E. J. Ourso College of Business
Department of Economics

DEPARTMENT OF ECONOMICS WORKING PAPER SERIES

Panel Parametric, Semi-parametric and Nonparametric Construction of Counterfactuals - California Tobacco Control Revisited

Cheng Hsiao
University of Southern California

Qiankun Zhou
Louisiana State University

Working Paper 2018-02
http://faculty.bus.lsu.edu/papers/pap18_02.pdf

Department of Economics
Louisiana State University
Baton Rouge, LA 70803-6306
http://www.bus.lsu.edu/economics/

Panel Parametric, Semi-parametric and Nonparametric Construction of Counterfactuals - California Tobacco Control Revisited*

Cheng Hsiao† Qiankun Zhou‡

This version April 15, 2018

Abstract

We consider panel parametric, seme-parametric and nonparametric methods of constructing counterfactuals. Through extensive simulations, no method is able to dominate other methods. In general, we find that if the observed data are stationary, the panel semi-parametric method appears capable of generating counterfactuals close to the (true) data generating process in a wide array of situations. If the data are nonstationary, then the panel nonparametric method appears to dominate the parametric or semiparametric approaches. We also suggest a model averaging method as a robust method to generate counterfactuals. We compare the different estimates of the impact of California Tabcco Control Program on per capita cigarette consumption.

Keywords: Treatment effects, Synthetic control method, Panel data analysis, Interactive effects, Tobacco control, Healthcare expenditure

JEL classification: C01, C21, C31, I18

*We would like to thank Yimeng Xie for able assistance on the simulation. Partial research support by China NSF #71103004 and #71631004 to the first author is also gratefully acknowledged.

†Department of Economics, University of Southern California, University Park, Los Angeles, California 90089, Email: chsiao@usc.edu, Department of Quantitative Finance, National Tsing Hua University and WISE, Xiamen University, China.

‡Department of Economics, Louisiana State University, Baton Rouge, LA 70803. Email: qzhou@lsu.edu.

1 Introduction

The measurement of treatment effect for the i-th unit at time t in a panel data set is the difference between the outcome under the treatment y_{it}^1, and the outcome in the absence of treatment, y_{it}^0, $\Delta_{it} = y_{it}^1 - y_{it}^0$. However, y_{it}^1 and y_{it}^0 are not simultaneously observed. The major challenge in the treatment literature is to construct counterfactuals for the missing y_{it}^1 or y_{it}^0 (e.g., Heckman and Vytlacil (2007a, 2007b)). Panel data containing time series information for a number of individuals may provide information about some individuals before and after intervention and the information about the differences between treated and untreated individuals at a given time. These information can be combined to get around the issues of selection on observables (e.g., Rosenbaum and Rubin (1983))and selection on unobservables (e.g., (Heckman (1979))) without taking a measurement with theory approach or nonconfoundness in nonparametric approach. In this paper, we consider the panel parametric, semiparametric and nonparametric approach for constructing the counterfactuals.

In Section 2, we consider the parametric approach. Section 3 and 4 consider the semiparametric and nonparametric approach, respectively. Section 5 considers a model averaging approach to generate robust estimates of counterfactuals. Section 6 compares the relative performance of three approaches under a variety of data generating processes. Section 7 reexamine the impact of 1988 California Proposition 99 on per capital cigarette consumption and personal healthcare expenditure. Concluding remarks are in Section 8.

2 Panel Parametric Approach

Suppose there are observations $(y_{it}, \mathbf{x}_{it})$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Let the dummy variable d_{it} indicate the ith unit's treatment status at time t with $d_{it} = 1$ if under the treatment and $d_{it} = 0$ if not. The observed data takes the form,

$$y_{it} = d_{it} y_{it}^1 + (1 - d_{it})\, y_{it}^0. \tag{2.1}$$

For ease of exposition, 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.

We assume y_{it}^0 is a function of k observable variables, $\mathbf{x}_{it}$,

$$y_{it}^0 = \mathbf{x}_{it}'\boldsymbol{\beta} + v_{it}, \quad 1 \leq t \leq T, \tag{2.2}$$

where $\boldsymbol{\beta}$ is a $k \times 1$ vector of constants and v_{it} represents the impact of unobserved variables. We decompose v_{it} as the sum of the impacts of those that are due to idiosyncratic factors, u_{it},

and those that are due to r common factors across individuals, $\mathbf{f}_t$. We allow the impacts of $\mathbf{f}_t$ to be different for different individuals, $\boldsymbol{\gamma}_i = (\gamma_{i1}, \ldots, \gamma_{ir})'$, an $r \times 1$ vector of constants over time, $i = 1, 2, \ldots, N$. Thus v_{it} is written as

$$v_{it} = \boldsymbol{\gamma}_i' \mathbf{f}_t + u_{it}, \tag{2.3}$$

Putting the unobserved individual-specific effects, $\boldsymbol{\gamma}_i$, and the common time-specific effects, $\mathbf{f}_t$, in the multiplication form has the advantage over the traditional additive form (e.g., Hsiao (2014)) in that it allows the impact of "globe shock at time t" to be different for different individuals due to the difference in natural endowment or distinct social or technological background. Moreover, the traditional additive form is nested within the multiplicative form (Bai (2009), Hsiao (2017)). The individual-specific effects, $\boldsymbol{\gamma}_i$, and the time-specific effects, $\mathbf{f}_t$, can be treated either as fixed constants (e.g., Bai (2009), Pesaran (2006)) or random (e.g., Sarafidis and Wansbeek (2012)). When $\boldsymbol{\gamma}_i$ and $\mathbf{f}_t$ are treated as constants, there is no need to specify the data generating processes of $\boldsymbol{\gamma}_i$ and $\mathbf{f}_t$, nor their relations with $\mathbf{x}_{it}$. When $\boldsymbol{\gamma}_i$ and $\mathbf{f}_t$ are treated as random variables, specific assumptions about their data generating process and their relations with observed $\mathbf{x}_{it}$ have to be made, although the inference on $\boldsymbol{\beta}$ in general is more efficient for unconditional inference than the conditional inference on $\boldsymbol{\beta}$ conditional on $\boldsymbol{\gamma}_i$ and/or $\mathbf{f}_t$ (e.g., Hsiao (2017)).

Following the tradition, we assume the idiosyncratic effects, u_{it}, are random variables with mean zero and constant variance. We also assume:

Assumption 1: The observed explanatory variables, $\mathbf{x}_{it}$, and the unobserved $(\boldsymbol{\gamma}_i, \mathbf{f}_t)$, are strictly exogenous with respect to u_{it}, i.e.,

$$E\left(u_{it} | \mathbf{x}_{is}, \mathbf{f}_s, \boldsymbol{\gamma}_i\right) = 0. \tag{2.4}$$

Assumption 2: u_{it} is independently distributed over t with $E\left(u_{it} u_{jt}\right) = \sigma_{ij}$ and $\lim_{N\to\infty} \sup_i \frac{1}{N} \sum_{j=1}^{N} |\sigma_{ij}| < \infty$.

Assumption 2 allows u_{it} to be weakly cross-sectional dependent (e.g., Chamberlain and Rothschild (1983), Chudik and Pesaran (2015)). When $\boldsymbol{\beta}$, $\boldsymbol{\gamma}_i$ and $\mathbf{f}_t$ are unknown constants, under assumptions 1 and 2, Bai (2009)'s least squares method are consistent if both N and T are large. Let $\mathbf{X}_i = (\mathbf{x}_{i1}', \ldots, \mathbf{x}_{iT}')'$, under the assumption that

Assumption 3: $f\left(\mathbf{X}_i | d_{it}\right) = f\left(\mathbf{X}_i\right)$ for all $i = 1, 2, \ldots, N$,

we can generate counterfactuals following the steps suggested by Xu (2017):

Step 1: estimate model (2.2) using the observation from the control group only, and obtain $\hat{\boldsymbol{\beta}}_{Bai}$, $\hat{\boldsymbol{\gamma}}_i$ $(i = 2, \ldots, N)$ and $\hat{\mathbf{f}}_t$ $(t = 1, \ldots, T)$.

Step 2: estimate the factor loadings for the treated unit by minimizing the mean squared

error of the predicted treated outcome in the pretreated periods,

$$\begin{aligned}\hat{\boldsymbol{\gamma}}_1 &= \arg\min_{\boldsymbol{\gamma}_1}\left(\mathbf{y}_1-\mathbf{X}_1\hat{\boldsymbol{\beta}}_{Bai}-\hat{\mathbf{F}}\boldsymbol{\gamma}_1\right)'\left(\mathbf{y}_1-\mathbf{X}_1\hat{\boldsymbol{\beta}}_{Bai}-\hat{\mathbf{F}}\boldsymbol{\gamma}_1\right)\\ &= \left(\hat{\mathbf{F}}'\hat{\mathbf{F}}\right)^{-1}\hat{\mathbf{F}}'\left(\mathbf{y}_1-\mathbf{X}_1\hat{\boldsymbol{\beta}}_{Bai}\right),\end{aligned}$$

where $\mathbf{y}_1=(y_{11},\ldots,y_{1T_0})'$, $\hat{\mathbf{F}}=\left(\hat{\mathbf{f}}_1,\ldots,\hat{\mathbf{f}}_{T_0}\right)'$ and $\mathbf{X}_1=(\mathbf{x}_{11},\ldots,\mathbf{x}_{1T_0})'$.

Step 3: estimate counterfactuals based on $\hat{\boldsymbol{\beta}}_{Bai}$, $\hat{\mathbf{F}}$ and $\hat{\boldsymbol{\gamma}}_1$ as

$$\hat{y}_{1t}^0=\mathbf{x}_{1t}'\hat{\boldsymbol{\beta}}_{Bai}+\hat{\boldsymbol{\gamma}}_1'\hat{\mathbf{f}}_t,\quad T_0+1\le t\le T. \tag{2.5}$$

Consequently, an estimator for the treatment effects on the treated unit (ATT) at time t is

$$ATT_t=y_{1t}-\hat{y}_{1t}^0,\quad \text{for } t>T_0.$$

3 Semi-parametric Method

For the idiosyncratic error u_{1t}, we note that although $E(u_{1t})=0$, $Var(u_{1t})\ge Var(u_{1t}|\tilde{\mathbf{u}}_t)$ where $\tilde{\mathbf{u}}_t=(v_{2t},\ldots,v_{Nt})'$. Let

$$u_{1t}=\boldsymbol{\sigma}_{1\tilde{u}}'\Sigma_{\tilde{u}}^{-1}\tilde{\mathbf{u}}_t+\eta_t, \tag{3.1}$$

where $\boldsymbol{\sigma}_{1\tilde{u}}=E(u_{1t}\tilde{\mathbf{u}}_t)$ and $\Sigma_{\tilde{u}}=E(\tilde{\mathbf{u}}_t\tilde{\mathbf{u}}_t')$. Noticing that

$$\tilde{\mathbf{u}}_t=\tilde{\mathbf{y}}_t-\tilde{\mathbf{X}}_t\boldsymbol{\beta}-\tilde{\Gamma}\mathbf{f}_t, \tag{3.2}$$

where $\tilde{\mathbf{y}}_t=(y_{2t},\ldots,y_{Nt})'$, $\tilde{\mathbf{X}}_t=(\mathbf{x}_{2t},\ldots,\mathbf{x}_{Nt})'$ and $\tilde{\Gamma}=(\boldsymbol{\gamma}_2,\ldots,\boldsymbol{\gamma}_N)'$.

Under the assumption that

Assumption 4, $rank\left(\tilde{\Gamma}\right)=r$,

and

Assumption 5, $\mathbf{y}_i\perp d_{1t}$ for $i=2,\ldots,N$, where $\mathbf{y}_i=(y_{i1},\ldots,y_{iT})'$,

then $\mathbf{f}_t$ can be represented as

$$\mathbf{f}_t=\left(\tilde{\Gamma}'\tilde{\Gamma}\right)^{-1}\tilde{\Gamma}'\left(\tilde{\mathbf{y}}_t-\tilde{\mathbf{X}}_t\boldsymbol{\beta}-\tilde{\mathbf{u}}_t\right). \tag{3.3}$$

Substituting (3.1) and (3.3) into (2.2) for the first unit yields

$$y_{1t}^0=\mathbf{x}_{1t}'\boldsymbol{\beta}+\left(\boldsymbol{\gamma}_1'\left(\tilde{\Gamma}'\tilde{\Gamma}\right)^{-1}\tilde{\Gamma}'+\boldsymbol{\sigma}_{1\tilde{u}}'\Sigma_{\tilde{u}}^{-1}\left(\mathbf{I}_{N-1}-\tilde{\Gamma}\left(\tilde{\Gamma}'\tilde{\Gamma}\right)^{-1}\tilde{\Gamma}'\right)\right)\left(\tilde{\mathbf{y}}_t-\tilde{\mathbf{X}}_t\boldsymbol{\beta}\right)+\eta_t^*, \tag{3.4}$$

where

$$\eta_t^*=\eta_t-\left(\boldsymbol{\gamma}_1'+\boldsymbol{\sigma}_{1\tilde{u}}'\Sigma_{\tilde{u}}^{-1}\tilde{\Gamma}\right)\left(\tilde{\Gamma}'\tilde{\Gamma}\right)^{-1}\tilde{\Gamma}'\tilde{\mathbf{u}}_t. \tag{3.5}$$

Instead of estimating (3.4), we suggest a data driven method to approximate y_{1t}^0,

$$y_{1t}^0 = \mathbf{x}_{1t}'\boldsymbol{\beta} + \mathbf{w}'\left(\tilde{\mathbf{y}}_t^* - \tilde{\mathbf{X}}_t^*\boldsymbol{\beta}\right) + \mu + \epsilon_{1t}, \tag{3.6}$$

where $\left(\tilde{\mathbf{y}}_t^* - \tilde{\mathbf{X}}_t^*\boldsymbol{\beta}\right)$ could be a subset of $\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\boldsymbol{\beta}\right)$ and $(\mu, \mathbf{w}')$ are unrestricted. Thus, we suggest the following steps to generate counterfactuals:

Step 1: Use pretreatment data and Bai (2009) or Pesaran (2006) method to estimate $\boldsymbol{\beta}$, denoted by $\tilde{\boldsymbol{\beta}}$.

Step 2: Conditional on $\tilde{\boldsymbol{\beta}}$, obtain μ and $\mathbf{w}$ by minimizing

$$\min_{\mu,\mathbf{w}} \sum_{t=1}^{T_0} \left[y_{1t} - \mathbf{x}_{1t}'\tilde{\boldsymbol{\beta}} - \mathbf{w}'\left(\tilde{\mathbf{y}}_t^* - \tilde{\mathbf{X}}_t^*\boldsymbol{\beta}\right) - \mu\right]^2, \tag{3.7}$$

When N is large, the subset $\left(\tilde{\mathbf{y}}_t^* - \tilde{\mathbf{X}}_t^*\boldsymbol{\beta}\right)$ can be chosen using some model selection criterion as in Hsiao et al. (2012).

Step 3: Generate counterfactuals by

$$\tilde{y}_{1t}^0 = \mathbf{x}_{1t}'\tilde{\boldsymbol{\beta}} + \hat{\mathbf{w}}'\left(\tilde{\mathbf{y}}_t^* - \tilde{\mathbf{X}}_t^*\boldsymbol{\beta}\right) + \hat{\mu}, \quad T_0 + 1 \le t \le T. \tag{3.8}$$

Remark 3.1 *The parametric approach needs both N and T large to get reliable estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}_1$ and $\mathbf{f}_t$, which is a luxury to meet in many applications. In many applications, T is finite, but N could be large. In this case, we can consider using the Pesaran (2006) common correlated effects (CCE) estimator to get consistent estimator of $\boldsymbol{\beta}$, namely, $\hat{\boldsymbol{\beta}}_{CCE}$. Under the assumption that $\mathbf{x}_{it}$ is generated by the same set of unobservable common factors and certain rank condition (e.g., Pesaran (2006)), the CCE estimator $\boldsymbol{\beta}$ is give by*

$$\hat{\boldsymbol{\beta}}_{CCE} = \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\bar{\mathbf{Z}}}\mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\bar{\mathbf{Z}}}\mathbf{y}_i, \tag{3.9}$$

where $\mathbf{M}_{\bar{\mathbf{Z}}} = \mathbf{I}_{T_0} - \bar{\mathbf{Z}}\left(\bar{\mathbf{Z}}'\bar{\mathbf{Z}}\right)^{-1}\bar{\mathbf{Z}}'$ with $\bar{\mathbf{Z}} = (\bar{\mathbf{z}}_1, \ldots, \bar{\mathbf{z}}_{T_0})'$ and $\bar{\mathbf{z}}_t = \frac{1}{N}\sum_{j=1}^{N}\mathbf{z}_{jt} = \frac{1}{N}\sum_{j=1}^{N}(y_{it}, \mathbf{x}_{it}')'$. Following either Pesaran (2006) or Zhang and Zhou (2016), it can be shown that the CCE estimator of $\boldsymbol{\beta}$ is consistent as long as $N \to \infty$.

Remark 3.2 *The advantage of generating counterfactuals by (3.8) is that we can use the regression method to obtain μ and $\mathbf{w}$ to avoid the infinite dimensional issues associated with "distributional free" method (e.g., Li and Racine (2007)). It is simple to implement and easy to construct the confidence band for "treatment effects" (e.g., Fujiki and Hsiao (2015)).*

4 Nonparametric Method

4.1 Synthetic Control Method (SCM)

Abadie et al. (2010) has proposed to predict y_{1t} by

$$\hat{y}_{1t}^{*} = \mathbf{w}'\tilde{\mathbf{y}}_t = \sum_{i=2}^{N} w_i y_{it}, \quad T_0 + 1 \leq t \leq T, \tag{4.1}$$

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

$$\sqrt{(\mathbf{M}_1 - \mathbf{M}_0\mathbf{w})' \mathbf{V} (\mathbf{M}_1 - \mathbf{M}_0\mathbf{w})}, \tag{4.2}$$

subject to

$$y_{1t} = \sum_{i=2}^{N} w_i y_{it}, \quad 1 \leq t \leq T_0, \tag{4.3}$$

$$\bar{\mathbf{x}}_{1k} = \sum_{i=2}^{N} w_i \bar{\mathbf{x}}_{ik}, \quad 1 \leq k \leq K, \tag{4.4}$$

and

$$w_i \geq 0 \ \text{ and } \ \sum_{i=2}^{N} w_i = 1, \tag{4.5}$$

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.

4.2 Panel Data Approach (PDA)

Under the assumption that y_{jt} and $\mathbf{x}_{it}$ are independent of d_{1t}, Hsiao et al. (2012) propose to predict y_{1t}^0 by

$$\tilde{y}_{1t}^{*} = \mu + \boldsymbol{\delta}'\mathbf{z}_t^{*}, \quad T_0 + 1 \leq t \leq T, \tag{4.6}$$

where $\mathbf{z}_t^{*}$ includes any $(y_{jt}, \mathbf{x}_{1t})$ that helps to predict y_{1t}^0.

Remark 4.1 *The emphasis of the 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 (constraints (4.3)-(4.5)). On the other hand, the PDA puts the emphasis on the accuracy of predicting the outcomes of the treatment unit. It does not require the predictors to be similar to the treatment units as long as they help prediction. The only requirement is that the control units*

are independent of the treatment (Assumption 3 and 4)[1]*. In the regression framework, the PDA is an unconstrained regression while the SCM is a constrained regression. When the constraints are valid, SCM is a more efficient method. When the constraints are not valid, SCM could lead to biased prediction of counterfactuals (Wan et al. (2018)).*

5 Model Averaging

The above sections discussed parametric, semi-parametric and nonparametric ways of constructing counterfactuals. Each has its advantages and disadvantages. Unfortunately, neither the conventional hypothesis testing approach, nor the predictive approach (e.g., Diebold and Mariano (1995), White (2000)) appears feasible to assess the appropriateness of which method is more likely to generate more accurate counterfactuals in a given situation because counterfactuals are unobservable. Bates and Granger (1969) have argued that even the most complicated model is likely to be misspecificed and combining forecasting across different models is a way to make the forecast more robust against misspecification biases and measurement errors in the data. Many authors have suggested different methods to combine forecasts (e.g., see the survey by Timmermann (2006)). Most of these methods depend on the relation between the actuals and forecasts while in our case the actuals are unobservable. On the other hand, the simulation and an empirical example conducted by Hsiao and Wan (2012) appear to indicate that no method is able to yield more accurate forecasts uniformly over time in a wide array of situations. A mean or a mean and scale correlated simple average appear to be a robust way to combine forecasts. Therefore, we suggest the following two ways to combine the different methods of generating counterfactuals.

Let $\bar{y}_t = \frac{1}{M}\sum_{j=1}^{M}\hat{y}_{jt}$ where $\hat{y}_{jt}$ denote the within sample or post-sample predicted value of y_{1t} based on j-th method.

(M1) Mean corrected simple average method

$$\hat{y}_{1t} = a + \bar{y}_t, \quad t = T_0 + 1, \ldots, T, \tag{5.1}$$

where a

$$a = \frac{1}{T_0}\sum_{t=1}^{T_0}\left(y_{1t} - \bar{y}_t\right), \quad t = 1, \ldots, T_0. \tag{5.2}$$

(M2) Mean and scaled corrected simple average method

$$\hat{y}_{1t} = a + b\bar{y}_t, \quad t = T_0 + 1, \ldots, T, \tag{5.3}$$

[1]The SCM also needs these two assumptions to ensure the generated predictions are ubiased if (4.5) holds.

where a and b are the least squares estimates of y_{1t} on the constant and $\bar{y}_t$ for the pretreatment period.

6 Simulation Studies

The treatment effects is measured as the difference between y_{it}^1 and the predicted y_{it}^0. Since the true data generating process (DGP) is unknown, the only way to consider which method is more likely to yield more accurate y_{it}^0 in a wide array of situations is through computer simulations. In the DGPs below, we assume the common factors f_{1t}, f_{2t} and f_{3t} are $iidN(0,1)$ unless they are specified in the DGP, the factor loadings $\gamma_{1,i}$, $\gamma_{2,i}$ and $\gamma_{3,i}$ are also $iidN(0,1)$. The coefficients are set as $\beta_1 = 1$, and $\beta_2 = 2$. The DGPs are designed as follows,

DGP1: Model with exogenous variables and common factors

$$y_{it} = x_{1,it}\beta_1 + x_{2,it}\beta_2 + \gamma_{1,i}f_{1t} + \gamma_{2,i}f_{2t} + \gamma_{3,i}f_{3t} + u_{it}. \tag{6.1}$$

The covariates $x_{k,it}$ $(k = 1, 2)$ are (positively) correlated with a subset of factors as follows

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

where $\rho_{k,i} \sim iidU\left(0.1, 0.9\right)$, c_{1i} and c_{2i} are $iidU\left(1, 2\right)$ and the error term η_{it} is $iid\left(\chi^2(1) - 1\right)$.

DGP2: Model with exogenous variables and common factors

$$y_{it} = x_{1,it}\beta_1 + x_{2,it}\beta_2 + \gamma_{1,i}f_{1t} + \gamma_{2,i}f_{2t} + u_{it}. \tag{6.2}$$

The covariates $x_{k,it}$ $(k = 1, 2)$ are (positively) correlated with the factors and extra factors as follows

$$x_{k,it} = 1 + \rho_{ki}x_{k,it-1} + \sum_{j=1}^{3} c_j f_{jt} + \eta_{k,it}, \quad k = 1, 2,$$

where $\rho_{k,i} \sim iidU\left(0.1, 0.9\right)$, $c_j \sim iidU\left(1, 2\right)$ and the error term η_{it} is $iid\left(\chi^2(1) - 1\right)$.

DGP3: Model with exogenous variables and common factors

$$y_{it} = x_{1,it}\beta_1 + x_{2,it}\beta_2 + \gamma_{1,i}f_{1t} + \gamma_{2,i}f_{2t} + u_{it}. \tag{6.3}$$

The covariates $x_{k,it}$ $(k = 1, 2)$ follow an ARMA process as

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

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

DGP4: DGP1 with two cointergrated factors

$$\begin{cases} f_{1t} = 0.5 f_{2t} + \xi_{1t} \\ f_{2t} = f_{2t-1} + \xi_{2t} \end{cases}, \tag{6.4}$$

where ξ_{1t} and ξ_{2t} are $iidN(0,1)$.

DGP5: Cointegrated models

$$\begin{aligned} y_{1t} &= \sum_{j=2}^{N} y_{jt}\beta_j + u_{1t}, \\ y_{jt} &= y_{jt-1} + v_{jt}, \\ \beta_j &\sim iidU\left(0,1\right). \end{aligned} \tag{6.5}$$

DGP6: Pure factor model

$$y_{it} = \gamma_{1,i} f_{1t} + \gamma_{2,i} f_{2t} + u_{it}. \tag{6.6}$$

with f_{1t} and f_{2t} are $iidN(0,1)$.

DGP7: Pure factor model (6.6) with f_{1t} and f_{2t} following random walk processes.

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

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

where σ_i^2 being random draw from $0.5\left(\chi^2\left(1\right)+1\right)$ and we let $b = 1$.

DGP 1 and 2 are parametric specifications with the DGP for x satisfying the Pesaran (2006) CCE condition, i.e., $\bar{z}_t$ captures the variation of f_{1t} and f_{2t}. DGP 3 is a parametric specification with the DGP for x independent of f_t (i.e., the Pesaran CCE does not capture the cross-sectional dependence). DGP 4 is also a parametric specification with one of the two factors following a unit root process. DGP 5 is a cointegrated model. DGP 6 is a factor model with stationary factors. DGP 7 is also a factor model with factors following random walk processes. However, since $r = 2$, y_{it} are cross-sectional cointegrated.

The treatment and control groups consist of 1 and $N-1$ units. The treatment on unit 1 starts at time $T_0 + 1$. The other $(N-1)$ units are not subject to treatment. We let $N-1$ be $30, 50$ and the pretreatment time $T_0 = 30, 50$, and post treatment periods $T - T_0 = 10$, i.e., $T = 40, 60$. The number of replication is set at $R = 1000$.

We consider several estimators for the above DGPs[2,3],

[2]In the simulation and estimation of (generalized) synthetic control method and PDA approach, we use the "`gsynth`" package by Xu and Liu (2017), and the "`pampe`" package by Vega-Bayo (2015) in R, respectively.

[3]We do not consider SCM in the simulation and estimation because our data generating processes do not

(**E1**) **PCA**: Estimate model (2.2) by Bai (2009)'s least squares method, then generate $\hat{y}_{1t}$ by (2.5).

(**E2**) **PCCE**: Estimate model (2.2) by Pesaran (2006)'s CCE method, then estimate $\boldsymbol{\gamma}_i$ and $\mathbf{f}_t$ by principle component (PC) method, and generate $\hat{y}_{1t}$ by (2.5).

(**E3**) **CCE**: Using the CCE method to estimate $\boldsymbol{\beta}$, then generate $\hat{y}_{1t}$ by using (3.8) through (3.7).

(**E4**) **CPDA**: Using the CCE method to estimate $\boldsymbol{\beta}$, then use the PDA approach to select a subset of $\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\boldsymbol{\beta}\right)$ as predictors.

(**E5) PDA:** Use Hsiao et al. (2012) PDA method to generate predictor for $T_0+1,\ldots,T$ using the observations on $\tilde{\mathbf{y}}_t$ $(t=1,\ldots,T_0)$ only.

(**E6) PDAX:** Use Hsiao et al. (2012) PDA method to generate predictor for $T_0+1,\ldots,T$ using the observations on $\mathbf{x}_{1t}$ and $\tilde{\mathbf{y}}_t$ $(t=1,\ldots,T_0)$.

(**E7) M1AV6:** Use the mean corrected simple average of (E1)-(E6).

(**E8) M1AV5:** Use the mean corrected simple average of (E1) and E(3)-(E6).

(**E9) M2AV6:** Use the mean and scaled corrected simple average of (E1)-(E6).

(**E10) M2AV5:** Use the mean and scaled corrected simple average of (E1) and E(3)-(E6).

For the above parametric estimators E1 and E2, we assume the dimension of $\mathbf{f}_t$, r, is unknown. We estimate r following Xu (2017)'s cross-validation method, then implement E1 and E2. For CPDA, PDA and PDAX, we use a model selection procedure to select a subset of control units to generate counterfactuals. Model selelction procedure performs well if $T_0 > N$. If N is greater than T_0, it does not work well (e.g., Wan et al (2018)). Therefore, when N is large relative to T_0 ($N = 50$), we suggest the following two methods to select control units as predictors:

(**Ma**) Combining stepwise regression method with model selection method to select predictors:

Step 1: Run the regression of y_{1t}^0 on each of the control units. Select the one that yields the highest R^2, say y_{jt}.

Step 2: Run the regression of y_{1t}^0 on y_{jt} and each of the remaining control units. Select the pair that yields the highest R^2, say y_{jt} and y_{lt}. Use the model selection criterion to decide whether one should just use y_{jt} as the predictor or one should use both y_{jt} and y_{lt} as predictors. If the model with y_{jt} only is selected, then stop the process.

Step 3: If (y_{jt}, y_{lt}) are chosen as predictors, run the regression of y_{1t}^0 on y_{jt} and y_{lt} and each of the remaining control units, choose the triplet (y_{jt}, y_{lt}, y_{mt}) as the predictors and compare the

satisfy (4.3)-(4.5). Also, the simulations conducted by Wan et al. (2018) appear to favor PDA in a wide array of situations.

performance of the model using (y_{jt}, y_{lt}, y_{mt}) with the model using (y_{jt}, y_{lt}) only as predictors. If the former model is selected, stop the process. If the latter is selected, continue the process until the model consists of k predictors is chosen over the model consists of $(k+1)$ predictors by some model selection criterion.

(**Mb**) Random Splitting Method

Step 1: Randomly split up the control units into G mutually exclusive subset: $\mathbf{Z}_1, \ldots, \mathbf{Z}_G$.

Step 2 For each subsets, use some model selection method, say AIC or AICC, to select the best predictors within each group, say, $\mathbf{Z}_1^*, \ldots, \mathbf{Z}_G^*$.

Step 3: Pool $\mathbf{Z}_1^*, \ldots, \mathbf{Z}_G^*$ to form a new set of control units, $\tilde{\mathbf{Z}}$. If the dimension of $\tilde{\mathbf{Z}}$ is smaller than T_0, then apply some model selection criterion to select the best predictors from $\tilde{\mathbf{Z}}$. If the dimension of $\tilde{\mathbf{Z}}$ is still close to T_0, repeat the steps on $\tilde{\mathbf{Z}}$.

For the DGP considered here, we find method (Ma) on average performs marginally better than method (Mb) when y_{it} is stationary. When y_{it} is nonstationary, method (Mb) performs substantially better.[4]

We consider three criterion for comparison: the mean of absolute error for the true observation and the counterfactuals at each post treatment date point (*MAB*), the mean of the sum of squared error for the true observation and the counterfactuals at each post treatment date point (*MSE*), and the mean of the ratio of absolute counterfactuals and absolute true outcomes at each date point in the treatment period (*MAP).* We consider the performance of constructing the counterfactuals of y_{1t} $(t = T_0 + 1, \ldots, T)$ by using the approaches E1-E10. For $N = 50$, when y_{it} is stationary, we use the stepwise method to select predictors. When y_{it} is nonstationary, we use the random split method by first splitting the sample randomly into two groups $(G = 2)$. The simulation results are summarized in Table 1-7. We also plot the root mean square errors of different methods for DGP 1-7 when $N = 30$ and $T = 40$ in Figure 1-7. In general, we find that

(i) When using parametric form (2.5) as predictors, the method based on Bai (2009) outperforms the Pesaran (2006) CCE method of first estimating the slope coefficient of $\mathbf{x}$, then applying the PC method to obtain $\boldsymbol{\gamma}_i$ (PCCE) .

(ii) When using the semi-parametric form (3.6) to generate counterfactuals, using the complete set of $\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\boldsymbol{\beta}\right)$ yields less accurate predictors than using a subset of $\left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t\boldsymbol{\beta}\right)$ through some model selection criterion (CCE vs CPDA).

(iii) When using the nonparametric form to generate counterfactuals, applying Hsiao et al. (2012) method based on the extended control pool consisting of both y_{jt}, and the treatment unit strictly exogenous variables (PDAX) outperforms the application of Hsiao et al. (2012)

[4]Details available upon request.

method with control pool consists of y_{jt} only.

(iv) When data are stationary and $T_0 \geq N$, then the semiparametric predictor (3.8) generates more accurate counterfactuals than both the parametric and nonparametric predictors, often by substantial margin except in the case when $N > T_0$. When $(y_{1t}, \ldots, y_{Nt})$ are non-stationary and cross-sectionally cointegrated, the nonparametric method appears to dominate. However, even in this case, the semi-parametric predictor (3.8) performs reasonably well and still dominates the parametric predictor (2.5).

(v) No method is able to dominate all other methods in all different data generating processes all the times. Model averaging method appears to be a robust way to construct counterfactuals for a wide array of situations.

7 California's Tobacco Control Program (CTCP) Revisited

In November 1988, California passed the Proposition 99 that increased California's cigarettes tax by 25 cents per pack and earmarked the tax revenue to health and anti-smoking measures. Proposition 99 also trigged a wave of local clean-air ordinances in California. Abadie et al. (2010) have shown that there was a significant impact of California Tobacco control program (CTCP) on per capita cigarette consumption for the period 1970-2000 and its impact continued to enhance over time using the synthetic control method (SCM). In this section, we revisit the effectiveness of CTCP on per capita cigarettes consumption and on personal healthcare expenditure using the panel parametric, semi-parametric and nonparametric methods discussed in this paper.

The observable factors that are likely to affect cigarettes consumption or personal healthcare expenditure are per capita GDP, cigarettes price, demographic composition, etc.. However, cigarettes price etc., could be affected by cigarettes tax, so we shall only use per capita GDP as observable exogenous factor and treat all others as unobservable factors in the counterfactual analysis.

For the observed per capita cigarette consumption or healthcare expenditure, we follow Abadie et al. (2010) to exclude the four states (Massachusetts, Arizona, Oregon, and Florida) that adopted some other large-scale tobacco control program and seven states (Alaska, Hawaii, Maryland, Michigan, New Jersey, New York, Washington) that have raised their state cigarette taxes by 50 cents or more over the 1989 to 2000 period. The District of Columbia is also excluded from our sample. Thus, our control group includes 38 states. The data is collected between 1970-2000 for cigarette consumption and 1980-2000 for personal healthcare expenditure.[5]

[5] We didn't include the extra variables as in Abadie et al (2010) because they contain lots of observations are

In our sample, $N\,(=38)$ is greater than T_0 ($T_0 = 19$ for cigarette consumption, $T_0 = 9$ for personal healthcare expenditure). Technically, it is hard to implement CCE. For CPDA, PDA and PDAX, we use the random split method discussed in Section 6 to select the subset of units to generate counterfactuals.

Table 8 provides the actual and estimated counterfactuals for cigarettes consumption per capita based on (E1)-(E10) excluding PCCE, M1AV6 and M2AV6.[6] Figure 8 plots the difference between the actual outcomes and the counterfactuals excluding those from PCCE, and replace M1AV6 and M2AV6 by M1AV5 and M2AV5, which are averages of PCA, CCE, CPDA, PDA and PDAX. As one can see, in general, CCE provides the lower bounds of the counterfactuals, the PDAX and PDA provide the upper bound, and the CPDA (the method that appears to yield counterfactuals closest to the unknown data generating process in our simulation) yields counterfactuals are in the middle. Thus, according to CCE, there was a small discouraging impact of 1988 tax increase and its impact tends to stay constant over time. On the other hand, based on PDA and PDAX, the CTCP has the expected discouraging effect on cigarette consumption and its impact appears to aggravate over time. The average of different methods' estimates indicates that the CTCP does have a negative impact on cigarette consumption. However, the negative effects appear to be much smaller absolute magnitude than those obtained by PDA, PDAX or SCM of Abadie et al (2010) (e.g., often less than half).

Since counterfactuals are unobservable, it is hard to know which estimates are close to the true treatment effects. We therefore propose to use indirect evidence to gauge which estimates could be closer to the true treatment effects. From the established medical link between cigarette consumption and lung cancer, and the link between cancer treatment and healthcare expenditure, we assume studying California residents' personal healthcare expenditure could shed light on this. To avoid the outcomes being contaminated by the treatment, again, like the study of cigarette consumption, we only use the personal healthcare expenditure for the 39 states from 1980-2000 and include per capita income as an additional control variable.[7]

Table 9 and Figure 9 provides the actual and estimated counterfactuals for personal healthcare expenditure based on (E1)-(E10) and their plots during the treatment period. As we can see from Table 9, all different methods yield similar counterfactuals. They appear to show that there is no significant decrease in personal healthcare expenditure.

missing and we are unable to recover the missing observations. The data of personal healthcare expenditure is collected from National Health Service (NHS) website.

[6] The reason of exclusion of these three method is that the PCCE yields unbelivably high estimate of the coefficient of per capita income. The detailed results of PCCE, M1AV6 and M2AV6 are available upon request.

[7] We exclude sales of cigarette and retail prices of cigarettes as conventionally assumed because they are likely to be affected by treatment.

Using the healthcare expenditure information as corroborating evidence, it appears that although there appears a discouraging effect of the CTCP on per capita cigarette consumption, contrary to the common belief, their impact and their absolute magnitude could be less than conventionally estimated.

8 Concluding Remarks

Using observed data to empirically estimate the treatment effects is challenging. Techniques of statistical analysis are based on certain *maintained* hypothesis. In this paper, we have considered panel parametric, seme-parametric and nonparametric methods of constructing counterfactuals. Our simulation results show that if the observed data are stationary, the panel semi-parametric method appears capable of generating counterfactuals close to the (true) data generating process in a wide array of situations. If the data are nonstationary, then the panel nonparametric method appears to dominate the parametric or semiparametric approaches. However, no method appears capable of dominating other methods in different data generating processes and different sample configurations of cross-sectional dimension N and pre-treatment time dimension T_0. Since the true data generating process is usually unknown and statistical findings could be very different for different situations, we have also suggested a model averaging method as a robust method to generate counterfactuals.

We have reexamined the impact of California Tobacco Control Program on per capita cigarette consumption using the methods studied in this paper. Our results show that the estimates are sensitive to the method used. However, using the model averaging method and the study of per capita cigarette consumption as corroborating evidence, it appears that the impact of the California 1988 proposition could be over estimated. There was a moderate discouraging effects, but less than half of the conventionally estimated.

Statistical analysis is not a proof of reality. Information contained in the data may be limited. Statistical inference could be fragile and sensitive to inferential procedures. We have to think of as many consequences of the underlying assumptions as possible. We must report our findings with a great deal of humility. We have to realize we are still in the process of groping toward the truth, not discovering the truth.

References

[1] Abadie A, Diamond A, and J. Hainmueller, 2010, Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. American Statistical Association 105: 493–505.

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

[3] Bates, J.M., and C. Granger, 1969, The Combination of Forecasts. Journal of the Operational Research Society 20, 451-468.

[4] Chamberlain, G., and M. Rothschild, 1983, Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets, Econometrica 51, 1281-1304.

[5] Chudik, C., and M.H. Pesaran, 2015, Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors, Journal of Econometrics 188, 393-420.

[6] Diebold, F. and R. Mariano, 2002, MarianoComparing Predictive Accuracy, Journal of Business & Economic Statistics 20, 134-144.

[7] Fujiki, H., and C. Hsiao, 2015, Disentangle Multiple Treatment Effects-Measuring the Net Economic Impact of the 1995 Great Hanshin-Awaji Earthquake, Journal of Econometrics 186, 66-73.

[8] Gardeazabal J, and A. Vega-Bayo, 2016. An empirical comparison between the synthetic control method and Hsiao et al.'s panel data approach to program evaluation. Journal of Applied Econometrics. DOI: 10.1002/jae.2557.

[9] Heckman, J., 1979, Sample selection bias as a specification error. Econometrica 47, 153-161.

[10] 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,

[11] 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,

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

[13] Hsiao, C., 2017, Panel Models with Interactive Effects, Journal of Econometrics, forthcoming.

[14] 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.

[15] Li, Q. and J. Racine, Nonparametric Econometrics: Theory and Practice, Princeton University Press, NJ, USA

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

[17] Rosenbaum, P., and D. Rubin, 1983, The Central Role of the Propensity Score in Observational Studies for Causal Effects, Biometrika 70, 41-55.

[18] Sarafidis, V., and T. Wansbeek, 2012, Cross-Sectional Dependence in Panel Data Analysis, Econometric Reviews 31, 483-531.

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

[20] Timmermann, A., 2006, Forecast Combinations, Chapter 04 in Handbook of Economic Forecasting Vol 1, 135-196, Elliott, G., Granger, C. and Timmermann, A. eds., Elsevier.

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

[22] Wan, S-K., Y. Xie and C. Hsiao, 2018, Panel Data Approach vs Synthetic Control Method, Economics Letters 164, 121-123.

[23] White, B., 2000, The Importance of Climate Variability and Seasonal Forecasting to the Australian Economy. In: Hammer G.L., Nicholls N., Mitchell C. (eds) Applications of Seasonal Climate Forecasting in Agricultural and Natural Ecosystems. Atmospheric and Oceanographic Sciences Library, vol 21. Springer, Dordrecht.

[24] 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.

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

[26] 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.

Table 1: Simulation results of different approaches to construct the counterfactuals for DGP 1

(T, N)		$N = 30$									
		PCA	PCCE	CCE	CPDA	PDA	PDAX	M1AV6	M1AV5	M2AV6	M2AV5
$T = 40$	MAB	2.1594	2.9902	5.3646	1.7136	5.4522	3.8404	2.3202	2.5244	2.3592	2.5589
	MSE	2.9233	3.8801	9.1313	2.2764	7.1026	5.1881	3.1055	3.3938	3.1732	3.4480
	MAP	1.0391	1.1255	1.1519	1.0440	1.0888	1.0398	1.0838	1.0658	1.0683	1.0597
$T = 60$	MAB	1.7495	2.5973	1.2075	1.1147	3.8675	2.4164	1.4498	1.4631	1.3841	1.4609
	MSE	2.3484	3.3473	1.5579	1.4442	4.9177	3.1257	1.9156	1.9138	1.8309	1.9079
	MAP	1.0022	1.0757	1.0115	1.0182	1.0827	1.0290	1.0494	1.0356	1.0314	1.0272
(T, N)		$N = 50$									
		PCA	PCCE	CCE	CPDA	PDA	PDAX	M1AV6	M1AV5	M2AV6	M2AV5
$T = 40$	MAB	2.2059	2.9252	2.8803	3.2505	7.3639	4.7273	2.4805	2.7147	2.4872	2.7268
	MSE	2.8860	3.7125	3.6931	4.2956	9.5806	6.1993	3.1786	3.4766	3.1927	3.4943
	MAP	0.9874	1.0064	1.0135	1.0134	1.0655	1.0254	1.0293	1.0265	1.0302	1.0269
$T = 60$	MAB	1.8563	2.5538	7.1890	2.1333	7.1706	4.6746	2.5557	2.9126	2.6154	2.9582
	MSE	2.3993	3.2239	14.191	2.9258	9.8993	6.6541	3.6059	4.1701	3.7168	4.2479
	MAP	0.9924	1.0074	1.0686	1.0079	1.0651	1.0191	1.0215	1.0191	1.0189	1.0179

Notes: 1. "PCA" to "M2AV4" refers to different estimators described as in (E1) to (E10), respectively.

2. "MAB" refers to mean of absolute bias for the true observation and the counterfactuals at each post treatment date point, "MSE" refers to the mean square of errors , and "MAP" refers to the mean of the ratio of absolute counterfactuals and absolute true outcomes at each date point after treatment.

Table 2: Simulation results of different approaches to construct the counterfactuals for DGP 2

(T,N)		$N=30$									
		PCA	PCCE	CCE	CPDA	PDA	PDAX	M1AV6	M1AV5	M2AV6	M2AV5
$T=40$	MAB	1.8283	2.5999	5.6001	1.3881	5.5254	3.8082	2.2656	2.5089	2.3018	2.5295
	MSE	2.4086	3.3067	13.741	1.8107	7.1577	5.0350	3.4079	3.8833	3.4759	3.9128
	MAP	1.6901	1.5653	3.0124	1.5356	3.3079	2.7070	1.7385	1.8556	1.7905	1.8692
$T=60$	MAB	1.7777	2.5337	1.0912	1.0147	4.0983	2.4653	1.4769	1.4733	1.4321	1.4629
	MSE	2.3781	3.2460	1.4229	1.3197	5.2025	3.1850	1.9454	1.9295	1.8975	1.9195
	MAP	2.1231	2.0609	1.7176	1.5778	3.5656	2.2387	1.6310	1.7428	1.6757	1.7576
(T,N)		$N=50$									
		PCA	PCCE	CCE	CPDA	PDA	PDAX	M1AV6	M1AV5	M2AV6	M2AV5
$T=40$	MAB	3.0818	3.3400	3.5854	3.5966	8.5529	6.0759	3.2355	3.4679	3.2602	3.4893
	MSE	4.3098	4.5848	4.6565	4.7874	11.0838	7.9870	4.2985	4.5530	4.3253	4.5767
	MAP	1.7252	1.7302	1.7989	1.7697	3.3645	2.3567	1.6248	1.6583	1.6556	1.6727
$T=60$	MAB	2.7742	3.0783	6.9442	2.3114	8.0545	5.6193	3.0023	3.2850	3.0548	3.3269
	MSE	4.0076	4.2832	11.339	3.1789	11.0607	7.9240	4.0648	4.4352	4.1348	4.4915
	MAP	2.0742	2.2412	3.9865	1.6253	4.3983	2.8383	1.9759	2.1414	2.0423	2.1808

See notes of Table 1.

Table 3: Simulation results of different approaches to construct the counterfactuals for DGP 3

(T,N)		$N=30$									
		PCA	PCCE	CCE	CPDA	PDA	PDAX	M1AV6	M1AV5	M2AV6	M2AV5
$T=40$	MAB	1.7023	2.4421	4.8152	1.2741	6.2052	4.3310	2.3080	2.5837	2.4214	2.6618
	MSE	2.3058	3.1399	8.7165	1.6938	8.0470	5.7521	3.0857	3.4828	3.2861	3.5996
	MAP	2.0434	2.7093	3.5339	1.8037	5.4261	3.4799	2.3061	2.4319	2.3808	2.4686
$T=60$	MAB	1.7331	2.4701	0.9501	0.9219	4.7997	2.8989	1.5632	1.5994	1.5106	1.5815
	MSE	2.2804	3.1384	1.2278	1.1871	6.1362	3.7211	2.0121	2.0442	1.9423	2.0255
	MAP	2.2391	2.5461	1.5780	1.4862	4.7677	2.9501	1.9490	1.9922	1.9144	1.9879
(T,N)		$N=50$									
		PCA	PCCE	CCE	CPDA	PDA	PDAX	M1AV6	M1AV5	M2AV6	M2AV5
$T=40$	MAB	2.3313	2.6858	3.2060	3.5063	9.5799	6.2329	2.9421	3.2830	3.0366	3.3572
	MSE	3.0395	3.4429	4.1296	4.6981	12.3682	8.1857	3.7589	4.1935	3.8816	4.2890
	MAP	2.2570	2.5199	1.8502	2.5400	6.2506	3.4896	2.5828	2.6695	2.5503	2.6704
$T=60$	MAB	2.5829	2.9635	7.1067	2.1277	9.6644	5.9604	3.0626	3.4507	3.1889	3.5625
	MSE	3.4348	3.842	13.547	2.8468	13.3117	8.4081	4.1781	4.7494	4.3951	4.9392
	MAP	1.8519	2.1403	6.1154	1.7339	4.9554	5.4531	2.8713	3.1070	2.9659	3.1693

See notes of Table 1.

Table 4: Simulation results of different approaches to construct the counterfactuals for DGP 4

(T,N)		$N=30$									
		PCA	PCCE	CCE	CPDA	PDA	PDAX	M1AV6	M1AV5	M2AV6	M2AV5
$T=40$	MAB	2.0238	12.6658	6.0039	1.7866	6.6651	4.9586	2.9557	2.8490	2.7144	2.8413
	MSE	2.7388	16.167	9.9213	2.4038	8.9456	6.7667	3.9042	3.8076	3.6029	3.8050
	MAP	1.1299	1.2182	1.4795	1.1091	1.5049	1.3218	1.1564	1.1942	1.1877	1.1999
$T=60$	MAB	1.6509	16.147	1.4038	1.1996	4.3058	2.7410	2.0409	1.4037	1.3212	1.3571
	MSE	2.2454	20.250	1.8242	1.5590	5.4945	3.5786	2.7957	1.9649	1.8671	1.9015
	MAP	1.0597	1.0452	1.0476	1.0393	1.1925	1.0979	0.9117	0.9119	0.9110	0.9117
(T,N)		$N=50$									
		PCA	PCCE	CCE	CPDA	PDA	PDAX	M1AV6	M1AV5	M2AV6	M2AV5
$T=40$	MAB	2.1421	12.544	2.8936	3.2506	8.7836	5.3328	3.0062	2.9713	2.7334	2.9744
	MSE	2.8162	16.006	3.7388	4.2928	11.773	7.0569	3.8917	3.8607	3.5439	3.8621
	MAP	1.0700	1.0044	1.1088	1.1480	1.4258	1.2620	1.0979	1.1261	1.1152	1.1276
$T=60$	MAB	1.8557	15.049	5.6595	2.1460	8.0664	4.8040	3.0059	2.2947	2.5928	2.8163
	MSE	2.4157	19.148	8.7089	2.9244	11.167	6.9167	3.8958	3.7385	3.4400	3.7731
	MAP	1.0752	1.0076	1.2550	1.0753	1.4361	1.1311	1.0924	1.1276	1.1156	1.1285

See notes of Table 1.

Table 5: Simulation results of different approaches to construct the counterfactuals for DGP 5

		$N = 30$									
		PCA	PCCE	CCE	CPDA	PDA	PDAX	M1AV6	M1AV5	M2AV6	M2AV5
$T = 40$	MAB	17.568	45.085	43.919	16.414	9.1401	9.4078	13.595	12.6865	13.909	12.8799
	MSE	23.565	57.929	77.752	21.613	12.497	12.855	19.138	19.5357	20.377	19.9163
	MAP	3.9978	0.6593	11.801	3.3842	1.9581	1.9762	3.1566	3.3122	3.5220	3.3587
$T = 60$	MAB	17.636	61.926	11.786	15.617	5.1520	5.1789	10.1937	7.4002	7.6573	7.4123
	MSE	23.507	77.960	16.226	20.187	6.9207	6.9480	12.8855	9.6187	9.9355	9.6549
	MAP	1.7736	0.1414	1.4130	1.6763	1.1704	1.1775	1.1271	1.2173	1.2157	1.2408
(T, N)		$N = 50$									
		PCA	PCCE	CCE	CPDA	PDA	PDAX	M1AV6	M1AV5	M2AV6	M2AV5
$T = 40$	MAB	39.2555	76.798	29.117	37.248	34.575	35.018	25.191	25.001	25.765	25.342
	MSE	52.514	98.930	40.007	50.140	45.996	46.578	32.664	32.858	33.902	33.331
	MAP	5.0456	1.6110	5.7428	6.1922	5.3704	5.3858	5.1132	5.0579	5.3127	5.1315
$T = 60$	MAB	36.8981	106.71	65.284	29.647	25.294	25.260	26.135	24.137	25.063	24.460
	MSE	48.643	135.93	103.97	38.913	32.943	33.034	34.248	32.829	33.565	33.338
	MAP	2.2548	0.3117	3.3094	1.8914	1.7856	1.8074	1.5522	1.6602	1.6849	1.6984

See notes of Table 1.

Table 6: Simulation results of different approaches to construct the counterfactuals for DGP 6

(T,N)		$N=30$									
		PCA	PCCE	CCE	CPDA	PDA	PDAX	M1AV6	M1AV5	M2AV6	M2AV5
$T=40$	MAB	1.7007	2.1233	5.1215	1.0919	1.0153	1.0674	1.5261	1.4901	1.7370	1.5852
	MSE	2.2518	2.7371	10.437	1.4321	1.3484	1.4173	2.2875	2.4176	2.7795	2.6601
	MAP	4.8415	2.0484	15.6809	3.4623	3.5486	4.2368	4.1868	4.7001	5.5567	5.2829
$T=60$	MAB	1.6810	2.0459	0.9798	0.8859	0.7599	0.7700	0.8484	0.7478	0.7423	0.7417
	MSE	2.2592	2.6702	1.2798	1.1445	0.9800	0.9952	1.1454	0.9963	0.9734	0.9702
	MAP	4.6510	0.7827	3.6500	3.1421	2.5878	2.6389	1.9413	2.2450	2.4506	2.4713
(T,N)		$N=50$									
		PCA	PCCE	CCE	CPDA	PDA	PDAX	M1AV6	M1AV5	M2AV6	M2AV5
$T=40$	MAB	2.7392	3.3675	3.3997	3.6748	1.7561	1.7807	2.1136	2.0751	2.1063	2.0963
	MSE	3.8953	4.5383	4.4080	4.9385	2.3896	2.4119	2.9503	2.8569	2.8575	2.8456
	MAP	4.6831	3.8364	11.062	7.3856	5.3068	5.5410	3.1177	3.9670	4.2791	4.4964
$T=60$	MAB	2.0652	2.7266	7.2913	2.5920	1.0218	1.0318	1.8498	1.9480	2.1121	2.1097
	MSE	2.6848	3.4307	12.542	3.3445	1.3319	1.3469	2.6494	2.9056	3.2120	3.2242
	MAP	4.7519	2.2001	20.667	5.9369	3.2115	3.2099	4.8654	5.9564	6.7192	6.8544

See notes of Table 1.

Table 7: Simulation results of different approaches to construct the counterfactuals for DGP 7

(T,N)		$N=30$									
		PCA	PCCE	CCE	CPDA	PDA	PDAX	M1AV6	M1AV5	M2AV6	M2AV5
$T=40$	MAB	1.8496	1.1783	6.0405	1.8852	1.1090	1.1783	1.8154	1.7025	1.7890	1.7686
	MSE	2.4259	9.1449	10.249	2.5118	1.4878	1.5933	2.4770	2.4759	2.5713	2.5881
	MAP	2.0495	1.2443	4.7408	2.2173	1.5272	1.5697	1.7706	1.9317	2.0555	2.0267
$T=60$	MAB	1.7413	9.7521	1.4522	1.2826	0.7900	0.8109	1.3799	0.8849	0.9231	0.8985
	MSE	2.3298	12.371	1.9057	1.6693	1.0169	1.0431	1.7675	1.1727	1.2220	1.1876
	MAP	1.8684	1.4484	1.5154	1.4609	1.3214	1.3267	1.4688	1.2836	1.3410	1.2161
(T,N)		$N=50$									
		PCA	PCCE	CCE	CPDA	PDA	PDAX	M1AV6	M1AV5	M2AV6	M2AV5
$T=40$	MAB	2.9419	10.443	3.6281	4.1084	1.9610	1.9936	2.5811	2.2673	2.3307	2.2833
	MSE	4.0960	13.158	4.7022	5.4612	2.6981	2.7333	3.4716	3.0913	3.1412	3.0902
	MAP	2.2444	1.2105	2.9485	3.2074	1.9578	2.0207	1.9109	2.0631	2.1672	2.1738
$T=60$	MAB	2.1246	12.201	6.0026	2.7568	1.0456	1.0589	2.2460	1.7640	2.0013	1.8255
	MSE	2.7753	15.799	9.0640	3.5801	1.3643	1.3837	2.9139	2.3841	2.6812	2.4745
	MAP	2.6238	2.1966	6.7861	2.8206	1.5903	1.5928	2.1201	2.6272	2.9548	2.8318

See notes of Table 1.

Figure 1: Plot of RMSE of different methods for DGP1 when $N = 30$ and $T = 40$

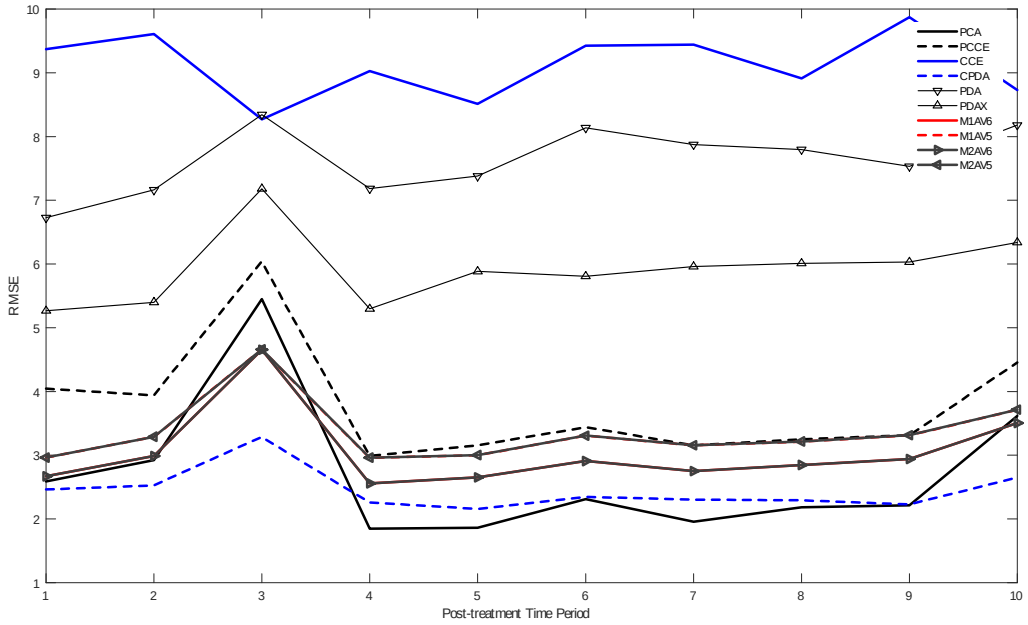

Figure 2: Plot of RMSE of different methods for DGP2 when $N = 30$ and $T = 40$

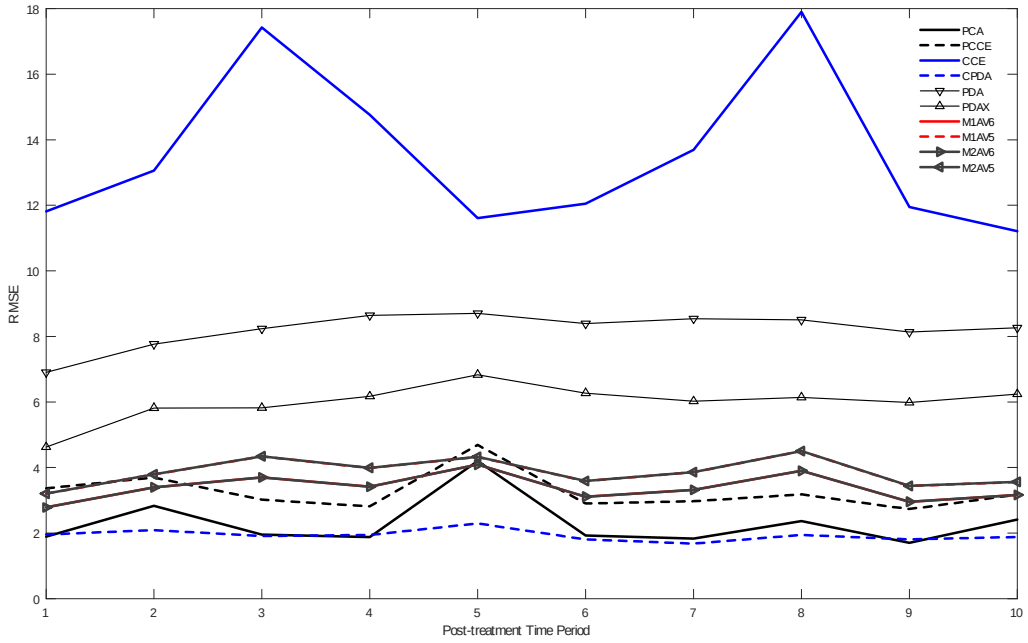

Figure 3: Plot of RMSE of different methods for DGP3 when $N = 30$ and $T = 40$

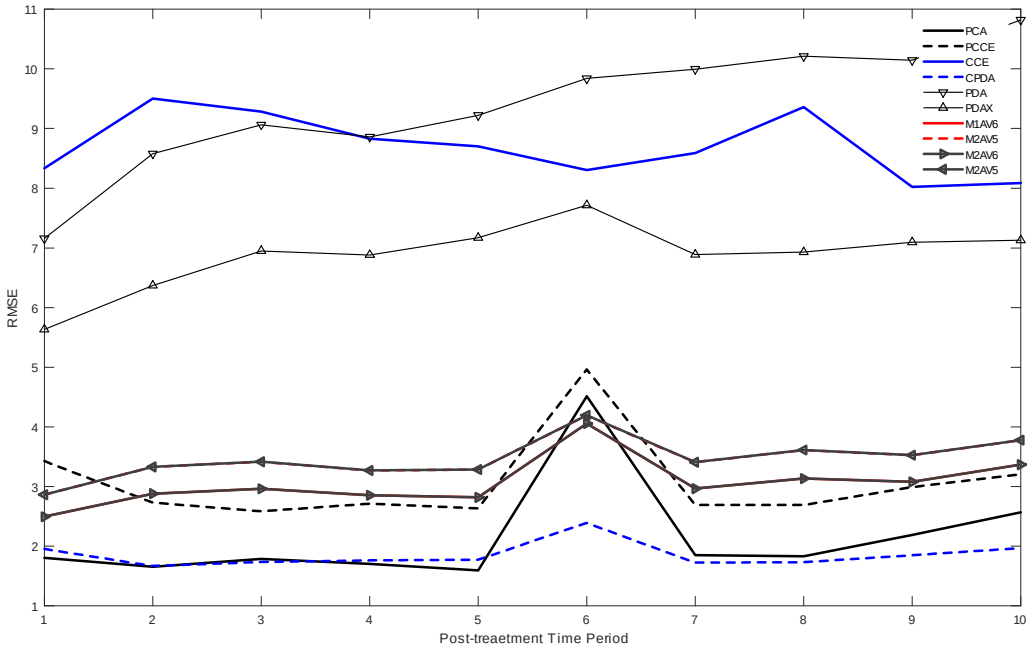

Figure 4: Plot of RMSE of different methods for DGP4 when $N = 30$ and $T = 40$

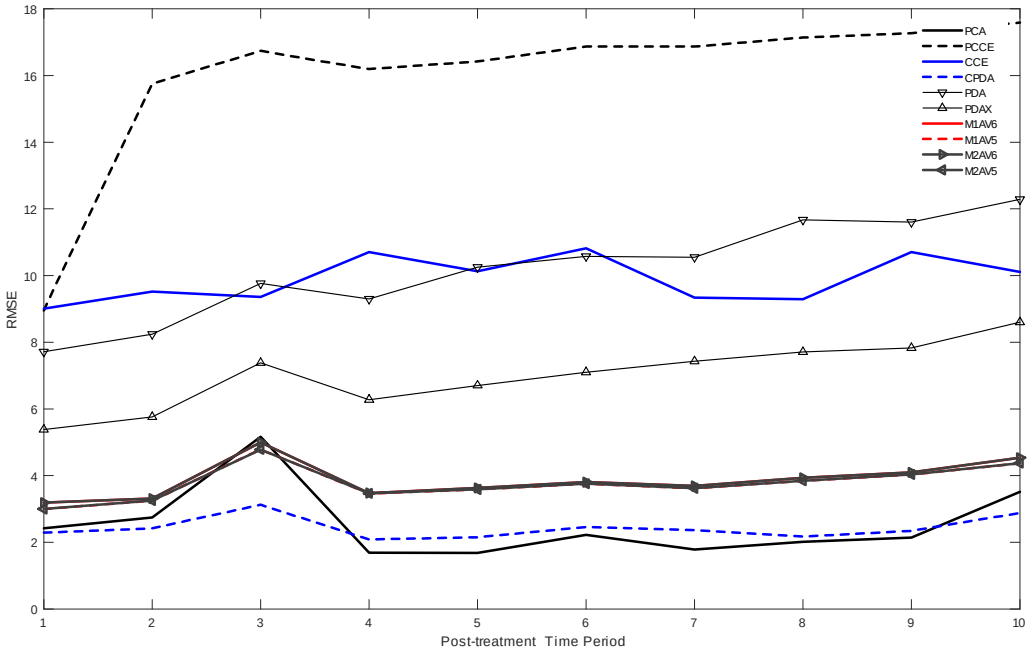

Figure 5: Plot of RMSE of different methods for DGP5 when $N = 30$ and $T = 40$

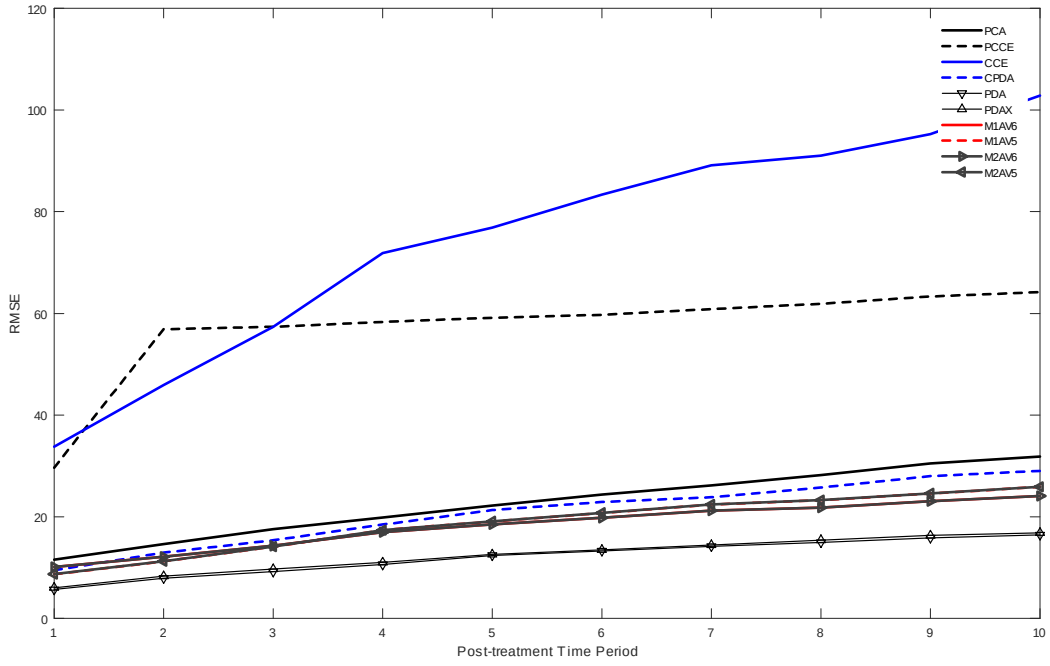

Figure 6: Plot of RMSE of different methods for DGP6 when $N = 30$ and $T = 40$

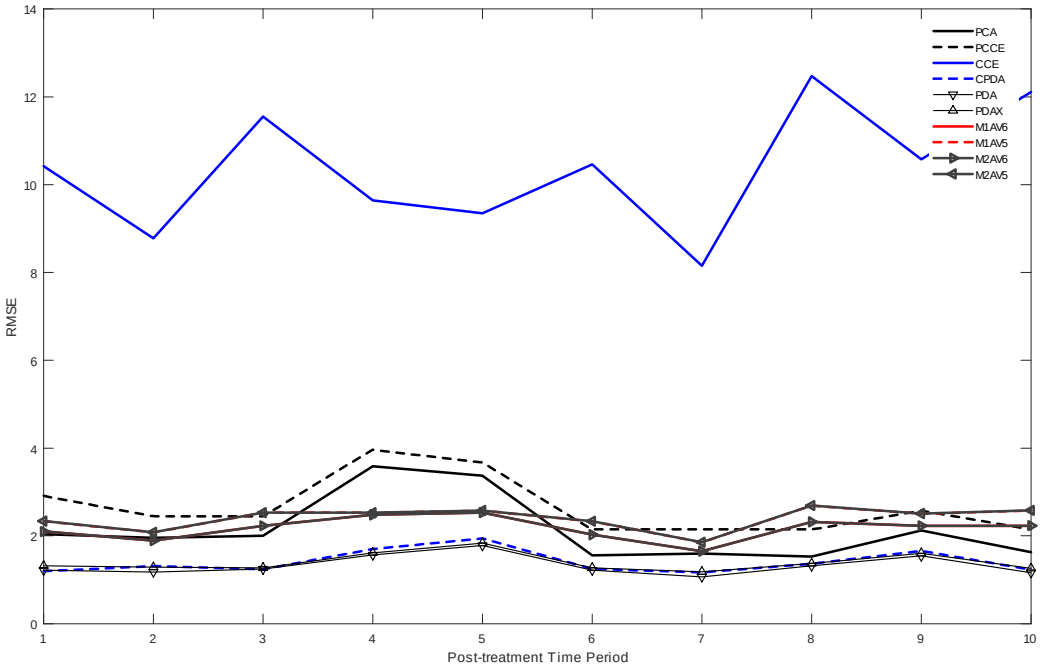

Figure 7: Plot of RMSE of different methods for DGP7 when $N = 30$ and $T = 40$

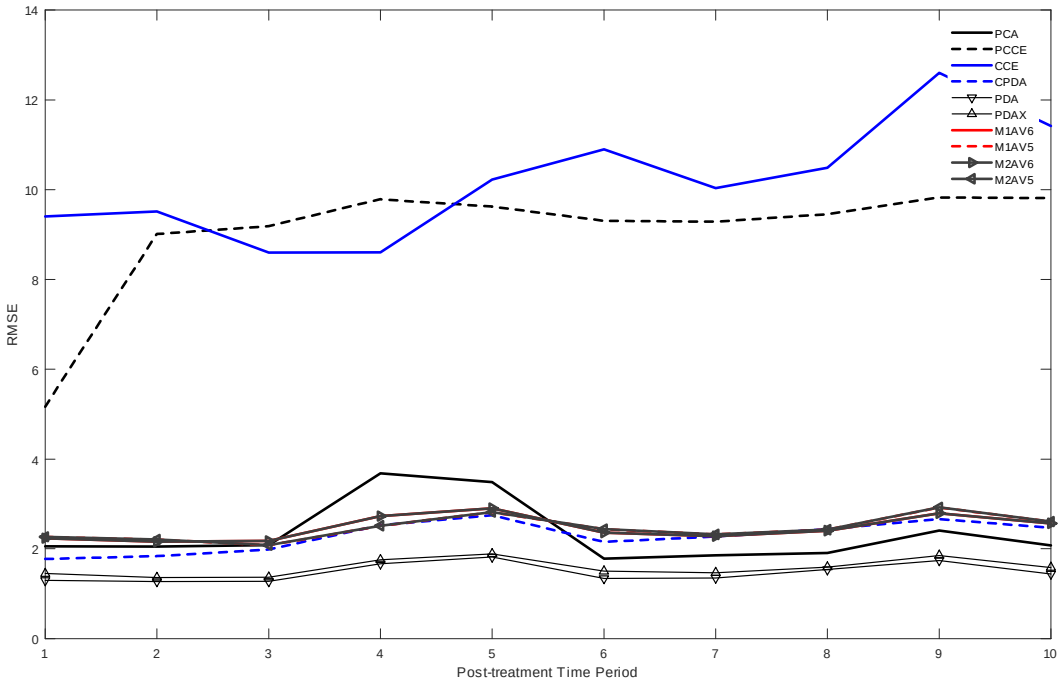

Table 8: Comparison of actual and counterfacutal cigarettes consumption

Year	Actual	SCM	PCA	CCE	CPDA	PDA	PDAX	M1AV5	M2AV5
1989	82.4	88.8	83.1	84.6	87.9	89.3	89.3	86.8	86.9
1990	77.8	86.7	82.3	79.9	83.0	86.4	86.4	83.6	83.6
1991	68.7	81.8	68.5	71.9	74.6	80.9	80.9	75.4	75.4
1992	67.5	81.3	69.4	70.1	73.8	79.9	79.9	74.6	74.6
1993	63.4	81.1	66.1	67.5	77.8	84.0	84.0	75.8	75.8
1994	58.6	80.7	63.6	62.6	76.4	82.6	82.6	73.8	73.8
1995	56.4	78.0	63.6	62.4	71.4	78.4	78.4	70.8	70.8
1996	54.5	77.1	64.6	62.8	74.2	81.1	81.1	72.7	72.7
1997	53.8	77.3	66.5	62.3	74.7	81.9	81.9	73.5	73.5
1998	52.3	73.7	62.7	60.7	71.4	76.5	76.5	69.5	69.5
1999	47.2	73.1	57.6	58.9	71.5	76.5	76.5	68.2	68.2
2000	41.6	66.8	52.3	54.8	71.2	75.6	75.6	65.9	65.9
MAB		18.5	6.42	6.20	15.3	20.7	20.7	13.8	13.8

Notes: MAB is defined in Table 1 and "SCM" refers to the counterfactals replicated from Abadie et al (2010).

Table 9: Comparison of actual and counterfacutal personal healthcare expenditure

Year	Actual	PCA	CCE	CPDA	PDA	PDAX	M1AV5	M2AV5
1989	8.9944	9.0115	8.9835	8.9974	8.9936	8.9936	8.9959	8.9960
1990	9.1201	9.1523	9.0945	9.1144	9.1099	9.1099	9.1162	9.1162
1991	9.2415	9.3033	9.1866	9.2117	9.2075	9.2075	9.2233	9.2218
1992	9.3173	9.3835	9.2522	9.2629	9.2841	9.2841	9.2931	9.2932
1993	9.3826	9.4569	9.3103	9.3201	9.3508	9.3508	9.3578	9.3578
1994	9.4416	9.5255	9.3606	9.3662	9.4067	9.4067	9.4132	9.4132
1995	9.5064	9.5850	9.4341	9.4285	9.4791	9.4791	9.4812	9.4812
1996	9.5681	9.6480	9.4933	9.4631	9.5401	9.4791	9.5369	9.54369
1997	9.6352	9.7163	9.5642	9.5299	9.6052	9.6052	9.6042	9.6042
1998	9.6581	9.6989	9.6341	9.5998	9.6781	9.6781	9.6578	9.6578
1999	9.6934	9.7005	9.7042	9.6627	9.7388	9.7388	9.7090	9.7090
2000	9.7556	9.7416	9.7923	9.7359	9.8072	9.8072	9.7768	9.7769
MAB		0.0531	0.0500	0.0522	0.0288	0.0288	0.0187	0.0187

Notes: MAB is the defined in Table 1.

Figure 8: Difference between actual and counterfactual cigarette consumption in treatment period

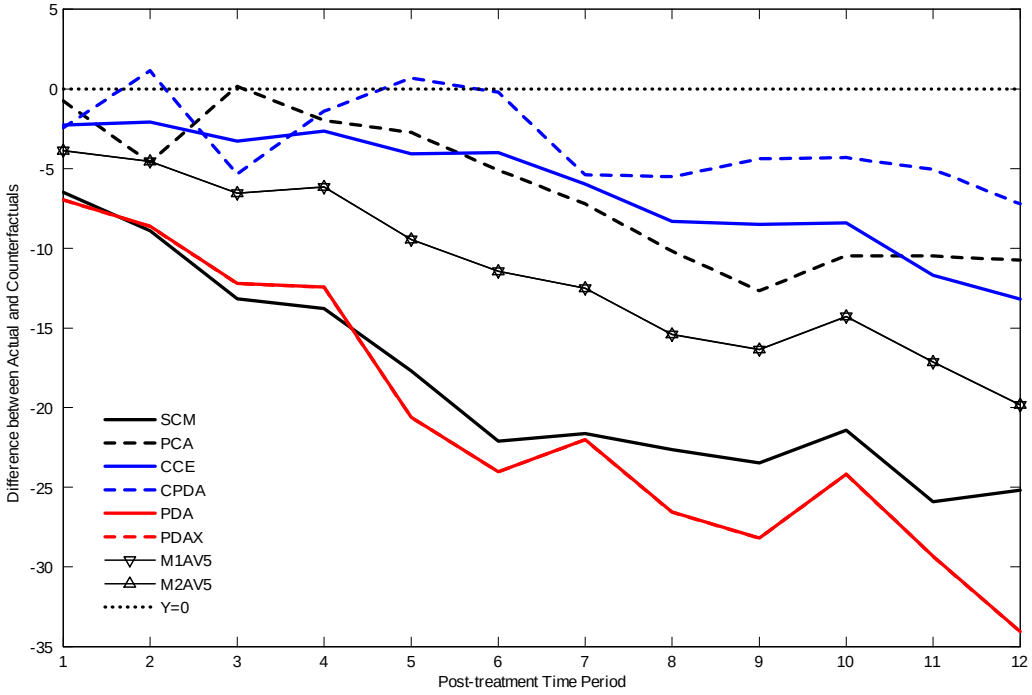

Figure 9: Actual and counterfacutal healthcare expenditure in treatment period