Estimation and Inference of Treatment Effects using a New Panel Data Approach with Application to the Impact of US SYG Law on State Level Murder Rate*

Huayan Geng[a], Qiankun Zhou[b†]

[a]Department of Economics, State Univeristy of New York at Binghamton

[b]Department of Economics, Louisiana State University

This Version December 13, 2018

Abstract

This paper proposes a new panel data approach to measure the impact of social policy. We consider a classical panel model with interactive fixed effects (IFE), which allows the cross-sectional dependence through the presence of some (unobserved) common factors. The new approach combines the idea of Pesaran (2006) to estimate panel model with IFE and Hsiao et al. (2012) to construct counterfactuals. For the new approach, instead of estimating the unobserved factors, we propose to use observed data. Compared to the existing methods such as Synthetic Control Method (SCM) (Abadie et al. (2010)) and the Generalized SCM (GSCM) (Xu (2017)), our new approach has the advantages of: (1) there is no need to impose constraints on both observables and unobservables; (2) the number of parameters to be estimated in the model is greatly reduced. Moreover, we establish the asymptotic properties for the average treatment effect (ATE) over post-treatment periods, which can be used to obtain statistical inference for the significance of ATE or to construct confidence band for the treatment effects in the post-treatment periods. Monte Carlo simulations show that our approach works remarkably well and has very desirable finite sample performance in terms of estimation bias, mean square of errors, and empirical rejection frequency. We apply our method to study the impact of US Stand Your Ground (SYG) law on the state-level murder rate, and we find, in general, the SYG has increased the murder rate for the states adapting SYG law.

Keywords: Panel data model, Interactive fixed effects, Treatment effects, Program evaluation, Stand Your Ground Law

JEL Classification: C12, C13, C14, C21, C23

---

*We would like to thank David Slitcher for helpful comments. All errors are the authors' sole responsibilities.

†Address correspondence to: Qiankun Zhou, Department of Economics, Louisiana State University, Baton Rouge, LA, USA. Email: qzhou@lsu.edu.

1 Introduction

When social scientists evaluate the economic impact of a policy intervention by using nonexperimental panel data, the major challenge in existing literature is to construct counterfactuals of the outcome in the absence of treatment, y_{it}^0, because the outcome under the treatment, y_{it}^1, and y_{it}^0 cannot be simultaneously observed in reality (e.g., Heckman and Vytlacil (2007a, 2007b)), but the treatment effect is measured as $y_{it}^1 - y_{it}^0$. Several approaches have been proposed in the literature to construct the counterfactuals and to estimate the treatment effects for the policy impact. To name a few, the synthetic control method (SCM) by Abadie et al. (2010) and the generalized synthetic control method (GSCM) by Xu (2017), among others.[1] Intuitively, the idea of SCM is to find control units that are similar to the treatment unit, then take a weighted average of such control units to generate counterfactuals. These weights are calculated in such a way that both the weighted outcomes and weighted control variables are close to the outcome and control variables for the treated unit in the pretreated period, respectively. Within the regression framework, the SCM constitutes a constrained regression. As argued by Wan et al. (2018), when the constraints are valid, SCM is an efficient method. When the constraints are not valid, SCM could lead to biased prediction of counterfactuals. On the other hand, the GSCM relies on the parametric specification of the model, and considers the estimation of all unknown parameters in the model. In general, one would expect the parametric approach to be the most efficient when the model is specified correctly. However, if the dimension of unobserved factors is unknown, then, first, there is the issue of identifying the dimension of the unobserved factors from a finite sample. Second, even if the dimension of the unobserved factors is known, the parametric model could involve estimating too many unknown parameters relative to the sample size. Furthermore, if the model is misspecified, then the resulting inference could be misleading.

To overcome the aforementioned difficulties, in this paper, we propose a simple-to-implement panel data method to evaluate the impacts of social policy. This new panel data approach, which is called PDX, is based on the classical linear fixed effects models with interactive fixed effects (IFE). The PDX approach does not rely on the knowledge of the dimension of the unobserved factors. Nor does it need to estimate the factor loading matrix. The number of unknown parameters involved could considerably less than the number involved in the parametric GSCM approach. Moreover, the PDX approach doesn't need to impose certain constraints on the outcomes and control variables between the treated units and control units.

Essentially, the PDX approach combines the idea of Pesaran (2006) to estimate panel model with IFE and Hsiao et al. (2012) to construct counterfactuals. On the one hand, the PDX approach estimates the common slope coefficient in the model by Pesaran (2006)'s common correlated effects (CCE) method for large cross-sectional units in a fixed time period, it can be shown that the CCE estimation is consistent (Zhou and Zhang (2016)). On the other hand, once the common slope

[1] When the model of interest is a pure factor model (i.e., no exogenous regressors in the model), Hsiao et al. (2012) propose to contruct the counterfactuals by the so called PDA approach. See Gardeazabal and Vega-Bayo (2017) for comparison between SCM and PDA approaches.

coefficient of the control variables are consistently estimated, then the resulting panel model will approximately be a pure factor model, and thus one can use the approach of Hsiao et al. (2012) to construct counterfactuals for the treated units. Intuitively, the PDX approach can be viewed as a semi-parametric approach since we propose to use observed data instead of trying to estimate the unobserved factors in the model.

Our approach contributes to the literature in the following ways. First, compared with the pure factor model considered by Hsiao et al (2012), our approach allows the impact of exogenous control covariates. Second, compared with the SCM, we don't put any constraints on the outcomes and control variables between the treated units and control units. Third, compared with the parametric approach such as GSCM, our method doesn't rely on the knowledge of the dimension of unobserved factors, and has greatly reduced the number of parameters to be estimated in the model. Finally, as the main contribution, we establish the asymptotics for the average treatment effects (ATE) over post-treatment periods. The asymptotic property allow researchers to obtain statistical inference about the significance of the ATE and to construct the confidence band for the treatment effects of the post-treatment periods.

In order to examine the finite sample properties of the PDX approach, we conduct a variety set of Monte Carlo simulations. Through the simulation studies, we can observe that the PDX approach outperforms both the SCM and GSCM approaches under all different data generating processes and different sample configurations of cross-sectional dimensions and pre-treatment time dimensions. In general, the counterfactuals of PDX have less bias and MSE than those obtained from SCM and GSCM approaches. On the other hand, we can also observe that the statistical inference obtained from the PDX is also valid, and the empirical rejection frequency is quite close to the nominal value for significance test. Empirical application of the PDX approach to measure the impact of the US Stand Your Ground (SYG) law on state level murder rate also highlights the necessity of using our new approach. The counterfactuals from our PDX approach in general is quite close to the pretreated actual murder rate, while the counterfactuals from both SCM and GSCM deviate from the actuals quite often. Based on the results of PDX approach, we can observe that the SYG law has certain positive effect on the state-level murder rate for states adapting SYG law, but the average impact is usually not very significant.

The rest of this paper is organized as follows. Section 2 sets up the model and proposes estimation steps of PDX. Asymptotics of the ATE constructed from PDX is provided in Section 3. Section 4 reports simulation results by comparing the relative performance of SCM, GSCM and PDX under a variety set of data generating processes. An application of the impact of the US SYG law on state level murder rate and conclusion are provided in Section 5 and Section 6, respectively. All mathematical proofs are relegated to the Appendix.

2 Model and Estimation

2.1 The Model

Suppose there are observations $(y_{it}, \mathbf{x}_{it})$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where y_{it} is the outcome of interest of unit i at time t, $\mathbf{x}_{it}$ is a $k \times 1$ vector of covariates and T is time periods for which all units are observed. Let T_0 be the number of pretreatment periods, which is first exposed to the treatment at time $T_0 + 1$. Let the dummy variable d_{it} indicate the ith unit's treatment status at time t. The treatment indicator $d_{it} = 1$ if unit i has been exposed to the treatment at time t and $d_{it} = 0$ otherwise, i.e. $d_{it} = 1$ for i is treated unit and $t > T_0$ and $d_{it} = 0$ otherwise. The observed data takes the form,

$$y_{it} = d_{it} y_{it}^1 + (1 - d_{it})\, y_{it}^0. \tag{1}$$

For simplification, we assume $d_{1t} = 0$ for $t = 1, \ldots, T_0$ and $d_{1t} = 1$ for $t = T_0 + 1, \ldots, T$, while $d_{it} = 0$ for $i = 2, \ldots, N$, and $t = 1, \ldots, T$, i.e., we assume only the first unit is intervened by the treatment. The method to be discussed can be generalized to more than one treated units.

We assume y_{it}^0 is a function of k observables strictly exogenous factors, $\mathbf{x}_{it}$,

$$y_{it}^0 = \mathbf{x}_{it}' \boldsymbol{\beta} + v_{it}, \quad 1 \leq t \leq T, \tag{2}$$

where $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown parameters. The error term v_{it} is decomposed as the sum of the impacts of r unobserved common factors across individuals, $\mathbf{f}_t = [f_{1t}, \ldots f_{rt}]'$, and the idiosyncratic error term, u_{it} with zero mean,

$$v_{it} = \boldsymbol{\gamma}_i' \mathbf{f}_t + u_{it}, \tag{3}$$

where $\boldsymbol{\gamma}_i = (\gamma_{i1}, \ldots, \gamma_{ir})'$ is an $r \times 1$ vector of unknown factor loadings indicating the impact $\mathbf{f}_t$ on the ith unit.[2]

Combining (1)-(3) yields

$$y_{it} = \delta_{it} d_{it} + \mathbf{x}_{it}' \boldsymbol{\beta} + \boldsymbol{\gamma}_i' \mathbf{f}_t + u_{it}, \tag{4}$$

where δ_{it} is the treated effect of unit i at time t. The format of factor component covers a wide range of unobserved heterogeneities. For example, if $\mathbf{f}_{1t} = 1$, $\mathbf{f}_{2t} = \xi_t$, $\gamma_{1i} = \alpha_i$ and $\gamma_{2i} = 1$, then the factor component of the model, $\boldsymbol{\gamma}_i' \mathbf{f}_t = \alpha_i + \xi_t$, stands for two-way fixed effect. Intuitively, putting the unobserved individual-specific factors and the common time-specific factor loadings in the multiplicative form also has the advantage over the traditional additive form (e.g., Hsiao (2014)) that allows "globe shocks at time t" to be different for different individuals due to the differences in natural endowment or distinct social or technological background. Moreover, the traditional additive form is nested within the multiplicative form (Bai (2009), Hsiao (2018)).

[2]For macroeconomics, if y_{it} is output or output growth for country i, then common factor $\mathbf{f}_t$ represents common shocks such as technological shocks and financial crisis and factor loadings stands for heterogeneous impact of common shocks, while, for microeconomics, if y_{it} represents the wage rate for individual i with age t, then $\boldsymbol{\gamma}_i$ represents a vector of unobservable characteristics or unmeasured skills, such as ability or motivations, and common factor is a vector of prices for factor loadings (Bai (2009)).

We shall make the following assumptions for the above model.

Assumption A1. $E\left(u_{it}|\mathbf{x}_{it}, \mathbf{f}_t, \boldsymbol{\gamma}_i, d_{1t}\right) = 0, \forall i, t,\ E\left(u_{it}u_{jt}\right) = 0$ and with finite fourth moment.

Assumption A2. $\mathbf{y}_i \perp d_{1t}$ for $i = 2, \ldots, N$, where $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$.

Assumption A3. Let $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})$, then $\mathbf{X}_i \perp d_{1t}$ for $i = 1, \ldots, N$.

Several remarks can be made for the above assumptions. Assumption A1 assumes both the observed explanatory variables, $\mathbf{x}_{it}$, and the unobserved common factors and factor loadings $(\mathbf{f}_t, \boldsymbol{\gamma}_i)$ are strictly exogenous with respect to the idiosyncratic errors u_{it}, and it can be relaxed to allow u_{it} exhibits weak cross-sectional dependence as in Hsiao and Zhou (2018). Assumption A2 assumes only the treated units are affected by the policy shock, while the control group units should not be influenced by the treatment. Assumption A3 restricts $\mathbf{X}_i$ being independent of the treatment. These assumptions are quite standard in the treatment effects literature using panel data, such as Hsiao et al (2012), Xu (2017), and Li and Bell (2017), among others.

2.2 The New Panel Data Approach

It worth pointing out that the GSCM method of Xu (2017) is a parametric approach for constructing counterfactuals, and the first step, which is known as the Principle Component Analysis (PCA) of Bai (2009), requires both N and T_0 to be large to obtain consistent and reliable estimation of $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}_i$ and $\hat{\mathbf{f}}_t$. Furthermore, there is the issue of identifying the dimension of the unobserved factors from a finite sample. Even if the dimension of the unobserved factors is known, the parametric model could involve estimating too many unknown parameters relative to the sample size. In many applications, especially for microeconomics data, the pretreatment period T_0 is usually finite, but the cross-sectional units N could be large. Consequently, we consider generating counterfactuals through the following the following approach using panel data with exogenous regressors (we name it PDX).

For model (2)-(3), instead of using the PCA approach of Bai (2009) to estimate the common slope coefficient $\boldsymbol{\beta}$, we can consider Pesaran's (2006) common correlated effects (CCE) estimation. The CCE estimator for $\boldsymbol{\beta}$ using the pretreated data has the form of

$$\hat{\boldsymbol{\beta}}_{CCE} = \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M}_{\bar{\mathbf{Z}}} \mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M}_{\bar{\mathbf{Z}}} \mathbf{y}_i, \tag{5}$$

where $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT_0})'$, $\mathbf{M}_{\bar{\mathbf{Z}}} = \mathbf{I}_{T_0} - \bar{\mathbf{Z}}\left(\bar{\mathbf{Z}}'\bar{\mathbf{Z}}\right)^{-1}\bar{\mathbf{Z}}'$ with $\bar{\mathbf{Z}} = (\bar{\mathbf{z}}_1, \ldots, \bar{\mathbf{z}}_{T_0})'$ and $\bar{\mathbf{z}}_t = \frac{1}{N}\sum_{j=1}^{N} \mathbf{z}_{jt} = \frac{1}{N}\sum_{j=1}^{N} \left(y_{it}, \mathbf{x}_{it}'\right)'$ .It is shown by Zhou and Zhang (2016) that the CCE estimator (5) is consistent as long as $N \to \infty$.

Given the consistent estimator of $\boldsymbol{\beta}$, we note that

$$\begin{aligned}
\mathbf{e}_t &= \mathbf{y}_t - \mathbf{X}_t' \hat{\boldsymbol{\beta}}_{CCE} = \Lambda \mathbf{f}_t + \mathbf{u}_t + \mathbf{X}_t \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{CCE}\right) \\
&= \Lambda \mathbf{f}_t + \mathbf{u}_t + O_p\left(N^{-1/2}\right),\ t = 1, \ldots, T_0,
\end{aligned} \tag{6}$$

is proximately a pure factor model, where $\Lambda = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_N)'$. Thus the Hsiao et al. (2012)'s approach can be applied to eliminate unobserved factors $\mathbf{f}_t$ and to construct counterfactuals. To this end, following Hsiao et al. (2012), we let $\mathbf{a}$ be a vector lying in the null space of Λ, $N(\Lambda)$, such that $\mathbf{a}'\Lambda = 0$. For ease of notation, we normalize the first element of $\mathbf{a}$ to be 1 and denote $\mathbf{a}' = (1, -\tilde{\mathbf{a}}')'$.

Multiplying both sides of (6) by $\mathbf{a}'$ yields

$$e_{1t} = \tilde{\mathbf{a}}'\tilde{\mathbf{e}}_t + u_{1t} - \tilde{\mathbf{a}}'\tilde{\mathbf{u}}_t + O_p\left(N^{-1/2}\right),\ t = 1, \ldots, T_0, \tag{7}$$

where $\tilde{\mathbf{e}}_t = (e_{2t}, \ldots, e_{Nt})'$ and $\tilde{\mathbf{u}}_t = (u_{2t}, \ldots, u_{N,t})'$.

For model (7), since $\mathbf{e}_t = \mathbf{y}_t - \mathbf{X}_t'\hat{\boldsymbol{\beta}}_{CCE}$ and $E(u_{1t} - \tilde{\mathbf{a}}'\tilde{\mathbf{u}}_t) = 0$, then we can run OLS to estimate $\tilde{\mathbf{a}}$. The OLS estimator of $\tilde{\mathbf{a}}$ is given by[3]

$$\widehat{\tilde{\mathbf{a}}} = \left(\sum_{t=1}^{T_0} \tilde{\mathbf{e}}_t\tilde{\mathbf{e}}_t'\right)^{-1} \sum_{t=1}^{T_0} \tilde{\mathbf{e}}_t e_{1t}. \tag{8}$$

Given the estimator of $\tilde{\mathbf{a}}$, we can construct the estimated counterfactual of y_{1t}^0 as

$$\begin{aligned} \hat{y}_{1t}^0 &= \mathbf{x}_{1t}'\hat{\boldsymbol{\beta}}_{CCE} + \widehat{\tilde{\mathbf{a}}}'\tilde{\mathbf{e}}_t \\ &= \mathbf{x}_{1t}'\hat{\boldsymbol{\beta}}_{CCE} + \widehat{\tilde{\mathbf{a}}}'\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\hat{\boldsymbol{\beta}}_{CCE}\right),\ t = T_0 + 1, \ldots, T. \end{aligned} \tag{9}$$

where $\tilde{\mathbf{y}}_t = (y_{2t}, \ldots, y_{Nt})'$ and $\tilde{\mathbf{X}}_t = (\mathbf{x}_{2t}, \ldots, \mathbf{x}_{N,t})'$ denote the observations from the control units.

Formally, the PDX procedure to generate counterfactuals can be reached in the following steps.

Step 1: Use all pretreated data and Pesaran's (2006) CCE method to estimate $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}_{CCE}$.

Step 2: Conditional on $\hat{\boldsymbol{\beta}}_{CCE}$, obtain $\tilde{\mathbf{a}}$ by minimizing[4]

$$\min_{\tilde{\mathbf{a}}} \sum_{t=1}^{T_0} \left[\left(y_{1t} - \mathbf{x}_{1t}'\hat{\boldsymbol{\beta}}_{CCE}\right) - \tilde{\mathbf{a}}'\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\hat{\boldsymbol{\beta}}_{CCE}\right)\right]^2, \tag{10}$$

using observations from the pretreated periods. This step can be estimated by ordinary least square (OLS), model selection as in Hsiao et al (2012) or LASSO by Li and Bell (2017) if N is moderate

[3] In principle, any choice of $\mathbf{a}$ that satisfies the condition that $\mathbf{a}'\Lambda = 0$ will be fine for constructing couterfactuals in (9) (see, e.g., the discussion in Hsiao et al. (2012)). However, the prediction error variance depends on $Var(u_{1t} - \tilde{\mathbf{a}}'\tilde{\mathbf{u}}_t)$. Therefore, we suggest choosing the element of $\mathbf{w}$ throught the optimization procedure using the pretreatment observations

$$\min_{\tilde{\mathbf{a}}} \sum_{t=1}^{T_0} \left[y_{1t} - \mathbf{x}_{1t}'\hat{\boldsymbol{\beta}}_{CCE} - \tilde{\mathbf{a}}'\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t'\hat{\boldsymbol{\beta}}_{CCE}\right)\right]^2.$$

[4] An intercept can be included in the optimization to avoid the scenario when $\mathbf{a}'\Lambda$ is not exactly equal to zero and hence to impove the approximation performance.

large.

Step 3: Generate counterfactuals by

$$\hat{y}_{1t}^{0} = \mathbf{x}_{1t}'\hat{\boldsymbol{\beta}}_{CCE} + \widehat{\tilde{\mathbf{a}}}'\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\hat{\boldsymbol{\beta}}_{CCE}\right), \quad T_0 + 1 \le t \le T. \tag{11}$$

Given the estimator (11), the treatment effect at time t is estimated by

$$\hat{\delta}_{1t} = y_{1t} - \hat{y}_{1t}^{0},\ t = T_0 + 1 \le t \le T, \tag{12}$$

and the average treatment effect (ATE) over post-treatment period is estimated by

$$\hat{\Delta}_1 = \frac{1}{T_1}\sum_{t=T_0+1}^{T}\hat{\delta}_{1t}, \tag{13}$$

where $T_1 = T - T_0$ denotes the length post-treatment time periods.

Intuitively, the above PDX approach can be viewed as an extension of the approach of Hsiao et al. (2012) to model with exogenous regressors, since the original Hsiao et al (2012)'s approach doesn't include the covariates $\mathbf{x}_{it}$. However, once the coefficient of $\mathbf{x}_{it}$ is given as prior or can be consistently estimated, then the e_{it} of (6) is known, and (6) approximately reduces to the pure factor model, thus the Hsiao et al (2012)'s approach can be applied to generated counterfactuals.

Remark 1 *The advantage of the PDX approach in generating counterfactuals using (11) is that we do not need to know the dimension of the unobserved factors. Also, there are significantly fewer parameters involved than in the original GSCM approach, which could be important in the cases with finite samples.*

3 Asymptotics of the ATE

In this section, we establish the asymptotic property of the ATE of (13). To begin with, let $\Delta_1 = E\left(\Delta_{1t}\right) = \frac{1}{T_1}\sum_{t=T_0+1}^{T}\left(y_{1t} - y_{1t}^{0}\right)$ be the average treatment effect for the first unit.

For estimated ATE (13) across post-treatment periods, we have

Proposition 1 *Under assumptions A1-A3, as $(N, T_0, T_1) \to \infty$, we have*

$$\hat{\Delta}_1 - \Delta_1 = O_p\left(N^{-1/2} + T_0^{-1/2} + T_1^{-1/2}\right).$$

Proof is provided in the Appendix.

Remark 2 *We note that from the above results, as long as N is large, T_0 and T_1 are large, then the ATE $\hat{\Delta}_1$ is a consistent estimator of Δ_1. It should be noted that, compared to the approach of Hsiao et al (2012), where the convergence rate of $\hat{\Delta}_1$ is $O_p\left(T_0^{-1/2}\right) + O_p\left(T_1^{-1/2}\right)$ (Li and Bell*

(2017)), while the convergence rate of ATE $\hat{\Delta}_1$ in our PDX approach is $O_p\left(N^{-1/2}\right)+O_p\left(T_0^{-1/2}\right)+O_p\left(T_1^{-1/2}\right)$. This is because we need to estimate the slope coefficient $\boldsymbol{\beta}$ first, once $\boldsymbol{\beta}$ is consistently estimated with convergence rate of $O_p\left(N^{-1/2}\right)$, then one can use the Hsiao et al. (2012)'s approach based on the residuals $e_{it} = y_{it} - \mathbf{x}_{it}'\hat{\boldsymbol{\beta}}_{CCE}$, and the ATE is of order $O_p\left(N^{-1/2}\right) + O_p\left(T_0^{-1/2}\right) + O_p\left(T_1^{-1/2}\right)$.

Now let's turn to the asymptotic distribution of the ATE $\hat{\Delta}_1$, which is summarized in the following proposition.

Proposition 2 *Under assumptions A1-A3, as $(N, T_0, T_1) \to \infty$, and $\frac{T_1}{N} \to \kappa$ where $0 < \kappa < \infty$,*[5] *we have*

$$\sqrt{T_1}\left(\hat{\Delta}_1 - \Delta_1\right) \xrightarrow{d} N\left(0, \sigma_{\Delta_1}^2\right),$$

where $\sigma_{\Delta_1}^2 = \sigma_{u1}^2 + \kappa^{-1}\left(\Sigma_x'\Sigma_{\boldsymbol{\beta}}\Sigma_x + \Sigma_f'\Sigma_{\tilde{\mathbf{a}}}\Sigma_f\right)$ is the asymptotic variance of $\sqrt{T_1}\left(\hat{\Delta}_1 - \Delta_1\right)$, σ_{u1}^2, $\Sigma_{\boldsymbol{\beta}}$, Σ_x, Σ_f and $\Sigma_{\tilde{\mathbf{a}}}$ are given in the Appendix.

Proof is provided in the Appendix.

Given the above derivation, a consistent estimator for $\sigma_{\Delta_1}^2$ can be obtained by replacing the probability limit with sample analogous. For instance, $\Sigma_{\boldsymbol{\beta}}$ can be estimated by $\hat{\Sigma}_{\boldsymbol{\beta}} = \hat{\mathbf{D}}^{-1}\hat{\mathbf{V}}\hat{\mathbf{D}}^{-1}$ with $\hat{\mathbf{D}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\bar{\mathbf{Z}}}\mathbf{X}_i^{-1}$ and $\hat{\mathbf{V}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\bar{\mathbf{Z}}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{M}_{\bar{\mathbf{Z}}}\mathbf{X}_i$ where $\hat{\mathbf{u}}_i = \mathbf{M}_{\bar{\mathbf{Z}}}\mathbf{y}_i - \mathbf{M}_{\bar{\mathbf{Z}}}\mathbf{X}_i\hat{\boldsymbol{\beta}}_{CCE}$ (Zhou and Zhang (2016)). Σ_x can be estimated by $\hat{\Sigma}_x = \frac{1}{T_1}\sum_{t=T_0+1}^{T}\mathbf{x}_{1t}$. The consistent estimator of $\Sigma_{\tilde{\mathbf{a}}}$, denoted as $\hat{\Sigma}_{\tilde{\mathbf{a}}}$, can be obtained from Cattaneo et al. (2018a, 2018b) even when $\dim(\tilde{\mathbf{a}}) = N-1$. For the estimation of Σ_f. We first note that

$$\tilde{\mathbf{y}}_t = \tilde{\mathbf{X}}_t\boldsymbol{\beta} + \tilde{\Lambda}\mathbf{f}_t + \tilde{\mathbf{u}}_t,$$

denotes the model for the control units, i.e., $\tilde{\mathbf{y}}_t = (y_{2t}, \ldots, y_{Nt})'$. Also, $\hat{\boldsymbol{\beta}}_{CCE} = \boldsymbol{\beta} + O_p\left(N^{-1/2}\right)$, thus

$$\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\hat{\boldsymbol{\beta}}_{CCE} = \tilde{\Lambda}\mathbf{f}_t + \tilde{\mathbf{u}}_t + O_p\left(N^{-1/2}\right),$$

averaging over post-treatment periods yields $\frac{1}{T_1}\sum_{t=T_0+1}^{T}\tilde{\Lambda}\mathbf{f}_t = \frac{1}{T_1}\sum_{t=T_0+1}^{T}\tilde{\mathbf{e}}_t + O_p\left(T_1^{-1/2}\right) + O_p\left(N^{-1/2}\right)$ with $\tilde{\mathbf{e}}_t = \tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\hat{\boldsymbol{\beta}}_{CCE}$. As a result, when $(N, T_1) \to \infty$, Σ_f can be consistently estimated by $\hat{\Sigma}_f = \frac{1}{T_1}\sum_{t=T_0+1}^{T}\tilde{\mathbf{e}}_t$. Finally, for the estimation σ_{u1}^2, we note that from (A.2) that $y_{1t}^0 - \hat{y}_{1t}^0 = u_{1t} + O_p\left(N^{-1/2}\right)$, then following the argument of Li and Bell (2017) and noticing that $\hat{\delta}_{1t} = y_{1t} - \hat{y}_{1t}^0$, a potential estimator of σ_{u1}^2 could be

$$\hat{\sigma}_{u1}^2 = \frac{1}{T_1}\sum_{t=T_0+1}^{T}\sum_{t=T_0+1, |t-s|<l}^{T}\left(\hat{\delta}_{1t} - \hat{\Delta}_1\right)\left(\hat{\delta}_{1s} - \hat{\Delta}_1\right), \tag{14}$$

[5] Similar restriction has also been imposed in Li and Bell (2017).

where $l \to \infty$ as $T_1 \to \infty$ but $\frac{l}{T_1} \to 0$. For instance, one can choose $l = O\left(T_1^{1/4}\right)$ (Newey and West, 1987).

In all, using the above arguments, a consistent estimator for $\sigma^2_{\Delta_1}$ can be obtained by

$$\sigma^2_{\Delta_1} = \hat{\sigma}^2_{u1} + \kappa^{-1}\left(\hat{\Sigma}'_x \hat{\Sigma}_{\boldsymbol{\beta}} \hat{\Sigma}_x + \hat{\Sigma}'_f \hat{\Sigma}_{\tilde{\mathbf{a}}} \hat{\Sigma}_f\right), \tag{15}$$

where $\kappa = \frac{T_1}{N}$.

Given the consistent estimator of $\sigma^2_{\Delta_1}$, denoted as $\hat{\sigma}^2_{\Delta_1}$, we can construct the usual t-statistics as

$$t_{\Delta_1} = \frac{\sqrt{T_1}\left(\hat{\Delta}_1 - \Delta_1\right)}{\hat{\sigma}_{\Delta_1}}, \tag{16}$$

which can be used to test hypothesis whether the treatment is significant, i.e., $H_0 : \Delta_1 = 0$. On the other hand, if H_0 is not rejected, then there is no significant treatment effects on the treated units for the policy shock. Otherwise, there is significant treatment effects on the treated units. Furthermore, given $\hat{\sigma}_{\Delta_1}$, we can also construct the 95% confidence interval for the treatment effects in the post-treatment periods as

$$\hat{\delta}_{1t} \pm 1.96 \frac{se\left(\hat{\Delta}_1\right)}{\sqrt{T_1}}, \quad \text{for } t = T_0 + 1, \ldots, T.$$

Remark 3 *For the above asymptotic distribution of the ATE, $\hat{\Delta}_1$, compared to the results of Li and Bell (2017), we can observe that there is one extra term in the asymptotic variance of $\hat{\Delta}_1$, which is caused by the existence of exogenous regressor $\mathbf{x}_{it}$. As argued above, once $\boldsymbol{\beta}$ is consistently estimated and the effects of $\mathbf{x}_{it}$ is controlled, then our model is approximately identical to the model considered by Hsiao et al. (2012). Hence, the results of Li and Bell (2017) can be applied here.*

4 Simulation Studies

4.1 Data Generation Processes

Since the true data generating process (DGP) is unknown, the only way to consider which method is more likely to yield more accurate y^0_{it} in various array of DGPs is through computer simulations. We generate four types of DGPs to obtain the "true" counterfactuals which can never be observed in reality and compare the true counterfactuals generating by DGPs with estimated counterfactuals obtained by different methods. In the DGPs below, we assume the factors f_{1t}, f_{2t} and f_{3t} are $iidN(0, 1)$, the factor loadings $\gamma_{1,i}$, $\gamma_{2,i}$ and $\gamma_{3,i}$ are also $iidN(0, 1)$, unless they are specified otherwise. The coefficients are set at $\beta_1 = 1$, and $\beta_2 = 2$. The specific DGPs are designed as follows.

DGP1: Model with exogenous variables and common factors

$$y_{it} = x_{1,it}\beta_1 + x_{2,it}\beta_2 + \gamma_{1,i} f_{1t} + \gamma_{2,i} f_{2t} + \gamma_{3,i} f_{3t} + u_{it}, \tag{17}$$

where the covariates $x_{k,it}$ $(k = 1, 2)$ are correlated with common factors as

$$x_{k,it} = 1 + \rho_{ki} x_{k,it-1} + c_{1i}\gamma_{k,i} + c_{2i} f_{kt} + \varepsilon_{k,it}, \quad k = 1, 2,$$

where $\rho_{k,i} \sim iidU(0.1, 0.9)$, c_{1i} and c_{2i} are $iidU(1, 2)$ and the error term $\varepsilon_{k,it}$ is $iidN(0, 1)$.

DGP2: Model with exogenous variables and common factors

$$y_{it} = x_{1,it}\beta_1 + x_{2,it}\beta_2 + \gamma_{1,i} f_{1t} + \gamma_{2,i} f_{2t} + \gamma_{3,i} f_{3t} + u_{it}, \tag{18}$$

where f_{1t} and f_{2t} are $iidN(0, 1)$. The covariates $x_{k,it}$ $(k = 1, 2)$ follow an ARMA process as

$$x_{k,it} = 1 + \rho_{ki} x_{k,it-1} + \eta_{k,it} + \rho_{\eta i}\eta_{k,it-1}, \quad k = 1, 2,$$

where $\rho_{k,i}$ and $\rho_{\eta i}$ are $iidU(0.1, 0.9)$ and the error term $\eta_{k,it}$ is $iidN(0, 1)$.

DGP3: The DGP is similar to DGP1 except now we assume

$$\begin{aligned} f_{1t} &= f_{1,t-1} + \xi_{1t}, \\ f_{2t} &= 0.5 f_{2,t-1} + \xi_{2t}, \\ f_{3t} &= 0.8 f_{3,t-1} + \xi_{3t}, \end{aligned} \tag{19}$$

where ξ_{kt} is $iidN(0, 1)$.

DGP4: Model with pure factor structure

$$y_{it} = \gamma_{1,i} f_{1t} + \gamma_{2,i} f_{2t} + \gamma_{3,i} f_{3t} + u_{it}. \tag{20}$$

and $x_{k,it}$ is the same as in DGP1.

For these DGPs, we assume the error term u_{it} are weakly cross-sectionally dependent (Stock and Watson (2002)), i.e.,

$$\begin{aligned} u_{it} &= 2v_{it} + v_{i+1,t} + v_{i-1,t}, \\ v_{it} &\sim iidN(0, \sigma_i^2), \end{aligned} \tag{21}$$

where the σ_i^2 are randomly drawn from $0.5\left(\chi^2(1) + 1\right)$.

We note that DGP 1 and 3 satisfy the rank condition required for the implementation of Pesaran (2006)'s CCE method (e.g., Hsiao (2014)), while DGP 4 is a pure factor model, so the rank condition for the CCE estimation is invalid.

The treatment and control groups consist of 1 and $N - 1$ units, respectively. The treatment starts to affect the treated units at time $T_0 + 1$. For these four DGPs, we assume that the control unit to be $N - 1 = 30, 50$ and the pretreatment time $T_0 = 30, 50$, and post treatment periods $T - T_0 = 10$, i.e., $T = 30, 50$. The number of replication is set at $R = 1000$.

We are also interested in testing whether the ATEs are significant using the asymptotic properties of the ATE in the above section. Taking DGP 1 as the base model, we consider two cases:

Case 1: No Treatment

For the treated unit (the first unit), we assume $\delta_{1t} = 0$ for $t = T_0 + 1, \ldots, T$, such that $\Delta_1 = 0$.

Case 2: Significant Treatment

For the treated unit (the first unit), we assume $\delta_{1t} = 2 + \frac{t}{T}$ for $t = T_0 + 1, \ldots, T$, such that Δ_1 is different from zero and has an increasing time trend.

For these two cases, we note that there is no treatment effect in Case 1 and there is significant treatment effect for Case 2. For the simulation of the significance test of treatment effects, we let N, T_0, and T_1 be the combination of 30, 50, and the standard error for the t-statisitics is calculated using (15)

4.2 Simulation Results

We consider several estimators for the above DGPs,[6]

(**E1**) **SCM:** Generate $\hat{y}_{1t}^0$ using Abadie et al.'s (2010) SCM method for model (2).[7]

(**E2**) **GSCM**: Estimate model (2) by Bai (2009)'s PCA method.[8]

(**E3**) **PDX1**: Estimate model (2) by the PDX method through ordinary least square (OLS), then generate $\hat{y}_{1t}^0$ by (11).

[6]We use R software for simulation and estimation. We also use the "synth" package by Hainmueller and Diamond (2015) for SCM, the "gsynth" package by Xu and Liu (2017) for GSCM, the "pampe" package by Vega-Bayo (2015) for model selection using AICc, and the "glmnet" package by Friedman et al (2018) for LASSO, respectively.

[7]To be more specific, the SCM predict y_{1t} by

$$\hat{y}_{1t}^* = \mathbf{w}'\tilde{\mathbf{y}}_t = \sum_{i=2}^{N} w_i y_{it}, \quad T_0 + 1 \leq t \leq T, \tag{22}$$

where $\mathbf{w} = (w_2, \ldots, w_N)'$ are obtained by minimizing the distance,

$$\sqrt{(\mathbf{M}_1 - \mathbf{M}_0\mathbf{w})' \mathbf{V} (\mathbf{M}_1 - \mathbf{M}_0\mathbf{w})}, \tag{23}$$

subject to

$$y_{1t} = \sum_{i=2}^{N} w_i y_{it}, \quad 1 \leq t \leq T_0, \bar{\mathbf{x}}_{1k} = \sum_{i=2}^{N} w_i \bar{\mathbf{x}}_{ik}, \quad 1 \leq k \leq K, \tag{24}$$

and

$$w_i \geq 0 \text{ and } \sum_{i=2}^{N} w_i = 1, \tag{25}$$

where $\mathbf{M}_1$ and $\mathbf{M}_0$ are $(T_0 + k) \times 1$ vector and $(T_0 + k) \times (N - 1)$ matrix of preintervention observations of $(y_{1t}, \bar{\mathbf{x}}_1)'$ and $(y_{jt}, \bar{\mathbf{x}}_j)$, respectively, $\bar{\mathbf{x}}_j$ denotes the time series mean of k covariates, $\mathbf{x}_{it}$, and $\mathbf{V}$ is a positive definite matrix.

[8]To be more specific, the GSCM contains the following steps.

Step 1: Use all NT observations to estimate $\boldsymbol{\beta}$, $\boldsymbol{\gamma}_i$ and $\mathbf{f}_t$, $i = 2, \ldots, N$ and $t = 1, \ldots, T_0$, as $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}_i$ and $\hat{\mathbf{f}}_t$.

Step 2: Estimate γ_1 by using variables of treated unit for pretreatment period, $t = 1, \ldots, T_0$, as $\hat{\gamma}_1$ by

$$\min_{\gamma_1} \sum_{t=1}^{T_0} \left(y_{it} - \mathbf{x}_{1t}'\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}_1'\hat{\mathbf{f}}_t \right)^2. \tag{26}$$

Step 3: Generate the estimated counterfactual of y_{1t}^0 by

$$\hat{y}_{1t}^0 = \mathbf{x}_{1t}'\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\gamma}}_1'\hat{\mathbf{f}}_t, \quad t = T_0 + 1, \ldots, T. \tag{27}$$

(**E4**) **PDX2**: Estimate model (2) by the PDX method and the model selection criterion to select control units from $\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t'\hat{\boldsymbol{\beta}}_{CCE}\right)$ in step 2 of (10).

(**E5**) **PDX3**: Estimate model (2) by the PDX method and the LASSO method to select control units from $\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t'\hat{\boldsymbol{\beta}}_{CCE}\right)$ in step 2 of (10).

For this estimation E2, since the number of common factors $\mathbf{f}_t$ is unknown in the estimation, we follow Xu (2017)'s cross-validation method to estimate the number of common factors, while the CCE approach for E3-E5 does not require the knowledge of number of common factors.

We consider three criterion for comparison: *MAB*, *MSE* and *MAP*. The *MAB*[9] is the mean of absolute bias for the true outcome and the counterfactuals at each post treatment date point. The *MSE*[10] is the square root of mean of sum of squared bias for the true observation and the counterfactuals at each post treatment date point, and the *MAP*[11] is the mean of the ratio of absolute counterfactuals and absolute true outcomes at each date point after treatment.

We consider the performance of constructing the counterfactuals of y_{1t} ($t = T_0 + 1, \ldots, T$) by using the approaches E1-E4 using *MAB*, *MSE* and *MAP*. The simulation results are summarized in Table 1-4 for DGP 1-4, respectively. We also draw the figure of RMSE in Figure 1-4 for different approaches at each post-treatment period. Based on the asymptotic results we obtained in the previous section, we can use the usual t-statistics (16) to test whether the treatment effects is significant for Case 1 and Case 2. We consider two critical values for double sided test: 1% and 5%, and calculate the empirical rejection frequency for the (16) using these two critical values. These results are summarized in the Table 5-6, respectively.

Several interesting results can be found in Table 1-4 and Figure 1-4. First and most important, we can observe that the PDX1-PDX3 works remarkably well across different DGPs and different configuration of N and T. In general, the counterfactuals of PDX have less bias and MSE than those obtained from SCM and GSCM approaches. Second, either SCM or GSCM is quite sensitive to the true DGPs, i.e., when the data is stationary (e.g., DGP1 and DGP2), GSCM works reasonably well, and when the model is a pure factor model (DGP4), the SCM also works reasonable well. Finally, for the plot of RMSE in Figure 1-4, we can also find that PDX works much better than SCM and GSCM across different DGPs.

---

[9] *MAB* is mearsured as

$$MAB = \frac{1}{RT}\sum_{r=1}^{R}\sum_{t=T_0+1}^{T}\left|y_{1t}^{0}(r) - \hat{y}_{1t}^{0}(r)\right|$$

, which represents the average distance between true counterfactuals and estimated counterfactuals by my method. Thus, the smaller the *MAB* is, the better performance the method is.

[10] *MSE* is calculated by

$$MSE = \sqrt{\frac{1}{RT}\sum_{r=1}^{R}\sum_{t=T_0+1}^{T}\left(y_{1t}^{0}(r) - \hat{y}_{1t}^{0}(r)\right)^2}$$

which is similar to *MAB*. The smaller it is, the better performance the method is.

[11] *MAP* is measured as

$$MAP = \frac{1}{RT}\sum_{r=1}^{R}\sum_{t=1}^{T}\frac{\left|\hat{y}_{1t}^{0}(r)\right|}{\left|y_{1t}^{0}(r)\right|}$$

The closer to 1, the better performance the method is.

Similar findings can be found in Table 5-6 for the significance test using the t-statistics of (16). When there is no treatment effect in Case 1, we find that the empirical rejection frequency is quite close to the nominal value (e.g., 1% or 5%), i.e., there is no evidence to reject the null hypothesis of no significant treatment effects. When there is significant treatment in Case 2, the empirical rejection frequency increases quite rapidly with the increase of either N or T_0, and most of the cases the empirical rejection frequency is close to 100%, i.e., we can reject the null of no significant treatment. In all, the simulation results in Table 1-6 and Figure 1-4 show that our PDX approach works remarkably well in terms of bias, RMSE and validity of statistical inference.

Table 1. Simulation results of GSCM and PDX for DGP 1

(T, N)		$N = 30$					$N = 50$				
		SCM	GSCM	PDX1	PDX2	PDX3	SCM	GSCM	PDX1	PDX2	PDX3
$T = 30$	MAB	4.887	2.619	5.481	1.491	1.497	16.06	2.477	2.875	2.076	1.962
	MSE	6.004	3.382	9.956	1.954	1.991	16.81	3.187	3.682	2.703	2.563
	MAP	0.251	1.074	1.058	1.022	1.027	0.594	1.083	1.033	1.033	1.040
$T = 50$	MAB	4.931	2.388	1.209	1.135	1.105	16.24	2.227	7.062	1.610	1.584
	MSE	5.978	3.061	1.566	1.492	1.453	16.95	2.840	12.68	2.085	2.037
	MAP	0.256	1.081	1.009	1.021	1.016	0.594	1.032	1.038	1.007	1.007

Notes: "GSCM" to "PDX3" refers to different estimators described as in (E1)-(E5) respectively.

Table 2: Simulation results of GSCM and PDX for DGP 2

(T, N)		$N = 30$					$N = 50$				
		SCM	GSCM	PDX1	PDX2	PDX3	SCM	GSCM	PDX1	PDX2	PDX3
$T = 30$	MAB	4.393	2.634	5.249	1.396	1.439	3.923	2.492	2.818	2.039	1.952
	MSE	5.465	3.398	9.596	1.827	1.922	4.903	3.204	3.605	2.659	2.548
	MAP	5.021	5.093	5.180	2.470	2.780	1.214	1.620	1.439	1.337	1.370
$T = 50$	MAB	4.292	2.400	1.061	1.048	1.003	3.779	2.240	6.638	1.581	1.574
	MSE	5.323	3.075	1.379	1.373	1.312	4.740	2.856	11.98	2.050	2.016
	MAP	6.141	3.091	1.542	1.513	1.522	1.208	1.523	1.789	1.147	1.179

Notes: "GSCM" to "PDX3" refers to different estimators described as in (E1)-(E5) respectively.

Table 3: Simulation results of GSCM and PDX for DGP 3.

(T, N)		$N = 30$					$N = 50$				
		SCM	GSCM	PDX1	PDX2	PDX3	SCM	GSCM	PDX1	PDX2	PDX3
$T = 30$	MAB	7.216	3.769	6.014	1.685	1.636	15.73	3.954	3.045	2.290	2.168
	MSE	8.674	4.759	10.04	2.211	2.165	17.11	4.994	3.909	2.981	2.818
	MAP	1.284	1.795	2.633	1.249	1.161	10.58	5.753	3.140	5.457	4.152
$T = 50$	MAB	7.992	4.045	1.366	1.258	1.179	15.77	4.308	6.380	1.807	1.715
	MSE	9.585	5.144	1.779	1.639	1.543	17.13	5.451	10.23	2.324	2.193
	MAP	2.005	1.718	1.302	1.435	1.230	2.028	1.388	1.703	1.148	1.118

Notes: "GSCM" to "PDX3" refers to different estimators described as in (E1)-(E5) respectively.

Table 4: Simulation results of GSCM and PDX for DGP 4

(T, N)		$N = 30$					$N = 50$				
		SCM	GSCM	PDX1	PDX2	PDX3	SCM	GSCM	PDX1	PDX2	PDX3
$T = 30$	MAB	2.460	2.619	5.481	1.491	1.497	2.388	2.477	2.875	2.076	1.962
	MSE	3.190	3.382	9.596	1.954	1.991	3.084	3.187	3.682	2.703	2.563
	MAP	1.815	2.143	12.66	3.252	2.440	1.542	0.492	5.613	3.711	2.739
$T = 50$	MAB	2.248	2.388	1.209	1.135	1.105	2.125	2.227	7.062	1.610	1.584
	MSE	2.897	3.061	1.566	1.492	1.453	2.713	2.840	12.68	2.085	2.037
	MAP	3.285	0.468	3.923	3.848	4.980	2.306	0.617	23.63	4.846	4.215

Notes: "GSCM" to "PDX3" refers to different estimators described as in (E1)-(E5) respectively.

Figure 1: RMSE for different approaches when $N = 30$ and $T = 60$ for DGP 1

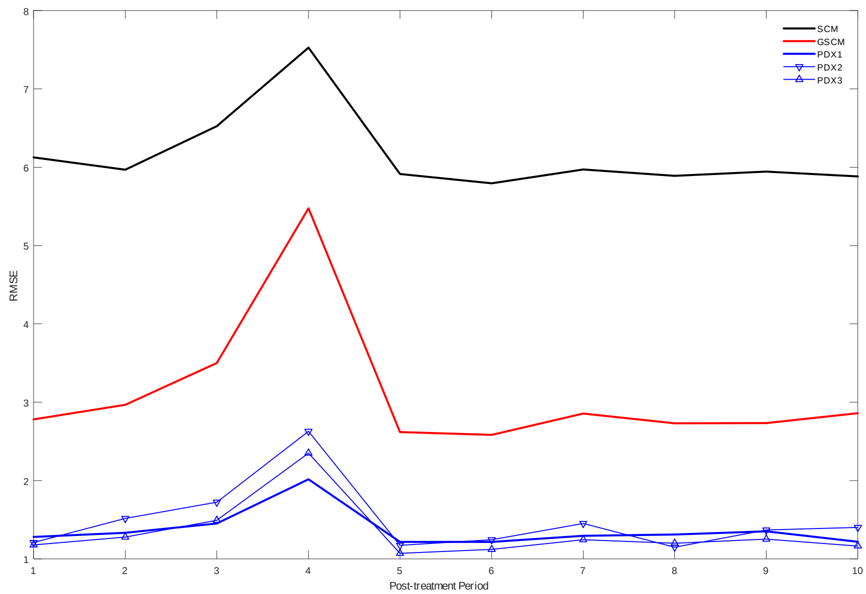


Figure 2: RMSE for different approaches when $N = 30$ and $T = 60$ for DGP 2

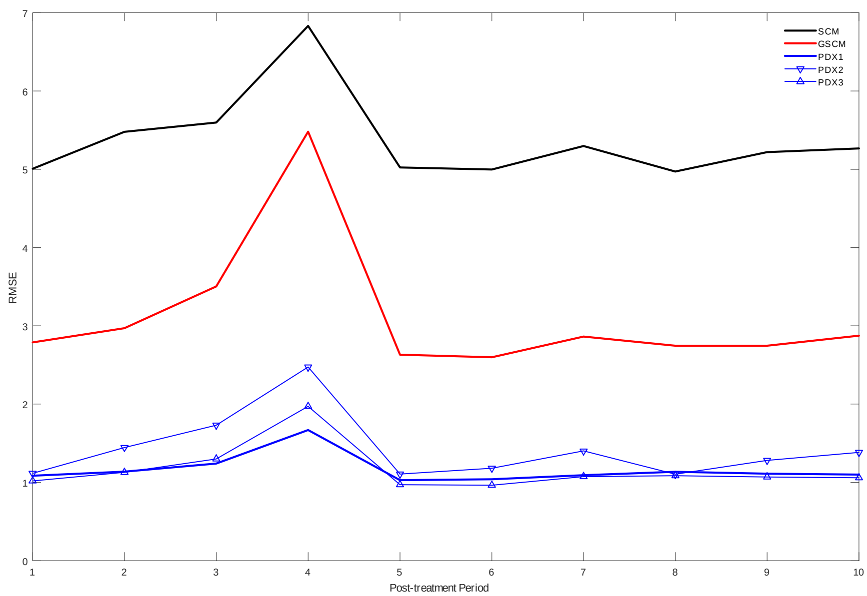


Figure 3: RMSE for different approaches when $N = 30$ and $T = 60$ for DGP 3

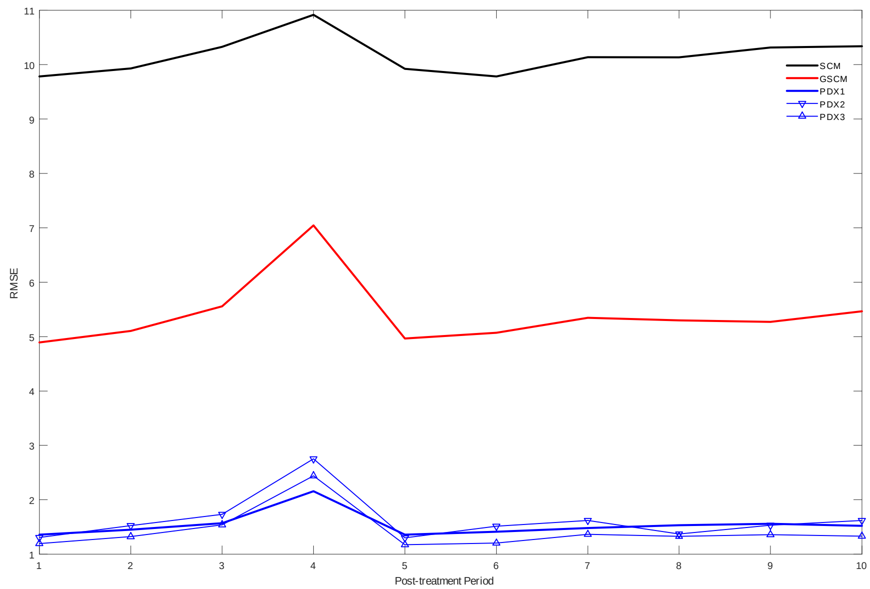


Figure 4: RMSE for different approaches when $N = 30$ and $T = 60$ for DGP 4

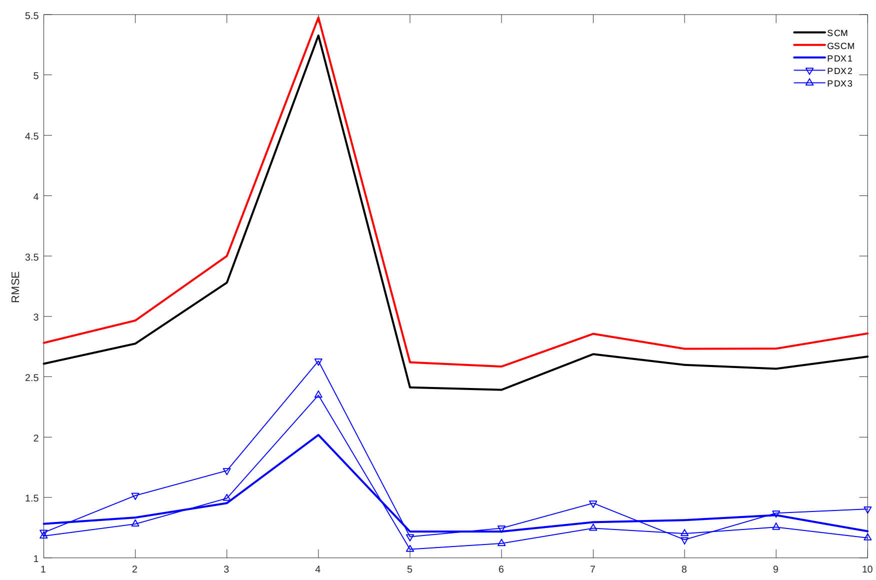


Table 5: Empirical Rejection Frequency for Case 1.

		1%						5%					
		$N = 30$			$N = 50$			$N = 30$			$N = 50$		
T_0	T_1	PDX1	PDX2	PDX3	PDX1	PDX2	PDX3	PDX1	PDX2	PDX3	PDX1	PDX2	PDX3
30	30	0.8%	1.1%	0.4%	0.3%	0.5%	0.6%	6.0%	5.2%	3.7%	6.9%	6.6%	4.6%
	50	0.3%	0.7%	0.8%	1.5%	0.4%	0.4%	5.4%	6.0%	5.0%	6.9%	5.8%	6.2%
50	30	0.4%	1.7%	1.1%	0.7%	0.4%	0.8%	1.7%	4.0%	4.2%	9.8%	4.1%	6.4%
	50	0.4%	0.5%	0.5%	0.8%	0.9%	0.6%	5.4%	6.7%	6.2%	6.0%	6.2%	5.4%

Table 6: Empirical Rejection Frequency for Case 2.

		1%						5%					
		$N = 30$			$N = 50$			$N = 30$			$N = 50$		
T_0	T_1	PDX1	PDX2	PDX3	PDX1	PDX2	PDX3	PDX1	PDX2	PDX3	PDX1	PDX2	PDX3
30	30	13.8%	56.1%	59.2%	16.1%	28.8%	25.5%	52.4%	97.4%	99.1%	77.1%	89.0%	91.1%
	50	30.3%	88.8%	92.9%	53.2%	70.6%	74.9%	64.3%	99.6%	99.9%	92.4%	98.1%	99.0%
50	30	69.2%	71.7%	75.3%	8.7%	48.0%	50.7%	98.6%	99.8%	99.9%	44.5%	98.6%	98.8%
	50	93.7%	97.7%	98.2%	21.9%	84.4%	88.4%	99.6%	100%	99.9%	58.7%	99.7%	99.9%

In general, the simulation results show that (i) using the improved GSCM to generate counterfactuals outperforms the GSCM method based on Bai (2009)'s approach, the findings are consistent for data with uncorrelated and correlated common factors, (ii) using the OLS to estimate $\mathbf{w}$ and μ will generate more accurate counterfactuals.

5 Application to Measure the Impact of SYG Law on the State Level Murder Rate

In this section, we illustrate the new panel data approach for evaluating the effects of the Stand Your Ground (SYG) Law on the US Murder Rate. Since 2005, a wave of U.S. states have passed laws expanding the circumstances under which individuals have the right to use deadly force to defend against a threat to their life or property. Such laws increase the number of situations where citizens are permitted to use deadly force against others. The following Table 7 provides a summary of when and which state passes the SYG laws.[12]

[12] We don't consider the state of Alaska, which passes the SYG law on 2014, and also delete Utah as the treated state since Utah passed the similar law on 1994.

Table 7. State Effective Date for the SYG Law

State	Year	State	Year
Alabama	2007	Arizona	2011
Florida	2006	Georgia	2007
Indiana	2007	Kansas	2011
Kentucky	2007	Louisiana	2007
Michigan	2007	Mississippi	2007
Montana	2010	Nevada	2012
New hampshire	2012	North Carolina	2012
Oklahoma	2007	Pennsylvinia	2012
South Carolina	2007	South Dakota	2007
Tennessee	2008	Texas	2008
West Virginia	2009		

Since the very beginning, there has been numerous components argue that such measures can be expected to deter criminal activity while opponents argue that such laws are likely to increase homicide rates. A number of recent studies consider different approaches to identify the effect of SYG laws on homicide rates and find, in general, a positive effect (McClellan and Tekin 2016). In our paper, we utilize the new pane data approach to reexamine the effects of SYG law on the state level murder rate. The data is collected from a variety of public sources and merged it with annual state-level murder rates from 1970 to 2015.[13] The control covariates include per capita income (in logarithm), poverty rate and education attainment, state-level population (in logarithm). It can be noted that these control variables are likely unaffected by the effectiveness of SYG. We let all states that don't possess the SYG law as control states, and consider the treatment effects for the states that passed the law before 2010 (e.g., Alabama, Florida, Georgia, Indiana, Kentucky, Louisiana, Michigan, Mississippi, Oklahoma, South Carolina, South Dakota, Tennessee, Texas and West Virginia).

In order to measure the effect of whether SYG increase/decrease murder rate for the treated states, we consider several approaches (E(1)-E(5)) to construct the counterfactuals and the treatment effects. We also consider the average treatment effects over post-treatment period and interested in testing whether the ATE is significant for the treated states. The estimation results are provided in Table 8-10 as well in Figure 5-7 for state of Florida, Mississippi and Louisiana. The results for other treated states are provided in the Appendix.

From these estimation results, we can find that: (1) The PDX approach (with or without model selection/Lasso) provides the most accurate prediction in the pretreated periods, while both the SCM and GSCM perform quite bad in the prediction of pretreated periods; (2) For post-treatment periods, one can observe that SCM and PDX (with or without model selection/Lasso) provide the similar conclusion that the SYG law in general has a positive effect on the state-level murder

[13]We would like to thank Anton Strezhnev for sharing his own SYG data.

rate. However, the magnitude of the impact varies across different methods. For instance, for the state of Florida, the ATE for SCM is 0.87, and is 3.51, 1.43 and 1.46 for PDX1, PDX2 and PDX3, respectively; (3) The GSCM approach provides quite opposite conclusion for most of the states with SYG law, except for the state of Louisiana; (4) Using the asymptotics we obtained for the ATE calculated from PDX, we note that even if the ATEs for PDX (with or without model selection/Lasso) are positive, while most of the ATEs are insignificant from zero (except for the ATEs calculated from PDX1 for the states of Florida, Louisiana and Mississippi).

In conclusion, based on the results in Table 8-10, Figure 5-7 and the results in the Appendix, we can find that the SYG law in general increases the state-level murder rate, while these impact could be insignificant based on the t-statistics.

Table 8: Actual and Counterfactual Murder Rate for **Florida** in the Post-treatment Period

Year	Actual	SCM	GSCM	PDX1	PDX2	PDX3
2006	6.2	5.5	9.5	3.9	5.0	5.0
2007	6.6	5.0	9.8	1.3	3.4	3.8
2008	6.3	5.0	10.0	1.4	4.8	4.5
2009	5.5	4.8	10.5	3.4	4.1	4.7
2010	5.2	4.4	10.8	0.1	4.4	4.1
2011	5.2	4.4	10.5	1.4	3.5	4.0
2012	5.2	4.3	10.6	3.6	4.9	4.6
2013	5.0	4.1	10.5	1.8	4.1	3.8
2014	4.9	4.1	11.0	1.3	3.0	2.9
2015	5.1	4.4	10.8	2.0	3.8	3.4
ATE		0.87	-4.92	3.51 (0.04)	1.43 (0.13)	1.46 (0.12)

Note: ATE is calculated over Post-treatment Periods. The number in parenthesis shows p-value of the significance test of the ATE.

Figure 5. Actual and Counterfactual Murder Rate for **Florida**

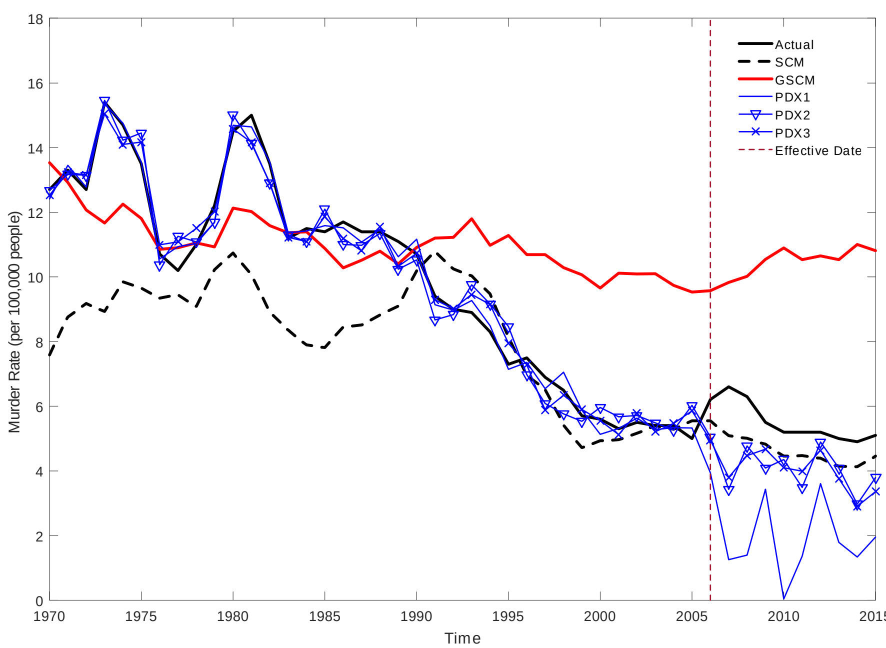


Table 9: Actual and Counterfactual Murder Rate for **Mississippi** in the Post-treatment Period

Year	Actual	SCM	GSCM	PDX1	PDX2	PDX3
2007	7.1	8.2	10.7	2.7	7.3	6.2
2008	8.0	7.5	9.4	7.6	5.6	7.5
2009	6.6	9.9	10.8	5.7	5.4	6.2
2010	6.9	6.8	10.6	0.7	4.8	4.9
2011	7.8	7.6	9.1	0.7	4.3	4.4
2012	7.1	5.6	10.4	3.7	4.2	4.3
2013	7.3	5.9	9.6	4.4	6.5	5.9
2014	8.7	4.8	10.4	3.7	4.2	4.6
2015	8.5	5.6	9.5	5.9	4.5	5.7
ATE		0.68	-2.54	3.66 (0.08)	2.35 (0.13)	2.01 (0.17)

Note: ATE is calculated over Post-treatment Periods. The number in parenthesis shows p-value of of the significance test of the ATE.

Figure 6. Actual and Counterfactual Murder Rate for **Mississippi**

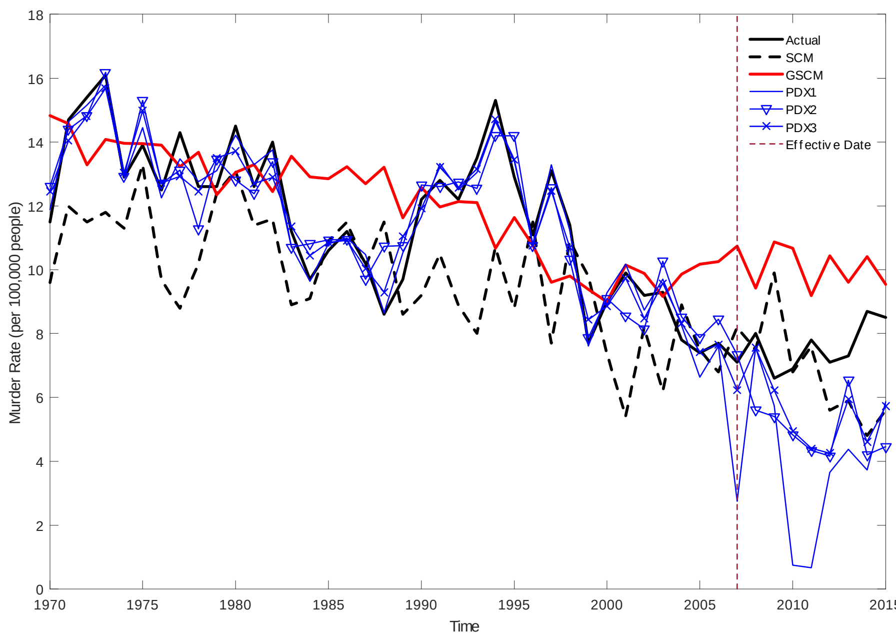
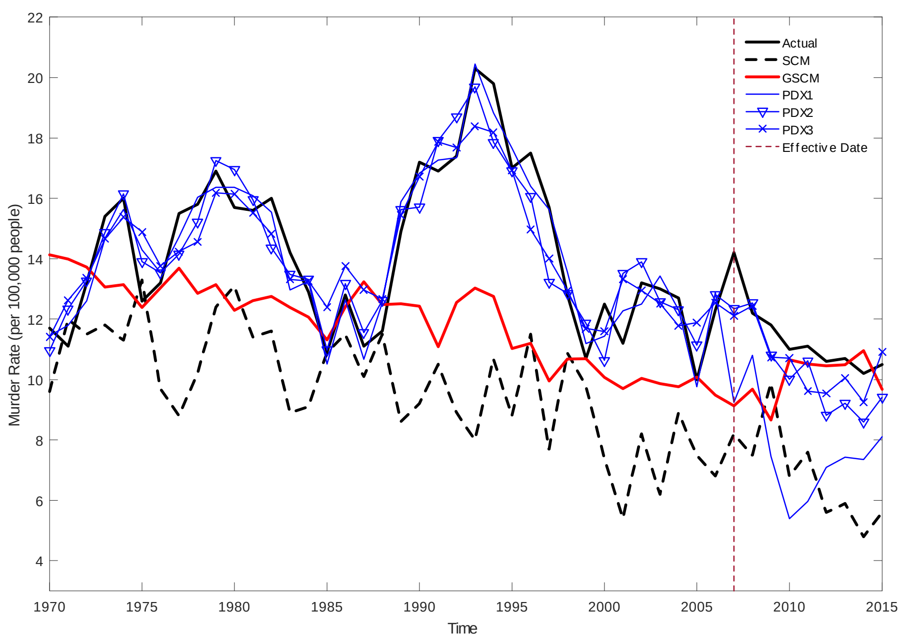


Table 10: Actual and Counterfactual Murder Rate for **Louisiana**

Year	Actual	SCM	GSCM	PDX1	PDX2	PDX3
2007	14.2	8.1	9.1	9.2	12.4	12.1
2008	12.2	7.4	9.6	10.8	12.5	12.4
2009	11.8	9.8	8.6	7.5	108	10.7
2010	11.0	6.7	10.6	5.4	10.0	10.7
2011	11.1	7.5	10.5	6.0	10.6	9.6
2012	10.6	5.5	10.4	7.1	8.8	9.5
2013	10.7	5.8	10.4	7.4	9.2	10.1
2014	10.2	4.7	10.9	7.4	8.6	9.3
2015	10.5	5.5	9.6	8.1	9.4	10.9
ATE		4.49	1.34	3.72 (0.06)	1.11 (0.37)	0.77 (0.51)

Note: ATE is calculated over Post-treatment Periods. The number in parenthesis shows p-value of of the significance test of the ATE.

Figure 7. Actual and Counterfactual Murder Rate for **Louisiana**



6 Conclusion

This paper proposes a new panel data approach for program evaluation of social policy. This new method unifies the idea of Pesaran's (2006) CCE and Hsiao et al.'s (2012) approach for panel with IFE models. It provides a semiparametric and data-driven approach of constructing counterfactuals for the treatment effect at time t of the treated unit. Different from the SCM, the PDX approach doesn't put any constraint on the outcomes and control covariates between the treated units and control units. Unlike the GSCM approach, the PDX approach doesn't rely on the unknown dimension of unobserved factors, and thus the number of unknown parameters involved could considerably less than the number involved in the parametric approach of GSCM. It is evident in the simulation that PDX approach generally outperforms that the GSCM method in a variety setup of DGPs, regardless whether N is large or T is large. It is also clear that the statistical inference is also convincing since the t-statistics based on asymptotics of the ATE using PDX approach is able to discriminate whether there is significant treatment effect for the treated units.

We apply the new PDX approach to study the impact of US SYG law on state-level murder rate. We find that, in general, the SYG law increases the state-level murder rate, while these impacts could be insignificant based on the t-statistics. We should also point out that the conclusion is quite different across different methods, e.g., the conclusion from GSCM could be on the opposite. In all, since all methods are based on certain maintained hypotheses, thus the results need to be interpreted with caution.

References

[1] Abadie, A. and Gardeazabal, J., 2003. The economic costs of conflict: A case study of the Basque Country. American economic review, 93(1), pp.113-132.

[2] Abadie, A., Diamond, A. and Hainmueller, J., 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), pp.493-505.

[3] Abadie, A., Diamond, A. and Hainmueller, J., 2015. Comparative politics and the synthetic control method. American Journal of Political Science, 59(2), pp.495-510.

[4] Athey, S. and Imbens, G.W., 2017. The state of applied econometrics: Causality and policy evaluation. Journal of Economic Perspectives, 31(2), pp.3-32.

[5] Bai, J., 2009, Panel Data Models With Interactive Fixed Effects, Econometrica 77, 1229–1279.

[6] Cattaneo, M., M. Jansson and W. Newey, 2018a, Inference in linear regression models with many covariates and heteroskedasticity, Journal of the American Statistical Association 113, 1350-1361.

[7] Cattaneo, M., M. Jansson and X. Ma, 2018b, Two-step Estimation and Inference with Possibly Many Included Covariates, Review of Economic Studies, forthcoming.

[8] Friedman, F., T. Hastie, R. Tibshirani, N. Simon, B. Narasimhan and J. Qian, Package "glmnet" for Lasso and Elastic-Net Regularized Generalized Linear Models, link: https://cran.r-project.org/web/packages/glmnet/index.html.

[9] Gardeazabal, J. and Vega-Bayo, A., 2017. An Empirical Comparison Between the Synthetic Control Method and HSIAO et al.'s Panel Data Approach to Program Evaluation. Journal of Applied Econometrics, 32(5), pp.983-1002.

[10] Gobillon, L. and Magnac, T., 2016. Regional policy evaluation: Interactive fixed effects and synthetic controls. Review of Economics and Statistics, 98(3), pp.535-551.

[11] Hainmueller, J, and A. Diamond, Package "synth" for Sythetic Control Method, link: https://cran.r-project.org/web/packages/Synth/index.html.

[12] Heckman, J. and E. Vytlacil, 2007a, Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation, in Handbook of Econometrics vol. 6B, Elsevier,

[13] Heckman, J. and E. Vytlacil, 2007b, 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, Elsevier,

[14] Hsiao, C., 2014, Analysis of Panel Data, Cambridge U. Press.

[15] Hsiao, C., 2018, Panel Models with Interactive Effects, Journal of Econometrics, forthcoming.

[16] Hsiao C, S. Ching, and S.K. Wan, 2012, A panel data approach for program evaluation: Measuring the benefits of political and economic integration of Hong Kong with mainland China. Journal of Applied Econometrics 27, 705–740.

[17] Newey, W., West, K., 1987. A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 55, 703–708.

[18] Pesaran, M.H., 2006, Estimation and Inference in Large Heterogeneous Panels with Cross Section Dependence, Econometrica, 74, 967-1012.

[19] Vega-Bayo, A., 2015, Package 'pampe' for PDA approach, link: https://cran.r-project.org/web/packages/pampe/index.html

[20] Wan, S-K., Y. Xie and C. Hsiao, 2018, Panel Data Approach vs Synthetic Control Method, working paper, Economic Letters

[21] White, H., 2001, Asymptotic Theory for Econometricians: Revised Edition, Emerald Group Publishing Limited.

[22] Xu, Y., 2017, Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Model, Political Analysis, Vol. 25, Iss. 1, January 2017, pp. 57-76.

[23] Xu, Y. and L. Liu, 2017, Package 'gsynth' for Generalized Synthetic Control Method, link: https://cran.r-project.org/web/packages/gsynth/index.html

[24] Zhang, Y. and Q. Zhou. 2016, Common Correlated Effects Estimation of Unbalanced Panel Data Models with Cross-Sectional Dependence, Journal of Economic Theory and Econometrics 27, 25-45.

Appendix

The appendix provides the derivations in the main paper and additional estimation results for the impact of SYG law on state level murder rate.

A Mathematical derivations

Proof of Proposition 1

By definition of $\hat{\Delta}_1$ and Δ_1, and since $\delta_{1t} = y_{1t} - y_{1t}^0$, we have

$$\begin{aligned}
\hat{\Delta}_1 &= \frac{1}{T_1}\sum_{t=T_0+1}^{T} \hat{\delta}_{1t} = \frac{1}{T_1}\sum_{t=T_0+1}^{T} \left(y_{1t} - y_{1t}^0 + y_{1t}^0 - \hat{y}_{1t}^0\right) \\
&= \frac{1}{T_1}\sum_{t=T_0+1}^{T} \left(y_{1t} - y_{1t}^0\right) + \frac{1}{T_1}\sum_{t=T_0+1}^{T} \left(y_{1t}^0 - \hat{y}_{1t}^0\right) \\
&= \frac{1}{T_1}\sum_{t=T_0+1}^{T} \left(y_{1t} - y_{1t}^0\right) + \frac{1}{T_1}\sum_{t=T_0+1}^{T} \left(y_{1t}^0 - \hat{y}_{1t}^0\right) \\
&= \Delta_1 + \frac{1}{T_1}\sum_{t=T_0+1}^{T} \left(y_{1t}^0 - \hat{y}_{1t}^0\right). \qquad \text{(A.1)}
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
y_{1t}^0 - \hat{y}_{1t}^0 &= \mathbf{x}_{1t}'\boldsymbol{\beta}_0 + \boldsymbol{\gamma}_1'\mathbf{f}_t + u_{1t} - \left(\mathbf{x}_{1t}'\hat{\boldsymbol{\beta}}_{CCE} + \widehat{\tilde{\mathbf{a}}}'\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\hat{\boldsymbol{\beta}}_{CCE}\right)\right) \\
&= u_{1t} + \mathbf{x}_{1t}'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{CCE}\right) + \boldsymbol{\gamma}_1'\mathbf{f}_t - \widehat{\tilde{\mathbf{a}}}'\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\hat{\boldsymbol{\beta}}_{CCE}\right) \\
&= u_{1t} + \mathbf{x}_{1t}'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{CCE}\right) + \boldsymbol{\gamma}_1'\mathbf{f}_t - \tilde{\mathbf{a}}'\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\hat{\boldsymbol{\beta}}_{CCE}\right) + \left(\tilde{\mathbf{a}} - \widehat{\tilde{\mathbf{a}}}\right)'\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\hat{\boldsymbol{\beta}}_{CCE}\right) \\
&= u_{1t} + \mathbf{x}_{1t}'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{CCE}\right) + \left(\boldsymbol{\gamma}_1' - \tilde{\mathbf{a}}'\tilde{\Lambda}_t\right)\mathbf{f}_t + \left(\tilde{\mathbf{a}} - \widehat{\tilde{\mathbf{a}}}\right)'\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\hat{\boldsymbol{\beta}}_{CCE}\right) \\
&= u_{1t} + \mathbf{x}_{1t}'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{CCE}\right) + \left(\tilde{\mathbf{a}} - \widehat{\tilde{\mathbf{a}}}\right)'\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\hat{\boldsymbol{\beta}}_{CCE}\right). \qquad \text{(A.2)}
\end{aligned}$$

Then under assumption A1-A2 and as $(N, T_0, T_1) \to \infty$, we obtain

$$\begin{aligned}
&\frac{1}{T_1}\sum_{t=T_0+1}^{T} \left(y_{1t}^0 - \hat{y}_{1t}^0\right) \\
&= \frac{1}{T_1}\sum_{t=T_0+1}^{T} u_{1t} + \frac{1}{T_1}\sum_{t=T_0+1}^{T} \mathbf{x}_{1t}'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{CCE}\right) + \left(\tilde{\mathbf{a}} - \widehat{\tilde{\mathbf{a}}}\right)'\frac{1}{T_1}\sum_{t=T_0+1}^{T}\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\hat{\boldsymbol{\beta}}_{CCE}\right) \\
&= O_p\left(T_1^{-1/2}\right) + \frac{1}{T_1}\sum_{t=T_0+1}^{T} \mathbf{x}_{1t}'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{CCE}\right) + \frac{1}{T_1}\sum_{t=T_0+1}^{T}\left(\tilde{\mathbf{a}} - \widehat{\tilde{\mathbf{a}}}\right)'\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\hat{\boldsymbol{\beta}}_{CCE}\right) \\
&= O_p\left(T_1^{-1/2}\right) + O_p\left(N^{-1/2}\right) + O_p\left(T_0^{-1/2}\right), \qquad \text{(A.3)}
\end{aligned}$$

where the last equation holds since

$$\hat{\boldsymbol{\beta}}_{CCE} = \hat{\boldsymbol{\beta}} + O_p\left(N^{-1/2}\right), \tag{A.4a}$$

from Zhang and Zhou (2016), and

$$\hat{\tilde{\mathbf{a}}} = \tilde{\mathbf{a}} + O_p\left(T_0^{-1/2}\right), \tag{A.5}$$

by using standard results for OLS estimation with many regressors, i.e., $\dim(\mathbf{w}) = N - 1$ is (moderate) large (e.g., Cattaneo et al (2018a, 2018b)).

Substituting (A.3) into (A.1) yields

$$\hat{\Delta}_1 = \Delta_1 + O_p\left(T_1^{-1/2}\right) + O_p\left(N^{-1/2}\right) + O_p\left(T_0^{-1/2}\right),$$

as required.

Proof of Proposition 2

Using the previous notations, we have

$$\begin{aligned}
&\sqrt{T_1}\left(\hat{\Delta}_1 - \Delta_1\right) \\
= &\ \frac{1}{\sqrt{T_1}}\sum_{t=T_0+1}^{T}\left(y_{1t}^0 - \hat{y}_{1t}^0\right) \\
= &\ \frac{1}{\sqrt{T_1}}\sum_{t=T_0+1}^{T} u_{1t} + \frac{1}{\sqrt{T_1}}\sum_{t=T_0+1}^{T}\mathbf{x}_{1t}'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{CCE}\right) + \left(\tilde{\mathbf{a}} - \hat{\tilde{\mathbf{a}}}\right)'\frac{1}{\sqrt{T_1}}\sum_{t=T_0+1}^{T}\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\hat{\boldsymbol{\beta}}_{CCE}\right)
\end{aligned} \tag{A.6}$$

For the first term of (A.6), using standard argument for CLT (e.g., White (2001)), we obtain

$$\frac{1}{\sqrt{T_1}}\sum_{t=T_0+1}^{T} u_{1t} \xrightarrow{d} N\left(0, \sigma_{u1}^2\right), \tag{A.7}$$

as $T_1 \to \infty$, with $\sigma_{u1}^2 = Var\left(\frac{1}{\sqrt{T_1}}\sum_{t=T_0+1}^{T} u_{1t}\right)$ under assumption A1.

For the second term, under assumption A4, as $(N, T_1) \to \infty$, we have

$$\begin{aligned}
\frac{1}{\sqrt{T_1}}\sum_{t=T_0+1}^{T}\mathbf{x}_{1t}'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{CCE}\right) &= \left(\kappa^{-1/2}\frac{1}{T_1}\sum_{t=T_0+1}^{T}\mathbf{x}_{1t}'\right)\sqrt{N}\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{CCE}\right) \\
&\xrightarrow{d} \kappa^{-1/2}\Sigma_x' N\left(0, \Sigma_{\boldsymbol{\beta}}\right),
\end{aligned} \tag{A.8}$$

where $\Sigma_x = plim_{T_1\to\infty}\frac{1}{T_1}\sum_{t=T_0+1}^{T}\mathbf{x}_{1t}$ and $\Sigma_{\boldsymbol{\beta}} = Var\left(\hat{\boldsymbol{\beta}}_{CCE}\right) = \mathbf{D}^{-1}\mathbf{V}\mathbf{D}^{-1}$ with $\mathbf{D} = plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\bar{\mathbf{Z}}}\mathbf{X}_i^{-1}$ and $\mathbf{V} = plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\bar{\mathbf{Z}}}\mathbf{u}_i\mathbf{u}_i'\mathbf{M}_{\bar{\mathbf{Z}}}\mathbf{X}_i$ (Zhou and Zhang (2016)).

The last term converges to

$$\begin{aligned}
\left(\tilde{\mathbf{a}}-\widehat{\tilde{\mathbf{a}}}\right)' \frac{1}{\sqrt{T_1}} \sum_{t=T_0+1}^{T} \left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \hat{\boldsymbol{\beta}}_{CCE}\right) &= \left(\kappa^{-1/2} \frac{1}{T_1} \sum_{t=T_0+1}^{T} \left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \hat{\boldsymbol{\beta}}_{CCE}\right)\right)' \sqrt{N}\left(\tilde{\mathbf{a}}-\widehat{\tilde{\mathbf{a}}}\right) \\
&= \left(\kappa^{-1/2} \frac{1}{T_1} \sum_{t=T_0+1}^{T} \left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta}_0\right)\right)' \sqrt{N}\left(\tilde{\mathbf{a}}-\widehat{\tilde{\mathbf{a}}}\right) + O_p\left(N^{-1/2}\right) \\
&\xrightarrow{d} \kappa^{-1/2} \Sigma_f N\left(0, \Sigma_{\tilde{\mathbf{a}}}\right), \qquad \text{(A.9)}
\end{aligned}$$

where $\Sigma_{\tilde{\mathbf{a}}} = Var\left(\sqrt{N}\left(\tilde{\mathbf{a}}-\widehat{\tilde{\mathbf{a}}}\right)\right)$ (which can be derived following Cattaneo et al (2018a, 2018b)) and $\Sigma_f = plim_{T_1\to\infty} \frac{1}{T_1}\sum_{t=T_0+1}^{T} \tilde{\Lambda}\mathbf{f}_t$ with $\tilde{\Lambda} = (\boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_N)'$.

Furthermore, we note that the covariance between the second and third term is given by

$$\begin{aligned}
& Cov\left(\frac{1}{\sqrt{T_1}} \sum_{t=T_0+1}^{T} \mathbf{x}_{1t}'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{CCE}\right), \left(\tilde{\mathbf{a}}-\widehat{\tilde{\mathbf{a}}}\right)' \frac{1}{\sqrt{T_1}} \sum_{t=T_0+1}^{T} \left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \hat{\boldsymbol{\beta}}_{CCE}\right)\right) \\
= \;& \frac{1}{T_1} Cov\left(\sum_{t=T_0+1}^{T} \mathbf{x}_{1t}'\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{CCE}\right), \left(\tilde{\mathbf{a}}-\widehat{\tilde{\mathbf{a}}}\right)' \sum_{t=T_0+1}^{T} \tilde{\Lambda}\mathbf{f}_t\right) + O\left(N^{-1/2}\right) \\
= \;& \frac{1}{T_1} Cov\left(\sum_{t=T_0+1}^{T} \mathbf{x}_{1t}' O_p\left(N^{-1/2}\right), O_p\left(T_0^{-1/2}\right) \sum_{t=T_0+1}^{T} \tilde{\Lambda}\mathbf{f}_t\right) + O\left(N^{-1/2}\right) \\
= \;& O\left(N^{-1/2}\right) + O\left(T_0^{-1/2}\right), \qquad \text{(A.10)}
\end{aligned}$$

by using the facts of (A.4a) and (A.5).

Consequently, substituting (A.7)-(A.10) into (A.6) yields

$$\sqrt{T_1}\left(\hat{\Delta}_1 - \Delta_1\right) \xrightarrow{d} N\left(0, \sigma^2_{\Delta_1}\right),$$

where $\sigma^2_{\Delta_1} = \sigma^2_{u1} + \kappa^{-1}\left(\Sigma_x' \Sigma_{\boldsymbol{\beta}} \Sigma_x + \Sigma_f' \Sigma_{\mathbf{w}} \Sigma_f\right)$. This completes the proof.

B Additional Results for the impact of SYG law on state level murder rate

This section provides additional empirical results for the treated states for the impact of SYG law on state level murder rate.

Table A1: Actual and Counterfactual Murder Rate for **Alabama** in the Post-treatment Period

Year	Actual	SCM	GSCM	PDX1	PDX2	PDX3
2007	8.9	6.5	8.7	7.6	7.1	7.2
2008	7.6	5.6	8.6	7.5	5.4	6.2
2009	6.8	6.2	9.3	5.7	5.0	6.0
2010	5.7	4.7	9.4	6.3	5.5	5.9
2011	6.2	5.4	8.9	5.5	4.7	5.5
2012	7.1	5.5	9.1	6.6	5.3	5.9
2013	7.2	5.0	9.7	6.0	5.6	5.9
2014	5.7	5.4	9.5	5.8	4.9	5.5
2015	7.2	5.7	9.0	8.0	7.3	7.1
ATE		1.33	-2.23	0.38 (0.70)	1.30 (0.27)	0.80 (0.45)

Note: ATE is calculated over Post-treatment Periods. The number in parenthesis shows p-value of the significance test of the ATE.

Figure A1. Actual and Counterfactual Murder Rate for **Alabama**

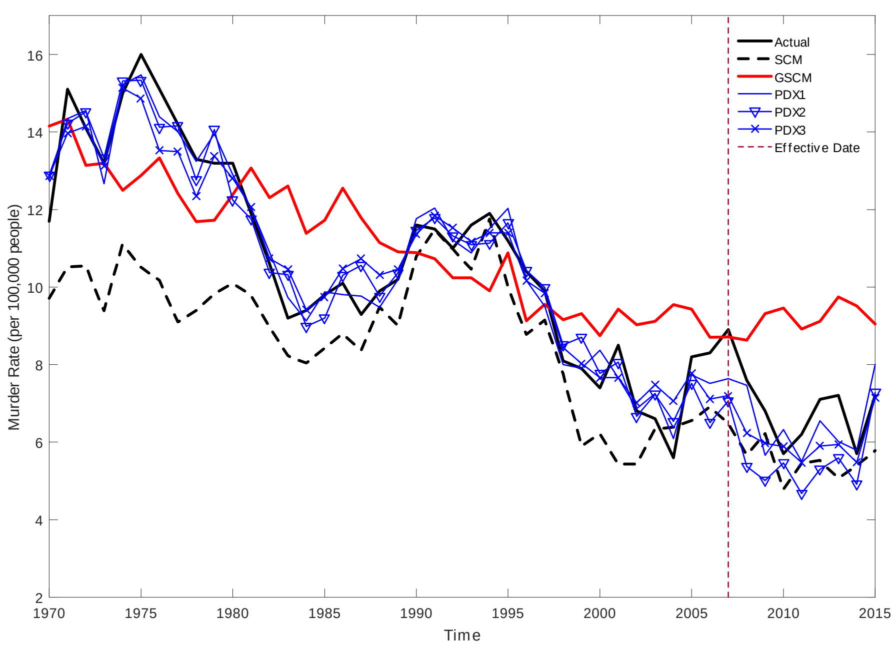


Table A2: Actual and Counterfactual Murder Rate for **Georgia** in the Post-treatment Period

Year	Actual	SCM	GSCM	PDX1	PDX2	PDX3
2007	7.5	5.5	9.4	5.8	6.1	6.1
2008	6.7	5.2	10.0	8.2	6.6	6.5
2009	5.8	5.2	10.8	6.0	5.4	5.7
2010	5.7	4.3	10.7	3.8	5.4	5.1
2011	5.6	4.7	10.7	3.6	4.4	4.5
2012	5.9	4.9	10.7	4.7	5.0	4.8
2013	5.6	4.5	10.8	4.2	5.0	4.6
2014	6.0	4.7	10.2	3.3	3.7	3.7
2015	6.1	5.2	10.6	6.8	5.6	5.4
ATE		1.14	-4.41	0.94 (0.32)	0.86 (0.30)	0.95 (0.27)

Note: ATE is calculated over Post-treatment Periods. The number in parenthesis shows p-value of the significance test of the ATE.

Figure A2. Actual and Counterfactual Murder Rate for **Georgia**

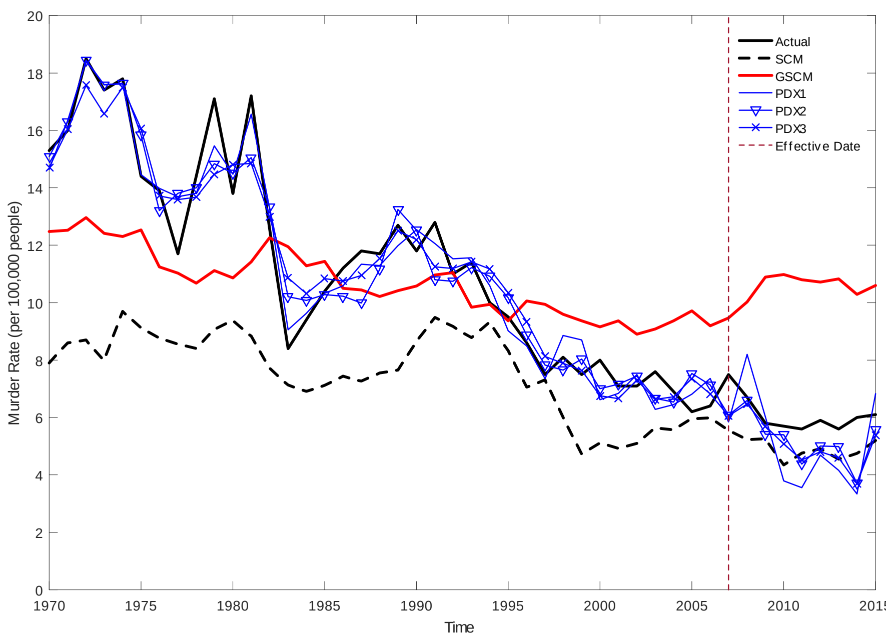


Table A3: Actual and Counterfactual Murder Rate for **Indiana** in the Post-treatment Period

Year	Actual	SCM	GSCM	PDX1	PDX2	PDX3
2007	5.6	5.0	8.5	6.3	6.5	5.7
2008	5.0	4.9	9.1	5.6	5.3	6.0
2009	4.9	4.9	9.7	5.3	5.4	5.4
2010	4.1	4.1	9.7	5.3	5.5	5.3
2011	4.7	4.5	9.4	5.7	6.6	5.4
2012	4.7	4.5	9.3	4.8	5.6	5.0
2013	5.4	4.3	9.6	5.9	5.9	5.3
2014	5.0	4.4	9.0	4.3	5.0	4.9
2015	5.6	4.9	8.7	4.7	5.2	5.7
ATE		0.35	-4.26	-0.33 (0.69)	-0.66 (0.45)	-0.40 (0.61)

Note: ATE is calculated over Post-treatment Periods. The number in parenthesis shows p-value of the significance test of the ATE.

Figure A3. Actual and Counterfactual Murder Rate for **Indiana**

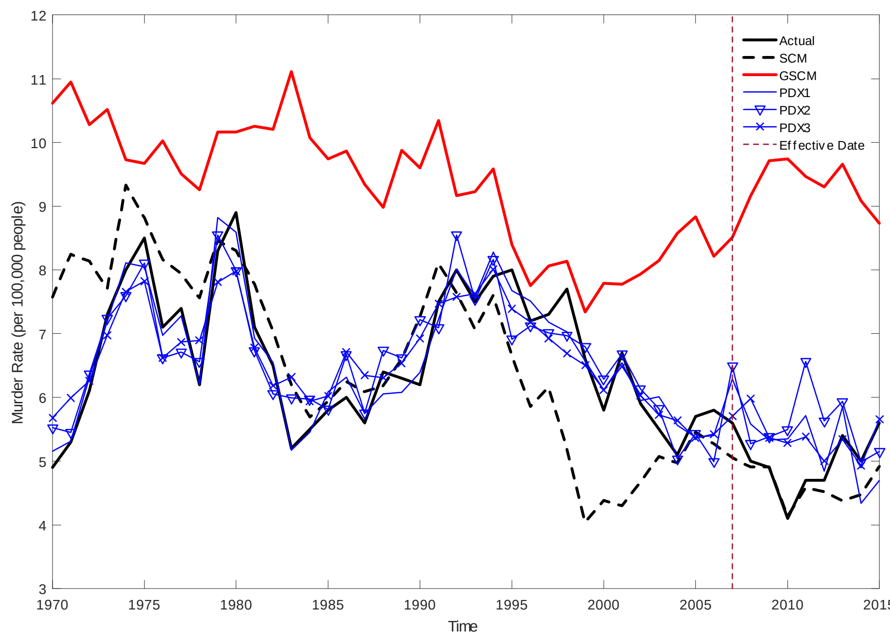


Table A4: Actual and Counterfactual Murder Rate for **Kentucky** in the Post-treatment Period

Year	Actual	SCM	GSCM	PDX1	PDX2	PDX3
2007	4.8	5.9	9.1	3.3	3.2	3.9
2008	4.7	5.3	9.5	3.2	5.5	5.6
2009	4.3	5.6	9.5	4.1	4.1	3.8
2010	4.3	4.4	9.7	4.4	4.0	3.7
2011	3.5	5.0	9.1	3.0	3.5	3.8
2012	4.6	5.3	9.6	2.4	3.6	3.7
2013	3.9	4.8	10.8	3.6	3.4	3.3
2014	3.7	5.2	10.2	3.5	4.2	4.0
2015	4.9	5.7	10.0	3.1	4.7	4.7
ATE		-0.99	-5.45	0.90 (0.34)	0.27 (0.75)	0.24 (0.77)

Note: ATE is calculated over Post-treatment Periods. The number in parenthesis shows p-value of the significance test of the ATE.

Figure A4. Actual and Counterfactual Murder Rate for **Kentucky**

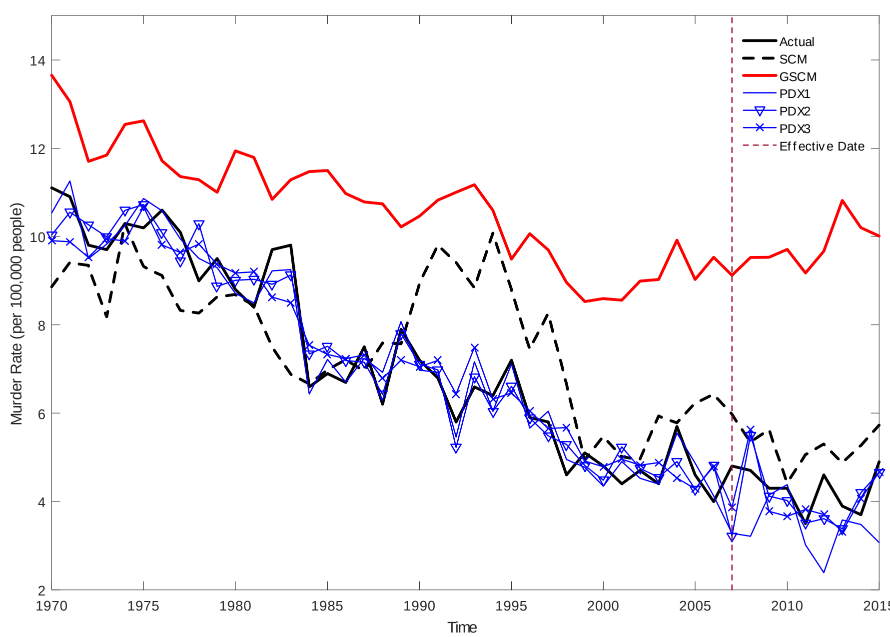


Table A5: Actual and Counterfactual Murder Rate for **Michigan** in the Post-treatment Period

Year	Actual	SCM	GSCM	PDX1	PDX2	PDX3
2007	6.7	5.0	8.7	6.5	5.7	5.8
2008	5.5	4.9	9.3	8.0	5.8	6.4
2009	6.2	4.5	9.6	6.8	6.4	6.1
2010	5.9	4.3	10.1	7.0	6.3	6.5
2011	6.2	4.2	9.8	6.0	5.9	5.7
2012	7.1	4.2	9.4	5.5	6.2	5.7
2013	6.3	3.9	9.1	5.6	6.1	5.8
2014	5.5	3.9	9.6	5.6	6.1	5.5
2015	5.9	4.4	9.0	7.2	6.3	6.0
ATE		1.74	-3.30	-0.32 (0.74)	0.05 (0.96)	0.20 (0.82)

Note: ATE is calculated over Post-treatment Periods. The number in parenthesis shows p-value of the significance test of the ATE.

Figure A5. Actual and Counterfactual Murder Rate for **Michigan**

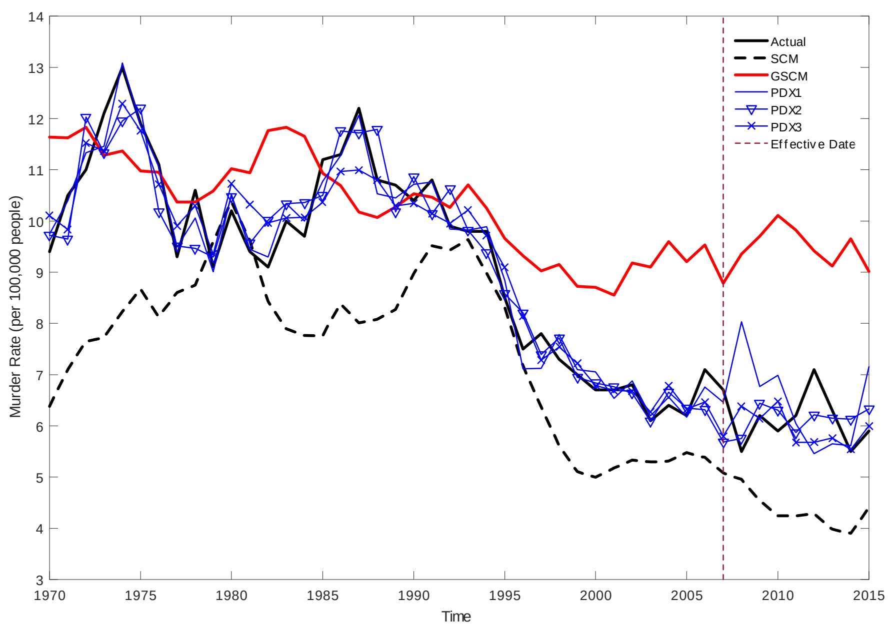


Table A6: Actual and Counterfactual Murder Rate for **Oklahoma** in the Post-treatment Period

Year	Actual	SCM	GSCM	PDX1	PDX2	PDX3
2007	6.1	5.3	8.1	4.9	5.2	5.7
2008	5.8	4.9	8.1	3.4	4.6	5.7
2009	6.3	5.1	8.0	1.6	3.2	5.1
2010	5.2	4.2	9.0	2.9	3.6	5.1
2011	5.6	4.6	8.2	0.8	3.4	5.1
2012	5.8	4.6	9.3	3.2	4.3	5.2
2013	5.1	4.3	10.1	4.4	4.5	5.4
2014	4.6	4.4	8.9	2.9	3.6	5.1
2015	6.1	4.9	8.1	5.4	4.2	5.4
ATE		0.90	-3.07	2.34 (0.12)	1.55 (0.16)	0.30 (0.73)

Note: ATE is calculated over Post-treatment Periods. The number in parenthesis shows p-value of the significance test of the ATE.

Figure A6. Actual and Counterfactual Murder Rate for **Oklahoma**

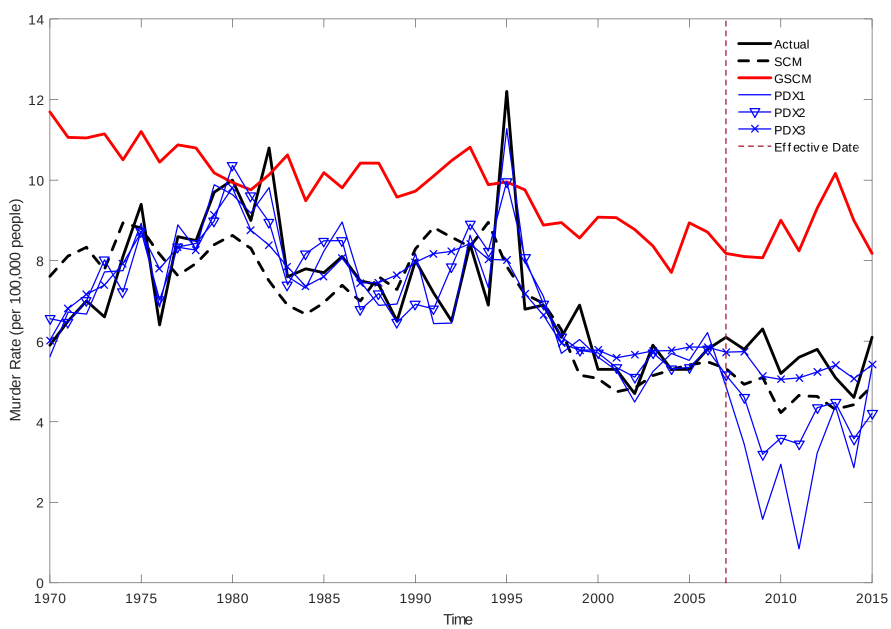


Table A7: Actual and Counterfactual Murder Rate for **South Carolina** in the Post-treatment Period

Year	Actual	SCM	GSCM	PDX1	PDX2	PDX3
2007	8.0	6.6	8.7	6.9	6.7	7.1
2008	6.8	5.7	8.7	9.1	8.8	7.2
2009	6.7	6.0	8.7	7.5	5.8	6.2
2010	5.7	4.6	9.6	6.1	5.4	6.0
2011	6.8	5.3	10.1	6.5	4.6	5.6
2012	7.0	5.7	9.4	6.7	5.4	5.4
2013	6.4	5.1	10.1	5.5	4.5	5.1
2014	6.7	5.6	9.3	7.5	5.9	5.5
2015	8.3	6.0	8.6	9.2	8.9	7.3
ATE		1.25	-2.35	-0.31 (0.70)	0.72 (0.42)	0.77 (0.36)

Note: ATE is calculated over Post-treatment Periods. The number in parenthesis shows p-value of the significance test of the ATE.

Figure A7. Actual and Counterfactual Murder Rate for **South Carolina**

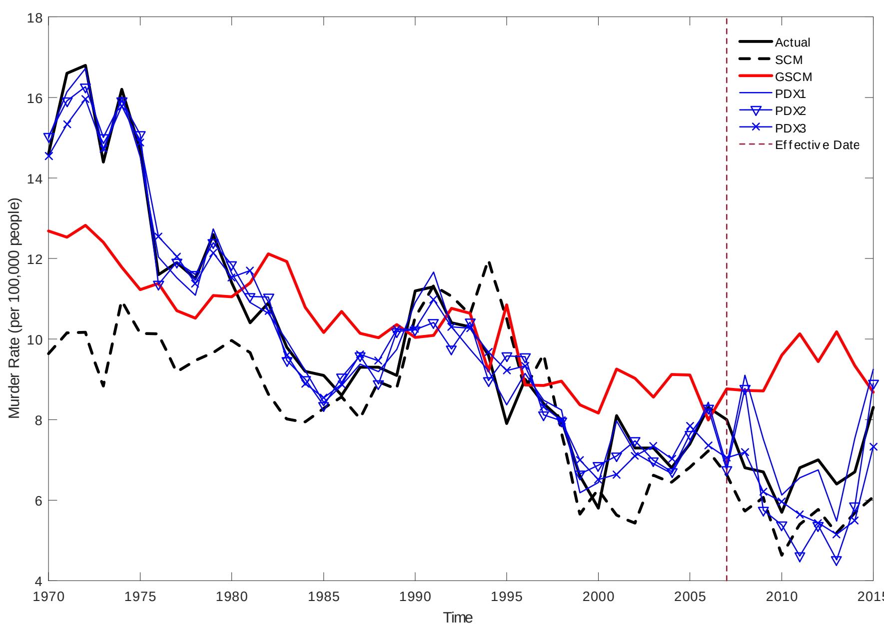


Table A8: Actual and Counterfactual Murder Rate for **South Dakota** in the Post-treatment Period

Year	Actual	SCM	GSCM	PDX1	PDX2	PDX3
2007	2.1	2.4	4.9	2.1	1.3	1.5
2008	4.6	1.3	5.9	0.1	2.2	2.0
2009	3.7	2.3	6.2	0.8	1.4	1.7
2010	2.8	1.8	6.0	3.5	2.0	1.8
2011	2.4	3.6	6.2	1.4	1.9	2.0
2012	2.8	3.8	5.7	1.1	2.0	1.9
2013	2.1	2.5	5.9	2.8	1.9	1.8
2014	2.7	3.3	5.7	2.1	2.2	1.9
2015	3.8	3.1	6.0	2.3	2.8	2.2
ATE		0.28	-2.89	1.20 (0.30)	1.04 (0.32)	1.14 (0.28)

Note: ATE is calculated over Post-treatment Periods. The number in parenthesis shows p-value of the significance test of the ATE.

Figure A8. Actual and Counterfactual Murder Rate for **South Dakota**

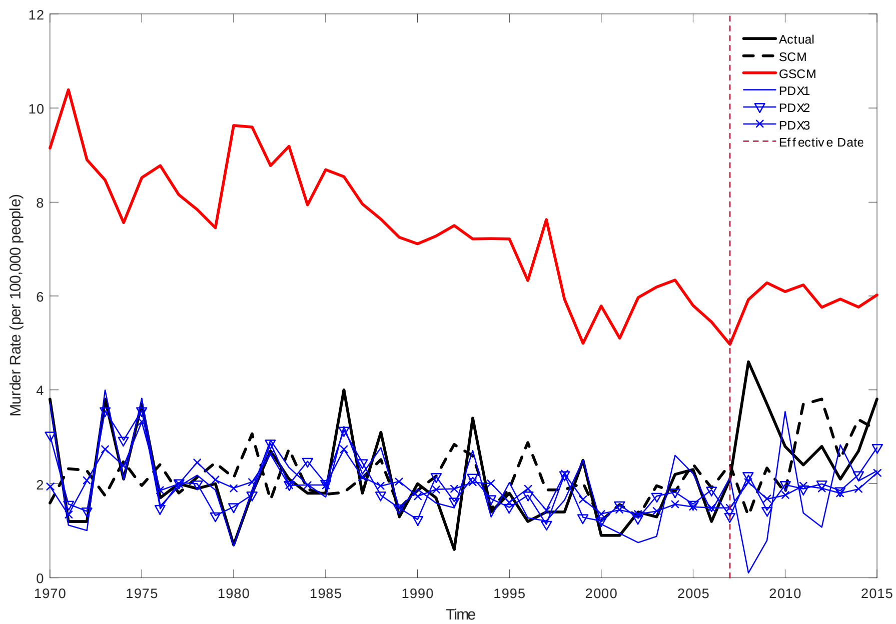


Table A9: Actual and Counterfactual Murder Rate for **Tennessee** in the Post-treatment Period

Year	Actual	SCM	GSCM	PDX1	PDX2	PDX3
2008	6.6	5.6	9.3	4.3	5.5	6.1
2009	7.4	5.7	9.7	4.9	5.2	5.6
2010	5.6	4.5	9.8	2.9	5.5	5.6
2011	6.0	5.1	9.6	3.9	4.7	5.2
2012	6.2	5.4	10.2	4.9	5.5	5.4
2013	5.2	4.9	9.3	5.0	6.0	5.5
2014	5.6	5.3	9.8	3.9	5.2	4.9
2015	6.3	5.7	9.0	6.0	6.7	6.6
ATE		0.78	-3.52	1.62 (0.20)	0.57 (0.57)	0.49 (0.61)

Note: ATE is calculated over Post-treatment Periods. The number in parenthesis shows p-value of the significance test of the ATE.

Figure A9. Actual and Counterfactual Murder Rate for **Tennessee**

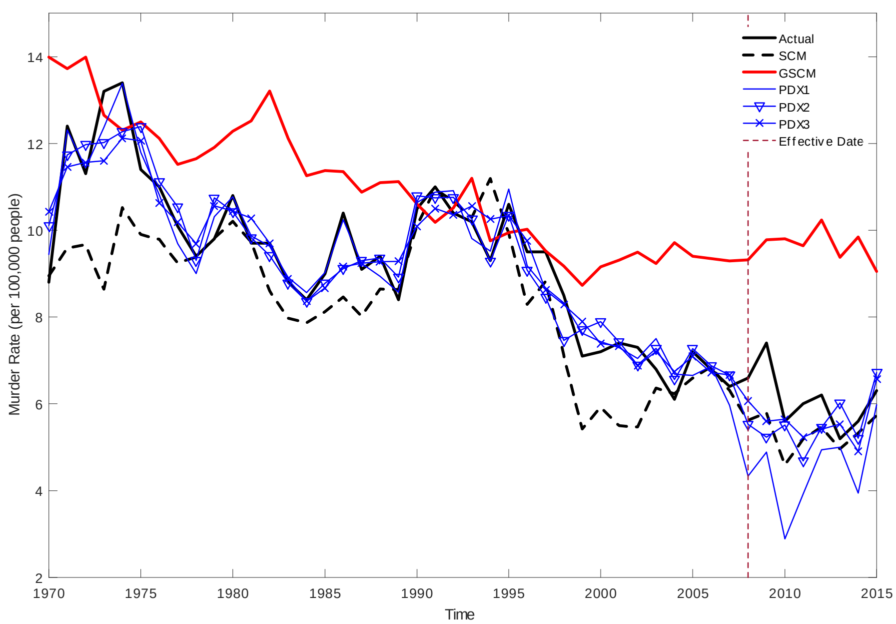


Table A10: Actual and Counterfactual Murder Rate for **Texas** in the Post-treatment Period

Year	Actual	SCM	GSCM	PDX1	PDX2	PDX3
2008	5.6	5.8	11.1	6.1	7.1	6.1
2009	5.4	5.4	11.6	6.6	6.5	6.0
2010	4.9	4.7	11.9	5.8	7.6	6.1
2011	4.4	4.9	11.6	7.5	7.5	6.4
2012	4.4	5.1	11.4	7.7	7.5	6.1
2013	4.3	4.6	11.4	5.6	7.0	5.7
2014	4.4	4.7	11.2	6.3	7.1	5.3
2015	4.8	5.1	10.7	6.0	6.1	4.6
ATE		-0.32	-6.63	-1.70 (0.23)	-2.28 (0.15)	-0.99 (0.43)

Note: ATE is calculated over Post-treatment Periods. The number in parenthesis shows p-value of the significance test of the ATE.

Figure A10. Actual and Counterfactual Murder Rate for **Texas**

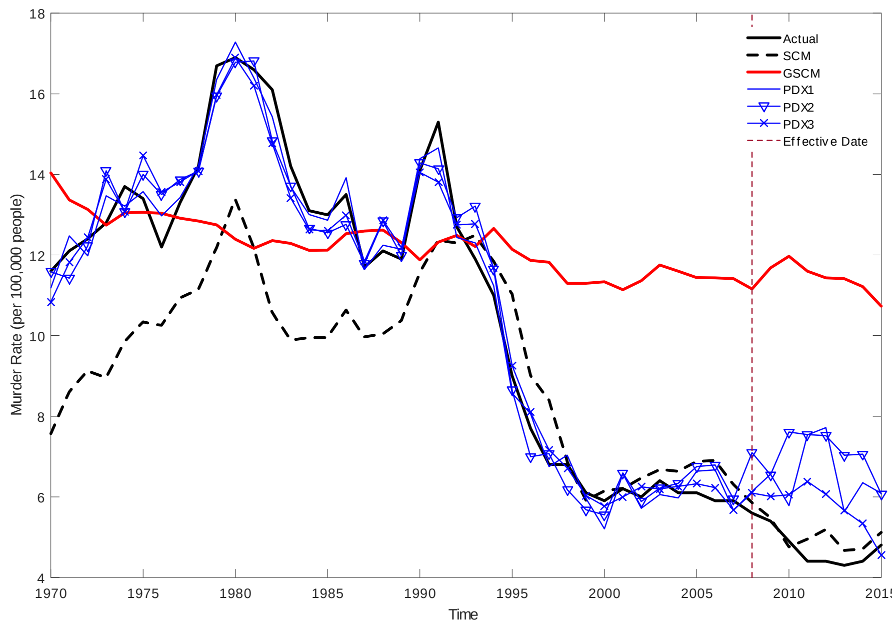


Table A11: Actual and Counterfactual Murder Rate for **West Virginia** in the Post-treatment Period

Year	Actual	SCM	GSCM	PDX1	PDX2	PDX3
2009	4.6	5.3	8.1	3.1	3.7	4.0
2010	3.1	3.9	8.3	2.7	4.0	4.2
2011	4.7	5.1	8.5	3.3	3.3	3.9
2012	3.8	5.4	8.2	2.8	3.4	3.7
2013	3.3	4.6	9.1	2.5	3.3	3.7
2014	4.5	5.3	9.3	3.1	3.6	3.7
2015	4.6	5.6	7.5	4.3	4.3	4.3
ATE		-0.98	-4.38	0.96 (0.51)	0.43 (0.75)	0.16 (0.91)

Note: ATE is calculated over Post-treatment Periods. The number in parenthesis shows p-value of the significance test of the ATE.

Figure A11. Actual and Counterfactual Murder Rate for **West Virginia**