American Economic Review 2019, 109(9): 3307–3338
https://doi.org/10.1257/aer.20180609

Pre-Event Trends in the Panel Event-Study Design†

By Simon Freyaldenhoven, Christian Hansen, and Jesse M. Shapiro*

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 2*SLS*, *and causal inference is valid even when endogeneity leads to pre-event trends* (*"pre-trends"*) *in the outcome. Alternative approaches perform poorly in our simulations.* (*JEL* C23, C26)

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 2007).

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.

To us, this approach seems incomplete. If pre-trends are not detected, it may be either that there are no pre-trends or that pre-trends are present but undetected due to

*Freyaldenhoven: Research Department, Federal Reserve Bank of Philadelphia, 10 Independence Mall, Philadelphia, PA 19106 (email: simon.freyaldenhoven@phil.frb.org); Hansen: Booth School of Business, University of Chicago, 5807 S. Woodlawn Ave., Chicago, IL 60637 (email: chansen1@chicagobooth.edu); Shapiro: Department of Economics, Brown University, Providence, RI 02906, and NBER (email: jesse_shapiro_1@brown.edu). Thomas Lemieux was the coeditor for this article. We thank Isaiah Andrews, Matias Cattaneo, Raj Chetty, Jeff Clemens, Amy Finkelstein, Josh Gottlieb, Pat Kline, David Neumark, Matt Notowidigdo, Emily Oster, Jonathan Roth, Bryce Steinberg, and seminar participants at Brown University, Hebrew University, the NBER Summer Institute, Stanford University, and Tel Aviv University for helpful comments. We thank Justine Hastings for sharing regression output. We acknowledge financial support from the National Science Foundation under grant 1558636 and grant 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 Federal Reserve Bank of Philadelphia or the Federal Reserve System.

†Go to https://doi.org/10.1257/aer.20180609 to visit the article page for additional materials and author disclosure statements.

[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, Dube, and Reich (2011) provide a plot of pre-trends.

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 x_{it} responds to labor demand η_{it} but plausibly 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 $t = 1, \ldots, T$ and the policy z_{it} in periods $t = 1 - L, \ldots, T + L$ for some $L \geq 1$ for units $i = 1, \ldots, N$. Say that

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

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

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

where (1) defines the causal model, (2) assumes 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. If the parameter γ is known to equal 0, then the confound does not affect the outcome, and identification of β is immediate.

Panel A of Figure 1 plots coefficients from a regression of y_{it} on $\{\Delta z_{i,t+l}\}_{l=-L}^{L}$ 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 estimates depicted are sometimes called "event-study estimates" (Hoynes and Schanzenbach 2009; Duggan, Garthwaite, and Goyal 2016).

Panel A shows a clear pre-trend in the outcome, indicating that $\gamma \neq 0$. Panel B 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 the

[2] Brown (1999, Table 3) describes 13 sets of models of the effect of the minimum wage on teenage or young adult unemployment that have been estimated using state-level panel data. In 9 of these, there is a control for the prime-age male unemployment rate. All of the rest include a control for the contemporaneous or past employment-to-population ratio, either for all workers or for males only.

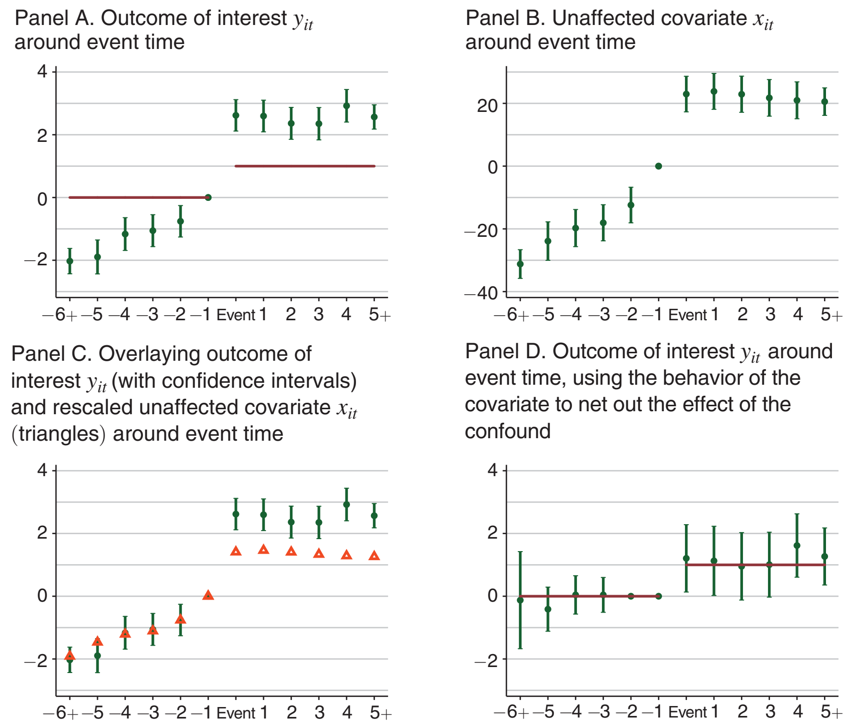

FIGURE 1. HYPOTHETICAL EVENT PLOTS

Notes: 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. Depicted are regression coefficients on indicators for time relative to policy change. Solid lines depict the true causal effect.

covariate (i.e., $\gamma = \lambda$). Extrapolating a trend in the outcome will be suitable only if the post-event behavior of the confound η_{it} can be inferred from its pre-event trend.

The alternative that we propose can be understood with reference to panel C of Figure 1. Here, we rescale the series in panel B so that it exactly matches that in panel A in the two periods immediately before the event. Under our maintained assumptions, comparing the two series in panel C 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 panel D removes the estimated effect of the pre-trend from panel A, 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 panel A of Figure 1 plots the reduced form for the outcome and panel B 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} 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 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 I 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 generalized method of moments (GMM) representation, from which standard results on estimation and inference (with large N and fixed T) are available.

Section II 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. We find that, 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}. We further find that our estimator outperforms the approach of extrapolating a linear trend from the pre-event period and the approach of conducting a test for pre-trends before proceeding with estimation.

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.) The choice of covariate should be guided by economic reasoning about the nature of the confound. In most applications we expect the number of plausible candidate covariates to be small, and we expect that our approach will work best when a small number of economic factors capture the important sources of endogeneity.

Section III 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, Shapiro, and Sinkinson 2011), and the effect of the minimum wage on youth employment (Neumark, Salas, and Wascher 2014; Allegretto et al. 2017). These applications illustrate a range of possibilities, including situations with clear pre-trends in the outcome, a situation without meaningful pre-trends, and a situation 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.

Section IV extends our model to cover the case of estimating a dynamic treatment effect and discusses issues of model testing and instrument selection when the model is overidentified.

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, Shapiro, and Sinkinson (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. Roth (2018) studies the bias introduced by pre-testing

[3] Specification (6) of Table B1 in Gentzkow, Shapiro, and Sinkinson (2011) uses a dynamic first stage analogous to panel B of Figure 1, and a static second stage analogous to (1). Gentzkow, Shapiro, and Sinkinson (2011, footnote 5) justify this estimator informally.

for pre-trends and shows how to correct for it.[4] Neither paper considers the use of covariates 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.[5] Replacement of η_{it} with x_{it} produces a factor model or 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.

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 such instruments are not readily available in many settings. 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, Shao, and Palta 2010). Our approach allows for arbitrary correlation in the measurement errors, but requires that the policy be strictly exogenous with respect to these errors. Yet another approach is to try to recover the time-series properties of η_{it}. Ashenfelter and Card (1985) estimate a model of the outcome y_{it} on data from a set of units unaffected by the policy, and use the estimated model to construct a counterfactual for the units affected by the policy under maintained assumptions about the determinants of the timing of policy change (see also Heckman and Robb 1985).

Another alternative approach to estimating policy effects in event-study settings are synthetic control methods (Abadie, Diamond, and Hainmueller 2010). These methods have been designed for settings with a small number of treated units, whereas our asymptotics are based on a large number of treated units. This makes direct formal comparison difficult. Intuitively, as synthetic control methods rely on using pre-event observable data to construct a counterfactual trend in the absence of the policy, conditions for valid inference tend to involve a form of exogeneity of treatment conditional on the observables used to construct the counterfactual (e.g., Chernozhukov, Wüthrich, and Zhu 2017; Ferman and Pinto 2017; Li 2017). Our approach instead allows for a latent trend that cannot be directly captured with observables, at the cost of additional economic structure.

[4] See also Kahn-Lang and Lang (forthcoming). More broadly, our recommendation to account for endogeneity in estimation, rather than pre-testing for it, 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 ordinary least squares (OLS) or instrumental variables (IV) estimation.

[5] See also Altonji, Elder, and Taber (2005) and Oster (2019), who propose an alternative way to use observed covariates to allow for unobservable confounds.

I. Setup and Proposed Estimator

A. *Model*

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. 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. The parameter of interest is β.

Vector q_{it} collects all observed exogenous variables (e.g., time period indicators) in the sense that we impose $E\left[\varepsilon_{it} | \{q_{it}\}_{t=1}^{T}\right] = E\left[u_{it} | \{q_{it}\}_{t=1}^{T}\right] = 0$ for all i and t. Vector q_{it} 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 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}.[6] Then, we can simplify (4) and (5) to obtain

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

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

REMARK 1: *The model in* (6) *is static in the sense that it features neither anticipatory effects nor dynamic treatment effects. We pursue these as an extension in Section IVA.*

REMARK 2: *We assume throughout that the causal effect β of the policy on the outcome is homogeneous across units i. A recent literature explores properties* (*and failures*) *of the two-way fixed effects estimator under heterogeneous causal effects* (*Abraham and Sun 2018, Athey and Imbens 2018, de Chaisemartin and D'Haultfoeuille 2018, Goodman-Bacon 2018*). *We expect similar issues to arise in our setting.*

[6] We also use this convention for leads and lags of a variable, so, for example, $\tilde{k}_{i,t-m} = k_{i,t-m} - \frac{1}{T}\sum_{s=1}^{T} k_{i,s-m}$. Although we simplify our model in terms of within-transformed variables, our analysis would apply to first-differenced variables, with corresponding changes in the interpretation of the assumptions.

B. *Identifying Assumptions*

We now state two assumptions that suffice to identify β.

ASSUMPTION 1 (Orthogonality Conditions): *There exists a set of nonnegative integers* $\mathcal{L} = \{0, 1, \ldots, L\}$ *such that*

(i) $E[\tilde{z}_{i,t+l}\tilde{\varepsilon}_{it}] = 0, \forall l \in \mathcal{L}$.

(ii) $E[\tilde{z}_{i,t+l}\tilde{u}_{it}] = 0, \forall l \in \mathcal{L}$.

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

(i) $\operatorname{rank}(\Lambda) = R$.

(ii) $\operatorname{rank}(H) = (R+1)$.

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(i), which allows $\tilde{z}_{it}$ and its leads to be correlated with $\tilde{y}_{it}$ only through $\tilde{\eta}_{it}$ or through the causal effect of the policy. Strict exogeneity of z_{it} in the first-stage relationship (5) implies Assumption 1(ii), which allows $\tilde{z}_{it}$ and its leads to be correlated with $\tilde{x}_{it}$ only through $\tilde{\eta}_{it}$. Strict exogeneity thus rules out a causal effect of the policy z_{it} on the covariate x_{it} and any correlation between the policy z_{it} and the measurement error u_{it}. We do not require the orthogonality of $\tilde{\varepsilon}_{it}$ and $\tilde{u}_{it}$, and thus the covariates x_{it} may be correlated with the outcome y_{it} through channels other than the confound η_{it}.

Assumption 2(i) imposes that the covariates x_{it} contain information about all of the latent factors η_{it}. Assumption 2(ii) 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}. On the other hand, because we allow for $\gamma = 0$, our assumptions do not imply a pre-trend in y_{it}.

REMARK 3: *If* $\operatorname{rank}(E(w_{it}[\tilde{z}_{it}, \tilde{\eta}'_{it}])) = (R+1)$, *then Assumption* 2$(ii)$ *follows from Assumption* 1, *Assumption* 2(i), *and* (7). *That is, a pre-trend in the confound implies a pre-trend in the covariate.*

REMARK 4: *Suppose that* $z_{it} = \mathbf{1}(\{\exists t^* \leq t : \eta_{it^*} > \eta^*\})$: *the policy* z_{it} *changes when* η_{it} *crosses some threshold* η^*. *Then, Assumption* 2(ii) *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, Shapiro, and Sinkinson (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 following passage of the law.*

REMARK 5: *It is also useful to consider examples of economic settings in which Assumption 2 will fail. These include:*

- Randomized controlled trial. *Suppose that the policy z_{it} is randomly assigned and therefore statistically independent of all unobservables. Then there is no pre-trend in the confound, and we would expect Assumption 2(ii) to fail.*
- Poor covariate. *Suppose that the covariate x_{it} is unrelated to the confound η_{it}. Then $\Lambda = 0$ and Assumption 2(i) fails. A related issue is that if Λ is nearly rank-deficient, then the model is only weakly identified. We illustrate these issues in simulations in Section II.*

REMARK 6: *In order to keep notation and statements simple, we treat z_{it} as univariate. It is straightforward to allow the dimension of z_{it} to be greater than 1. We allow for $R > 1$ throughout. We note, however, that the rank condition in Assumption 2(ii) is likely to become increasingly demanding as R grows, and in our simulations and applications we consider only cases with $R = 1$. A model with $R = 1$ may be thought of as an approximation to a model with $R > 1$ in which the different suspected confounds have similar dynamics around the event time. In this sense, we think our proposed approach will be most useful in settings in which the confound can be well approximated by a small number of economic factors.*

C. *GMM Representation and 2SLS Estimator*

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

$$\tilde{v}_{it} \equiv \tilde{\varepsilon}_{it} - \tilde{u}_{it}'\tilde{\Gamma} \tag{8}$$

$$= \tilde{y}_{it} - \beta\tilde{z}_{it} - \tilde{x}_{it}'\tilde{\Gamma}, \tag{9}$$

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

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

Assumption 2(ii) 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.[7] We will use this 2SLS estimator in our simulations and applications.

REMARK 7: *In principle, any functions of the leads of the event,* $\tilde{z}_{i,t+l}$, $l \in \mathcal{L}$, *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, which results in an exactly identified model. This is in line with the way we implement our estimator in both our simulations and applications. We discuss issues of overidentification and instrument selection in Section IVB.*

REMARK 8: *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* $\tilde{x}_{it}$.

REMARK 9: *If, in a given economic setting, we are concerned about failures of the exclusion restrictions in Assumption 1, we may apply existing approaches to adjust inference for plausible violations of moment conditions (e.g., Conley, Hansen, and Rossi 2012; Andrews, Gentzkow, and Shapiro 2017). Because Assumption 1(i) follows from the common assumption of strict exogeneity, we expect that in most cases Assumption 1(ii) will be the more controversial component of Assumption 1. In Section III, we discuss the economic content of Assumption 1(ii) in the context of our applications.*

II. Simulations

This section presents results from Monte Carlo simulations. These allow us to compare the performance of alternative estimators and to assess the adequacy

[7] For example, in the case where $R = K = L = 1$, this approach is equivalent to estimating a linear instrumental variables model with structural equation $\tilde{y}_{it} = \beta\tilde{z}_{it} + \tilde{\Gamma}\tilde{x}_{it} + \tilde{v}_{it}$ and corresponding first-stage equation $\tilde{x}_{it} = \pi_0\tilde{z}_{it} + \pi_1\tilde{z}_{i,t+1} + \tilde{\zeta}^x_{it}$, where π_0 and π_1 are parameters, $\tilde{\zeta}^x_{it}$ is the first-stage error, and $\tilde{\Gamma}$ and $\tilde{v}_{it}$ are defined as above.

of standard asymptotic approximations of the finite-sample distributions of our estimator.

A. *Data-Generating Processes and Estimators*

DEFINITION 1 (Data-Generating Processes): *Throughout this section, we consider the following data-generating processes (DGPs):*

- $\eta_{it} = \rho\eta_{i,t-1} + \zeta_{it}$, *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.*[8]
- $K = 1$ *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 IIB 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 10 time periods $t \in \{6,7,\ldots,15\}$ as the sample for estimation.

Section IIC 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 1,000 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 $N = 1{,}000$ units, of which approximately 200 experience an event.[9] As online Appendix Figure 3 illustrates, these designs feature a mean-reverting confound as in, for example, Ashenfelter (1978).

[8] Specifically, $\eta^* = \mathbf{1}(\rho \leq 0.8)(1.96 + 0.2\rho) + \mathbf{1}(\rho = 0.9)1.85 + \mathbf{1}(\rho = 1)4$. Online Appendix Figure 1 shows how the performance of our estimator changes in a specification where z_{it} is determined by η_{it} and an additional noise variable that allows us to vary the importance of η_{it} in determining z_{it}.

[9] Online Appendix Figure 2 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, at least 99.4 percent of draws have between 160 and 240 units with an event. 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 different 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 the 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 1, x_{it} is a perfect proxy, and the best possible control for η_{it} is x_{it}. As this R^2 approaches 0, the proxy x_{it} provides no signal about the latent variable η_{it}, and identification fails.

We consider five 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 4 presents corresponding results using the Bayesian information criterion (BIC) to choose the number of first-stage leads. The fourth estimator attempts to account for the confound by extrapolating a linear trend from the three periods immediately preceding the event (*extrapolating a linear trend*).[10]

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 implement this estimator, we first compute the typical event-study estimates, normalized so that the coefficient on $z_{i,t+1}$ is equal to 0. We then perform a conventional test that the coefficient on $z_{i,t+2}$ is equal to 0 at the 5 percent level. If we fail to reject the hypothesis, we conclude that there is no pre-trend and proceed with the analysis as in *failing to control*.[11] 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 6 summarizes the rejection frequency of the pre-test.

B. *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

[10] See Dobkin et al. (2018) for a recent article using this approach and Hausman and Rapson (2018) for a discussion of related estimators. Online Appendix Figure 5 reports findings from an estimator that includes unit-specific deterministic linear trends as a control (Jacobson, LaLonde, and Sullivan 1993).

[11] 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, Caprettini, and Ponticelli (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 VB). 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.

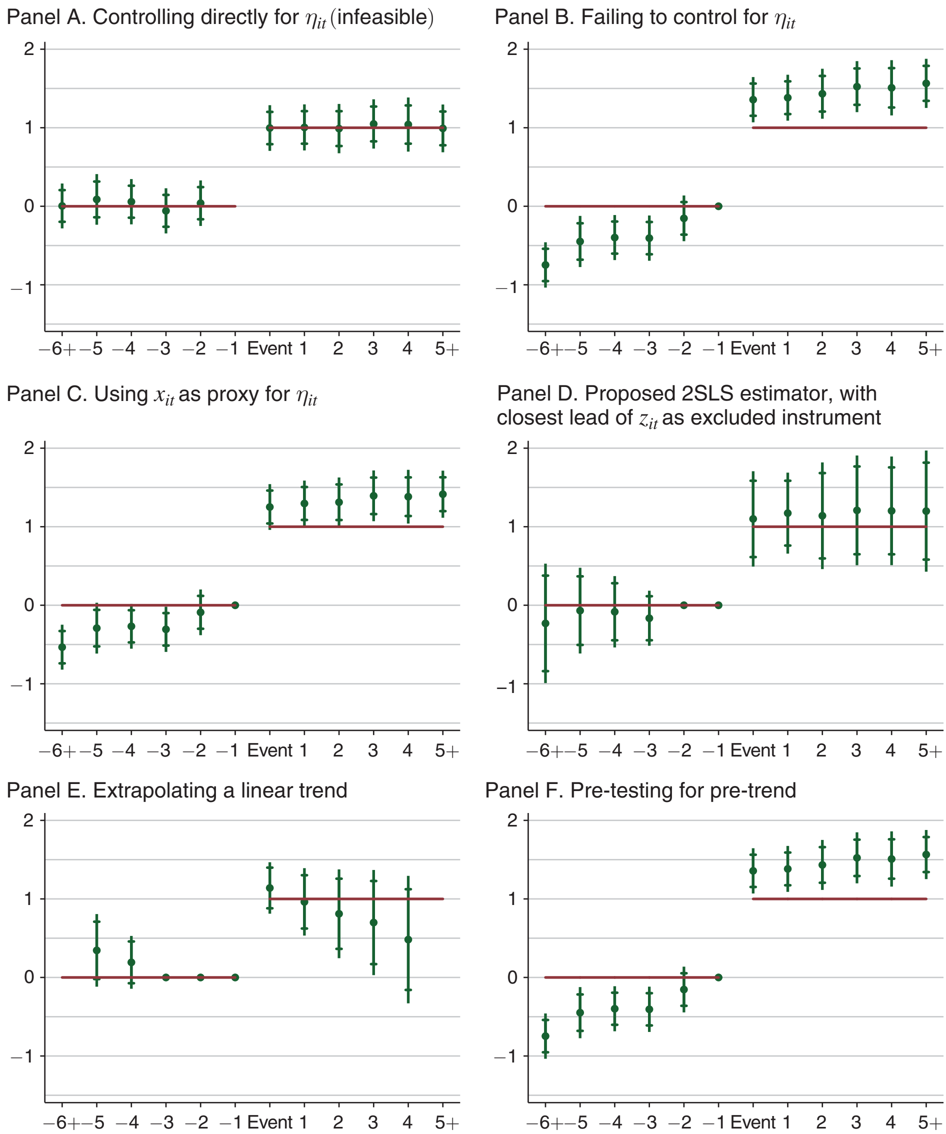

FIGURE 2. EXEMPLARY EVENT PLOTS IN THE PRESENCE OF A CONFOUNDING FACTOR USING SIMULATED DATA

Notes: All plots are based on a single draw from the benchmark DGP defined in Section IIA with a true causal effect of $\beta = 1$, represented by the solid line. Each plot shows estimates of the coefficients δ_k from (13) using either the infeasible estimator or one of the five feasible estimators defined in Section IIA. Inner confidence sets as indicated by the dashes correspond to 95 percent pointwise confidence intervals, while outer confidence sets are the uniform 95 percent sup-t bands (with critical values obtained via simulation). Standard errors are clustered at the individual level.

coefficients δ_k from a different method of estimating the parameters of the following model:

$$(13)\quad y_{it} = \delta_{-6+}(1 - z_{i,t+5}) + \delta_{5+} z_{i,t-5} + \sum_{k=-4}^{5} \delta_{-k} \Delta z_{i,t+k} + \omega_t + \alpha_i + \eta_{it} + \varepsilon_{it},$$

where ω_t are time effects, $\left(1 - z_{i,t+5}\right)$ 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 percent confidence intervals and uniform 95 percent sup-t confidence bands (Olea and Plagborg-Møller 2019). Applied papers commonly include pointwise confidence intervals in event plots.[12] 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 percent of the time, and are therefore arguably more useful for giving readers a sense of what kinds of pre-trends are consistent with the data.

Panel A of Figure 2 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 1, the true value.

Panel B reports estimates without any control for η_{it} and shows both strong pre-trends and substantial bias in the estimated effects of the policy. Panel C 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 panel B.

Panel D 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 $z_{i,t+1}$.[13] 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 panel A. As we discuss in Section IVB, inspection of pre-trends in this corrected plot may be thought of as a visual test of overidentification in a model with multiple exogenous leads.

Panel E extrapolates a linear trend from the three periods immediately preceding the event.[14] Let $p_{it} = \left(t+1\right) - \min\left\{t' : z_{it'} = 1\right\}$ be the "event time" of period t for unit i, normalized to be 0 in the period before the policy change. Then we estimate

$$(14) \qquad y_{it} = \delta_{-6+}\left(1 - z_{i,t+5}\right) + \delta_{5+} z_{i,t-5} + \Omega p_{it} \mathbf{1}\left(-4 \leq p_{it} \leq 5\right) + \sum_{k=4}^{5} \delta_{-k} \Delta z_{i,t+k} + \sum_{k=-4}^{0} \delta_{-k} \Delta z_{i,t+k} + \omega_t + \alpha_i + \varepsilon_{it}.$$

The coefficient Ω is the slope of the trend, and each δ_k represents the deviation of y_{it} relative to the trend when $-4 \leq p_{it} \leq 5$. Panel E of Figure 2 depicts the

[12] Of the 9 articles in the 2016 *AER* that include an event plot, 7 include confidence intervals on the plot, of which all 7 are pointwise.

[13] Using $z_{i,t+1}$ as an instrument means that we need to normalize δ_k for an additional k. In panel D of Figure 2, we set $\delta_{-2} = 0$. The 2SLS specification appears noisy in the plot because there is only a modest pre-trend in x_{it} in this specification. Since instrument strength varies across normalizations, the precision of our estimator will also vary across normalizations. Online Appendix Figure 7 depicts how our proposed 2SLS estimator depends on the choice of normalization. We discuss instrument choice in Section IVB.

[14] Online Appendix Figure 8 shows results from extrapolating from the two, three, four, or five periods immediately preceding the event.

corresponding δ_k. In this realization, the trend understates the role of the confound in the immediate post-event period, and overstates it in subsequent periods.

Panel F of Figure 2 reports estimates after pre-testing.[15] As no pre-trend is detected in this particular realization, this plot is identical to panel B.

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 estimator is centered at the true value. In online Appendix Table 1, we show the median bias, median absolute deviation, and coverage of the 95 percent confidence intervals for an estimate of the causal parameter β from a static analogue of the dynamic specifications depicted in Figure 3. The proposed 2SLS estimator exhibits the lowest median bias and median absolute deviation among the feasible estimators.

C. *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

$$y_{it} = \beta z_{it} + \omega_t + \alpha_i + \eta_{it}\gamma + \varepsilon_{it}, \tag{15}$$

where ω_t are time effects. We consider the one infeasible and five feasible estimators defined in Section IIA. Each estimator takes a different approach to addressing the confound η_{it}. In the case of the linear extrapolation estimator, we use $\hat{\beta} = \frac{1}{5}\sum_{k=-4}^{0}\hat{\delta}_{-k}$, with δ_k from equation (14), as an estimate for the causal effect β.[16]

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} (panel B). 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} (panel C). Also in line with our expectations, the median of our proposed 2SLS estimator is close to the true value across most of the parameter space (panel D). 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}. Extrapolating a linear trend from the pre-event period (panel E) produces biased estimates across the parameter space. Finally, pre-testing for pre-trends leads to little improvement relative to no controls at all (panel F). As online Appendix Figure 6 illustrates, even when $\rho = 0.9$, we reject the null of no pre-trend in less than 30 percent of simulations, even though they are always present in population.

Figure 5 depicts the median absolute deviation of each estimator from the true parameter value. The sampling distributions of estimators other than our proposed 2SLS estimator are dominated by bias. Therefore, for these estimators, the plots

[15] Online Appendix Figure 9 shows results from pre-testing based on the hypothesis that δ_k is equal to 0 for multiple periods preceding the event, and from pre-testing based on the hypothesis that the slope coefficient Ω on event time in (14) is equal to 0.

[16] Online Appendix Figure 10 shows results when we instead use $\hat{\beta} = \hat{\delta}_0$ as an estimate for the causal effect β.

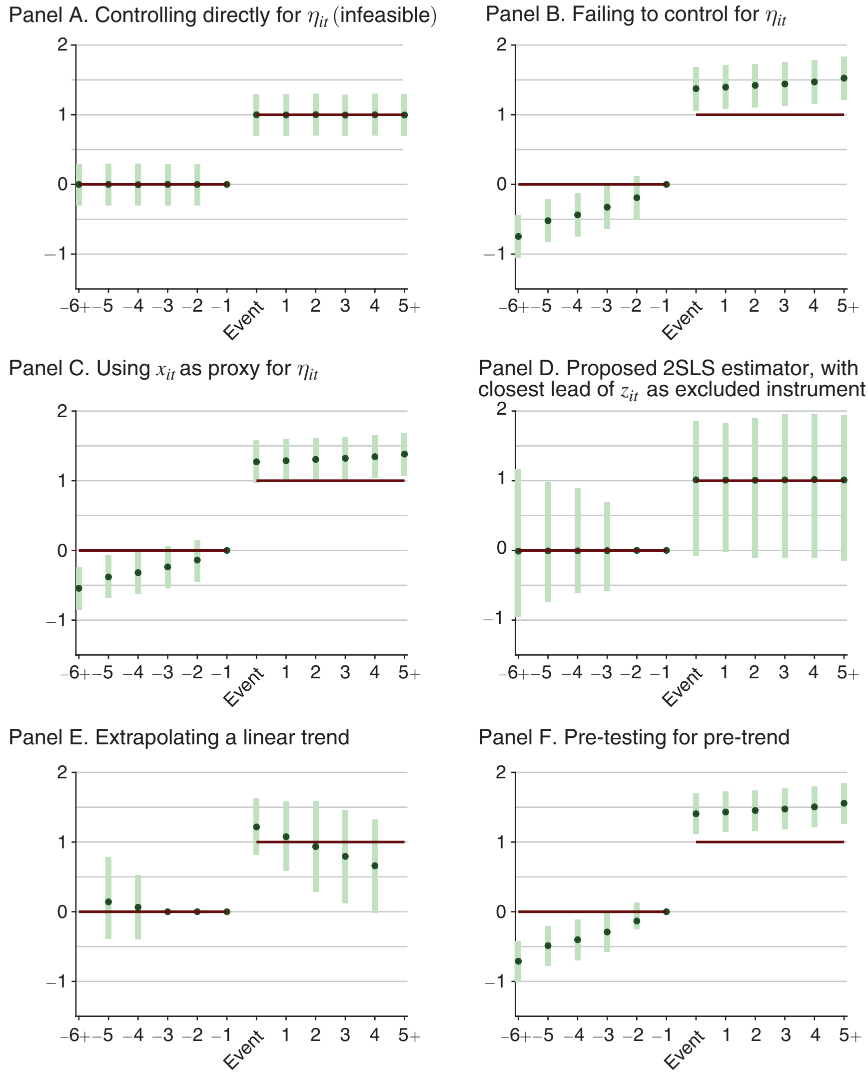

FIGURE 3. DISTRIBUTION OF EVENT PLOTS UNDER THE PRESENCE OF A CONFOUNDING FACTOR USING SIMULATED DATA

Notes: All plots are based on 5,000 simulations of the benchmark DGP defined in Section IIA with a true causal effect of $\beta = 1$, represented by the solid line. Each plot shows estimates of the coefficients δ_k from (13) using either the infeasible estimator or one of the five feasible estimators defined in Section IIA. The dots in the center represent the median estimate across all realizations, while the shaded areas depict a uniform 95 percent confidence band: 95 percent 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 2,930 realizations in which we do not detect a pre-trend.

in Figure 5 closely resemble those in Figure 4. In contrast, our proposed estimator (panel D) performs well except in regions of the parameter space in which identification is weak.

Figure 6 depicts the coverage of the 95 percent confidence interval for each estimator constructed from the usual asymptotic approximation assuming the underlying sampling distribution is approximately normal and correctly centered. Failing

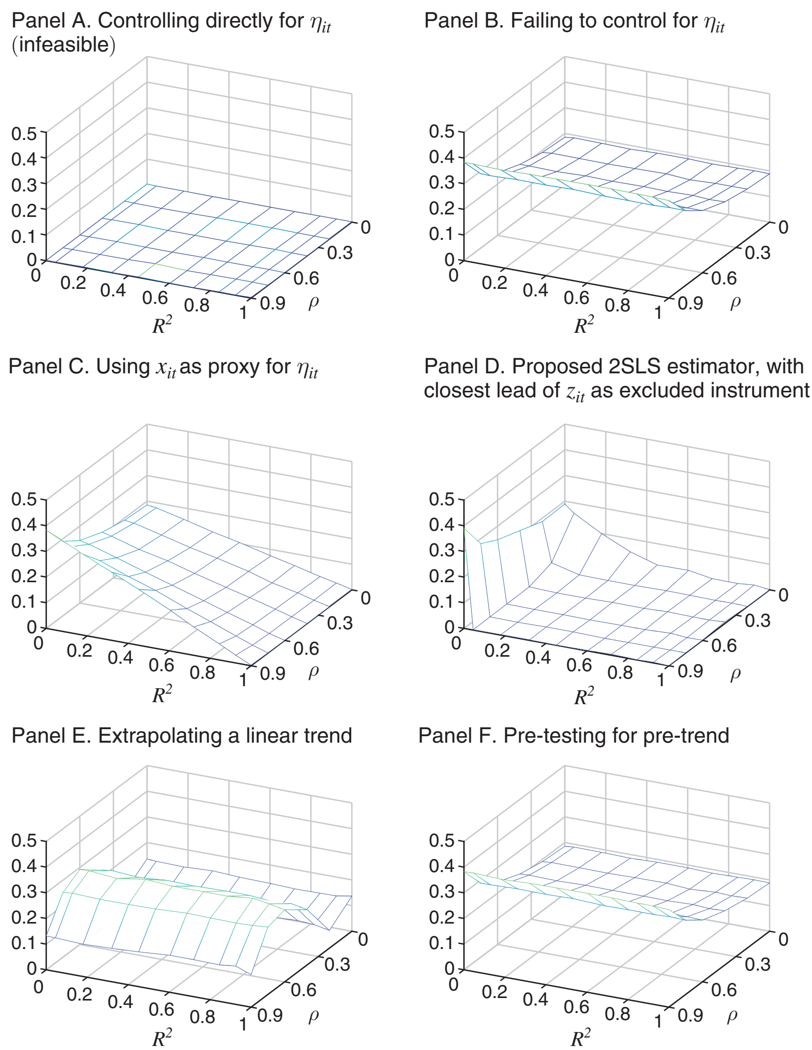

Figure 4. Median Bias for Each Estimator Defined in Section IIA

Notes: Each point represents the median bias across 2,000 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).

to do anything to account for η_{it} results in severe size distortions across the entire parameter space (panel B). 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 (panel C). In contrast, empirical coverage for the 2SLS estimator is close to 95 percent throughout the parameter space, except where identification is weak (panel D).[17] Finally, both linear

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

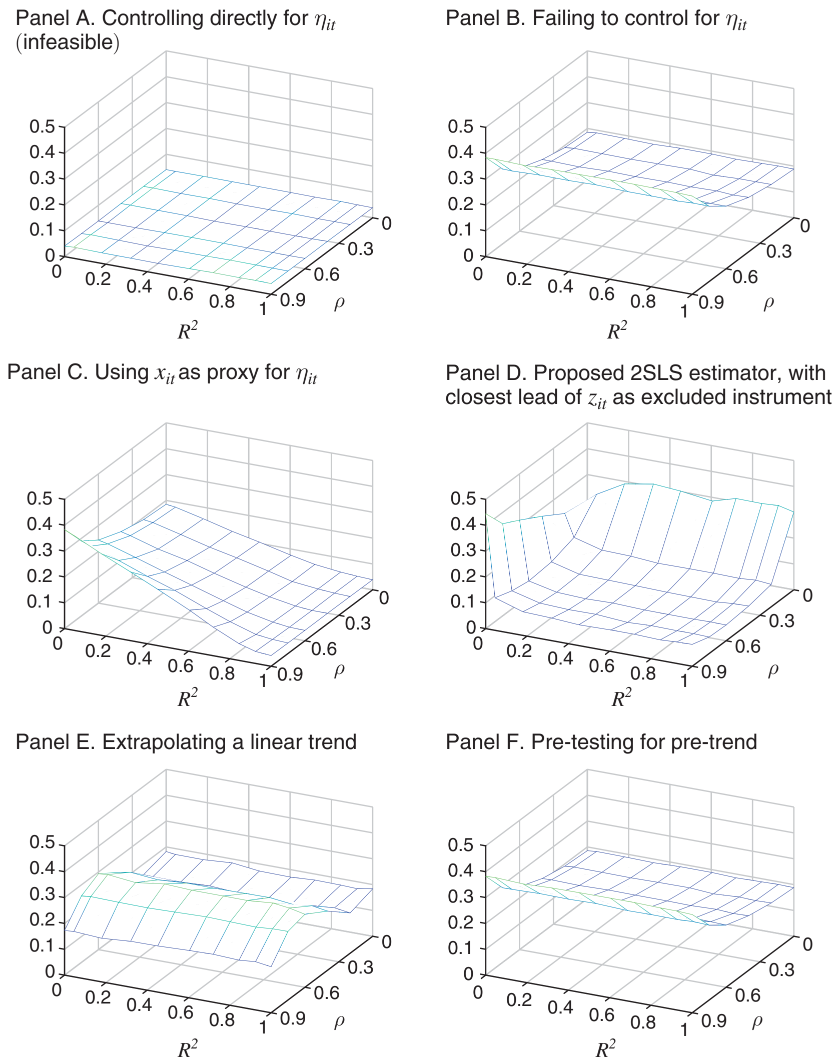

Figure 5. Median Absolute Deviation from the True Parameter Value for Each Estimator Defined in Section IIA

Notes: Each point represents the median absolute deviation across 2,000 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).

extrapolation and pre-testing result in uniformly poor coverage in this simulation design (panels E and F).[18]

[18] The observed coverage of the pre-test estimator (Figure 6, panel F) is a consequence of two offsetting patterns. When we reject the null of no pre-trend, coverage is necessarily equal to 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.

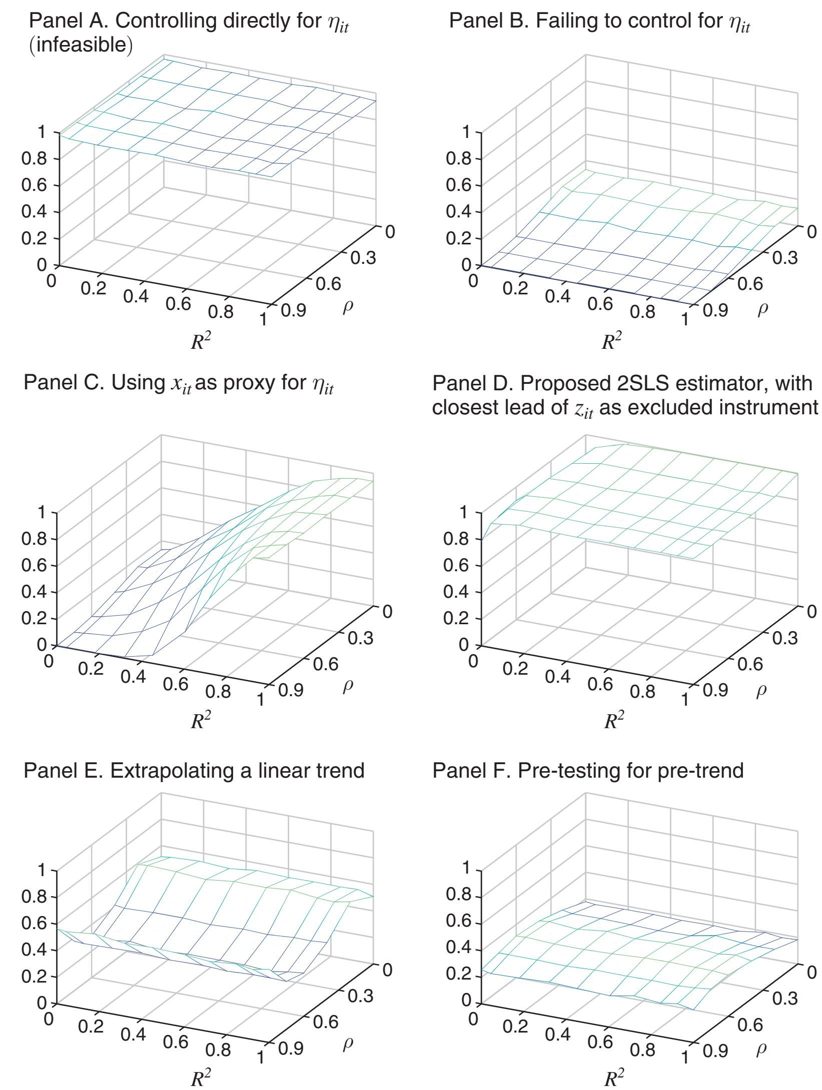

Figure 6. Coverage of the 95 Percent Confidence Interval for Each Estimator Defined in Section IIA

Notes: Each point represents the coverage of the 95 percent confidence interval across 2,000 simulation 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).

III. Applications

In this section, we apply our proposed estimator to empirical settings corresponding to the three examples discussed in Remark 4. Together, these capture many of the scenarios a practitioner might encounter:

- no pre-trend in the outcome variable and a clear pre-trend in the covariate (Section IIIA);

- a clear pre-trend in the outcome variable and a clear pre-trend in the covariate (Section IIIA);
- an unclear pre-trend in the outcome variable and a clear pre-trend in the covariate (Section IIIB);
- an unclear pre-trend in the outcome variable and an unclear pre-trend in the covariate (Section IIIC).

A. *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 include SNAP participation z_{it} and a measure x_{it} of household income, and separate data from a grocery retailer, which include SNAP participation z_{it} and the outcomes y_{it}.[19]

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

$$(16) \qquad x_{it} = \delta_{-5+}\big(1 - z_{i,t+4}\big) + \delta_{5+} z_{i,t-5} + \sum_{k=-4}^{4} \delta_{-k} \Delta z_{i,t+k} + \phi_t + \nu_i + u_{it},$$

where ϕ_t are time effects.

The patterns in Figure 7 are consistent with a model in which household income is a determinant of SNAP eligibility as in Remark 4. 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

$$(17) \qquad y_{it} = \delta_{-5+}\big(1 - z_{i,t+4}\big) + \delta_{5+} z_{i,t-5} + \sum_{k=-4}^{4} \delta_{-k} \Delta z_{i,t-k} + \gamma\eta_{it} + \alpha_i + \omega_t + \varepsilon_{it},$$

[19] 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. The measure x_{it} of household income is the monthly average of in-state earnings and UI benefits received by adults in the household.

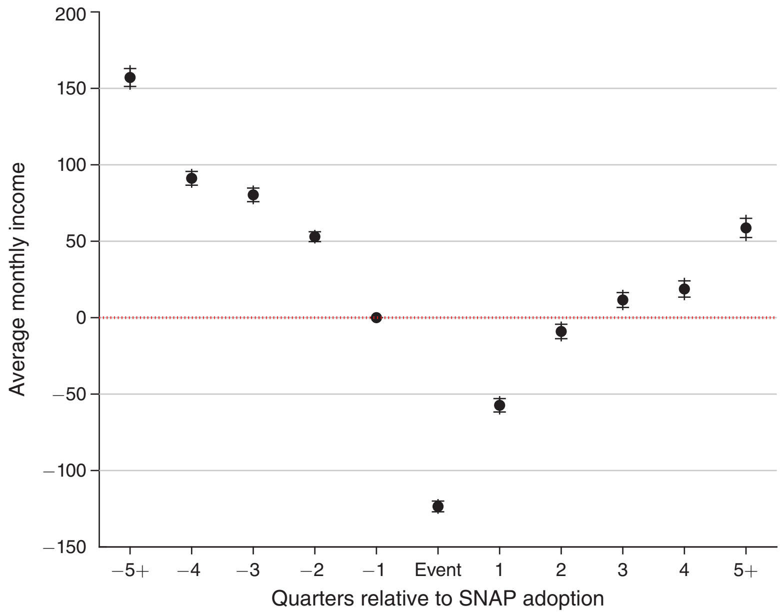

Figure 7. Estimated Changes in Household Income at Quarters around SNAP Adoption

Notes: 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 percent pointwise confidence intervals, while outer confidence sets are the uniform 95 percent 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 (Figure 8, panels A and B) or the store-brand share of food expenditures (Figure 8, panels C and D).

In panels A and C of Figure 8, the term in (17) involving η_{it} is ignored and so no attempt is made to control for confounds. Panel A 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, panel C of Figure 8 shows a 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."

Panels B and D of Figure 8 use our proposed estimator, with the closest lead of z_{it} serving as an excluded instrument for x_{it}.[20] In this setting, the exclusion restriction in Assumption 1(ii) requires that SNAP receipt does not directly affect cash income. This would fail if, for example, program entry or the anticipation of entry leads households to reduce labor supply. Assumption 1(ii) also rules out that SNAP receipt is correlated with the measurement error in the proxy x_{it} for income.

In the case of at-home food expenditures, panel B of Figure 8 shows that taking the income confound into account does not alter the conclusions from the

[20] 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).

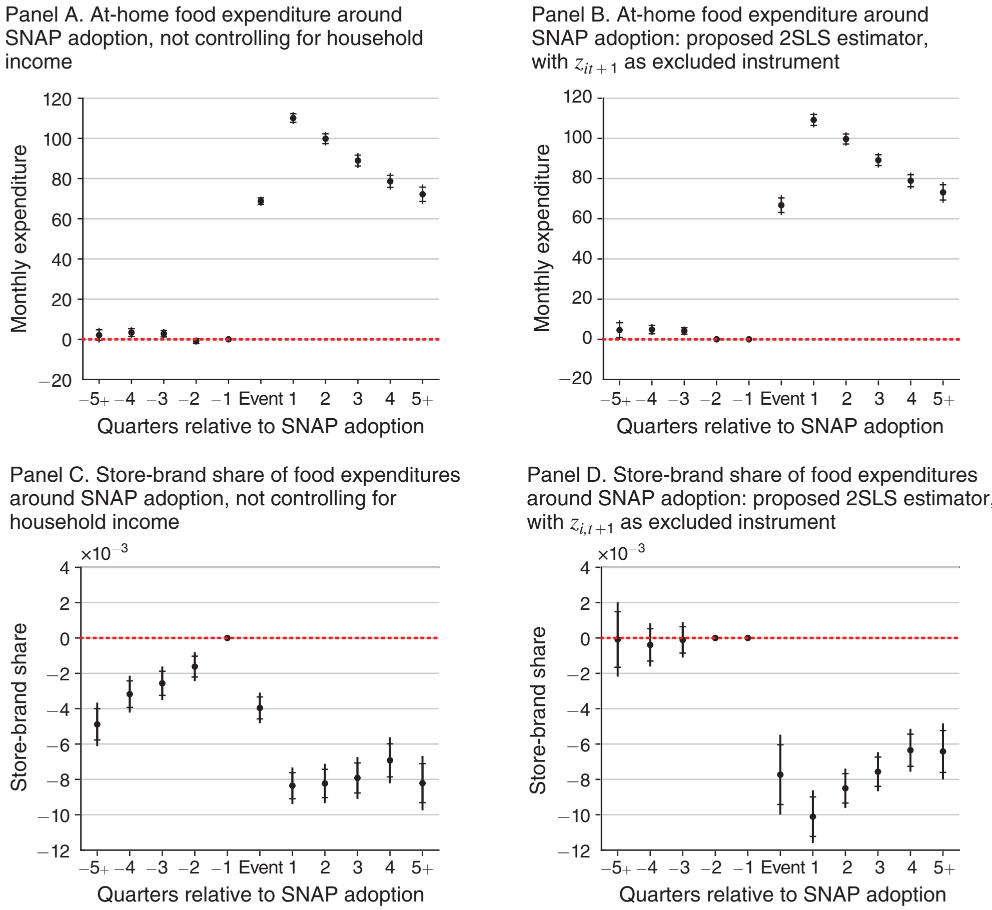

Figure 8. Estimated Changes in Outcomes at Quarters around SNAP Adoption

Notes: Each figure plots estimates of coefficients δ from (17), 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 percent pointwise confidence intervals, while outer confidence sets are the uniform 95 percent sup-t bands (with critical values obtained via simulation). Standard errors are clustered at the household level.

uncorrected plot in panel A. This is because the negligible pre-trend in expenditure implies a small response to changes in income.

By contrast, panel D of Figure 8 differs markedly from panel C because the relatively large pre-trend in store-brand share implies a significant response to changes in income. The 2SLS estimator 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 reflect households' gradual exit from the program following adoption.

Panel A of Figure 9 provides a geometric intuition for our proposed procedure. It combines a rescaled version of Figure 7 with panel C of Figure 8. 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 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 panel D of Figure 8, as an

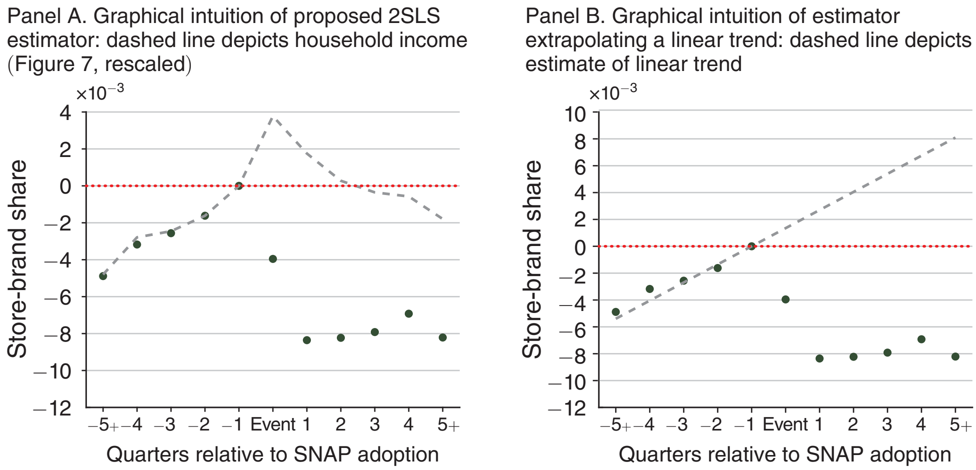

Figure 9. Geometric Illustration of Two Estimators

Notes: Dashed line depicts implicit counterfactual. Round markers depict store-brand share of food expenditures (Figure 8, panel C) around SNAP adoption. The vertical difference between the dashed line and the dots at event time or post-event is the implied estimate of the dynamic causal effect for each estimator.

approximation of the causal effect of SNAP adoption on the store-brand share. For comparison, panel B of Figure 9 shows the dynamics of the confound implied by linear extrapolation from the three periods immediately preceding the event. These dynamics differ markedly from those of the income proxy, and imply a small initial effect of SNAP that grows larger over time.

Figure 8 depicts 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 (17). 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, while SNAP adoption leads to a decrease in the store-brand share of 0.4 percentage points. 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, similar to the first row. On the other hand, the estimated effect on the store-brand share is a decrease of 0.7 percentage points, an increase in magnitude of almost 60 percent compared to the first row. As expected from Figure 7, the first stage is highly significant.

The pre-test estimator, though not depicted in Table 1, is also feasible in this context. For monthly expenditure, we cannot reject $\delta_{-2} = 0$ at the 5 percent level in (17) (see panel A of Figure 8), so the pre-test estimator is equivalent to using no

Table 1—Estimates of the Effect of SNAP Adoption

	Effect of SNAP adoption on		
Estimator	Monthly expenditure	Store-brand share	Coefficient on lead 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)

Notes: 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 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.

control for income. For store-brand share, we reject $\delta_{-2} = 0$ at the 5 percent level in (17) (see panel C of Figure 8), so the pre-test estimator suggests to give up.

B. *The Effect of Newspaper Entry and Exit on Electoral Politics*

Gentzkow, Shapiro, and Sinkinson (2011) study the effect of newspapers on voter turnout, exploiting variation generated by daily newspapers' entries and exits in local markets in the United States. 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 4, it is reasonable to expect the entry of a newspaper to coincide with an improvement in market profitability η_{it}. Because the state of the local economy could also affect voter turnout, market profitability is a potential confound. Gentzkow, Shapiro, and Sinkinson (2011) have proxies for profitability, including a measure x_{it} of the log of the voting-eligible population. Following Gentzkow, Shapiro, and Sinkinson (2011), we depict plots with first-differenced dependent variables.[21]

Figure 10 depicts estimates of the coefficients δ_k from

$$\Delta x_{it} = \sum_{k=-5}^{5} \delta_{-k} \Delta z_{i,t+k} + \Delta \phi_{st} + \Delta u_{it}, \tag{18}$$

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.

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

$$\Delta y_{it} = \sum_{k=-5}^{5} \delta_{-k} \Delta z_{i,t+k} + \Delta \omega_{st} + \gamma \Delta \eta_{it} + \Delta \varepsilon_{it}, \tag{19}$$

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

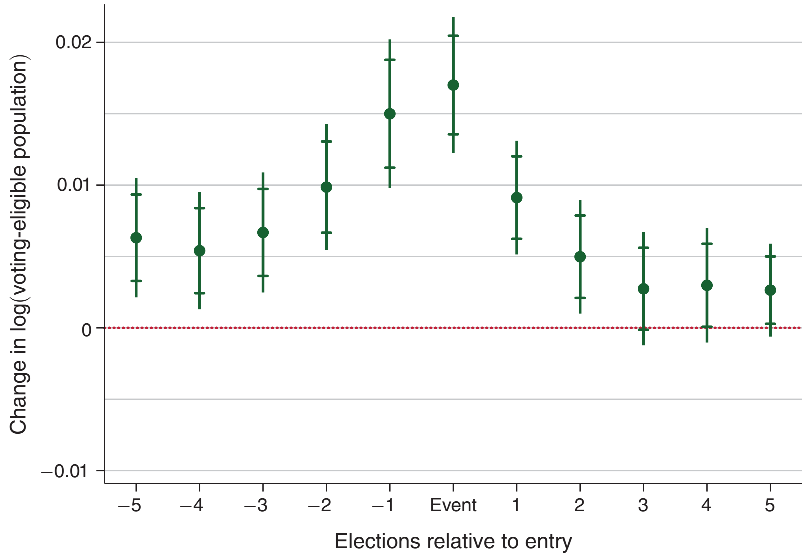

Figure 10. Estimated Changes in Population at Election Years around Newspaper Entries/Exits

Notes: The plot shows estimates of coefficients δ_k from (18). Inner confidence sets as indicated by the dashes correspond to 95 percent pointwise confidence intervals, while outer confidence sets are the uniform 95 percent sup-t bands (with critical values obtained via simulation). Standard errors are clustered at the county level.

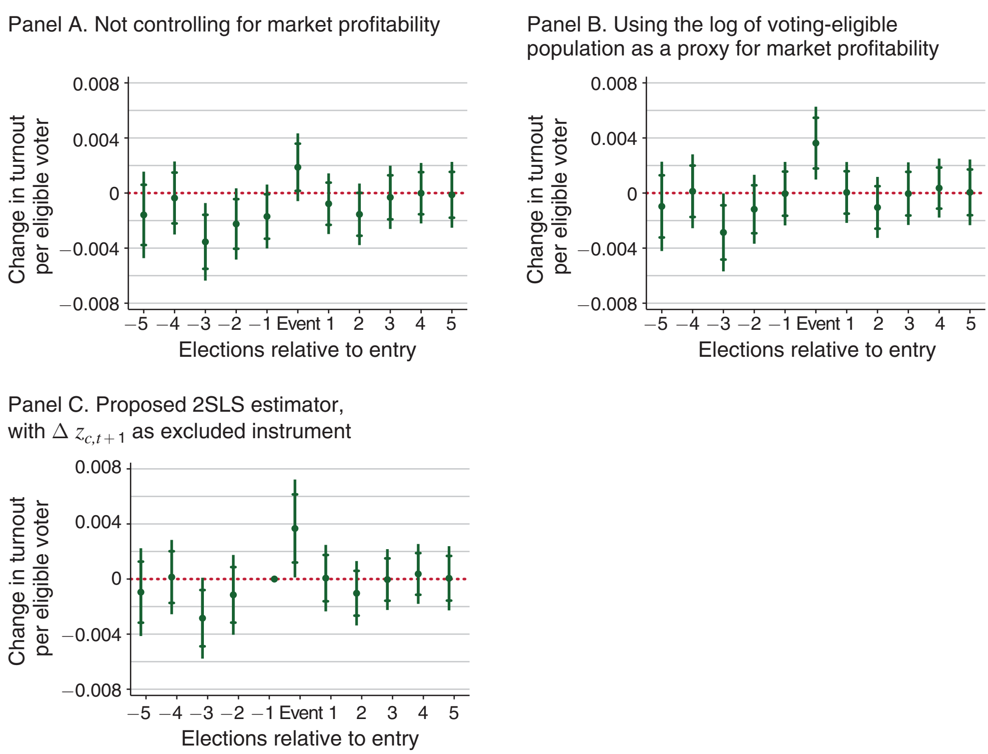

Figure 11. Estimated Effects on Voter Turnout in Presidential Election Years around Newspaper Entries/Exits

Notes: The plot shows estimates of coefficients δ_k from (19). Inner confidence sets as indicated by the dashes correspond to 95 percent pointwise confidence intervals, while outer confidence sets are the uniform 95 percent sup-t bands (with critical values obtained via simulation). Standard errors are clustered at the county level.

Table 2—Estimates of the Effect of Newspapers on Voter Turnout

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	0.0034	0.0128
(one lead)	(0.0013)	(0.0017)

Notes: 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 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.

where ω_{st} is a state-year fixed effect. We omit additional control variables but show in online Appendix Table 2 and online Appendix Figure 11 how their inclusion affects our results. In panel A of Figure 11, the term involving $\Delta \eta_{it}$ is omitted from (19). This specification therefore does not control for newspaper profitability. Panel B controls for market profitability by directly substituting the observed Δx_{it} for $\Delta \eta_{it}$ in (19). Panel C uses our proposed 2SLS estimator, with the closest lead of Δz_{it} serving as an excluded instrument for Δx_{it}. In this setting, the exclusion restriction in Assumption 1(ii) requires that newspaper entry and exit do not directly affect population or its relationship to newspaper profitability. Figure 11 shows that we obtain qualitatively similar results when controlling for x_{it} directly and when using our proposed estimator.

Table 2 presents estimates $\hat{\beta}$ from the static analogue of (19), 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 the 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 the second row.[22]

C. *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, Salas, and Wascher 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 2007). We proxy for labor

[22] 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.

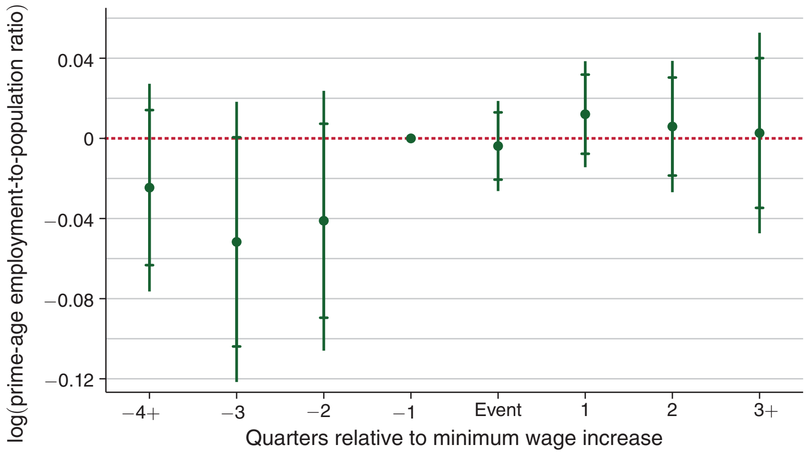

Figure 12. Prime-Age Employment at Quarters around Minimum Wage Increases

Notes: The plot shows estimates of coefficients δ from (20). Inner confidence sets as indicated by the dashes correspond to 95 percent pointwise confidence intervals, while outer confidence sets are the uniform 95 percent sup-t bands (with critical values obtained via simulation). Standard errors are clustered at the state level.

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 plausibly small compared to other sources of variation (Brown 1999), lending credibility to the exclusion restriction in Assumption 1(ii), which requires that the minimum wage does not affect prime-age employment or its relationship to the level of labor demand. 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.[23] We obtain data on z_{it} from David Neumark's Minimum Wage Dataset.[24] 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

$$(20) \quad x_{it} = \delta_{-4+}(1 - z_{i,t+3}) + \delta_{3+} z_{i,t-3} + \sum_{k=-2}^{3} \delta_{-k} \Delta z_{i,t+k} + q_{it}' \psi + \phi_t + \nu_i + u_{it}.$$

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.

[23] The Current Population Survey data are 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.

[24] The minimum wage data are 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.

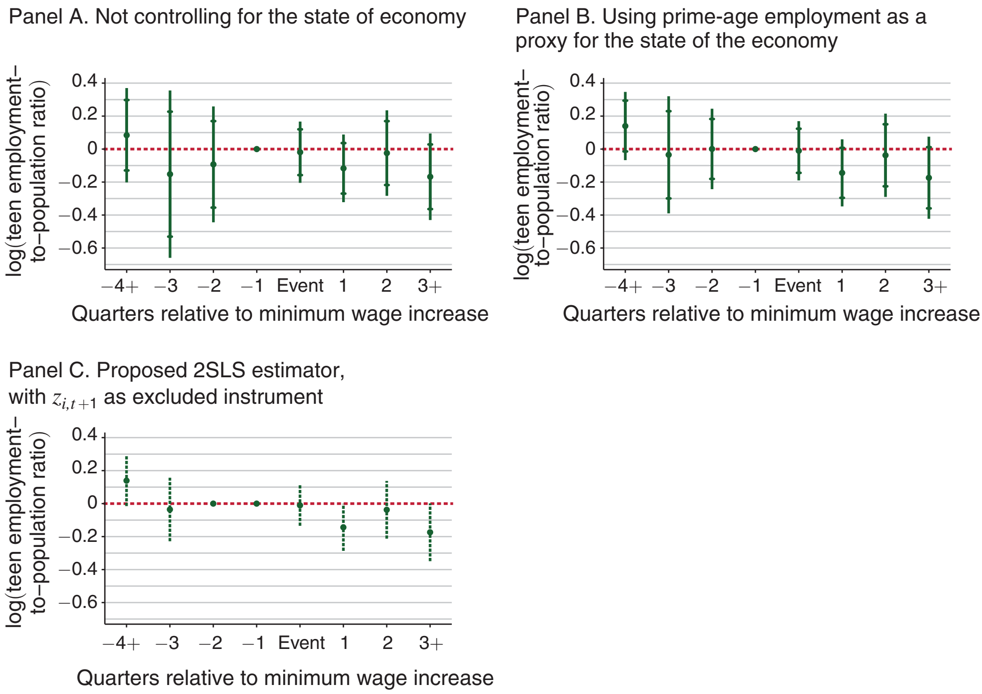

Figure 13. Teen Employment at Quarters around Minimum Wage Increases

Notes: The plot shows estimates of coefficients δ_k from (21). In the top row, inner confidence sets as indicated by the dashes correspond to 95 percent pointwise confidence intervals, while outer confidence sets are the uniform 95 percent sup-t bands (with critical values obtained via simulation). In panel C, dashed confidence intervals correspond to 95 percent pointwise confidence intervals, ignoring weak identification. Standard errors are clustered at the state level.

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:

$$(21) \quad y_{it} = \delta_{-4+}(1 - z_{i,t+3}) + \delta_{3+} z_{i,t-3} + \sum_{k=-2}^{3} \delta_k \Delta z_{i,t+k} + \gamma \eta_{it} + q_{it}'\theta + \omega_t + \alpha_i + \varepsilon_{it}.$$

In panel A of Figure 13, the term involving η_{it} is omitted from (21). This specification therefore does not control for the state of the labor market. Panel B uses prime-age employment x_{it} directly as a control. Panel C 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}. Because the first stage for this model is weak (cf. Table 3), the confidence set for the coefficient on x_{it} 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 panel C of Figure 13 also include the entire real line.

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

$$(22) \quad \Delta y_{it} = \beta \Delta z_{it} + \Delta q_{it}'\theta + \Delta \omega_t + \gamma \Delta \eta_{it} + \Delta \varepsilon_{it}.$$

Table 3—Estimates of the Effect of the Minimum Wage on Teen (16–19) Employment

	Effect of log(minimum wage)	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	0.0314
	(0.0668)	(0.0136)
	$[-\infty, \infty]$	

Notes: Dependent variable: log(employment/population). Each row corresponds to a different estimate $\hat{\beta}$ from the model in first differences given by (22). The first row uses no control for the state of the economy. The second row uses the prime-age employment-to-population ratio 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 the prime-age employment-to-population ratio. 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 the coefficient on the change in the log of the prime-age employment-to-population ratio. All regressions are weighted by teen population. Standard errors are clustered at the state level.

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 the coefficient on x_{it}, which consists of the real line.

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.

IV. Extensions

A. *Anticipatory and Dynamic Treatment Effects*

The model in (6) is static in the sense that the policy has only contemporaneous effects on the outcome. The following generalization allows for both anticipatory effects and dynamic treatment effects:

$$\tilde{y}_{it} = \sum_{m=-G}^{M} \beta_m \tilde{z}_{i,t-m} + \tilde{\eta}_{it}' \gamma + \tilde{\varepsilon}_{it}. \tag{23}$$

We assume that the number of periods G and M over which anticipatory and dynamic effects operate are known.

ASSUMPTION 1′ (Orthogonality Conditions—Dynamic Model): *There exists a positive integer L such that for* $\mathcal{L} = \{-M, \ldots, G-1, G, G+1, \ldots, G+L\}$,

$$(i)\ E[\tilde{z}_{i,t+l}\tilde{\varepsilon}_{it}] = 0 \text{ for all } l \in \mathcal{L}.$$

$$(ii)\ E[\tilde{z}_{i,t+l}\tilde{u}_{it}] = 0 \text{ for all } l \in \mathcal{L}.$$

ASSUMPTION 2′ (Rank Conditions—Dynamic Model): *Let* $w_{it} = (\tilde{z}_{i,t-M}, \ldots, \tilde{z}_{it}, \ldots, \tilde{z}_{i,t+G+L})'$ *and define a matrix H as* $H = E(w_{it}[\tilde{z}_{i,t-M}, \ldots, \tilde{z}_{i,t+G}, \tilde{x}'_{it}])$. *Then,*

$$(i)\ \operatorname{rank}(\Lambda) = R.$$

$$(ii)\ \operatorname{rank}(H) = (R+M+G+1).$$

Following the same reasoning as in Section I, Assumption 2′ (i) allows us to write

$$\tilde{v}_{it} \equiv \tilde{\varepsilon}_{it} - \tilde{u}'_{it}\tilde{\Gamma} = \tilde{y}_{it} - \sum_{m=-G}^{M} \beta_m \tilde{z}_{i,t-m} - \tilde{x}'_{it}\tilde{\Gamma}. \tag{24}$$

Assumption 1′ implies that $E[w_{it}\tilde{v}_{it}] = 0$, and Assumption 2′ (ii) guarantees that these moment conditions are sufficient to identify the causal effects β_m.

As in the static model, estimation may then proceed by GMM, for example through a 2SLS regression of $\tilde{y}_{it}$ on $\{\tilde{z}_{i,t-m}\}_{m=-G}^{M}$ and $\tilde{x}_{it}$, treating the covariates $\tilde{x}_{it}$ as mismeasured regressors and the leads of $\tilde{z}_{it}$ beyond period G as the excluded instruments.

B. *Overidentification and Moment Selection*

We have maintained throughout that $K = R$, and in our simulations and applications we use the R closest leads of $\tilde{z}_{it}$ as instruments. These choices result in an exactly identified model. In practical situations, the model may be overidentified, which raises issues of moment selection and specification testing.

More Leads than Confounds.—In some situations, the researcher may be interested in using more than R leads of $\tilde{z}_{it}$ as instruments. Because the number of potential leads 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 (Rao et al. 2001). In online Appendix Figure 4, we present simulation results from using BIC to select among potential leads.

If more than R leads are used, the model is overidentified and the usual principles for choosing an efficient GMM estimator will apply (Newey and McFadden 1994). Further, the overidentifying restrictions can be tested, and such a test is intuitively similar to looking at whether there are pre-trends in the event plots for our proposed 2SLS estimator. However, while such tests are valid tests of a null of correct specification, choosing a model based on such tests will create the usual pre-test biases (Leeb and Pötscher 2005, Guggenberger 2010).

More Covariates than Confounds.—If the number of covariates is larger than the number of confounds ($K > R$), standard methods for moment selection and combination may be used to find data-driven weights to use all covariates efficiently (Newey and McFadden 1994).

As in the case of multiple leads, tests for overidentification will be available in this case, and with the same caveats.

An alternative approach available in this case is to use the methods developed in the literature on measurement errors (Abbring and Heckman 2007, Heckman and Vytlacil 2007). For instance, suppose that $R = 1, K = 2$, and that the errors $\tilde{u}_{it}$ are uncorrelated across covariates and with $\tilde{\varepsilon}_{it}$. Then, under a rank condition analogous to Assumption 2(ii), the model is identified without the need to treat leads of the policy as excluded instruments. Of course, whether such covariance restrictions are appropriate will depend on the economic setting.[25]

V. Conclusion

We consider a linear panel data model with possible endogeneity. 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. Alternative approaches, such as estimation following a test for pre-trends, perform poorly in our simulations.

REFERENCES

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "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.

Abbring, Jaap H., and James 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 *Handbook of Econometrics*, Vol. 6B, edited by James J. Heckman and Edward E. Leamer, 5145–303. Amsterdam: North-Holland.

Abraham, Sarah, and Liyang Sun. 2018. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." arXiv Working Paper 1804.05785.

Aigner, Dennis J., Cheng Hsiao, Arie Kapteyn, and Tom Wansbeek. 1984. "Latent Variable Models in Econometrics." In *Handbook of Econometrics*, Vol. 2, edited by Zvi Griliches and Michael D. Intriligator, 1321–93. Amsterdam: North-Holland.

Allegretto, Sylvia A., Arindrajit Dube, and Michael Reich. 2011. "Do Minimum Wages Really Reduce Teen Employment? Accounting for Heterogeneity and Selectivity in State Panel Data." *Industrial Relations* 50 (2): 205–40.

Allegretto, Sylvia, Arindrajit Dube, Michael Reich, and Ben Zipperer. 2017. "Credible Research Designs for Minimum Wage Studies: A Response to Neumark, Salas, and Wascher." *ILR Review* 70 (3): 559–92.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113 (1): 151–84.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro. 2017. "Measuring the Sensitivity of Parameter Estimates to Estimation Moments." *Quarterly Journal of Economics* 132 (4): 1553–92.

[25] In the applications we have in mind, the number of plausible proxies is small. If there are instead many covariates that contain independent information about the unobserved confounds, one may alternatively adapt methods from the 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 (2019).

Andrews, Isaiah, and Anna Mikusheva. 2016. "A Geometric Approach to Nonlinear Econometric Models." *Econometrica* 84 (3): 1249–64.

Angrist, Joshua D., and Alan 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): 328–36.

Ashenfelter, Orley. 1978. "Estimating the Effect of Training Programs on Earnings." *Review of Economics and Statistics* 60 (1): 47–57.

Ashenfelter, Orley, and David Card. 1985. "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs." *Review of Economics and Statistics* 67 (4): 648–60.

Athey, Susan, and Guido W. Imbens. 2018. "Design-Based Analysis in Difference-In-Differences Settings with Staggered Adoption." NBER Working Paper 24963.

Bai, Jushan. 2003. "Inferential Theory for Factor Models of Large Dimensions." *Econometrica* 71 (1): 135–71.

Bai, Jushan, and Serena Ng. 2010. "Instrumental Variable Estimation in a Data Rich Environment." *Econometric Theory* 26 (6): 1577–1606.

Ball, Ray, and Philip Brown. 1968. "An Empirical Evaluation of Accounting Income Numbers." *Journal of Accounting Research* 6 (2): 159–78.

Besley, Timothy, and Anne Case. 2000. "Unnatural Experiments? Estimating the Incidence of Endogenous Policies." *Economic Journal* 110 (467): 672–94.

Borusyak, Kirill, and Xavier Jaravel. 2017. "Revisiting Event Study Designs." Unpublished.

Bronnenberg, Bart J., Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M. Shapiro. 2015. "Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium." *Quarterly Journal of Economics* 130 (4): 1669–1726.

Brown, Charles. 1999. "Minimum Wages, Employment, and the Distribution of Income." In *Handbook of Labor Economics*, Vol. 3B, edited by Orley Ashenfelter and David Card, 2101–63. Amsterdam: North-Holland.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli. 2016. "Agricultural Productivity and Structural Transformation: Evidence from Brazil." *American Economic Review* 106 (6): 1320–65.

Card, David, and Alan B. Krueger. 1995. *Myth and Measurement: The New Economics of the Minimum Wage*. Princeton, NJ: Princeton University Press.

Castner, Laura, and James Mabli. 2010. *Low-Income Household Spending Patterns and Measures of Poverty.* Washington, DC: Mathematica Policy Research.

Chernozhukov, Victor, Kaspar Wüthrich, and Yinchu Zhu. 2017. "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls." arXiv Working Paper 1712.09089.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi. 2012. "Plausibly Exogenous." *Review of Economics and Statistics* 94 (1): 260–72.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2018. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." arXiv Working Paper 1803.08807.

Dixit, Avinash K., and Robert S. Pindyck. 1994. *Investment under Uncertainty*. Princeton, NJ: Princeton University Press.

Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew J. Notowidigdo. 2018. "The Economic Consequences of Hospital Admissions." *American Economic Review* 108 (2): 308–52.

Duggan, Mark, Craig Garthwaite, and Aparajita Goyal. 2016. "The Market Impacts of Pharmaceutical Product Patents in Developing Countries: Evidence from India." *American Economic Review* 106 (1): 99–135.

Ferman, Bruno, and Christine Pinto. 2017. "Placebo Tests for Synthetic Controls." Munich Personal RePEc Archive Working Paper 78079.

Freyaldenhoven, Simon, Christian Hansen, and Jesse M. Shapiro. 2019. "Pre-Event Trends in the Panel Event-Study Design: Dataset." *American Economic Review.* https://doi.org/10.1257/aer.20180609.

Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson. 2011. "The Effect of Newspaper Entry and Exit on Electoral Politics." *American Economic Review* 101 (7): 2980–3018.

Goodman-Bacon, Andrew. 2018. "Difference-In-Differences with Variation in Treatment Timing." NBER Working Paper 25018.

Griliches, Zvi, and Jerry A. Hausman. 1986. "Errors in Variables in Panel Data." *Journal of Econometrics* 31 (1): 93–118.

Guggenberger, Patrik. 2010. "The Impact of a Hausman Pretest on the Asymptotic Size of a Hypothesis Test." *Econometric Theory* 26 (2): 369–82.

Hansen, Christian, and Yuan Liao. 2019. "The Factor-Lasso and *K*-Step Bootstrap Approach for Inference in High-Dimensional Economic Applications." *Econometric Theory* 35 (3): 465–509.

Hastings, Justine S., and Jesse M. Shapiro. 2018. "How Are SNAP Benefits Spent? Evidence from a Retail Panel." *American Economic Review* 108 (12): 3493–540.

Hausman, Catherine, and David S. Rapson. 2018. "Regression Discontinuity in Time: Considerations for Empirical Applications." *Annual Review of Resource Economics* 10: 533–52.

Heckman, James J., and Richard Robb. 1985. "Alternative Methods for Evaluating the Impact of Interventions." In *Longitudinal Analysis of Labor Market Data*, edited by James J. Heckman and Burton S. Singer, 156–246. Cambridge, UK: Cambridge University Press.

Heckman, James J., and Edward 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 *Handbook of Econometrics*, Vol. 6B, edited by James J. Heckman and Edward E. Leamer, 4875–5143. Amsterdam: North-Holland.

Hoynes, Hilary W., and Diane Whitmore Schanzenbach. 2009. "Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program." *American Economic Journal: Applied Economics* 1 (4): 109–39.

Inoue, Atsushi, and Gary Solon. 2010. "Two-Sample Instrumental Variables Estimators." *Review of Economics and Statistics* 92 (3): 557–61.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan. 1993. "Earnings Losses of Displaced Workers." *American Economic Review* 83 (4): 685–709.

Kahn-Lang, Ariella, and Kevin Lang. Forthcoming. "The Promise and Pitfalls of Differences-in-Differences: Reflections on '16 and Pregnant' and Other Applications." *Journal of Business and Economic Statistics*.

Leeb, Hannes, and Benedikt M. Pötscher. 2005. "Model Selection and Inference: Facts and Fiction." *Econometric Theory* 21 (1): 21–59.

Li, Kathleen T. 2017. "Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods." Unpublished.

MacKinlay, A. Craig. 1997. "Event Studies in Economics and Finance." *Journal of Economic Literature* 35 (1): 13–39.

Matzkin, Rosa L. 2007. "Nonparametric Identification." In *Handbook of Econometrics*, Vol. 6B, edited by James J. Heckman and Edward E. Leamer, 5307–68. Amsterdam: North-Holland.

McDonald, Robert, and Daniel Siegel. 1986. "The Value of Waiting to Invest." *Quarterly Journal of Economics* 101 (4): 707–27.

Neumark, David, J. M. Ian Salas, and William 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, David, and William L. Wascher. 2007. "Minimum Wages and Employment." *Foundations and Trends in Microeconomics* 3 (1–2): 1–182.

Newey, Whitney K., and Daniel McFadden. 1994. "Large Sample Estimation and Hypothesis Testing." In *Handbook of Econometrics*, Vol. 4, edited by Robert F. Engle and Daniel L. McFadden, 2111–245. Amsterdam: North-Holland.

Olea, José Luis Montiel, and Mikkel Plagborg-Møller. 2019. "Simultaneous Confidence Bands: Theory, Implementation, and an Application to SVARs." *Journal of Applied Econometrics* 34 (1): 1–17.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business and Economic Statistics* 37 (2): 187–204.

Pierce, Justin R., and Peter K. Schott. 2016. "The Surprisingly Swift Decline of US Manufacturing Employment." *American Economic Review* 106 (7): 1632–62.

Rao, C. R., Y. Wu, Sadanori Konishi, and Rahul Mukerjee. 2001. "On Model Selection." *Lecture Notes-Monograph Series* 38: 1–64.

Roth, Jonathan. 2018. "Pre-Test with Caution: Event-Study Estimates after Testing for Parallel Trends." Unpublished.

Stock, James H., and Mark W. Watson. 2002. "Forecasting Using Principal Components from a Large Number of Predictors." *Journal of the American Statistical Association* 97 (460): 1167–79.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business and Economic Statistics* 20 (4): 518–29.

Wansbeek, Tom. 2001. "GMM Estimation in Panel Data Models with Measurement Error." *Journal of Econometrics* 104 (2): 259–68.

Xiao, Zhiguo, Jun Shao, and Mari Palta. 2010. "Instrumental Variable and GMM Estimation for Panel Data with Measurement Error." *Statistica Sinica* 20 (4): 1725–47.