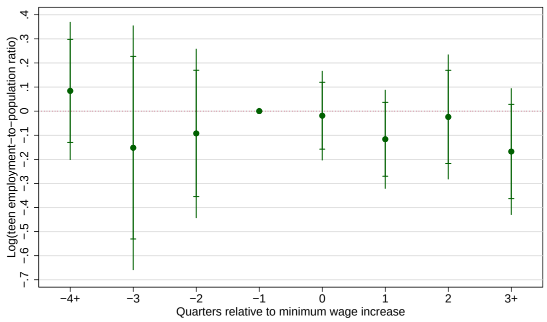
We construct data on y_{it}, x_{it}, and q_{it} from the CPS Outgoing Rotation Groups for the years $1985 - 2014$.[15] We obtain data on z_{it} from David Neumark's Minimum Wage Dataset.[16] All regressions in this section are weighted by teen population.

Figure 12 depicts the time path of our proxy, the log of prime-age employment, around minimum wage increases. Specifically, the figure depicts estimates of the coefficients δ_k from

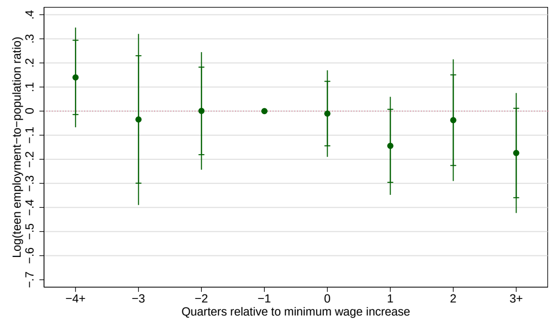
$$x_{it} = \delta_{-4+}(1 - z_{i,t+3}) + \delta_{3+} z_{i,t-3} + \sum_{k=-2}^{3} \delta_{-k} \Delta z_{i,t+k} + q_{it}' \psi + \phi_t + \nu_i + u_{it}. \tag{19}$$

[15]The Current Population Survey data is available at `http://www.nber.org/data/morg.html`. We construct the employment-to-population ratios as the proportion of individuals in the corresponding age category who self-report as either "Working" or "With a job, not at work." We weight individual observations using the final weight variable to obtain state-level aggregates.

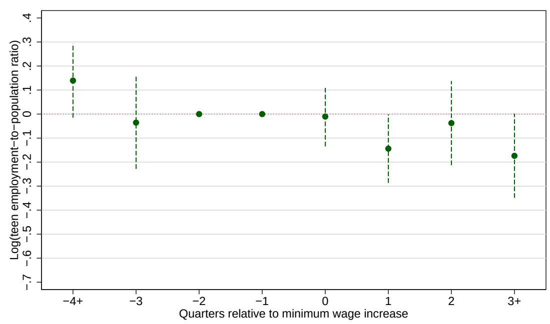
[16]The minimum wage data is available at `http://www.socsci.uci.edu/~dneumark/datasets.html` We use the higher of the federal or state minimum wage as the prevailing minimum wage.



(a) Not controlling for state of economy



(b) Using prime-age employment as a proxy for the state of the economy



(c) Proposed 2SLS estimator, with $z_{i,t+1}$ as excluded instrument

Figure 13: Teen employment at quarters around minimum wage increases. The plot shows estimates of coefficients δ_k from (20). In the top row, inner confidence sets as indicated by the dashes correspond to 95% pointwise confidence intervals, while outer confidence sets are the uniform 95% sup-t bands (with critical values obtained via simulation). In the bottom figure, dashed confidence intervals correspond to 95% pointwise confidence intervals, ignoring weak identification. Standard errors are clustered at the state level.

Here, we slightly abuse notation to define q_{it} to exclude time-period indicators. Consistent with our expectation, the point estimates indicate that increases in the minimum wage tend to occur following an increase in prime-age employment. However, the estimates are imprecise, and based on the uniform confidence intervals we cannot reject the hypothesis of no pre-trends.

Figure 13 depicts estimates $\hat{\delta}$ from three specifications of the equation:

$$y_{it} = \delta_{-4+}(1 - z_{i,t+3}) + \delta_{3+} z_{i,t-3} + \sum_{k=-2}^{3} \delta_k \Delta z_{i,t+k} + \gamma \eta_{it} + q_{it}' \theta + \omega_t + \alpha_i + \varepsilon_{it}. \tag{20}$$

In Figure 13a the term involving η_{it} is omitted from (20). This specification therefore does not control for the state of the labor market. Figure 13b uses prime-age employment x_{it} directly as a control. Figure 13c depicts the results from our proposed estimator, in which we use the closest lead of the policy, $z_{i,t+1}$, as an excluded instrument for x_{it}. Note that the first stage for this model is weak, and the confidence set for γ/λ based on inversion of the Anderson-Rubin (AR) test consists of the entire real line. A projection argument therefore implies that valid confidence sets in Figure 13c also include the entire real line.

Table 3 presents estimates $\hat{\beta}$ from the following static model, represented in first differences:

$$\Delta y_{it} = \beta \Delta z_{it} + \Delta q_{it}' \theta + \Delta \omega_t + \gamma \Delta \eta_{it} + \Delta \varepsilon_{it}. \tag{21}$$

The first row of Table 3 shows that with no controls we estimate a statistically insignificant elasticity of teen employment with respect to the minimum wage of -0.0114. The second row shows that controlling for adult employment leads the estimated elasticity to decline in absolute magnitude to -0.0094. This estimate remains statistically insignificant. The third row shows that using our proposed 2SLS estimator we estimate an elasticity of 0.0003. This estimate is statistically insignificant according both to conventional standard errors, and to a confidence interval constructed by projection based on inversion of the AR test for γ/λ, which has infinite length. The last row shows

	Effect of log(MW)	Coefficient on lead in first stage
No control	-0.0114 (0.0743)	
Controlling for prime-age employment	-0.0094 (0.0708)	
Proposed 2SLS estimator (one lead)	0.0003 (0.0668) $[-\infty, \infty]$	0.0314 (0.0136)
Pre-testing for pre-trend	-0.0114 (0.0743)	

Table 3: Estimates of the effect minimum wage on teen (16-19) employment. Dependent variable: log(employment/population). Each row corresponds to a different estimate $\hat{\beta}$ from the model in first differences given by (21). The first row uses no control for the state of the economy. The second row uses the prime-age employment as a proxy. The third row uses our proposed 2SLS estimator, treating the change in the first lead of the log of the minimum wage as an excluded instrument for the change in the log of prime-age employment. We present both conventional standard errors and a confidence interval (in square brackets) constructed by projection based on an inversion of the AR test for γ/λ. The last row depicts our pre-test estimator. Because we cannot reject $\delta_{-2} = 0$ at the 5% level in (20) (see Figure 13a), we proceed with estimation as in "No control". All regressions are weighted by teen population. Standard errors are clustered at the state level.

that, because we cannot reject the hypothesis of no pre-trends in y_{it}, the estimate from the pre-test procedure coincides with the one in the first row.

This last application demonstrates the limitations of our proposed estimator. Instrument relevance requires a strong pre-trend in the covariate x_{it}. Absent such a pre-trend, our proposed estimator is not strongly identified and our approach implies that the econometrician cannot learn about the parameter of interest. Arguably, however, that is a valid conclusion if we are concerned about a confound η_{it} and are not confident that x_{it} is a perfect proxy for that confound.

5 Conclusion

We consider a linear panel data model with possible endogeneity. The common approach of examining plots of event-study estimates, or formally testing for the presence of pre-trends, is inadequate. We show how to exploit a covariate related to the confound but unaffected by the policy of interest to perform causal inference in this setting. We validate our proposal in simulations from a range of data-generating processes, and apply it to three economic settings of interest.

References

Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, Chapter 72, pp. 5145–5303. Elsevier.

Aigner, D. J., C. Hsiao, A. Kapteyn, and T. Wansbeek (1984). Latent variable models in econometrics. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2, Chapter 23, pp. 1321–1393. Elsevier.

Allegretto, S., A. Dube, M. Reich, and B. Zipperer (2017). Credible research designs for minimum wage studies: A response to Neumark, Salas, and Wascher. *ILR Review 70*(3), 559–592.

Allegretto, S. A., A. Dube, and M. Reich (2011). Do minimum wages really reduce teen employment? Accounting for heterogeneity and selectivity in state panel data. *Industrial Relations: A Journal of Economy and Society 50*(2), 205–240.

Andrews, I. and A. Mikusheva (2016). A geometric approach to nonlinear econometric models. *Econometrica 84*(3), 1249–1264.

Angrist, J. D. and A. B. Krueger (1992). The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples. *Journal of the American Statistical Association 87*(418), 328–336.

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

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bai, J. and S. Ng (2010). Instrumental variable estimation in a data rich environment. *Econometric Theory 26*(6), 1577–1606.

Ball, R. and P. Brown (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research 6*(2), 159–178.

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

Borusyak, K. and X. Jaravel (2017). Revisiting event study designs. Working paper.

Bronnenberg, B. J., J.-P. Dubé, M. Gentzkow, and J. M. Shapiro (2015). Do pharmacists buy Bayer? Informed shoppers and the brand premium. *The Quarterly Journal of Economics 130*(4), 1669–1726.

Brown, C. (1999). Minimum wages, employment, and the distribution of income. In O. C. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3, Chapter 32, pp. 2101 – 2163. Elsevier.

Bustos, P., B. Caprettini, and J. Ponticelli (2016). Agricultural productivity and structural transformation: Evidence from Brazil. *American Economic Review 106*(6), 1320–65.

Card, D. E. and A. B. Krueger (1995). *Myth and Measurement: The New Economics of the Minimum Wage*. Princeton University Press.

Castner, L. and J. Mabli (2010). Low-income household spending patterns and measures of poverty. *Washington, DC: Mathematica Policy Research*.

Dixit, A. K. and R. S. Pindyck (1994). *Investment under uncertainty*. Princeton University Press.

Duggan, M., C. Garthwaite, and A. Goyal (2016). The market impacts of pharmaceutical product patents in developing countries: Evidence from India. *The American Economic Review 106*(1), 99–135.

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

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

Guggenberger, P. (2010). The impact of a Hausman pretest on the asymptotic size of a hypothesis test. *Econometric Theory 26*(02), 369–382.

Han, C. and P. C. B. Phillips (2006). GMM with many moment conditions. *Econometrica 74*(1), 147–192.

Hansen, C. and Y. Liao (2016). The factor-lasso and k-step bootstrap approach for inference in high-dimensional economic applications. arXiv preprint:1611.09420v2.

Hastings, J. S. and J. M. Shapiro (2018). How are SNAP benefits spent? Evidence from a retail panel. *The American Economic Review*. Forthcoming.

Heckman, J. J. and E. J. Vytlacil (2007). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, Chapter 71, pp. 4875–5143. Elsevier.

Hoynes, H. W. and D. W. Schanzenbach (2009). Consumption responses to in-kind transfers: Evidence from the introduction of the Food Stamp Program. *American Economic Journal: Applied Economics 1*(4), 109–139.

Inoue, A. and G. Solon (2010). Two-sample instrumental variables estimators. *The Review of Economics and Statistics 92*(3), 557–561.

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993). Earnings losses of displaced workers. *The American Economic Review 83*(4), 685–709.

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature 35*(1), 13–39.

Matzkin, R. L. (2007). Nonparametric identification. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, Chapter 73, pp. 5307 – 5368. Elsevier.

McDonald, R. and D. Siegel (1986). The value of waiting to invest. *The Quarterly Journal of Economics 101*(4), 707–728.

Neumark, D., J. I. Salas, and W. Wascher (2014). More on recent evidence on the effects of minimum wages in the United States. *IZA Journal of Labor Policy 3*(1), 24.

Neumark, D. and W. Wascher (2006). Minimum wages and employment: A review of evidence from the new minimum wage research. NBER working paper 12663.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 36, pp. 2111 – 2245. Elsevier.

Newey, W. K. and F. Windmeijer (2009). Generalized method of moments with many weak moment conditions. *Econometrica 77*(3), 687–719.

Olea, J. L. M. and M. Plagborg-Møller (2017). Simultaneous confidence bands: Theoretical comparisons and suggestions for practice. Working paper.

Pierce, J. R. and P. K. Schott (2016). The surprisingly swift decline of US manufacturing employment. *American Economic Review 106*(7), 1632–62.

Stock, J. H. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*(460), 1167–1179.

Stock, J. H., J. H. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics 20*(4), 518–529.

Wansbeek, T. (2001). GMM estimation in panel data models with measurement error. *Journal of Econometrics 104*(2), 259–268.

Xiao, Z., J. Shao, and M. Palta (2010). Instrumental variable and GMM estimation for panel data with measurement error. *Statistica Sinica 20*(4), 1725–1747.