Supplemental Information for: “How to Make Causal Inferences with Time-Series Cross-Sectional Data under Selection on Observables”

Matthew Blackwell*
Adam Glynn†

May 30, 2018

Contents

A Long Run Multiplier

Another quantity of interest in traditional TSCS models is the long-run multiplier (LRM), which is the effect of a one-unit change the equilibrium level of X_t on the equilibrium level of Y_t (Greene, 2012, pp. 422, De Boef and Keele, 2008).

*Department of Government and Institute for Quantitative Social Science, Harvard University, 1737 Cambridge St, MA 02138. web: http://www.mattblackwell.org email: mblackwell@gov.harvard.edu
†Department of Political Science, Emory University, 327 Tarbutton Hall, 1555 Dickey Drive, Atlanta, GA 30322 email: aglynn@emory.edu

We do not fully consider this quantity in this paper because its definition requires additional assumptions that, while relatively easy to discuss within the context of standard parametric TSCS models, are more complicated within the nonparametric approach. Most simply, our fixed time-window approach essentially precludes assessment of this quantity. However, in the interest in clarifying the differences between our approach and the econometric TSCS traditions, we provide a short discussion here. Equilibrium in the potential outcomes framework would be the long-run averages of the potential outcomes under a constant treatment history, if they exist. For instance, the equilibrium level of Y_{it} under treatment would be:

$$\lim_{t \to \infty} E[Y_{it}(1_t)].$$

The LRM, then, is the average causal effect with a comparison between always treated, $(1, 1, \dots)$, and never treated, $(0, 0, \dots)$ as we let t go to infinity:

$$LRM = \lim_{t \to \infty} E[Y_{it}(1_t) - Y_{it}(0_t)]. \qquad (1)$$

Identification of the LRM suffers from a few challenges. First, there is no guarantee that the limit in (1) exists. One of the principal reasons the time series literature focuses on the dynamics of the outcome is to ensure that the empirical processes are stable (or stationary) and that such limits exist. Identification, then, will depend on *some* assumptions about the distribution of the dependent variable. Second, even if the limit exists, the LRM cannot be nonparametrically identified without further restrictions since it depends on estimating the mean potential outcome after an infinite number of time periods.

B Consistent variance estimation

In this section we present a consistent estimator for the variance of the SNMM approach with linear models, a no time-varying interactions assumption, and time-constant impulse response. Let w_{it}^j be a $1 \times k_j$ vector of unit i covariates for estimating the IRF at lag j. In general, this vector will be some function of the treatment and the time-varying covariates $w_{it}^j = f_(z_{i,1}, x_{i,1}, \dots, z_{i,t-j}, x_{i,t-j})$. Some of these covariates, $\tilde{x}_{it}^j$, are those in the impulse response function and will be used to transform the outcome for the next lag. The remaining covariates, $\tilde{z}_{it}^j$, are covariates used to satisfy sequential ignorability. These two sets of covariates partition the vector, $w_{it}^j = (\tilde{x}_{it}^j, \tilde{z}_{it}^j)$.

We collect these vectors into a $T_j \times k_j$ matrix of covariates for unit i at lag j, W_{ij}, where the number of observations per unit, T_j, will depend on the covariates chosen. For instance, certain lagged covariates might be missing in earlier time periods since they would have occurred before baseline measurements. We define the matrices $\tilde{X}_{ij}$ and $\tilde{Z}_{ij}$ similarly. Let $V_i = (y_i, W_{i0}, \ldots, W_{iJ})$ be the observed data for unit i.

Let γ_j be a $k_j \times 1$ vector of coefficients for w_{it}^j and let β_j be the subvector of γ_j associated with the IRF covariates, $\tilde{x}_{it}^j$. The vector $\gamma = (\gamma_0', \gamma_1'''', \ldots, \gamma_J')'$ is the target of inference. Under sequential ignorability and a linear model with time-constant effects for $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})$, the system of equations must satisfy the following moment conditions:

$$E[W_{i0}'''(y_i - W_{i0}\gamma_0)] = 0 \tag{2}$$

$$E[W_{i1}'(y_i - \tilde{X}_{i0}\beta_0 - W_{i1}\gamma_1)] = 0 \tag{3}$$

$$E[W_{i2}'(y_i - \tilde{X}_{i0}\beta_0 - \tilde{X}_{i1}\beta_1 - W_{i2}\gamma_2)] = 0 \tag{4}$$

$$\vdots \qquad = 0$$

$$E[W_{iJ}'(y_i - (\sum_{j=0}^{J-1} \tilde{X}_{ij}\beta_j) - W_{iJ}\gamma_J)] = 0 \tag{5}$$

To simplify notation, we assume that y_i and $\tilde{X}_{ij}$ are properly truncated whenever appropriate so that they are conformable with the other matrices.

Let $g(V_i, \gamma)$ be the $K \times 1$ vector of estimating equations defined above, where $K = \sum_{j=1}^{J} k_j$ is the dimensionality of γ. Thus, we can compactly write the moment conditions as $E[g(V_i, \gamma^*)] = 0$, where γ^* is the true value of the parameters. The usual GMM approach here is to find $\widehat{\gamma}$ such that $(1/n)\sum_i g(V_i, \widehat{\gamma}) = 0$. Here we have as many moment conditions as parameters to estimate so there is an exact solution, which can easily be found with standard software by iterating through the lags, estimating $\widehat{\gamma}_j$ and using it to transform y_i to estimate $\widehat{\gamma}_{j+1}$. The point estimate from that approach will be identical to one from estimating all parameters jointly. The standard errors on $\widehat{\gamma}$, though, will be incorrect because they ignore the fact that estimates for one period depend on estimates from previous periods.

Standard theory on GMM estimators can help us derive asymptotically correct standard errors. Let γ^* be the true value of Define the $K \times K$ matrices $G \equiv E[\nabla_\gamma g(V_i, \gamma^*)]$ and $B \equiv E[g(V_i, \gamma^*)g(V_i, \gamma^*)']$. Then, under regularity conditions, $\widehat{\gamma}$ will be asymptotically Normal with asymptotic variance,

$$\text{Avar}(\widehat{\gamma}) = (G'G)^{-1}G'BG(G'G)^{-1}/N.$$

Let $\tilde{W}_{ij} = [\tilde{X}_{ij} \;\; 0]$ be the matrix of covariates at lag j with zeros replacing any covariates not included in the IRF. Then it is easy to show that with the above moment conditions, G will have the following form:

$$G = E\begin{bmatrix} W'_{i0}W_{i0} & 0 & 0 & 0 & \cdots & 0 \\ W'_{i1}\tilde{W}_{i0} & W'_{i1}W_{i1} & 0 & 0 & \cdots & 0 \\ W'_{i2}\tilde{W}_{i0} & W'_{i2}\tilde{W}_{i1} & W'_{i2}W_{i2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ W'_{iJ}\tilde{W}_{i0} & W'_{iJ}\tilde{W}_{i1} & W'_{iJ}\tilde{W}_{i2} & W'_{iJ}\tilde{W}_{i3} & \cdots & W'_{iJ}W_{iJ} \end{bmatrix}. \tag{6}$$

Let W_j be the stacked $NT_j \times k_j$ matrix of all W_{ij} and define $\tilde{W}_j$ similarly. Then, under the appropriate regularity conditions, a consistent estimator of G will be:

$$\widehat{G} = N^{-1}\begin{bmatrix} W'_0W_0 & 0 & 0 & 0 & \cdots & 0 \\ W'_1\tilde{W}_0 & W'_1W_1 & 0 & 0 & \cdots & 0 \\ W'_2\tilde{W}_0 & W'_2\tilde{W}_1 & W'_2W_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ W'_J\tilde{W}_0 & W'_J\tilde{W}_1 & W'_J\tilde{W}_2 & W'_J\tilde{W}_3 & \cdots & W'_JW_J \end{bmatrix}. \tag{7}$$

To estimate B it is useful to derive it for this specific context. Let $u_{ij}(\gamma) = y_i - \sum_{s=0}^{j} \tilde{X}_{is}\beta_s - W_{ij}\gamma_j$

be errors associated with lag j. Then we can write B in the following form:

$$B = E \begin{bmatrix} W_{i0}' u_{i0}(\gamma) u_{i0}(\gamma)' W_{i0} & W_{i0}' u_{i0}(\gamma) u_{i1}(\gamma)' W_{i1} & \cdots & W_{i0}' u_{i0}(\gamma) u_{iJ}(\gamma)' W_{iJ} \\ W_{i1}' u_{i1}(\gamma) u_{i0}(\gamma)' W_{i0} & W_{i1}' u_{i1}(\gamma) u_{i1}(\gamma)' W_{i1} & \cdots & W_{i1}' u_{i1}(\gamma) u_{iJ}(\gamma)' W_{iJ} \\ \vdots & \vdots & \ddots & \vdots \\ W_{iJ}' u_{iJ}(\gamma) u_{i0}(\gamma)' W_{i0} & W_{iJ}' u_{iJ}(\gamma) u_{i1}(\gamma)' W_{i1} & \cdots & W_{iJ}' u_{iJ}(\gamma) u_{iJ}(\gamma)' W_{iJ} \end{bmatrix}. \tag{8}$$

Letting $\widehat{u}_{ij} = u_{ij}(\widehat{\gamma})$ be the residuals from lag j, we can consistently estimate B with:

$$\widehat{B} = N^{-1} \sum_{i=1}^{N} \begin{bmatrix} W_{i0}' \widehat{u}_{i0} \widehat{u}_{i0}' W_{i0} & W_{i0}' \widehat{u}_{i0} \widehat{u}_{i1}' W_{i1} & \cdots & W_{i0}' \widehat{u}_{i0} \widehat{u}_{iJ}' W_{iJ} \\ W_{i1}' \widehat{u}_{i1} \widehat{u}_{i0}' W_{i0} & W_{i1}' \widehat{u}_{i1} \widehat{u}_{i1}' W_{i1} & \cdots & W_{i1}' \widehat{u}_{i1} \widehat{u}_{iJ}' W_{iJ} \\ \vdots & \vdots & \ddots & \vdots \\ W_{iJ}' \widehat{u}_{iJ} \widehat{u}_{i0}' W_{i0} & W_{iJ}' \widehat{u}_{iJ} \widehat{u}_{i1}' W_{i1} & \cdots & W_{iJ}' \widehat{u}_{iJ} \widehat{u}_{iJ}' W_{iJ} \end{bmatrix}. \tag{9}$$

Given these two consistent estimators, we can apply standard asymptotic theory to derive the following estimator which is consistent for $\text{Avar}(\widehat{\gamma})$:

$$\widehat{\text{Var}}[\widehat{\gamma}] = (\widehat{G}'\widehat{G})^{-1} \widehat{G}' \widehat{B} \widehat{G} (\widehat{G}'\widehat{G})^{-1}. \tag{10}$$

Note that this estimator is robust to heteroskedasticity and serial correlation. The asymptotic properties hold as $N \to \infty$ with both T and J fixed, so this estimator is likely to perform best if N is large relative to T and J. One could impose a system homoskedasticity assumption and estimate the variance under a feasible GLS approach, which might be more efficient if T and N are closer in size. Alternatively, there are several finite-sample corrections that can improve inference with T is large.

C Proof of Sequential g-estimation/ADL near equivalence

Suppose the vectors Y_t, Y_{t-1}, X_t and X_{t-1} have been centered, and define the X matrix $X = [X_{t-1} \quad X_t \quad Y_{t-1}]$ to be the combination of these column vectors. Let $\widehat{\beta}$ be the coefficient vector from the regression of Y_t on X so that $\widehat{\beta} = (X'X)^{-1} X' Y_t$ and has entries, $\widehat{\beta} = (\widehat{\beta}_2, \widehat{\beta}_1, \widehat{\alpha})'$. Note the lack of an intercept due to centering of all variables.

The SNMM approach can be accomplished by blipping down and regressing on X_{t-1}. This can also be re-written as the difference between the coefficient on X_{t-1} from the simple regression of Y_t on X_{t-1} and the coefficient on X_{t-1} from the simple regression of X_t on X_{t-1} times the coefficient on X_{t-1} from the multiple regression.

$$
\begin{aligned}
\widetilde{Y}_t &= Y_t - X_t\widehat{\beta}_1 \\
\widehat{\psi}_1 &= (X'_{t-1}X_{t-1})^{-1}X'_{t-1}\widetilde{Y}_t \\
&= (X'_{t-1}X_{t-1})^{-1}X'_{t-1}(Y_t - X_t\widehat{\beta}_1) \\
&= (X'_{t-1}X_{t-1})^{-1}X'_{t-1}Y_t - (X'_{t-1}X_{t-1})^{-1}X'_{t-1}X_t\widehat{\beta}_1
\end{aligned}
$$

We also know from the normal equations of the full multivariate regression that

$$
\begin{aligned}
(X'_{t-1}X_{t-1})\widehat{\beta}_2 + (X'_{t-1}X_t)\widehat{\beta}_1 + (X'_{t-1}Y_{t-1})\widehat{\alpha} &= (X'_{t-1}Y_t) \\
\widehat{\beta}_2 &= (X'_{t-1}X_{t-1})^{-1}X'_{t-1}Y_t \\
&\quad - (X'_{t-1}X_{t-1})^{-1}(X'_{t-1}X_t)\widehat{\beta}_1 - (X'_{t-1}X_{t-1})^{-1}(X'_{t-1}Y_{t-1})\widehat{\alpha} \\
\widehat{\beta}_2 + (X'_{t-1}X_{t-1})^{-1}(X'_{t-1}Y_{t-1})\widehat{\alpha} &= (X'_{t-1}X_{t-1})^{-1}X'_{t-1}Y_t - (X'_{t-1}X_{t-1})^{-1}(X'_{t-1}X_t)\widehat{\beta}_1 \\
&= \widehat{\psi}_1
\end{aligned}
$$

Note that $\widehat{\psi}_1 = \widehat{\beta}_2 + (X'_{t-1}X_{t-1})^{-1}(X'_{t-1}Y_{t-1})\widehat{\alpha}$ is close to the estimated impulse response from the ADL approach ($\widehat{\beta}_2 + \widehat{\beta}_1\widehat{\alpha}$). The difference is that the ADL approach uses the contemporaneous effect $\widehat{\beta}_1$ (the estimate of the effect of X_t on Y_t) while the sequential g-estimation approach uses $(X'_{t-1}X_{t-1})^{-1}(X'_{t-1}Y_{t-1})$ (the estimate of the effect of X_{t-1} on Y_{t-1}). Therefore, note that the approaches will only be equivalent when the effects of X on Y are constant across time.

D Simulation Details

For the simulations, we generated the baseline covariate as $Z_{i1} \sim \mathcal{N}(0.4, 0.1^2)$ and a time-constant omitted variable as $U_i \sim \mathcal{N}(0, 0.1^2)$. Then, in each period, we generated the data with the following specification:

$$\begin{aligned}
Y_{it}(1,1) &= 0.8 + \mu_{1.1} + \mu_{2.11} + 0.9 \cdot U_i + \mathcal{N}(0, 0.1^2) \\
Y_{it}(1,0) &= 0.8 + \mu_{1.1} + 0.9 \cdot U_i + \mathcal{N}(0, 0.1^2) \\
Y_{it}(0,1) &= 0.8 + \mu_{2.01} + 0.9 \cdot U_i + \mathcal{N}(0, 0.1^2) \\
Y_{it}(0,0) &= 0.8 + 0.9 \cdot U_i + \mathcal{N}(0, 0.1^2) \\
Z_{it}(1) &= \gamma_0 + \gamma_1 + 0.7 \cdot U_i + \mathcal{N}(0, 0.1^2) \\
Z_{it}(0) &= \gamma_0 + 0.1 \cdot U_i + \mathcal{N}(0, 0.1^2) \\
Z_{it} &= X_{i,t-1} Z_{it}(1) + (1 - X_{i,t-1}) Z_{it}(0) \\
\Pr[X_{it} = 1 | Z_{it}, Y_{i,t-1}] &= \text{inv.logit}(\alpha_0 + \alpha_1 \cdot Z_{it} + \alpha_2 \cdot Y_{i,t-1})
\end{aligned}$$

In the simulations in the paper, we set $\mu_{1.1} = \mu_{2.01} = \mu_{2.11} = -0.1$, $\alpha = (-1.3, 1.5, 2.5)$, and $\gamma_0 = 0.5$. In the two settings discussed in the paper, γ_1 was set to -0.5 when the time-varying confounder is affected by treatment and 0 when it is not. Note that for each time-series, i, the DGP does not depend on t and the DGP for period t only relies on data for periods t and $t-1$. Conditional on $X_{i,t-1:t}$, the vector $\{Y_{it}, Z_{it}\}$ is clearly stationary because its only remaining time-varying variation comes from i.i.d. errors. It is easy to show that after marginalizing over these vectors, the process X_{it} forms a time-homogeneous Markov chain, implying the overall DGP is stationary. We checked this via simulation by simulating 1000 time-series of length $T = 1000$ and found that the means and autocovariances of each process were constant over time. Furthermore, all process clearly rejected the unit-root null hypothesis of an augmented Dickey-Fuller test.

In addition to the results in the paper, we also conducted a second simulation study with misspecification in the time-varying covariates. In particular, we assume that Z_{it} was not directly observable, and instead we observe a non-linear transformation of that covariate, $Z^*_{it} = \exp(0.25 * (Z_{it} - 0.5)^3)$. If an analyst knew this deterministic transformation, then she could correctly specify the functional form as $\log(Z_{it})^{1/3}$. Theoretically, this misspecification two possible effects. First, it can increase omitted variable bias for the contemporaneous effect of treatment because we are not conditioning on the correct confounders. Second, it could actually reduce post-treatment bias for the lagged effect since we now conditioning on a noisier version of the post-treatment variable. Thus, we show four scenarios in figure 1 that vary whether or not we have the correct Z_{it} and investigating the RMSE of the contemporaneous and lagged effects. In these results, the Z_{it} is endogenous.

From these results, we can see that even under misspecification, the ADL model has significant bias for the lagged effect. The ADL and MSM models have worse performance in that setting, but they still outperform the ADL model. Likewise, the SNMM and MSM approaches also see increases in RMSE for the contemporaneous effect due to omitted variable bias and model misspecification. These results show that post-treatment bias of the ADL model can easily overwhelm any problems with misspecification of the proposed modeling strategies in this paper.

Finally, we also show the results how the performance of the estimators changes as the number of time periods grows keeping the number of units fixed. Figure 2 shows that the MSM and SNMM approaches perform well when Z_{it} is either endogenous or exogenous and their RMSE decreases as the number of time periods increases. The rank order of performance is similar to that of the setting where we increase N and doesn't change as $T \to \infty$. The story for the ADL approach is very similar to the setting in the main paper—badly biased at all values of T when Z_{it} is endogenous, and performs similarly to the SNMM approach when Z_{it} is exogenous. These results appear to confirm that the proposed estimator can perform well even when the sample size is fixed and the number of time periods is growing.

E Additional illustration: The effect of trade on taxation in OECD countries

In this section, we describe another empirical illustration that shows how the MSM/IPTW approach gives strikingly different results from the conventional TSCS approach when we apply each to the data from Swank and Steinmo (2002). These scholars estimate the effects of domestic economic policies on tax rates in advanced industrialized democracies. Here we focus on one of their explanatory variables, trade openness, and its effect on one of their outcomes, the effective tax rate on labor. In their models,

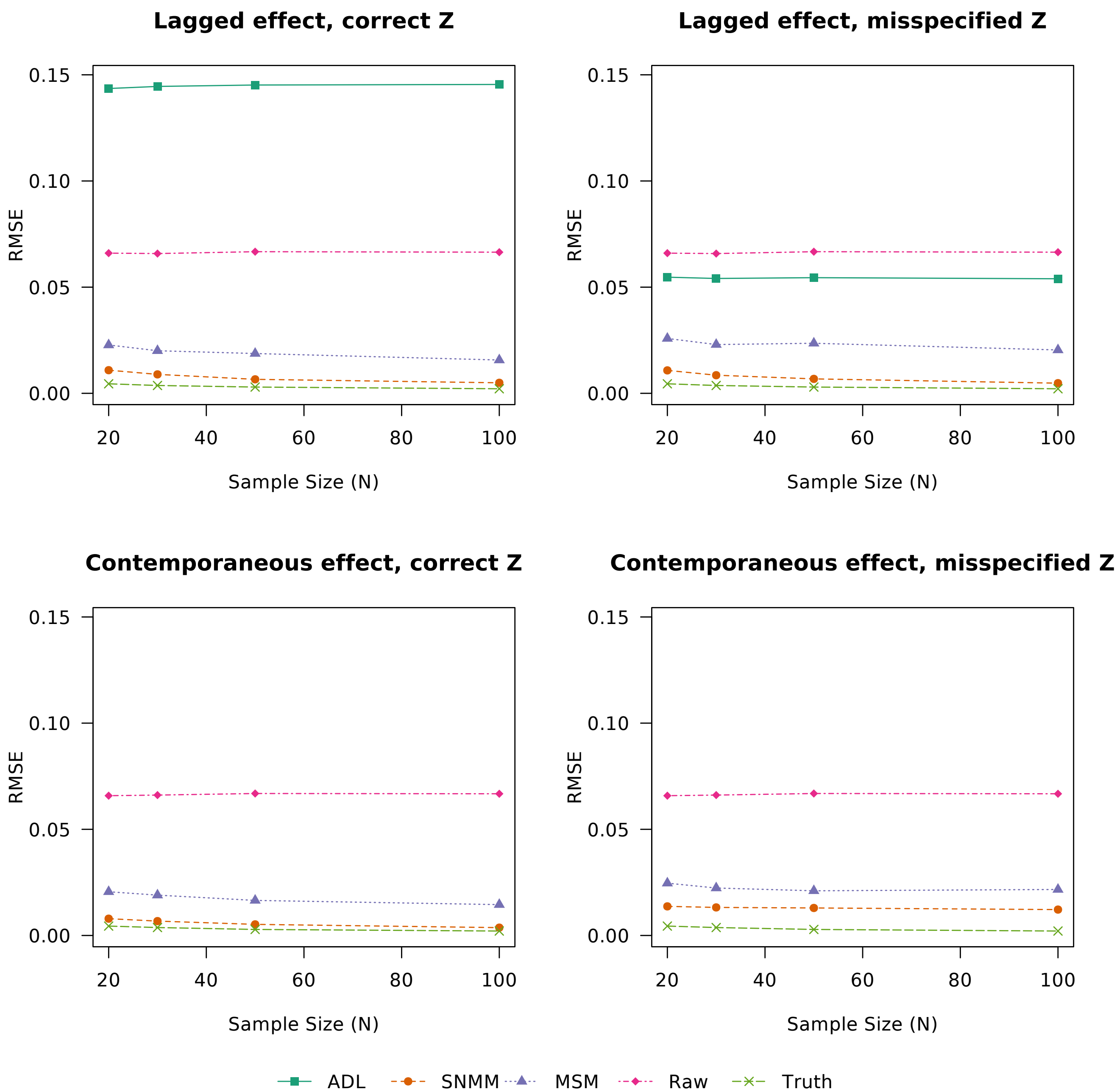

Figure 1: Simulation results when the time-varying confounder is correctly specified in all models (left column) and when it is incorrectly specified in all models (right column). Top row is the RMSE for the lagged effect of treatment and the bottom row is the contemporaneous effect of treatment. In the bottom row, the ADL results are identical to the SNMM. In these simulations, $T = 20$.

Swank and Steinmo find trade openness to have no statistically significant effect on these tax rates, but they only considered the effect of trade openness in the previous year. While Swank and Steinmo

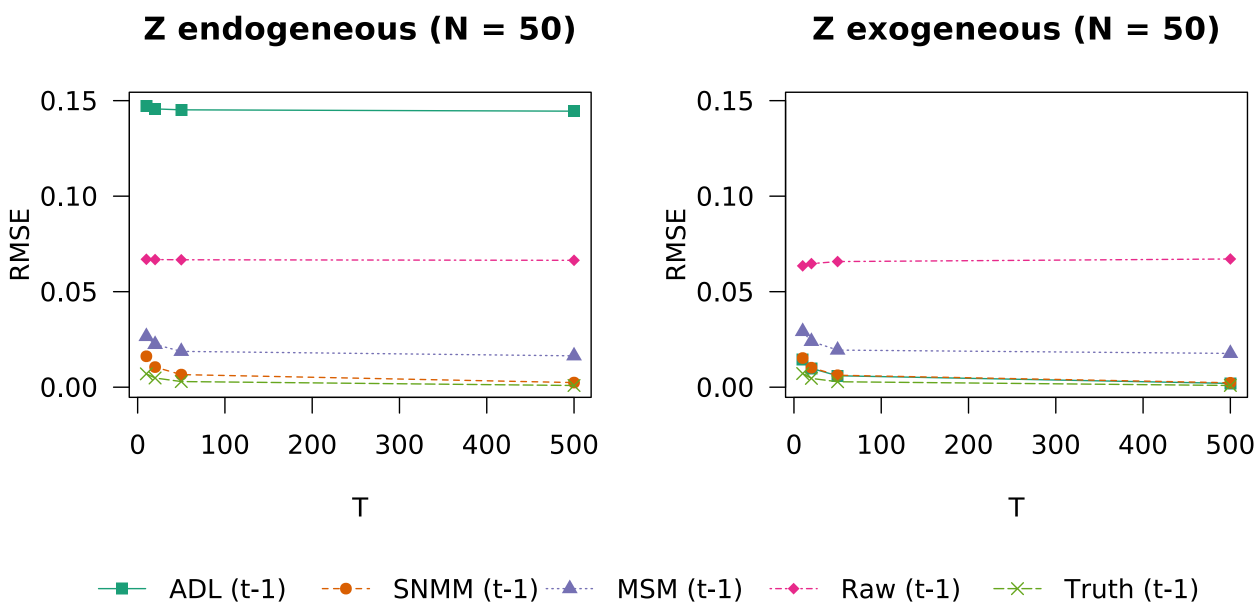

Figure 2: Simulation results fixing the number of units at $N = 50$ and allowing the number of time periods to grow.

discuss the long-run effects of economic policies, they only estimate the contemporaneous effect of this trade policy, leaving aside any effects of history.

Swank and Steinmo adhere to the guidance of previous methodological research on TSCS data (Beck and Katz, 1996). The authors regress the tax rate in a given year on economic and political features of each country from the previous year. In addition to trade openness ($X_{i,t-1}$), these attributes include liberalization of capital controls, unemployment, leftist share of the government, and importantly, a lagged measure of the dependent variable. We refer to the lagged dependent variable as $Y_{i,t-1}$ and the set of attributes (excluding trade openness) as $Z_{i,t-1}$. Thus we can write their main estimating equation as:

$$Y_{it} = \beta_0 + \beta_1 X_{i,t-1} + \beta_2 Y_{i,t-1} + \beta_3 Z_{i,t-1} + \varepsilon_{it}. \tag{11}$$

Keep in mind that β_1 only has a causal interpretation as the CET when sequential ignorability holds and when the effect of $X_{i,t-1}$ is constant across the covariates, $Y_{i,t-1}$ and $Z_{i,t-1}$, and across time.

To uncover any historical effects of trade openness on the labor tax rate, we expand the model of Swank and Steinmo beyond a single lag. We instead take the cumulative years of trade openness as our main independent variable:[1]

$$Y_{it} = \beta_0 + \beta_1 \left(\sum_{k=1}^{t-1} X_{i,k} \right) + \beta_2 Y_{i,t-1} + \beta_3 Z_{i,t-1} + \nu_{it}. \tag{12}$$

Unfortunately, post-treatment bias ruins the causal interpretation of the coefficient on our new measure, β_1. Earlier values of trade openness, such as $X_{i,t-2}$, might affect the lagged tax rate, for instance. To avoid this difficulty, we can take a second approach—omitting the time-varying confounders, $Y_{i,t-1}$ and $Z_{i,t-1}$, from our model. Here we would estimate the effect of trade openness only conditioning on a time trend:

$$Y_{it} = \widetilde{\beta}_0 + \widetilde{\beta}_1 \left(\sum_{k=1}^{t-1} X_{i,k} \right) + \widetilde{\beta}_2 t + \eta_{it}. \tag{13}$$

While this method avoids the issue of post-treatment bias entirely, it admits the possibility of omitted variable bias. If past values of the tax rate affect future trade openness, for instance, then excluding these lags of the dependent variable will produce bias in our estimated effects. Each approach has its drawbacks, but we can learn a great deal by comparing their results to our preferred weighting method.

What do these approaches discover about the effects of trade openness? As Figure 3 shows, both methods—omitting and controlling for time-varying confounders—lead to the same basic conclusion: there is no statistically significant effect of trade openness on tax policy.[2] These results are consistent with the findings of Swank and Steinmo (2002). An alternative to both of these approaches is the above weighting method. To implement IPTW in this case, we omit the time-varying confounders from the tax rate model and instead include those in a propensity score model to create weights as shown the main text. We then use those weights in a weighted GEE model. Instead of controlling for the time-varying confounders in our regression model, these weights adjust for the confounding in the time-varying covariates without inducing post-treatment bias. Figure 3 shows

[1] Here we need trade openness as a binary treatment, so we create a new trade openness variable which is 1 if the county-year had a score at or above the median of the entire sample. The results are substantively unchanged if we use continuous measures, though, as noted above, IPTW in those situations has much poorer properties (Goetgeluk, Vansteelandt and Goetghebeur, 2008).

[2] We estimate both of these models using a generalized estimating equations approach with robust standard errors, allowing for arbitrary correlation of observations within a country (Liang and Zeger, 1986).

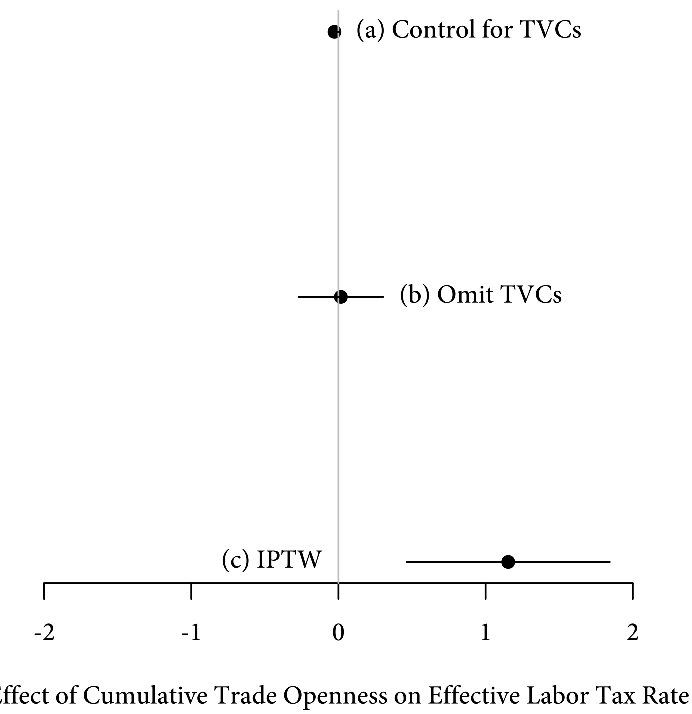

Figure 3: Estimated effect on labor tax rates of cumulative trade openness using three models. They represent the estimated effect and 95% confidence interval (a) when controlling for variables that trade openness affects, (b) when omitting those variables from the model, and (c) using the recommended IPTW approach.

the IPTW estimates are not only significant and positive, but also far larger in magnitude than either of the other approaches.

Bibliography

Beck, Nathaniel and Jonathan N. Katz. 1996. "Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models." *Political Analysis* 6(1):1–36.

De Boef, Suzanna and Luke Keele. 2008. "Taking Time Seriously." *American Journal of Political Science* 52(1):185–200.

Goetgeluk, Sylvie, Sijn Vansteelandt and Els Goetghebeur. 2008. "Estimation of Controlled Direct

Effects." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 70(5):1049–1066.

Greene, William H. 2012. *Econometric analysis*. 7 ed. Boston: Prentice Hall.

Liang, Kung-Yee and Scott L Zeger. 1986. "Longitudinal data analysis using generalized linear models." *Biometrika* 73(1):13–22.

Swank, Duane and Sven Steinmo. 2002. "The New Political Economy of Taxation in Advanced Capitalist Democracies." *American Journal of Political Science* 46(3):642–655.