SCIENCE CHINA
Mathematics

• ARTICLES • https://doi.org/10.1007/s11425-023-xxxx-x

Dynamic Synthetic Control Method for Semiparametric Time-varying Models

Shouxia Wang[1], Xiangyu Zheng[2] & Song Xi Chen[2,*]

[1]*School of Statistics and Data Science, Shanghai University of Finance and Economics, Shanghai,200433, China;*
[2]*Department of Statistics and Data Science, Tsinghua University, Beijing, 100084, China*

Email: wangshouxia@sufe.edu.cn, zhengxiangyu@tsinghua.edu.cn, songxichen@pku.edu.cn

Received January 1, 2026; accepted January 1, 2026; published online January 1, 2026

Abstract Motivated by evaluating the treatment effects of a policy on an outcome variable with non-linear time-varying confounding variables, we propose a dynamic synthetic control (DSC) method for semiparametric time-varying additive autoregression outcome models. The proposed method allows microlevel data with non-linear time-varying confounders, multiple treated units and spatial correlations. The DSC weights are constructed by the empirical likelihood, which guarantees a unique solution and a consistent estimation of the average treatment effect on the treated group. The semiparametric additive model provides more flexibility in modeling and estimation, making it more favorable when either a parametric form of the model is unknown or may be mis-specified. We develop a test for the unconfoundedness assumption based on the estimated effects in the pre-treatment period and a normalized placebo test to determine the significance of the estimated treatment effects. The proposed DSC method is evaluated by numerical simulations and a case study that evaluates the effects of air pollution alerts in Beijing.

Keywords Empirical likelihood, Synthetic control, Time-varying additive model, Treatment effects.

MSC(2020) 62D20,62G08

Citation: Shouxia Wang, Xiangyu Zheng, Song Xi Chen. Dynamic Synthetic Control Method for Semiparametric Time-varying Models. Sci China Math, 2026, 69, https://doi.org/10.1007/s11425-023-xxxx-x

1 Introduction

Evaluation of effects of a treatment or policy intervention is an important task in many scientific fields such as economic research and evaluation, medical studies especially clinical trials, social sciences and beyond. Examples include studying the effects of natural disasters on economy (Cavallo et al., 2013), health insurance ineligibility on access to medical care (Viviano and Bradic, 2023), single-unit intervention in epidemiology (Bonander et al., 2021), impact of vaccines or new drugs (Bruhn et al., 2017; Lin et al., 2018) and so on. When there is no random assignment, the synthetic control (SC) methods (Abadie et al., 2010, 2015; Abadie and Gardeazabal, 2003; Doudchenko and Imbens, 2016) are often applied to estimate the treatment effects of policy intervention. The SC method enables estimation of the causal effects of a particular treatment or policy by constructing synthetic controls that mimic the characteristics of the treated unit in the absence of the intervention, which has become a vital analytical tool in many scientific

* Corresponding author

fields for evaluating policy intervention. Athey and Imbens (2017) described the SC method as "arguably the most important innovation in the evaluation literature in the last 15 years."

Extensions of the SC have greatly expanded in recent years. Abadie and L'Hour (2021) proposed penalized SC methods for multiple treated units, Ben-Michael et al. (2021a) proposed the modified SC along the lines of bias-correction, and Ben-Michael et al. (2021b) considered the case where the treatment initiation time varies among units. Athey et al. (2021) and Bai and Ng (2021) applied matrix completion to estimate the causal effects for panel data. Ferman and Pinto (2021) proposed a demeaned version of the SC method when the pre-treatment fit was imperfect. Viviano and Bradic (2023) developed a non-parametric algorithm for detecting the average treatment effects over time in the context of the SC.

A major limitation of the SC method is its requirement that the synthetic control weights remain time-invariant, which prevents its use in dynamic situations. Specifically, the weights of the SC method are constructed by matching the outcomes and covariates in the pre-treatment period. Although Abadie et al. (2010) considered a dual dynamic linear model, the weights were still time-invariant and obtained by matching the covariates and outcomes only at the time point when the treatment was assigned. To accommodate time-varying covariates, Zheng and Chen (2024) developed a linear dynamic synthetic control (DSC) approach, which showed that the SC estimate would incur a larger variance under the dynamic linear models compared with the linear DSC estimates as the SC method does not control the covariates at each time point.

Both the SC and the DSC (Zheng and Chen, 2024) approaches assume linear models which may not be satisfied as the influence of the confounding variables may be non-linear and nonparametric. Semiparametric models provide more robust specifications, making them more favorable when either the parametric form is either not available or subject to mis-specification. Moreover, there are usually many variables with non-linear time-varying effects on the outcomes, in which case only matching the covariates and outcomes may lead to undesirable estimates of the treatment effect. As shown in our case study of the effects of the air pollution alerts, the variables like wind speed, humidity, and dew point temperature have evidently non-linear effects on the air quality.

To evaluate dynamic treatment effects under a general dynamic panel data model with time-varying confounding variables, we use the time-varying additive autoregression to model the outcome variable, where the unknown additive functions are estimated by the spline-back-fitted-kernel (SPBK) method (Wang and Yang, 2007). After attaining the estimated additive functions, we apply the empirical likelihood (EL) (Chen and Van Keilegom, 2009; Owen, 1988, 1990) to construct the synthetic weights. The use of the EL ensures unique solutions to the matching weights, which overcomes an issue of the SC, and leads to consistent estimates of the average treatment effect on the treated (ATT).

As the model contains lagged outcome variables that are unobservable in the post-treatment period, one has to impute them to evaluate any treatment effects. For a general bivariate function $m_y(\cdot,\cdot)$ of t and y_{t-1}, $m_y(\cdot,\cdot)$ may be non-linear in the lagged outcome $Y_{i,t-1}(0)$, then we need to impute each treated unit as $\mathrm{E}\left[m_y\left(t, Y_{i,t-1}(0)\right)\right] \neq m_y\left(t, \mathrm{E}\left[Y_{i,t-1}(0)\right]\right)$. For a special case with $m_y(\cdot,\cdot)$ linear in $Y_{i,t-1}(0)$, we only need to impute the mean of the lagged outcome variables in the absence of the treatment. We propose a statistical test to test whether $m_y(\cdot,\cdot)$ is linear in y_{t-1}. We also develop an unconfoundedness assumption assessment test and a normalized placebo test to determine the significance of the estimated treatment effects. Extensive simulation studies are conducted to show the correctness of the theoretical results and the finite sample performance of the proposed method. Finally, we apply the proposed methods to evaluate the treatment effects of air pollution alerts in Beijing.

The main contributions are summarized as follows. First, the weights were time-invariant in the existing synthetic control method, while we allow the weights to differ at different time points. Second, we allow the covariates $\mathbf{X}$ and the lagged term $Y_{it}(0)$ to have nonparametric time-varying effects on the outcomes, which are modeled by general bivariate additive functions $\sum_{k=1}^{p} m_k(t, X_k)$ and estimated by the SPBK method. In contrast, both the SC and the dynamic SC (Zheng and Chen, 2024) approaches assume linear models. As shown in our case study of air pollution alerts, variables like wind speed, humidity, and dew point temperature had evidently non-linear effects on air quality. Third, we impute the lagged outcome for each treated unit in the post-treatment period rather than directly estimating the mean of the counterfactual outcomes by synthetic controls due to the non-linearity in the lagged outcome and its unobservability for the treated group in the post-treatment period. We also develop a hypothesis test for the linearity in the lagged outcome, as the linearity in the lagged outcome would greatly reduce the complexity of the DSC process. The theoretical results for the proposed method are also established, including the theoretical properties of the SPBK estimation of additive functions, bounds on the weights constructed by the empirical likelihood and convergence of the estimated ATT under both perfect and imperfect matching.

The paper is organized as follows. Sections 2 and 3 introduce the DSC method for the time-varying additive autoregression model. The estimation of the additive functions is given in Section 4. Section 5 presents the theoretical results of the proposed method. Several hypothesis tests are shown in Section 6. Simulation results are reported in Section 7 and a real data application is given in Section 8, followed by a conclusion in Section 9. Technical details and additional results are given in the Supplementary Material (SM).

2 Semiparametric Model Setting

Let $N = N_{tr} + N_{co}$ denote the number of units observed for a duration T with N_{tr} and N_{co} units in the treated $\mathcal{T}$ and control groups $\mathcal{C}$, where $\mathcal{T}$ and $\mathcal{C}$ denote the sets of unit indexes in the treated and control groups, respectively. Units in $\mathcal{T}$ are exposed to a policy intervention after time T_0, which are called the treated units. Units in $\mathcal{C}$ remain unaffected by the intervention. The variable D_{it} indicates whether unit i is exposed to the intervention at time t ($D_{it} = 1$), or not ($D_{it} = 0$) and $D_i = D_{i,T_0+1}$ denotes whether unit i is exposed to the intervention after T_0. The outcome of interest is denoted by Y_{it} and $\mathbf{X}_{it}$ is the p-dimensional vector of covariates affecting the outcome Y_{it}. Following the notion of potential outcomes (Rubin, 1974), let $Y_{it}(0)$ and $Y_{it}(1)$ denote the potential outcomes observed for unit i at time t when $D_{it} = 0$ and $D_{it} = 1$, respectively. The observed outcome $Y_{it} = Y_{it}(D_{it})$ is contingent on whether the intervention affects unit i at time t or not. The goal is to estimate the average treatment effect on the treated group during the post-treatment period at time t,

$$\tau_t = \mathrm{E}\left(Y_{it}(1)|D_i = 1\right) - \mathrm{E}\left(Y_{it}(0)|D_i = 1\right) \text{ for } t > T_0. \tag{2.1}$$

While the first term $\mathrm{E}\left(Y_{it}(1)|D_i = 1\right)$ can be readily estimated by the sample average of the outcomes for the treated, namely $N_{tr}^{-1}\sum_{j\in\mathcal{T}} Y_{jt}$, an estimator for the second term $\mathrm{E}(Y_{it}(0)|D_i = 1)$ has to be specially constructed as the would-be outcomes in the absence of the treatment for the treated are not observed. We consider its estimation by the proposed semiparametric DSC method.

To accommodate the non-linear influence of the covariates on the outcomes, we consider the semiparametric time-varying additive autoregression model (2.2),

$$Y_{it}(0) = m_0(t) + \sum_{k=1}^{p} m_k(t, X_{it,k}) + m_y(t, Y_{i,t-1}(0)) + \varepsilon_{it}, \text{ for } 1 \leqslant i \leqslant N, \tag{2.2}$$

where $m_0(t)$ is an unknown baseline function of time t, $X_{it,k}$ is the k-th covariate of $\mathbf{X}_{it}$ for unit i at time t, and $m_y(t, Y_{i,t-1}(0))$ and $m_k(t, X_{it,k})$ are the unknown time-varying additive functions which are identifiable under the constraints $\int m_k(t, x_k) f_k(t, x_k)\, dx_k = 0$, for every $t, k = 1, \ldots, p+1$, where $f_k(t, \cdot)$ denotes the marginal density of the kth covariate at time t,with $X_{ij,p+1} = Y_{i(j-1)}(0)$ and $m_{p+1} = m_y$. The error term ε_{it} is assumed to be independent of $\{X_{it}, Y_{i,t-1}(0)\}$ with zero mean at each time t and is allowed to be spatially dependent such that $\sum_{0\leqslant|i-j|<N} \mathrm{E}(\varepsilon_{it}\varepsilon_{jt}) = O(N)$. Model (2.2) is more general than that treated in Zheng and Chen (2024) in that the nonparametric form is assumed on the influence of time and the covariates $\mathbf{X}_t, Y_{t-1}(0)$. This allows the lagged outcome $Y_{it-1}(0)$ and the covariates $\mathbf{X}_t$ to have non-linear and time-varying effects on the outcome.

The following assumption is needed in the analysis.

Assumption 2.1. *(i) For* $1 \leqslant t \leqslant T, 1 \leqslant i \leqslant N$, $Y_{it} = D_{it}Y_{it}(1) + (1 - D_{it})Y_{it}(0)$; *(ii)* $\mathrm{E}(Y_{it}(0)|Y_{i,t-1}(0), \mathbf{X}_{it}, D_i = 1) = \mathrm{E}(Y_{it}(0)|Y_{i,t-1}(0), \mathbf{X}_{it}, D_i = 0)$; *(iii)* $\mathrm{P}(D_i = 1) > 0$; $\mathrm{P}(D_i = 1|\mathbf{X}_{it}, Y_{i,t-1}(0)) < 1$ *with probability one.*

The consistency assumption, Assumption 2.1(i), guarantees each unit at each time has one of the potential outcomes observed. By Assumption 2.1(i), $\tau_t = \mathrm{E}(Y_{it}|D_i = 1) - \mathrm{E}(Y_{it}(0)|D_i = 1)$ for $t > T_0$. Assumption 2.1(ii), called the unconfoundedness assumption (Imbens and Wooldridge, 2009), is used to identify τ_t. It is a common assumption in treatment effects estimation (Abadie et al., 2010; Shi et al., 2022; Viviano and Bradic, 2023) and also relates to the conditional parallel trends assumption if lagged outcomes are included as covariates (Roth et al., 2023). Assumption 2.1(ii) requires that the conditional expectations of $Y_{it}(0)$ are the same for both the treated and control groups when the covariates ($Y_{i,t-1}(0)$ and $\mathbf{X}_{it}$) are controlled. It is equivalent to the conditional mean independence assumption that the conditional mean of $Y_{it}(0)$ given the covariates $Y_{i,t-1}(0), \mathbf{X}_{it}$ is independent of the treatment D_i, that is, $\mathrm{E}(Y_{it}(0)|Y_{i,t-1}(0), \mathbf{X}_{it}, D_i) = \mathrm{E}(Y_{it}(0)|Y_{i,t-1}(0), \mathbf{X}_{it})$, as assumed in Chapter 5 of Angrist and Pischke (2009). Assumption 2.1 (ii) is a weak version of the (full) conditional independence assumption (Imbens and Wooldridge, 2009; Shi et al., 2022) or the so-called ignorability conditional on the lagged dependent variable (Ding and Li, 2019; Viviano and Bradic, 2023). The ignorability conditional on the lagged dependent variable means that $Y_{it}(0)$ is independent of the treatment D_i given the covariates $Y_{i,t-1}(0), \mathbf{X}_{it}$, that is, $Y_{it}(0) \perp\!\!\!\perp D_i \mid (Y_{i,t-1}(0), \mathbf{X}_{it})$.

Under Assumption 2.1(i)-(ii), the conditional expectation of $Y_{it}(0)$ in the treated group is identifiable; namely, it can be expressed by the observed data, $\mathrm{E}(Y_{it}(0)|\mathbf{X}_{it}, Y_{i,t-1}(0), D_i = 1) = \mathrm{E}(Y_{it}|\mathbf{X}_{it}, Y_{i,t-1}, D_i = 0)$. Assumption 2.1(iii) implies that for any given $\mathbf{X}_{it}$ and $Y_{i,t-1}(0)$, there is always a proportion of untreated population that can serve as controls. Here, the reasons for including the lagged outcomes $Y_{i,t-1}(0)$ as covariates, as also seen in lagged-dependent-variable adjustment (Ding and Li, 2019), are that the lagged outcomes $Y_{i,t-1}(0)$ have big influence on the outcome, especially in our case study of the $\mathrm{PM}_{2.5}$ and that controlling for lagged outcomes can reduce bias from unobserved confounders (Ryan, 2018). In many real-world scenarios, the unobserved factors (e.g., managerial quality or latent ability) that create selection bias are 'sticky' and are reflected in the previous period's outcome. By conditioning on $Y_{i,t-1}(0)$, we are comparing treated and control units that started from the same baseline performance level.

3 Semiparametric Dynamic Synthetic Control

In this section, we introduce the semiparametric DSC method for estimating $\mu_t(0) = \mathrm{E}(Y_t(0)|D = 1)$ by establishing dynamic synthetic controls that mimic the counterfactual scenario of the treated group in the absence of the treatment for $t > T_0$. The linear DSC proposed in Zheng and Chen (2024) constructed the synthetic weights for each time step. However, as the effects of $\mathbf{X}_{it}$ and $Y_{i,t-1}(0)$ in Model (2.2) are non-linear, we cannot directly apply the linear DSC method by just matching $\mathbf{X}_{it}$ and $Y_{i,t-1}(0)$. Furthermore, the lagged outcome $Y_{i,t-1}(0)$ is unobserved for the treated group when $t > T_0 + 1$. The non-linearity in the lagged outcome, namely $\mathrm{E}\left[m_y\left(t, Y_{i,t-1}(0)\right)\right] \neq m_y\left(t, \mathrm{E}\left[Y_{i,t-1}(0)\right]\right)$, prevents the direct estimation of $\mu_t(0)$ by synthetic controls. To tackle this problem, we need to establish synthetic controls for each unit in the treated group, that is to find a set of non-negative weights $\{w_{it}^j\}_{i=1}^{N_{co}}$ for each $j \in \mathcal{T}$ at time t such that $\sum_{i=1}^{N_{co}} w_{it}^j = 1$ and the imputed value is denoted by $\widehat{Y}_{jt}(0) = \sum_{i \in \mathcal{C}} w_{it}^j Y_{it}(0)$. According to Model (2.2), for a fixed $j \in \mathcal{T}$ at time t,

$$Y_{jt}(0) = m_0(t) + \sum_{k=1}^{p} m_k\left(t, X_{jt,k}\right) + m_y\left(t, Y_{j,t-1}(0)\right) + \varepsilon_{jt}, \tag{3.1}$$

$$\sum_{i \in \mathcal{C}} w_{it}^j Y_{it}(0) = \sum_{i \in \mathcal{C}} w_{it}^j m_0(t) + \sum_{k=1}^{p} \sum_{i \in \mathcal{C}} w_{it}^j m_k\left(t, X_{it,k}\right) + \sum_{i \in \mathcal{C}} w_{it}^j m_y\left(t, Y_{i,t-1}(0)\right) + \sum_{i \in \mathcal{C}} w_{it}^j \varepsilon_{it}. \tag{3.2}$$

Although $Y_{jt}(0)$ in the treated group is unobservable for $t > T_0$, matching (3.2) with (3.1) leads to two matching conditions (3.4) and (3.5) below for the dynamic synthetic weights $\{w_{it}^j\}_{i=1}^{N_{co}}$, with $m_0(\cdot)$, $m_k(\cdot,\cdot)$ and $m_y(\cdot,\cdot)$ replaced by their respective consistent estimators $\widehat{m}_0(\cdot)$, $\widehat{m}_k(\cdot,\cdot)$ and $\widehat{m}_y(\cdot,\cdot)$. The details of these estimators are deferred to the next section.

The DSC weights $\{w_{it}^j\}_{i=1}^{N_{co}}$ are established via the EL (Chen and Van Keilegom, 2009; Owen, 1988, 1990) which ensures a unique solution for the weights. For a fixed $j \in \mathcal{T}$ at time t, the EL weights $\{w_{it}^j\}$ are obtained by maximizing the empirical likelihood $\prod_{i\in\mathcal{C}} w_{it}$ subject to the matching conditions (3.4) and (3.5). Specifically, the EL weights are given by

$$\mathbf{w}_t^j = (w_{1t}^j, \cdots, w_{N_{co}t}^j) = \underset{\mathbf{w}_t \in \mathcal{W}_t^j}{\arg\max} \prod_{i\in\mathcal{C}} w_{it}, \tag{3.3}$$

where $\mathcal{W}_t^j = \{(w_{1t}, \cdots, w_{N_{co}t}) : \sum_{i=1}^{N_{co}} w_{it} = 1, w_{it} \geqslant 0\}$, and are subject to

$$\textstyle\sum_{i\in\mathcal{C}} w_{it}\widehat{m}_k\left(t, X_{it,k}\right) = \widehat{m}_k\left(t, X_{jt,k}\right) \text{ for } k = 1, \cdots, p, \tag{3.4}$$

$$\sum_{i\in\mathcal{C}} w_{it}\widehat{m}_y\left(t, Y_{i,t-1}(0)\right) = \begin{cases} \widehat{m}_y(t, Y_{j,t-1}(0)), t \leqslant T_0+1, \\ \widehat{m}_y(t, \widehat{Y}_{j,t-1}(0) + \varepsilon_{jt}^*) = \widehat{m}_y(t, \sum_{i\in\mathcal{C}} w_{i,t-1}^j Y_{i,t-1}(0) + \varepsilon_{jt}^*), t > T_0+1, \end{cases} \tag{3.5}$$

where $\varepsilon_{jt}^* \sim \widehat{F}$ and $\widehat{F}$ is the empirical distribution of centered control residuals. The error ε_{jt}^* is used only in the next lag matching target. The ATT estimator continues to use $\widehat{Y}_{j,t-1}(0)$. Note that Constraints (3.4) and (3.5) are the results of matching the second and third terms on the right-hand side (RHS) of (3.2) with the corresponding terms in (3.1). Specifically, (3.4) matches the covariate functions between the treatment and control groups based on the p additive components of Model (2.2), while (3.5) matches the function of lagged values of the potential outcomes $Y_{i,t-1}(0)$ before and after the time point T_0+1. When $t \leqslant T_0+1$, the lagged outcomes $Y_{i,t-1}(0)$ are observed in both the control and treated groups. For $t > T_0+1$, as $Y_{j,t-1}(0)$ are unobserved for the treated group, we substitute $m_y\left(t, Y_{j,t-1}(0)\right)$ by $m_y(t, \widehat{Y}_{j,t-1}(0) + \varepsilon_{jt}^*) = \widehat{m}_y(t, \sum_{i\in\mathcal{C}} w_{i,t-1}^j Y_{i,t-1}(0) + \varepsilon_{jt}^*)$.

The EL weights in (3.3) can be obtained by the Lagrange multiplier method, which leads to unique solutions for the weights and facilitates asymptotic analysis of the weights. Let $\mathbf{Z}_{it,\mathcal{C}} = (m_1(t, X_{it,1}), \cdots, m_p(t, X_{it,p}), m_y(t, Y_{i,t-1}(0)))^\top$, $\widehat{\mathbf{Z}}_{it,\mathcal{C}} = (\widehat{m}_1(t, X_{it,1}), \cdots, \widehat{m}_p(t, X_{it,p}), \widehat{m}_y(t, Y_{i,t-1}(0)))^\top$ for $1 \leqslant i \leqslant N_{co}$, $1 \leqslant t \leqslant T$, and $\overline{\widehat{\mathbf{Z}}}_{jt,\mathcal{T}} = (\widehat{m}_1(t, X_{jt,1}), \cdots, \widehat{m}_p(t, X_{jt,p}), \widehat{m}_y(t, Y_{j,t-1}(0)))^\top$ for $j \in \mathcal{T}$, $t \leqslant T_0+1$, and for $j \in \mathcal{T}$, $t > T_0+1$, $\overline{\widehat{\mathbf{Z}}}_{jt,\mathcal{T}} = (\widehat{m}_1(t, X_{jt,1}), \cdots, \widehat{m}_p(t, X_{jt,p}), \widehat{m}_y(t, \sum_{i\in\mathcal{C}} w_{i,t-1}^j Y_{i,t-1}(0)) + \varepsilon_{jt}^*)^\top$. Then, the structural constraints (3.4) and (3.5) can be expressed by $\sum_{i\in\mathcal{C}} w_{it}^j \widehat{\mathbf{Z}}_{it,\mathcal{C}} = \overline{\widehat{\mathbf{Z}}}_{jt,\mathcal{T}}$. Note that the maximizing problem (3.3) is equivalent to minimizing

$$L(\mathbf{w}_t^j) = -\sum_{i=1}^{N_{co}} \log(w_{it}^j) \text{ over } w_{it}^j \geqslant 0, \text{ subject to } \sum_{i=1}^{N_{co}} w_{it}^j = 1, \text{ and } \sum_{i\in\mathcal{C}} w_{it}^j \widehat{\mathbf{Z}}_{it,\mathcal{C}} = \overline{\widehat{\mathbf{Z}}}_{jt,\mathcal{T}}.$$

Let $G_{tj}(\mathbf{w}_t^j, \lambda, \gamma) = \sum_{i=1}^{N_{co}} \log(w_{it}^j) - N_{co}\lambda \sum_{i=1}^{N_{co}} w_{it}^j \big(\widehat{\mathbf{Z}}_{it,\mathcal{C}} - \overline{\widehat{\mathbf{Z}}}_{jt,\mathcal{T}}\big) + \gamma(\sum_{i=1}^{N_{co}} w_{it}^j - 1)$, where λ and γ are Lagrange multipliers. Taking the partial derivatives of G_{tj} with respect to w_{it}^j, λ and γ, we get

$$\partial G_{tj}/\partial w_{it}^j = 1/w_{it}^j - N_{co}\lambda\big(\widehat{\mathbf{Z}}_{it,\mathcal{C}} - \overline{\widehat{\mathbf{Z}}}_{jt,\mathcal{T}}\big) + \gamma = 0, \quad \sum_{i=1}^{N_{co}} w_{it}^j \big(\widehat{\mathbf{Z}}_{it,\mathcal{C}} - \overline{\widehat{\mathbf{Z}}}_{jt,\mathcal{T}}\big) = 0, \quad \sum_{i=1}^{N_{co}} w_{it}^j = 1,$$

which implies $\sum w_{it}^j \partial G_{tj}/\partial w_{it}^j = 0$. Then, we can derive that the EL weights take the form

$$w_{it}^j = N_{co}^{-1}\big\{1 + \boldsymbol{\lambda}^\top\big(\widehat{\mathbf{Z}}_{it,\mathcal{C}} - \overline{\widehat{\mathbf{Z}}}_{jt,\mathcal{T}}\big)\big\}^{-1} \text{ for } i = 1, \cdots, N_{co},\, j \in \mathcal{T}, \tag{3.6}$$

where $\boldsymbol{\lambda}$ satisfies $\sum_{i=1}^{N}(\widehat{\mathbf{Z}}_{it,\mathcal{C}} - \overline{\widehat{\mathbf{Z}}}_{jt,\mathcal{T}})\{1 + \boldsymbol{\lambda}^\top(\widehat{\mathbf{Z}}_{it,\mathcal{C}} - \overline{\widehat{\mathbf{Z}}}_{jt,\mathcal{T}})\}^{-1} = 0$. Lemmas S.14-S.15 of the SM show that the weight solution to problem (3.3) (or (3.10)) uniquely exists. In our simulation studies and real data analysis, the EL weights were obtained by the R packages "LowRankQP" and "NlcOptim". We also check the 'convex hull' requirement of the EL. Overall, the algorithm was stable in our simulations and empirical analysis.

With $\{w_{it}^j\}_{i=1}^{N_{co}}$, we can impute $Y_{jt}(0)$ by $\widehat{Y}_{jt}(0) = \sum_{i\in\mathcal{C}} w_{it}^j Y_{it}(0)$. With all weights $\{\mathbf{w}_t^j\}_{j\in\mathcal{T}}$, we can estimate $\mu_t(0)$ and τ_t by $\widehat{\mu}_t(0) = N_{tr}^{-1}\sum_{j\in\mathcal{T}}\widehat{Y}_{jt}(0) = N_{tr}^{-1}\sum_{j\in\mathcal{T}}\sum_{i\in\mathcal{C}} w_{it}^j Y_{it}(0)$ and

$$\widehat{\tau}_t =: N_{tr}^{-1}\sum_{i\in\mathcal{T}} Y_{jt} - \widehat{\mu}_t(0) = N_{tr}^{-1}\sum_{j\in\mathcal{T}}\big(Y_{jt} - \sum_{i\in\mathcal{C}} w_{it}^j Y_{it}(0)\big). \tag{3.7}$$

For the special case where $m_y(t,\cdot)$ is linear in the lagged outcome so that $\mathrm{E}[m_y(t,Y_{i,t-1}(0))] = m_y\left(t, \mathrm{E}\left[Y_{i,t-1}(0)\right]\right)$, the estimation of τ_t is simpler. An example of the linear case is $m_y(t,Y_{i,t-1}(0)) = h(t)Y_{i,t-1}(0)$ for a temporal function $h(t)$. The linearity in the lagged outcome allows for the construction of synthetic controls using weighted averages of the outcomes in the control group directly for the estimation of $\mu_t(0)$ rather than for each treated unit, that is to find a set of non-negative weights $\{w_{it}^*\}_{i=1}^{N_{co}}$ that sum to 1 such that $\widehat{\mu}_t(0) = \sum_{i\in\mathcal{C}} w_{it}^* Y_{it}(0)$ at time t. This implies that the complexity of the above DSC process can be reduced due to the linearity in the lagged outcome. By (2.2), the weighted average of the control group and the average of the potential outcome over the treated group satisfy

$$\sum_{i\in\mathcal{C}} w_{it}^* Y_{it}(0) = \sum_{i\in\mathcal{C}} w_{it}^* m_0(t) + \sum_{k=1}^{p}\sum_{i\in\mathcal{C}} w_{it}^* m_k\left(t, X_{it,k}\right) + \sum_{i\in\mathcal{C}} w_{it}^* m_y\left(t, Y_{i,t-1}(0)\right) + \sum_{i\in\mathcal{C}} w_{it}^*\varepsilon_{it},$$
$$\sum_{i\in\mathcal{T}} \frac{Y_{it}(0)}{N_{tr}} = m_0(t) + \sum_{k=1}^{p}\sum_{i\in\mathcal{T}} \frac{m_k\left(t, X_{it,k}\right)}{N_{tr}} + \sum_{i\in\mathcal{T}} \frac{m_y\left(t, Y_{i,t-1}(0)\right)}{N_{tr}} + \sum_{i\in\mathcal{T}} \frac{\varepsilon_{it}}{N_{tr}}.$$

Matching the two equations above leads to two constraints

$$\textstyle\sum_{i\in\mathcal{C}} w_{it}\widehat{m}_k\left(t, X_{it,k}\right) = N_{tr}^{-1}\sum_{j\in\mathcal{T}}\widehat{m}_k\left(t, X_{jt,k}\right), \text{ for } k = 1,\cdots,p, \tag{3.8}$$

$$\sum_{i\in\mathcal{C}} w_{it}\widehat{m}_y\left(t, Y_{i,t-1}(0)\right) = \begin{cases} N_{tr}^{-1}\sum_{j\in\mathcal{T}}\widehat{m}_y\left(t, Y_{j,t-1}(0)\right), t \leqslant T_0+1, \\ \widehat{m}_y\left(t, \widehat{\mu}_{t-1}(0)\right) = \widehat{m}_y(t, \sum_{i\in\mathcal{C}} w_{i,t-1}^* Y_{i,t-1}(0)), t > T_0+1, \end{cases} \tag{3.9}$$

Subject to (3.8) and (3.9), the EL weights $\{w_{it}^*\}_{i=1}^{N_{co}}$ are given by

$$\mathbf{w}_t^* = (w_{1t}^*,\cdots,w_{N_{co}t}^*) = \arg\max_{\mathbf{w}_t\in\mathcal{W}_t}\prod_{i\in\mathcal{C}} w_{it} \tag{3.10}$$

where $\mathcal{W}_t = \{(w_{1t},\cdots,w_{N_{co}t}) : \sum_{i=1}^{N_{co}} w_{it} = 1, w_{it} \geqslant 0\}$. The EL weights are constructed by the EL in a recursive way similar to the general case with $\overline{\overline{\mathbf{Z}}}_{jt,\mathcal{T}}$ in (3.6) replaced by $\overline{\overline{\mathbf{Z}}}_{t,\mathcal{T}} = N_{tr}^{-1}\sum_{j\in\mathcal{T}}(\widehat{m}_1(t,X_{jt,1}),\cdots,\widehat{m}_p(t,X_{jt,p}),\widehat{m}_y(t,Y_{jt}(0)))^\top$ for $t \leqslant T_0+1$ and $\overline{\overline{\mathbf{Z}}}_{t,\mathcal{T}} = (N_{tr}^{-1}\sum_{j\in\mathcal{T}}\widehat{m}_1(t,X_{jt,1}),\cdots,\widehat{m}_p(t,X_{jt,p}),\widehat{m}_y(t,\widehat{\mu}_{t-1}(0)))^\top$ for $t > T_0+1$. The difference is that the weights are constructed for the average of the treated group instead of for each treated unit with $N_{tr}^{-1}\sum_{i\in\mathcal{T}} m_y\left(t, Y_{i,t-1}(0)\right)$ on the RHS of (3.9) for $t > T_0+1$ replaced by $m_y\left(t,\widehat{\mu}_{t-1}(0)\right)$ since $m_y(t,\cdot)$ is linear in $Y_{i,t-1}(0)$. With the weights $\{w_{it}^*\}$, the estimator of the ATT τ_t is

$$\widehat{\tau}_t = N_{tr}^{-1}\sum_{j\in\mathcal{T}} Y_{jt} - \widehat{\mu}_t(0) = N_{tr}^{-1}\sum_{j\in\mathcal{T}} Y_{jt} - \sum_{i\in\mathcal{C}} w_{it}^* Y_{it}, \quad t > T_0. \tag{3.11}$$

4 Estimation of Additive Functions

Before we present the theoretical properties of the semiparametric DSC estimator $\widehat{\tau}_t$, we provide details on the estimation of the additive functions in Model (2.2). We employ a two-step SPBK method based on all the observations of the subjects in the control group for the whole time T. Specifically, we utilize the

B-spline procedure to obtain the initial estimators of $m_0(\cdot)$, $m_k(\cdot,\cdot)$ and then get the refined estimators using a local linear estimation. Denote the observed sample as $\{(Y_{ij}, t_{ij}, X_{ij,1}, \ldots, X_{ij,p}) : i = 1, \ldots, n, j = 1, \ldots, T\}$, where $n = N_{co}$ is the total number of the control subjects and t_{ij} is the observation time.

Initial spline estimation of component functions. Following De Boor (1978), let $\mathbf{b}_k(x_k) = (b_{k,1}(x_k), \ldots, b_{k,K+q}(x_k))^\top$ be a q-order B-spline basis vector with K interior knots. We center and standardize the B-spline basis at each fixed time t and denote the resulting basis by $\mathbf{B}_k(x_k) = (B_{k,1}(x_k), \ldots, B_{k,K+q}(x_k))^\top$. Let K_1 and K_2 be the numbers of the interior knots for t and x_k, respectively. To simplify notation, let $X_{ij,p+1} = Y_{i(j-1)}(0)$, $m_{p+1}(t,\cdot) = m_y(t,\cdot)$ and $x_{p+1} = y_{t-1}$, and denote the $K_1 \times K_2$ dimensional tensor-product B-spline basis functions by $\mathbf{B}_k(t, x_k) = \mathbf{b}_0(t) \otimes \mathbf{B}_k(x_k)$. Then, the additive component functions in Model (2.2) can be approximated by $m_0(t) \approx \boldsymbol{\alpha}_0^\top \mathbf{b}_0(t)$ and $m_k(t, x_k) \approx \boldsymbol{\alpha}_k^\top \mathbf{B}_k(t, x_k)$, $k = 1, 2, \ldots, p+1$, and the coefficients $\boldsymbol{\alpha}_0 = (\alpha_{0,1}, \ldots, \alpha_{0,K_1+q})^\top$ and $\boldsymbol{\alpha}_k = (\alpha_{k,1}, \ldots, \alpha_{k,(K_1+q)(K_2+q)})^\top$ can be estimated by $\arg\min_{(\boldsymbol{\alpha}_0,\ldots,\boldsymbol{\alpha}_{p+1})} \sum_{i=1}^n \sum_{j=1}^T \{Y_{ij} - \boldsymbol{\alpha}_0^\top \mathbf{b}_0(t_{ij}) - \sum_{k=1}^{p+1} \boldsymbol{\alpha}_k^\top \mathbf{B}_k(t_{ij}, X_{ijk})\}^2$. Together with the mean-zero identification conditions, the initial spline estimators are $\widehat{m}_{0,\mathrm{I}}(t) = \widehat{\boldsymbol{\alpha}}_0^\top \mathbf{b}_0(t) + (nT)^{-1} \sum_{i=1}^n \sum_{j=1}^T \sum_{k=1}^{p+1} \widehat{\boldsymbol{\alpha}}_k^\top \mathbf{B}_k(t_{ij}, X_{ijk})$, and $\widehat{m}_{k,\mathrm{I}}(t, x_k) = \widehat{\boldsymbol{\alpha}}_k^\top \mathbf{B}_k(t, x_k) - (nT)^{-1} \sum_{i=1}^n \sum_{j=1}^T \widehat{\boldsymbol{\alpha}}_k^\top \mathbf{B}_k(t_{ij}, X_{ijk})$ for $k = 1, \ldots, p+1$.

Local linear kernel smoothing estimation of component functions. Let $\widehat{Y}_{ij,0} = Y_{ij} - \sum_{k=1}^{p+1} \widehat{m}_{k,\mathrm{I}}(t_{ij}, X_{ijk})$ and $\widehat{Y}_{ij,k} = Y_{ij} - \widehat{m}_{0,\mathrm{I}}(t_{ij}) - \sum_{\ell \neq k} \widehat{m}_{\ell,\mathrm{I}}(t_{ij}, X_{ij\ell})$. Then, the local linear estimators $\widehat{m}_{0,\mathrm{II}}(t) = \widehat{\beta}_0$ and $\widehat{m}_{k,\mathrm{II}}(t, x_k) = \widehat{\beta}_k$, respectively, are obtained by $\arg\min_{(\beta_0, b_0)} \sum_{i=1}^n \sum_{j=1}^T \big(\widehat{Y}_{ij,0} - \beta_0 - (t_{ij} - t)b_0/h_0\big)^2 K_{h_0}(t_{ij} - t)$ and $\arg\min_{(\beta_1, \mathbf{b}_k)} \sum_{i=1}^n \sum_{j=1}^T \big\{\widehat{Y}_{ij,k} - \beta_k - \big((t_{ij} - t)/h_{k,t}, (X_{ijk} - x_k)/h_{k,x}\big)\mathbf{b}_k\big\}^2 \times K_{h_{k,t}}(t_{ij} - t) K_{h_{k,x}}(X_{ijk} - x_k)$, where $K_h(u) = K(u/h)/h$, h_0 and $(h_{k,t}, h_{k,x})$ are bandwidths for $m_0(t)$ and $m_k(t, x_k)$, respectively. Then, the final estimators of the additive functions are $\widehat{m}_0(t) = \widehat{m}_{0,\mathrm{II}}(t)$ and $\widehat{m}_k(t, x_k) = \widehat{m}_{k,\mathrm{II}}(t, x_k), k = 1, \ldots, p+1$.

Lemma 4.1 reports the theoretical properties of the SPBK estimation of the additive functions, which requires some regular conditions Assumptions A1-A5 in Section S.1.3 of the SM. Define $\kappa_j = \int_{-\infty}^{\infty} u^j K(u) du$, $\nu_j = \int_{-\infty}^{\infty} u^j K^2(u) du$, $\mathrm{bias}_k(t, x_k) = \kappa_2 \{h_{k,t}^2 \partial^2 m_k(t, x_k)/\partial t^2 + h_{k,x}^2 \partial^2 m_k(t, x_k)/\partial x_k^2\}/2$, $k = 1, \ldots, p+1$ and $\mathrm{bias}_0(t) = \kappa_2 m_0''(t) h_0^2/2$. For two positive sequences a_n and b_n, $a_n \asymp b_n$ means that a_n/b_n is bounded away from zero and infinity.

Lemma 4.1. *Under Assumptions A1-A5 of the SM, if* $h_0 \asymp (nT)^{-1/5}$, $h_{k,t} \asymp h_{k,x} \asymp (nT)^{-1/6}$, $K_1 \asymp K_2 \asymp (nT)^{2/9} \log(nT)$, *then (i)* $\sqrt{nTh_0}\,\{\widehat{m}_0(t) - m_0(t) - \mathrm{bias}_0(t)\} \to N\left(0, \sigma^2 \nu_0 \gamma(t)\right)$; *(ii)* $\sqrt{nTh_{k,t}h_{k,x}}\,\{\widehat{m}_k\,(t, x_k) - m_k\,(t, x_k) - \mathrm{bias}_k(t, x_k)\} \to N(0, \sigma^2 \nu_0^2/f(t, x_k))$.

Lemma 4.1 indicates that the SPBK estimators of the univariate and bivariate functions, $\widehat{m}_0(t)$ and $\widehat{m}_k\,(t, x_k)$, converge to the normal distributions at rates $\sqrt{nTh_0}$ and $\sqrt{nTh_{k,t}h_{k,x}}$, respectively. If $h_0 \asymp (nT)^{-1/5}$, then $\widehat{m}_0(t)$ achieves the optimal convergence rate $(nT)^{-2/5}$ and $\widehat{m}_k\,(t, x_k)$ attains the optimal convergence rate $(nT)^{-1/3}$ for $h_{k,t} \asymp h_{k,x} \asymp (nT)^{-1/6}$.

5 Theoretical Properties

We first introduce some notations and assumptions. Let $\widetilde{\mathbf{X}}_{it} = \left(\mathbf{X}_{it}^\top, Y_{i,t-1}(0)\right)^\top$ for $1 \leqslant i \leqslant N, 1 \leqslant t \leqslant T$. Under Assumption 5.1, Lemmas S.14-S.15 of the SM show that the weight solution to problem (3.3) (or (3.10)) uniquely exists.

Assumption 5.1. *There exists a binary variable* $S_i \in \{0, 1\}$ *such that* $\mathbb{P}(S_i = 1 | D_i = 0) > 0$, $\mathrm{E}(\widetilde{\mathbf{X}}_{it} | D_i = 0, S_i = 1) = \mathrm{E}(\widetilde{\mathbf{X}}_{it} | D_i = 1)$, *and* $\mathrm{Var}(\widetilde{\mathbf{X}}_{it} | D_i = 1, S_i = 1)$ *is of full rank.*

If systematic differences exist in the covariates $\widetilde{\mathbf{X}}_{it}$ between the treated and control groups, i.e., $\mathrm{E}(\widetilde{\mathbf{X}}_{it} | D_i = 0) \neq \mathrm{E}(\widetilde{\mathbf{X}}_{it} | D_i = 1)$, direct comparisons between the treated and control groups are improper. Assumption 5.1 is essentially a common support or overlap condition, which is required to ensure that the control group always contains a comparable portion of $\widetilde{\mathbf{X}}_{it}$ with the same expectation as in the treated group. In practice, Assumption 5.1 means that the control group contains enough diversity to mimic the treated group. For instance, when we study the impact of air pollution alerts, this assumption requires that the control pool contains enough air pollution stations with similar environmental features so that a weighted average (a 'Synthetic Control') can be constructed.

Recall that the standard result of the EL (Owen, 2001) requires $\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 0) = \mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 1)$ to obtain $w_{it}^* = O_p\left(N_{co}^{-1}\right)$ for $i = 1, \cdots, N_{co}$ as $N_{co}, N_{tr} \to \infty$. However, the mean equivalence condition in $\widetilde{\mathbf{X}}_{it}$ between the treated and control groups is often violated in some SC situations, which prevents the weights on control units from achieving a uniform order of $O_p(N_{co}^{-1})$. For example, meteorological conditions like wind speed and humidity may differ in different cities. In such cases, the EL weights exhibit a 'singularity' where one or two units may capture a significant portion of the total mass. Thus, the following Assumption 5.2 adapted from Ghosh et al. (2023) is used to establish the bounds of w_{it}^* and w_{it}^j when $\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 0) \neq \mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 1)$. For a vector $\theta \in \Theta$ with $\|\theta\| = 1$, define $\xi_i(\theta) = \theta^\top [\mathbf{Z}_{it,\mathcal{C}} - \mathrm{E}(\mathbf{Z}_{it,\mathcal{C}})]$ for $1 \leqslant i \leqslant N_{co}$. Let $\xi_{\min}^-(\theta) = \min_{i\in\mathcal{C}}\{\xi_i(\theta)1_{\xi_i(\theta)<0}\}$, $\xi_{\max}^+(\theta) = \max_{i\in\mathcal{C}}\{\xi_i(\theta)1_{\xi_i(\theta)>0}\}$ and $\xi_{\min,2}^-(\theta) = \min_{i\in\mathcal{C}}\{\{\xi_i(\theta)1_{\xi_i(\theta)<0}\}\backslash\{\xi_{\min}^-(\theta)\}\}$.

Assumption 5.2. *For any $\theta \in \Theta$ with $\|\theta\| = 1$, $\mathrm{Var}(\xi_i(\theta)) > 0$ and $\sum_{i-j=-\infty}^{\infty} \mathrm{cov}(\xi_i(\theta), \xi_j(\theta)) < \infty$. There exist $\gamma > 0$, $0 < \delta < 1$ and a non-random sequence $b_{N_{co}} \to \infty$ such that (i) $b_{N_{co}}^{\gamma+2} = o(N_{co})$, $b_{N_{co}}^{4\gamma} = o(N_{co})$, $|\xi_{\min}^-(\theta)| = O_p\left(b_{N_{co}}\right) = |\xi_{\max}^+(\theta)|$; (ii) $\mathrm{P}(|\xi_{\min}^-(\theta) - \xi_{\min,2}^-(\theta)| \geqslant b_{N_{co}}^{-\gamma}) \to 1$ as $N_{co} \to \infty$ and (iii) $b_{N_{co}}^{\gamma+2}\mathrm{P}(|\xi_i(\theta)| > b_{N_{co}}^{1-\delta}) \to 0$ as $N_{co} \to \infty$.*

Assumption 5.3. *The number of knots satisfies $K \asymp (N_{co}T)^{2/9}\log(N_{co}T)$. The bandwidths satisfy $h_0 \asymp (N_{co}T)^{-1/5}$, $N_{tr}h_{k,t}^4 \to 0$, $N_{tr}h_{k,x}^4 \to 0$, $N_{co}h_{k,t}^4 \to 0$, $N_{co}h_{k,x}^4 \to 0$ and $h_{k,t} \asymp h_{k,x} \gtrsim (N_{co}T)^{-1/6}$.*

Assumption 5.2 puts constraints on the growth rate of the extreme order statistics $|\xi_{\min}^-(\theta)|$ and $|\xi_{\min}^-(\theta) - \xi_{\min,2}^-(\theta)|$ and the decay rate of the tail probability $\mathrm{P}(|\xi_i(\theta)| > b_{N_{co}}^{1-\delta})$. The $b_{N_{co}}$ bounds the maximum difference between the confounders and their mean in the control group. Intuitively, Assumption 5.2 ensures that the tail of the control group distribution is thick enough to support this singularity without numerical failure and also guarantees that the largest weight (the 'singular weight') grows at a predictable rate relative to the other weights, allowing for consistency even under misspecification. Many distributions of $\xi_i(\theta)$ would satisfy Assumption 5.2 by the standard results of extreme value theory (Leadbetter et al., 1983) such as the Gaussian, exponential, and Laplace distributions. For example, when $\{\xi_i(\theta)\}$ are Gaussian distributed, Assumption 5.2 is satisfied by taking $b_n = \sqrt{\log n}, \gamma > 1/2$, and $0 < \delta < 1/2$. Assumption 5.3 specifies the required bandwidths for negligible estimation errors of the unknown functions in weight estimation. We show in the SM that under Assumption 5.3, the estimation errors of the unknown functions are of order $o_p(N_{co}^{-1/2})$, which do not affect the leading estimation error of the EL weights.

Next, we derive the bounds of the weights w_{it}^j and w_{it}^* under $\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 0) = \mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 1)$ and $\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 0) \neq \mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 1)$.

Lemma 5.4. *Under Assumptions 2.1-5.1,5.3 and A1-A5, as N_{co} and $N_{tr} \to \infty$, (i) if $\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 0) = \mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 1)$, then $w_{it}^j = O_p\left(N_{co}^{-1}\right)$ uniformly for $i = 1, \cdots, N_{co}, j = 1, \cdots, N_{tr}$ and $w_{it}^* = O_p(N_{co}^{-1})$ uniformly for $i = 1, \cdots, N_{co}$; (ii) if $\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 0) \neq \mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 1)$, under additional Assumption 5.2, then $\max\{w_{it}^j : i \neq i_{\min}\} = O_p(b_{N_{co}}^{\gamma+1}N_{co}^{-1})$, $w_{i_{\min},t}^j = O_p(b_{N_{co}}^{-1})$ for $i = i_{\min}$, $\max\{w_{it}^* : i \neq i_{\min}\} = O_p(b_{N_{co}}^{\gamma+1}N_{co}^{-1})$ and $w_{i_{\min},t}^* = O_p(b_{N_{co}}^{-1})$ for $i = i_{\min}$, where $i_{\min}$ is the index of $\xi_{\min}(\theta^*)$.*

Lemma 5.4 suggests that the semiparametric estimation error of the additive functions does not affect the leading order of the estimated weights. When $\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 0) = \mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 1)$, the result is the same as that of the standard result of the EL (Qin and Lawless, 1994). If $\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 0) \neq \mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 1)$, the weights are larger than N_{co}^{-1} and the largest weight is of order $O_p(b_{N_{co}}^{-1})$. The following theorem establishes the convergence rate of $\widehat{\tau}_t$ to τ_t.

Theorem 5.5. *Under Assumptions 2.1-5.3, A1-A5 and Model (2.2), the matching equations (3.4) and (3.5) ((3.8) and (3.9)) can be attained with probability approaching 1 and there exists a unique solution for the EL weights $\left\{w_{it}^j\right\}_{i=1}^{N_{co}}$ ($\{w_{it}^*\}_{i=1}^{N_{co}}$). (i) If $\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 0) = \mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 1)$, then $\widehat{\tau}_t \xrightarrow{p} \tau_t$ as N_{tr} and $N_{co} \to \infty$, and $|\widehat{\tau}_t - \tau_t| = O_p(N_{co}^{-1/2} + N_{tr}^{-1/2})$. (ii) If $\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 0) \neq \mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 1)$, then under an additional assumption $b_{N_{co}}^{2\gamma+2} = o_p(N_{co})$, the DSC estimator $|\widehat{\tau}_t - \tau_t| = O_p(b_{N_{co}}^{\gamma+1}N_{co}^{-1/2} + b_{N_{co}}^{-1} + N_{tr}^{-1/2})$.*

Theorem 5.5 shows that $\widehat{\tau}_t$ is consistent with τ_t for a fixed t when $m_y(t,\cdot)$ is either linear or non-linear in the lagged outcome. When $\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i=0)=\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i=1)$, the estimation error in $\widehat{\tau}_t$ depends only on the sample sizes N_{tr} and N_{co}. When $\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i=0)\neq\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i=1)$, the estimation error in $\widehat{\tau}_t$ not only depends on the sample sizes N_{tr} and N_{co}, but also on $b_{N_{co}}$ which bounds the maximum difference between the confounding variables and their mean in the control group. In addition, the semiparametric estimation error of the additive functions does not affect the accuracy of $\widehat{\tau}_t$ in the leading order.

Remark 5.6. For large T, if $\overline{\lim}_{t\to\infty}|\rho(t)|<1$, where $\rho(t)=\sup_y \partial m_y(t,y)/\partial y$ or $\rho(t)=m_y(t)$ when $m_y(t,\cdot)$ is linear in the lagged outcome such that $m_y(t,Y_{i,t-1}(0))=m_y(t)Y_{i,t-1}(0)$, then the results in Theorem 5.5 still hold for $(T-T_0)\to\infty$. In the case of $\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i=0)=\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i=1)$, under (a) $\overline{\lim}_{t\to\infty}|\rho(t)|>1$, $(T-T_0)\log(T-T_0)=o(\log N_{co})=o(\log N_{tr})$ or (b) $\overline{\lim}_{t\to\infty}|\rho(t)|=1$, $O\big(\sum_{s=T_0+1}^{T-1}\prod_{t=s+1}^{T}\rho(t)\big)=o(N_{tr}^{1/2})=o(N_{co}^{1/2})$, we have $\widehat{\tau}_T\xrightarrow{p}\tau_T$ as $N_{tr},N_{co},(T-T_0)\to\infty$. In the case of $\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i=0)\neq\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i=1)$, under (c) $\overline{\lim}_{t\to\infty}|\rho(t)|>1$, $(T-T_0)\log(T-T_0)=o(\log N_{co})=o(\log b_{N_{co}})=o(\log N_{tr})$ or (d) $\overline{\lim}_{t\to\infty}|\rho(t)|=1$, $O\big(\sum_{s=T_0+1}^{T-1}\prod_{t=s+1}^{T}\rho(t)\big)=o(b_{N_{co}}^{-\gamma-1}N_{co}^{1/2})=o(b_{N_{co}})=o(N_{tr}^{1/2})$, we have $\widehat{\tau}_T\xrightarrow{p}\tau_T$ as $N_{tr},N_{co},(T-T_0)\to\infty$. Note that for $\overline{\lim}_{t\to\infty}|\rho(t)|=1$, the specific order of $\sum_{s=T_0+1}^{T-1}\prod_{t=s+1}^{T}\rho(t)$ depends on the form of $\rho(t)$. For example, the order is $O(T\log T)$ for $\rho(t)=1+1/t$ and $O((T-T_0)(T+T_0)/T)$ for $\rho(t)=1-1/t$.

6 Hypothesis Testing

This section considers tests for three key aspects of the semiparametric DSC approach: the linearity of $m_y(t,\cdot)$ in the lagged outcome, the unconfoundedness assumption and the placebo test for the significance of the estimated effect $\widehat{\tau}_t$.

6.1 Testing linearity of $m_y(t,\cdot)$

As introduced in Section 3, the linearity in the lagged outcome allows us to directly estimate $\mu_t(0)$ using weighted averages of control group outcomes. In contrast, in the non-linear lagged outcome case, we need to build the synthetic controls for each treated unit, which is much more involved than the linear case. Hence, the linearity in the lagged outcome would greatly reduce the complexity of the DSC process, and it is important to test for it first.

We formulate a test of whether $m_y(t,\cdot)$ is linear in the lagged outcome $Y_{i,t-1}(0)$ or not. Similar testing procedures can be applied to test whether the effect of the covariate $\mathbf{X}_t$ is linear or not, although they have no implications on the way the DSC weights are calculated. Denote the additive components by $\mathbf{m}(t,\widetilde{\mathbf{X}}_t)=m_0(t)+\sum_{k=1}^{p}m_k(t,x_{t,k})+m_y(t,y_{t-1})$ in Model (2.2). Then, the null linearity hypothesis is $H_0:\mathbf{m}(t,\widetilde{\mathbf{x}}_t)=m_0(t)+\sum_{k=1}^{p}m_k(t,x_{t,k})+m_y(t)\,y_{t-1}$ a.s. on $[0,1]\times\mathcal{X}$, where $\mathcal{X}$ is the support of covariates $\widetilde{\mathbf{X}}_t$. Let $\mathcal{Y}$ be the support of covariates $Y_{t-1}(0)$. The null hypothesis can be expressed as

$$H_0:\int_0^1\int_{\mathcal{Y}}\left[(m_0(t)+m_y(t,y_{t-1}))-\big(m_0(t)+m_y(t)y_{t-1}\big)\right]^2dy_{t-1}dt=0. \tag{6.1}$$

The inclusion of the baseline function $m_0(t)$ is due to the mean-zero identification condition of the additive model. A natural way to formulate the test statistic is to build a distance measure between the nonparametric estimates for $m_0(t),m_y(t,y_{t-1})$ and their counterpart under H_0. Let $\check{m}_0(t)$ and $\check{m}_y(t)$ be estimators of $m_0(t)$ and $m_y(t)$ under H_0. The L_2-distance-based test statistic is

$$\mathrm{ISE}_n=\int_0^1\int_{\mathcal{Y}}\left[\widehat{m}_y(t,y_{t-1})+\widehat{m}_0(t)-\big(\check{m}_0(t)+\check{m}_y(t)y_{t-1}\big)\right]^2dy_{t-1}dt. \tag{6.2}$$

The following proposition presents a central limit theorem for $\mathcal{I}_n=\mathrm{ISE}_n-\mathrm{E}_{\mathrm{H}_0}(\mathrm{ISE}_n)$, where $\mathrm{E}_{\mathrm{H}_0}(\mathrm{ISE}_n)$ is the expectation of ISE under H_0.

Algorithm 1 Bootstrap test for linear lagged outcome.

Input: Data $\{Y_{it}, \mathbf{X}_{it}, t = 1, \cdots, T, i = 1, \cdots, N_{co}\}$.

1. Perform the SPBK method to get the estimation of $m_0(\cdot)$ and $m_k(\cdot, \cdot)$ for $k = 1, \ldots, p + 1$ and calculate the residuals $\widehat{\varepsilon}_{ij} = Y_{ij} - \widehat{\mathbf{m}}(t_{ij}, \widetilde{\mathbf{X}}_{ij}) = Y_{ij} - \widehat{m}_0(t_{ij}) - \sum_{k=1}^{p+1} \widehat{m}\,(t_{ij}, X_{ij,k})$.
2. Under H_0, obtain the estimates $\check{m}_y(\cdot)$, $\check{m}_0(\cdot)$, $\check{m}_k(\cdot, \cdot)$ for $k = 1, \ldots, p$ and compute the ISE by (6.2).
   for $b = 1, \ldots, B$ **do**
   (a) Generate sample $\{Y^*_{ij}, \mathbf{X}_{ij}\}$ by $\varepsilon^*_{ij} = \widehat{\varepsilon}_{ij}\eta_{ij}$, $\eta_{ij} \overset{i.i.d.}{\sim} N(0, 1)$ and $Y^*_{ij} = \check{m}_y(t_{ij})Y^*_{ij-1} + \check{m}_0(t_{ij}) + \sum_{k=1}^{p} \check{m}_k(t_{ij}, X_{ij,k}) + \varepsilon^*_{ij}$.
   (b) Given $\{Y^*_{ij}, \mathbf{X}_{ij}\}$, compute the bootstrap test statistic by Steps 1-2 $\mathrm{ISE}^*_b = \int_0^1 \int_{\mathcal{Y}} [\widehat{m}^*_0(t) + \widehat{m}^*(t, y_{t-1}) - \check{m}^*_0(t) - \check{m}^*_y(t)y_{t-1}]^2 dy_{t-1}dt$.

   end
3. Compute the test p-value by $B^{-1} \sum_{b=1}^{B} I(\mathrm{ISE}^*_b \geqslant \mathrm{ISE})$. Reject H_0 when the p-value is smaller than a prespecified significant level α.

Output: The p-value.

Proposition 6.1. *Under Assumptions A1-A5, $h_{1,t} \asymp h_{1,y} \asymp (nT)^{-1/6} \log^{-1}(nT)$, and the null hypothesis H_0 (6.1), as $nT \to \infty$, then $nTh_{1,t}h_{1,y}\mathcal{I}_n \xrightarrow{d} N(0, V)$, where $V = 2\nu_0^2 \int_0^1 \int_{y_{t-1}\in\mathcal{Y}} [\sigma^2/f\,(t, y_{t-1})]^2 dy_{t-1}dt$.*

The asymptotic normality in Proposition 6.1 can be used to formulate an asymptotic test for H_0 after we estimate $\mathrm{E}_{\mathrm{H}_0}(\mathrm{ISE}_n)$ and V. However, the quality of the test requires larger sample sizes (Härdle and Mammen, 1993) and the asymptotic variance involves unknown quantities. To enhance the performance of the test in finite samples, we employ a bootstrap procedure given in Algorithm 1, which is shown to be consistent in Theorem S.3 of the SM. In the pollution alert study in Section 8, we applied the bootstrap procedure to test whether the effect of the lagged $\mathrm{PM}_{2.5}$ is linear or not, which rejected the linearity hypothesis.

6.2 Unconfoundedness Assumption Assessment

As $Y_{it}(0)$ are observed for both the treated and control groups before the treatment and we do not use the pre-treatment outcomes to estimate the ATT, we can use the pre-treatment data to evaluate a part of the unconfoundedness Assumption 2.1(ii). We extend the notion of τ_t to the pre-treatment period by defining $\tau_t = \mathrm{E}(Y_{it} - Y_{it}(0)|D_i = 1) = 0$ for $1 \leqslant t \leqslant T_0$. The DSC estimate $\widehat{\tau}_t$ can also be extended to cover the entire period, that is $\widehat{\tau}_t = N_{tr}^{-1} \sum_{j\in\mathcal{T}}(Y_{jt} - \sum_{i\in\mathcal{C}} w^j_{it}Y_{it})$ for general $m_y(t, \cdot)$, with w^j_{it} also obtained by solving (3.3).

The DSC estimator under Model (2.2) in the pre-treatment period should satisfy $\mathrm{E}\,(\widehat{\tau}_t) = 0$ for $t \leqslant T_0$. As a way to check the unconfoundedness assumption, we utilize the availability of data for both the treatment and control groups in the pre-treatment period to test $H^t_0 : \mathrm{E}\,(\widehat{\tau}_t) = 0$ for $t \leqslant T_0$. Denote the conditional variance by $\sigma^2_{\widehat{\tau}_t} = \mathrm{Var}\,(\widehat{\tau}_t|\mathcal{F}_t)$, where $\mathcal{F}_t$ is the σ-field generated by $\{\mathbf{X}_{i1}, \cdots, \mathbf{X}_{it}, Y_{i1}, \cdots, Y_{i,t-1}\}$. Then $\widehat{\tau}_t/\widehat{\sigma}_{\widehat{\tau}_t}$ is the test statistic we use, where $\widehat{\sigma}^2_{\widehat{\tau}_t}$ is an estimator of $\sigma^2_{\widehat{\tau}_t}$ as given in Section S.2.2 of the SM. Similar formulations for the case $m_y(t, \cdot)$ being linear in the lagged outcome are outlined in SM Section S.2.2. As we have to test H^t_0 for $t = 1, \cdots, T_0$ repeatedly, the issue of multiplicity in hypothesis testing arises. To mitigate the issue, we employ the FDR procedure (Benjamini and Yekutieli, 2001) to control the FDR; see Section 8 for a demonstration of the test process for the pollution alert study.

6.3 Placebo Test

To assess the significance of the estimated treatment effects, we utilize a method known as the placebo test similar to that of Abadie et al. (2010). It examines whether the estimated treatment effect for the treatment group is significantly different from the pseudo effect for a randomly chosen untreated group. If there are no treatment effects, the empirical distribution of the pseudo effects for repeatedly randomly sampled untreated groups can be used to estimate the distribution of $\widehat{\tau}_t$ under the null.

To conduct the placebo test, we repeatedly choose random subsets of size N_{tr} from the control group $\mathcal{C}$ to form placebo treatment groups. Denote the indexes of the selected placebo treated group in the k-th round of the placebo sampling by $\mathcal{T}^{(k)}$ and the control group $\mathcal{C}^{(k)} = \mathcal{C}\backslash\mathcal{T}^{(k)}$. Then, we apply the proposed DSC method on the placebo treatment group $\mathcal{T}^{(k)}$ and the control group $\mathcal{C}^{(k)}$. The placebo estimate for the pseudo-effect at time t for general $m_y(t,\cdot)$ is $\widehat{\tau}_t^{(k)} = N_{tr}^{-1}\sum_{j\in\mathcal{T}^{(k)}} Y_{jt} - N_{tr}^{-1}\sum_{j\in\mathcal{T}^{(k)}}\sum_{i\in\mathcal{C}^{(k)}} w_{it}^{j(k)} Y_{it}$, where $\{w_{it}^{j(k)}\}$ are the EL DSC weights for the j^{th} subject in the treatment group $\mathcal{T}^{(k)}$. Similar formulations for the linear lagged outcome ($m_y(t,\cdot)$ being linear in $Y_{i,t-1}(0)$) can be formulated whose details are available in S.2.3 of the SM.

We repeat the above process K times and obtain estimates for the placebo effects $\{\widehat{\tau}_t^{(k)}\}_{k=1}^K$. Note that the conditional variance of the ATT estimator $\widehat{\tau}_t$, denoted by σ_t^2, depends on the weights $\{w_{it}^j\}_{i=1,j=1}^{N_{co},N_{tr}}$ which would be different for different pairs of treatment/control groups in the placebo test. To ensure that the variance of the estimated pseudo-effects $\widehat{\tau}_t^{(k)}$ is the same as that of $\widehat{\tau}_t$, it is necessary to normalize the placebo estimates $\widehat{\tau}_t^{(k)}$ by the corresponding deviations $\sigma_t^{(k)}$. Let $z_t^{(k)} = \widehat{\tau}_t^{(k)} \times \left(\widehat{\sigma}_t^{(k)}\right)^{-1} \times \widehat{\sigma}_t$ be the normalized synthetic control estimators, where $\widehat{\sigma}_t^{(k)}$ and $\widehat{\sigma}_t$ are the estimators of the standard deviations $\sigma_t^{(k)}$ and σ_t, which can be referred to in Section S.2.3 of the SM. Then the distribution of $\widehat{\tau}_t$ under the null hypothesis can be approximated by the empirical distribution of $\{z_t^{(k)}\}_{k=1}^K$ at time t. If negative treatment effects are expected, then the alternative hypothesis is $\tau_t < 0$ and the approximate p-value is $K^{-1}\sum_{k=1}^K 1(z_t^{(k)} \leqslant \widehat{\tau}_t)$.

7 Simulation Study

We report results of simulation experiments which were designed to confirm the theoretical findings of the proposed semiparametric DSC approach. We experimented with a general case (7.1) where $m_y(t,\cdot)$ was non-linear in the lagged outcome and two special cases (7.2)-(7.3) with linear lagged outcome functions,

$$Y_{it}(0) = 5\left(u_t - 0.6\right)^2 + 2u_t \sin\left(0.1\pi Y_{i,t-1}(0)\right) + 2e^{u_t} X_{it,1} + 0.5e^{1-u_t} X_{it,2}^2 + \varepsilon_{it} \tag{7.1}$$

$$Y_{it}(0) = 5\left(u_t - 0.6\right)^2 + 0.5u_t^2 Y_{i,t-1}(0) + 2e^{u_t} X_{it,1} + 0.5e^{1-u_t} X_{it,2}^2 + \varepsilon_{it}, \tag{7.2}$$

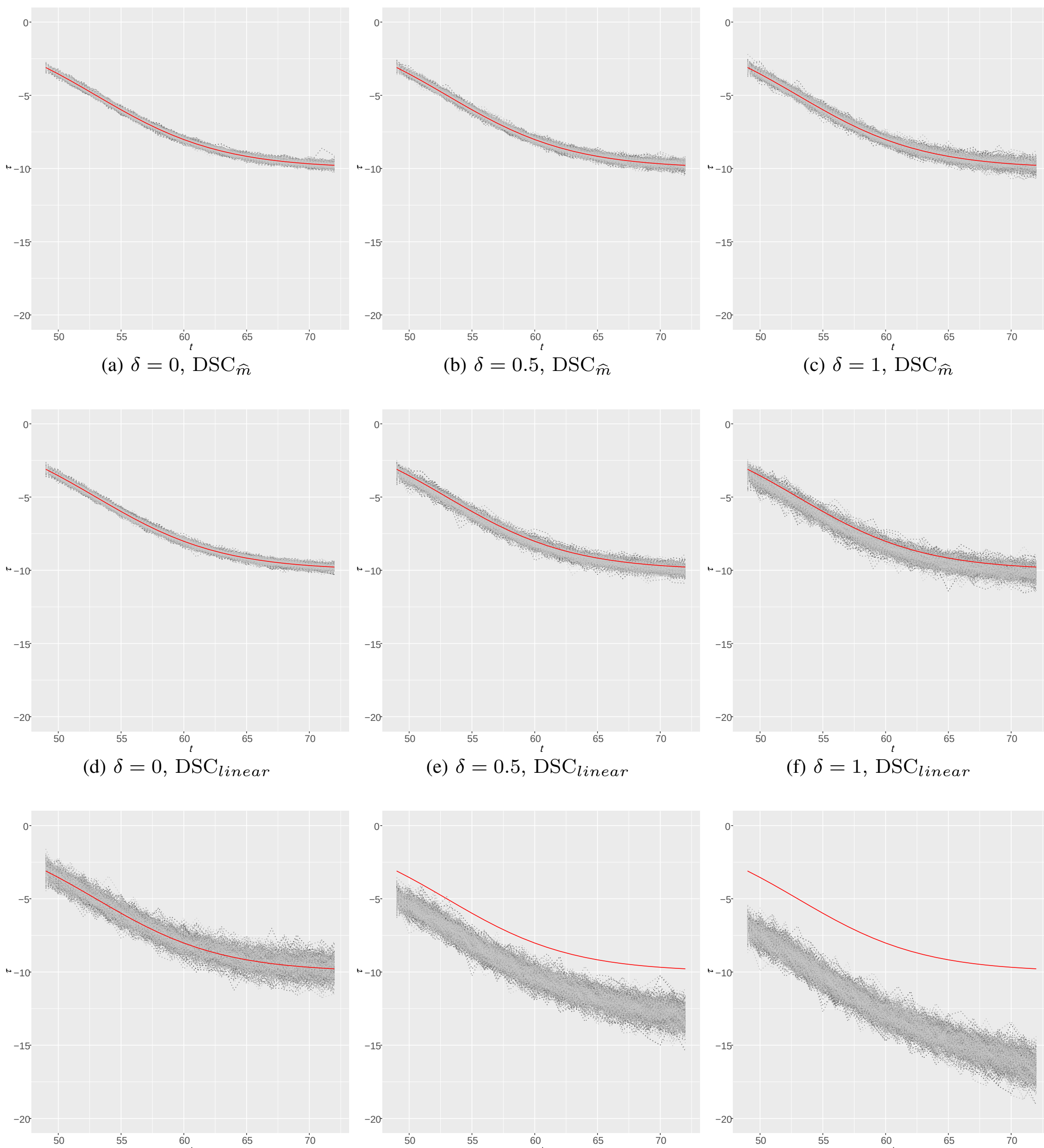
$$Y_{it}(0) = 5\left(u_t - 0.6\right)^2 + 0.5u_t^2 Y_{i,t-1}(0) + 2e^{u_t} X_{it,1}\sin(X_{it,1}) + 0.5e^{1-u_t}\cos(X_{it,2}^2) + \varepsilon_{it}, \tag{7.3}$$

$$Y_{it}(1) = Y_{it}(0) + \tau_{it},\ \varepsilon_{it} \sim N\left(0, \sigma_\varepsilon^2\right) \quad \text{and} \quad \tau_{it} \sim N\left(\tau_t, \sigma_\tau^2\right), \tag{7.4}$$

where $u_t = t/T$, $\tau_t = -10/\left[1 + \exp\left\{-0.2\left(t - T_0 - 5\right)\right\}\right]$ was the underlying time-varying treatment effects, $\mathbf{X}_{it} = (X_{it,1}, X_{it,2})^\top$ were the time-varying confounders with $p = 2$, and $\sigma_\varepsilon^2 = \sigma_\tau^2 = 0.5$. Then, the observed outcomes were generated by $Y_{it} = Y_{it}(0)I(t \leqslant T_0) + Y_{it}(1)I(t > T_0)$ for $i \in \mathcal{T}$ and $Y_{it} = Y_{it}(0)$ for $i \in C$ with pre-treatment span $T_0 = 48$ and total time $T = 72$. The treated and control sample sizes were $N_{tr} = 100, 400$ and $N_{co} = 100, 400$ for each experimental design, respectively. Note that $\mathbf{X}_{it}$ are non-linear for both the non-linear and linear lagged outcome cases, and Model (7.3) exhibited greater non-linearity in $\mathbf{X}_{it}$ compared to (7.2). An entirely linear model (S.1) case (both effects of $\mathbf{X}_{it}$ and $Y_{i,t-1}(0)$ are linear) was also considered later to compare fairly with the linear DSC (Zheng and Chen, 2024).

The $\mathbf{X}_{it}$ in the treated and control groups were set to have either identical or non-identical means, to mimic Cases (i) and (ii) of Lemma 5.4. Specifically, $\mathbf{X}_{it}$ followed $N(\boldsymbol{\mu}, \mathbf{I}_p)$ in the treated group and a mixture distribution $0.5N(\boldsymbol{\mu}, \mathbf{I}_p) + 0.5N(\boldsymbol{\mu} + \mathbf{d}, \mathbf{I}_p)$ in the control group, where $\boldsymbol{\mu} = (1,1)^\top$, $\mathbf{d} = \delta(1,1)^\top$ for $\delta = 0$, 0.5 and 1. In such a design, a subset of the control group shared the same mean of $\mathbf{X}_{it}$ as that in the treated group, meeting Assumption 5.1.

In each setting, we applied the proposed semiparametric DSC, the linear DSC (Zheng and Chen, 2024) and the naive approach of taking raw differences between the treated and control groups to estimate τ_t. Each setting was repeated 500 times. For Model (7.1)(non-linear lagged outcome), the proposed semiparametric DSC estimated τ_t by (3.7). For Models (7.2) and (7.3)(linear lagged outcome), the proposed DSC estimated τ_t by (3.11).



(a) $\delta = 0$, $\mathrm{DSC}_{\widehat{m}}$ (b) $\delta = 0.5$, $\mathrm{DSC}_{\widehat{m}}$ (c) $\delta = 1$, $\mathrm{DSC}_{\widehat{m}}$

(d) $\delta = 0$, DSC_{linear} (e) $\delta = 0.5$, DSC_{linear} (f) $\delta = 1$, DSC_{linear}

(g) $\delta = 0$, Raw-difference (h) $\delta = 0.5$, Raw-difference (i) $\delta = 1$, Raw-difference

Figure 1 The estimated ATT (the grey dashed lines) by the proposed semiparametric DSC ($\mathrm{DSC}_{\widehat{m}}$), the linear DSC ($\mathrm{DSC}_{linear}$) and the Raw-difference for Model (7.2) with $N_{tr} = N_{co} = 100$ and different disparity values (δ) in the means between the treatment and control groups based on 500 simulations. The true treatment effects are shown in the red solid curves.

Figure 1 displays the estimated ATT along the time in 500 replications by the semiparametric DSC, the linear DSC and the Raw-difference for Model (7.2). It shows that the true τ_t was at the center of all realized $\{\widehat{\tau}_t\}$ by the semiparametric DSC for both $\delta = 0$ and $\delta \neq 0$. However, the estimation of the

Raw-difference was severely biased and deviated further from the true τ_t as the disparity parameter δ (mean shift in $\mathbf{X}_{it}$) increased. It appears that the linear DSC provided biased ATT estimates when $\delta \neq 0$ and the bias got larger as δ increased. Figure S.1 of the SM shows similar results for Model (7.1).

Table 1 The average RMSEs of the estimated ATT over the post-treatment period ($T - T_0$) by the semiparametric DSC using the estimated m (DSC$_{\widehat{m}}$) and the true m (DSC$_m$), the linear DSC (DSC$_{linear}$; Zheng and Chen, 2024) and the Raw-difference (Raw-diff) for Models (7.1) (non-linear in the lagged outcome), (7.2) (linear in the lagged outcome) and (7.3) (linear in the lagged outcome with stronger non-linearity in $\mathbf{X}_t$) with different disparities (δ) in the means between the treatment and control groups.

δ	N_{co}	method	Model (7.1)		Model (7.2)		Model (7.3)	
			$N_{tr}=100$	$N_{tr}=400$	$N_{tr}=100$	$N_{tr}=400$	$N_{tr}=100$	$N_{tr}=400$
$\delta=0$	$N_{co}=100$	DSC$_{\widehat{m}}$	0.1524	0.1210	0.1390	0.1020	0.2249	0.1789
		DSC$_m$	0.1270	0.0932	0.1279	0.0930	0.1282	0.0931
		DSC$_{linear}$	0.1633	0.1280	0.1551	0.1216	0.3360	0.2589
		Raw-diff	0.5065	0.3879	0.5610	0.4241	0.5241	0.4085
	$N_{co}=400$	DSC$_{\widehat{m}}$	0.1530	0.1082	0.1118	0.0665	0.1378	0.0781
		DSC$_m$	0.1103	0.0764	0.1098	0.0647	0.1099	0.0649
		DSC$_{linear}$	0.1878	0.1582	0.1605	0.1151	0.3575	0.2667
		Raw-diff	0.3974	0.2581	0.4301	0.2674	0.4170	0.2593
$\delta=0.5$	$N_{co}=100$	DSC$_{\widehat{m}}$	0.1803	0.1607	0.1782	0.1490	0.2399	0.1656
		DSC$_m$	0.1225	0.0992	0.1248	0.0895	0.1281	0.0913
		DSC$_{linear}$	0.2710	0.2517	0.2983	0.2717	0.4590	0.4064
		Raw-diff	1.4966	1.4662	2.5080	2.4908	0.8508	0.7876
	$N_{co}=400$	DSC$_{\widehat{m}}$	0.1337	0.1163	0.1565	0.1316	0.1457	0.0981
		DSC$_m$	0.1117	0.0786	0.1205	0.0727	0.1119	0.0639
		DSC$_{linear}$	0.2990	0.2755	0.3183	0.3058	0.5062	0.4096
		Raw-diff	1.4295	1.3830	2.4985	2.4854	0.7963	0.7202
$\delta=1$	$N_{co}=100$	DSC$_{\widehat{m}}$	0.1911	0.1864	0.2413	0.2167	0.3479	0.2945
		DSC$_m$	0.1213	0.0966	0.1231	0.0843	0.1280	0.0914
		DSC$_{linear}$	0.3494	0.3448	0.5013	0.4865	0.6929	0.6485
		Raw-diff	2.9583	2.9468	5.1482	5.1227	0.9126	0.8466
	$N_{co}=400$	DSC$_{\widehat{m}}$	0.1468	0.1353	0.2349	0.2072	0.1867	0.0985
		DSC$_m$	0.1035	0.0746	0.1169	0.0721	0.1119	0.0640
		DSC$_{linear}$	0.3532	0.3493	0.5588	0.5438	0.6272	0.5555
		Raw-diff	2.9036	2.8740	5.1249	5.1065	0.8487	0.7825

To provide more details on the performance of the estimators, we report in Table 1 the averaged root mean square errors (RMSEs) of the estimated treatment effects over the post-treatment period by the semiparametric DSC using the estimated m (DSC$_{\widehat{m}}$) and true m (DSC$_m$), linear DSC (DSC$_{linear}$), and the Raw-difference, respectively, for Models (7.1), (7.2) and (7.3). It shows that the proposed semiparametric DSC method consistently yielded the smallest RMSEs in all settings regardless $m_y(t,\cdot)$ being non-linear or linear in the lagged outcome. The superiority of the semiparametric DSC was more apparent when $\mathbf{X}_{it}$ had different means between the treatment and control groups ($\delta \neq 0$), as the proposed DSC method still provided quite accurate estimations of the treatment effects when compared with the other methods, although the level of the estimation error was elevated as δ became larger. The Raw-difference estimator has much larger RMSEs than its two peers, which shows how bad things could be if proper treatment estimators were not used. The inferior performance of the linear DSC even in the linear lagged value case was due to the bias caused by the non-linear effect of the covariates $\mathbf{X}_{it}$ as both the SC and the linear DSC methods from Abadie et al. (2010) to Zheng and Chen (2024) were based on the entire linearity specification. Compared to the linear DSC, the superiority of the semiparametric DSC was more noticeable for Model (7.3) (linear in the lagged outcome), where the non-linearity of $\mathbf{X}_t$ was more pronounced.

Figure S.2 presents the heatmaps of the DSC weights $\{w_{it}^*\}_{i=1}^{N_{co}}{}_{t=1}^{T}$ and the average maximum weight for each post-treatment time t over 500 simulations under $\delta = 0, 0.5, 1$. When $\mathbf{X}_{it}$ had identical means ($\delta = 0$), the majority of the DSC weights appeared to be distributed quite evenly, which demonstrated Lemma 5.4(i). For $\mathbf{X}_{it}$ with non-identical means ($\delta = 0.5$ and 1), the maximum weights at each time were much larger than the other weights, which confirmed Lemma 5.4(ii). The larger order weight would lead to a larger variance of $\widehat{\tau}_t$, which is also the reason for the inferior performance of the DSC when $\delta = 0.5, 1$ in Table 1.

To make fair comparisons with the linear DSC (Zheng and Chen, 2024), we experimented with an entirely linear model (S.1) in SM Section S.3.1. The results summarized in Table S.1 suggest that the performances of the semiparametric DSC and the linear DSC were comparable for entirely linear models. Specifically, for $\delta = 0$ where $\mathbf{X}_{it}$ had identical means in the treated and control groups, the linear DSC was slightly better and vice versa for $\delta \neq 0$.

We also conducted experiments with the time-varying $\delta_t = 0.5 - 1.5(t/T)^2$ where we set $\mathbf{X}_{it}$ in the control group to follow a mixture distribution $0.5N(\boldsymbol{\mu}, \mathbf{I}_p) + 0.5N(\boldsymbol{\mu} + \delta_t(1,1)^\top, \mathbf{I}_p)$ under Models (7.1), (7.2) and (7.3) when there were non-linear effects of $\mathbf{X}_{it}$ or $Y_{i,t-1}(0)$ and the entirely linear model (S.1). The average RMSEs of the estimated ATT over the post-treatment period $(T - T_0)$ are shown in Table 2 and estimated ATT over the post-treatment period for 500 simulations is depicted in Figure 2, from which the performances were similar to those of the fixed δ case. Specifically, the proposed semiparametric DSC performed much better than the linear DSC (DSC$_{linear}$; Zheng and Chen, 2024) for Models (7.1), (7.2) and (7.3) when there were non-linear effects of $\mathbf{X}_{it}$ or $Y_{i,t-1}(0)$. For the entirely linear model (S.1) (linear in $\mathbf{X}_{it}$ and $Y_{i,t-1}(0)$), the performances of the proposed semiparametric DSC and the linear DSC (DSC$_{linear}$; Zheng and Chen, 2024) were similar, with slightly lower RMSEs of the proposed method due to $\delta_t \neq 0$, which was in accordance with the performance in Table S.3.1.

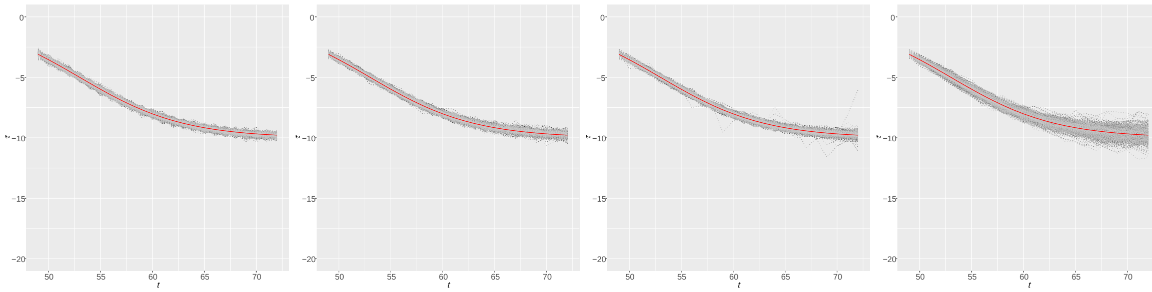
Table 2 The average RMSEs of the estimated ATT over the post-treatment period $(T - T_0)$ by the proposed semiparametric DSC (DSC$_{\widehat{m}}$), the linear DSC (DSC$_{linear}$; Zheng and Chen, 2024) and the Raw-difference (Raw-diff) for Model (7.1) (non-linear in the lagged outcome), (7.2) (linear in the lagged outcome), (7.3) (linear in the lagged outcome with stronger non-linearity in $\mathbf{X}_t$) and the entirely linear model (S.1) (linear in $\mathbf{X}_{it}$ and $Y_{i,t-1}(0)$) with disparity (δ_t) in the means between the treatment and control groups varying with time $\delta_t = 0.5 - 1.5(t/T)^2$ for $N_{co} = N_{tr} = 100$.

Methods	Model (7.1)	Model (7.2)	Model (7.3)	Model (S.1)
DSC$_{\widehat{m}}$	0.1681	0.1752	0.1765	0.3299
DSC$_{linear}$	0.4121	0.2506	0.4704	0.3526
Raw-diff	0.6462	2.9894	0.9982	3.7368

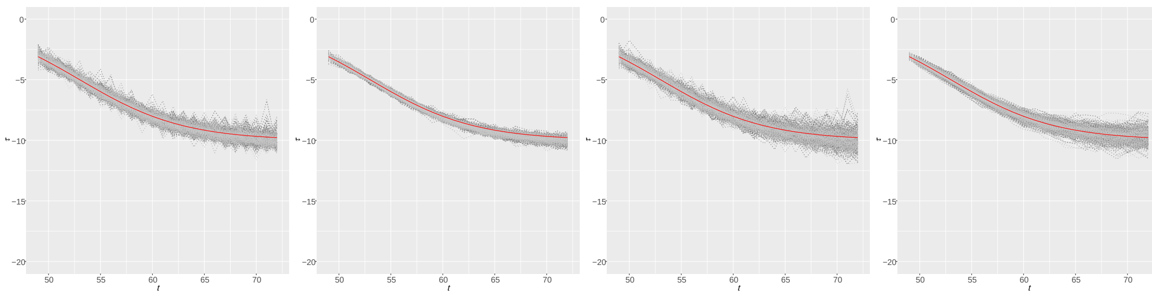
8 Case Study

In Beijing, air pollution alerts are categorized into blue, yellow, orange, and red in ascending severity and specific emergency measures would be implemented accordingly. We applied the proposed DSC method to evaluate the effects of two air pollution alerts, an orange alert beginning at 00:00 on November 17th, 2016 and a red alert at 20:00 on December 26th, 2016, on hourly $PM_{2.5}$ concentration over the 24 hours after the alerts. During the studied alerts, emission-reduction measures were implemented only in Beijing (the treated area) where we considered 20 air quality stations. The control units were 74 air quality stations in the neighboring Tianjin municipalities, Hebei, Shandong and Shanxi provinces due to the absence of air pollution alerts in these regions during the two concerned alerts. This made $N = 94$ air quality stations in total, with $N_{tr} = 20$ in the treated group in Beijing and $N_{co} = 74$ sites in the control group, respectively.

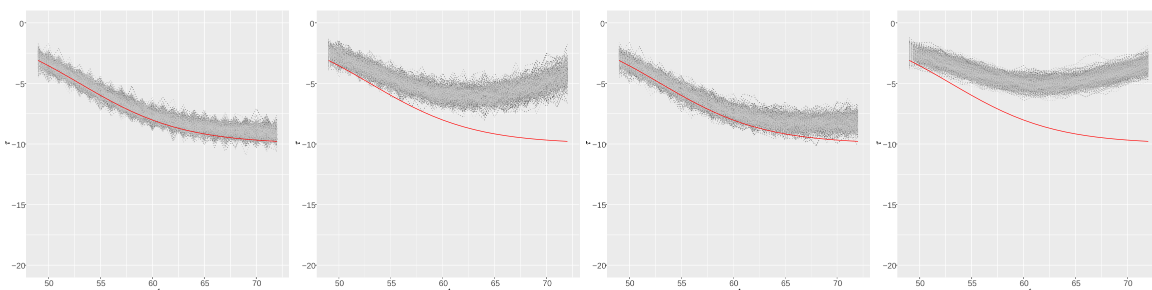
The outcome variable was hourly $PM_{2.5}$ concentration measured in $\mu g/m^3$ at each monitoring station. Four meteorological variables, wind speed (WSPM), humidity (HUMI), dew point temperature (DEWP) and air pressure (PRES), which are known to influence $PM_{2.5}$ levels (Liang et al., 2015), together with



(a) Model (7.1), $\mathrm{DSC}_{\widehat{m}}$ (b) Model (7.2), $\mathrm{DSC}_{\widehat{m}}$ (c) Model (7.3), $\mathrm{DSC}_{\widehat{m}}$ (d) Model (S.1), $\mathrm{DSC}_{\widehat{m}}$



(e) Model (7.1), DSC_{linear} (f) Model (7.2), DSC_{linear} (g) Model (7.3), DSC_{linear} (h) Model (S.1), DSC_{linear}



(i) Model (7.1), Raw-diff (j) Model (7.2), Raw-diff (k) Model (7.3), Raw-diff (l) Model (S.1), Raw-diff

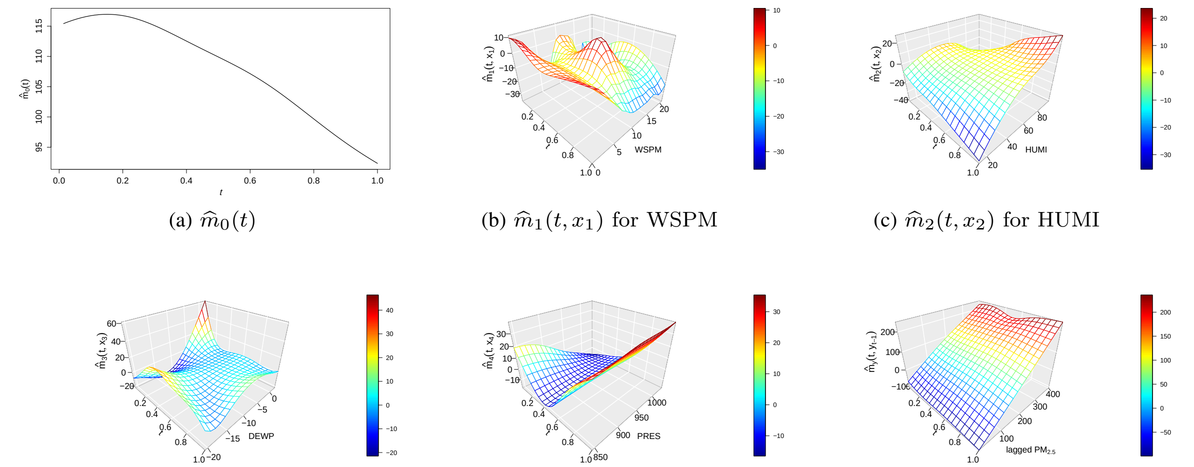
Figure 2 The estimated ATT (the grey dashed lines) for 500 simulations by the proposed semiparametric DSC ($\mathrm{DSC}_{\widehat{m}}$), the linear DSC ($\mathrm{DSC}_{linear}$) and the Raw-difference for Models (7.1) (non-linear in the lagged outcome), (7.2) (linear in the lagged outcome), (7.3) (linear in the lagged outcome with stronger non-linearity in $\mathbf{X}_t$) and the entirely linear model (S.1) (linear in $\mathbf{X}_{it}$ and $Y_{i,t-1}(0)$) with disparity (δ_t) in the means between the treatment and control groups varying with time $\delta_t = 0.5 - 1.5(t/T)^2$ for $N_{co} = N_{tr} = 100$. The true treatment effects are shown in the red solid curves.

the lagged $\mathrm{PM}_{2.5}$ were included in Model (2.2). To verify Assumption 2.1(ii), we chose $T_0 = 48$ hours before the alerts as the pre-treatment periods. We applied the proposed semiparametric DSC to estimate the effect of pollution alerts on the hourly $\mathrm{PM}_{2.5}$ for the 24 hours after the alerts.

Figure 3 displays the estimated additive functions of the time, the four meteorological variables and the lagged $\mathrm{PM}_{2.5}$ for the red alert. It suggests the lagged $\mathrm{PM}_{2.5}$'s effect was linear, which was empirically confirmed by the bootstrap test given in Section 6.1 with the p-value close to 1. Panels (b)-(e) of Figure 3 show that the estimated meteorological effects $\{\widehat{m}_k(\cdot,\cdot)\}_{k=1}^{3}$ were quite non-linear, implying that Model (2.2) was more suitable than the entirely linear models as assumed in Abadie et al. (2010) and the linear DSC (Zheng and Chen, 2024). Similar results for the orange alert were shown in Figure S.3 of the SM.

Figure 4 displays the estimated optimal DSC weights $\{w_{it}^{*}\}_{i=1,t=1}^{N_{co},T}$ by the EL for the two alerts, which shows that only a handful of sites had larger weights (red) with the rest of the stations given weights close to zero (light yellow) at each time t. For instance, at $t = 1$, during the orange alert, only three of the 72 weights exceeded 0.1, at 0.217, 0.384, and 0.399, while the remaining weights were all below 10^{-10}. This implies that the scenarios for the two alerts corresponded to the case of $\mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 0) - \mathrm{E}(\widetilde{\mathbf{X}}_{it}|D_i = 1) \neq 0$ in Lemma 5.4(ii).

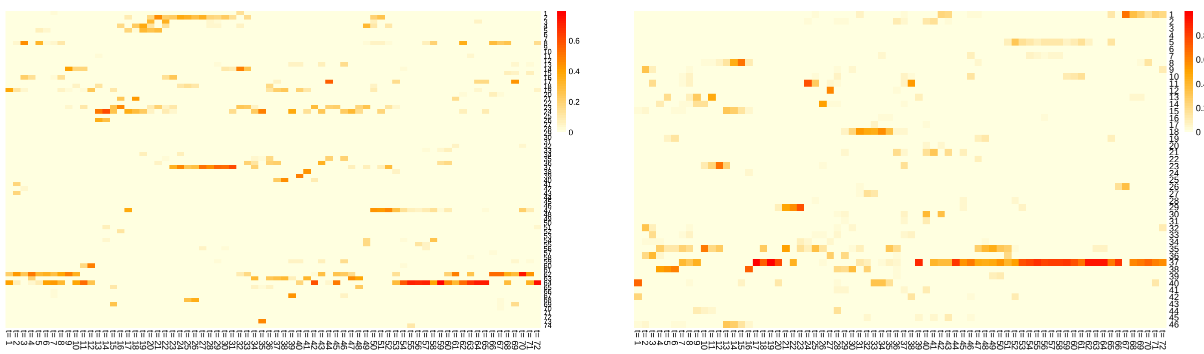
Figure 5 presents the estimated average potential outcomes and the ATT by different methods for the



(a) $\widehat{m}_0(t)$ (b) $\widehat{m}_1(t, x_1)$ for WSPM (c) $\widehat{m}_2(t, x_2)$ for HUMI

(d) $\widehat{m}_3(t, x_3)$ for DEWP (e) $\widehat{m}_4(t, x_4)$ for PRES (f) $\widehat{m}_y(t, y_{t-1})$ for lagged $\mathrm{PM}_{2.5}$

Figure 3 The estimated functions $m_0(t)$ and $m_k(t, \cdot)$ for the red alert for all covariates: time t, wind speed (WSPM), humidity (HUMI), dew point temperature (DEWP), air pressure (PRES) and lagged $\mathrm{PM}_{2.5}$.

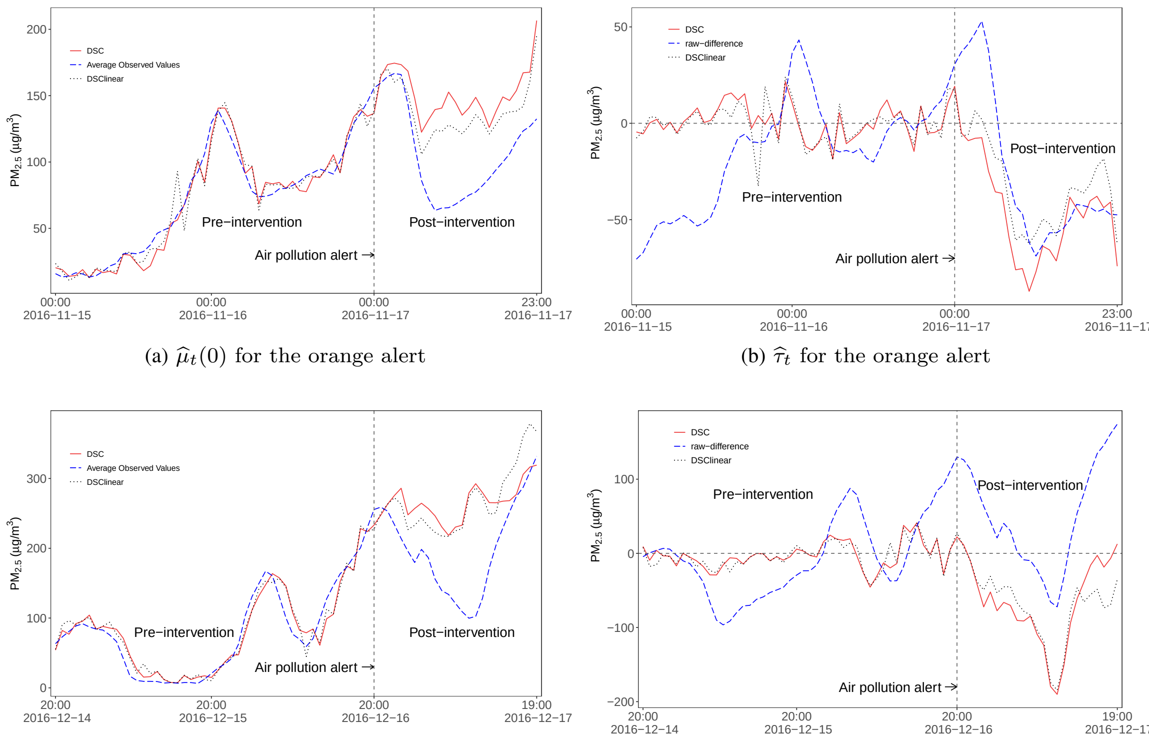


(a) $\{w^*_{it}\}_{i=1,t=1}^{N_{co},T}$ for the orange alert (b) $\{w^*_{it}\}_{i=1,t=1}^{N_{co},T}$ for the red alert

Figure 4 The DSC weights $\{w^*_{it}\}_{i=1,t=1}^{N_{co},T}$ at each time $t = 1, \cdots, 72$ for the orange and red alerts.

two pollution alerts. From Figure(a)(c), before the alerts, the DSC estimates closely tracked the trajectory of the observed concentration of $\mathrm{PM}_{2.5}$ for both alerts, which also implies that the estimated treatment effects for both alerts were around zero before the air pollution alerts as shown in Figure 5(b)(d). The RMSEs of the semiparametric DSC in the pre-intervention periods were $80.5\mu g/m^3$ and $324.8\mu g/m^3$ for the orange and red alerts, respectively, which were lower than their linear DSC counterparts at $104.5\mu g/m^3$ and $332.5\mu g/m^3$, respectively. Thus, the incorporation of non-linear effects enhanced the quality of the synthetic controls. In the post-intervention periods, both the semiparametric and linear DSC estimates of $\mu_t(0)$ showed substantially higher $\mathrm{PM}_{2.5}$ levels than the observed, suggesting that the air pollution alerts had caused the decline in the $\mathrm{PM}_{2.5}$ concentration, which was also reflected by the negative treatment effects after the alerts as seen in Figure 5(b)(d). In particular, the semiparametric DSC estimates of $\mu_t(0)$ tended to be above those of the linear DSC in the intervention period except toward the end of the red alert. These suggest that considering the non-linear meteorological effects in the DSC formulation can improve the estimation of the dynamic treatment effects for the air pollution alerts.

For the orange alert, the average observed $\mathrm{PM}_{2.5}$ over the post-treatment period was 105.28 $\mu g/m^3$, while the potential average $\mathrm{PM}_{2.5}$ without the alert was 152.24 $\mu g/m^3$ by the semiparametric DSC. This suggests the emission reduction measures could reduce the $\mathrm{PM}_{2.5}$ by 46.96 $\mu g/m^3$ on average (standard

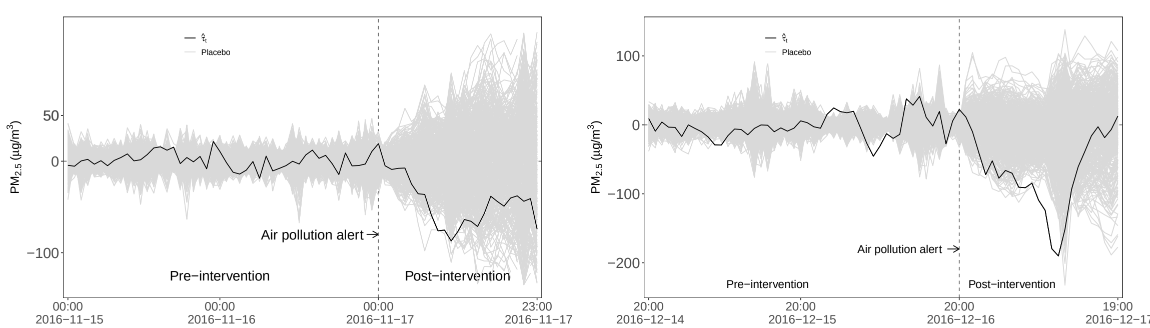


(a) $\widehat{\mu}_t(0)$ for the orange alert   (b) $\widehat{\tau}_t$ for the orange alert

(c) $\widehat{\mu}_t(0)$ for the red alert   (d) $\widehat{\tau}_t$ for the red alert

Figure 5 The estimated average potential outcomes $\widehat{\mu}_t(0)$ by the proposed semiparametric DSC (solid lines) and by the linear DSC (dotted lines) and the observed average treated values $\sum_{i\in\mathcal{T}} N_{tr}^{-1} Y_{it}$ (dashed lines), and the estimated treatment effects $\widehat{\tau}_t$ by the proposed semiparametric DSC, the linear DSC and Raw-difference for the orange and red alerts, respectively.

error=4.98 $\mu g/m^3$), accounting for 30.8% of the potential concentration. For the red alert, $PM_{2.5}$ was reduced by 67.64 $\mu g/m^3$ on average.



(a) $\widehat{\tau}_t$ for the orange alert   (b) $\widehat{\tau}_t$ for the red alert

Figure 6 The estimated effects $\widehat{\tau}_t$ (black lines) and the normalized differences $\{z_t^{(k)}\}_{k=1}^{500}$ for the placebo treatment group (grey lines) for the orange and red alerts.

We adopted the test in Section 6.2 to evaluate Assumption 2.1(ii) by verifying $\mathrm{E}(\widehat{\tau}_t) = 0$ for $t \leqslant T_0$. The sorted p-values for both alerts at each time are shown in Figure S.4 of the SM and we could not reject the null hypotheses while maintaining the FDR control at level 0.05, which provided support for Assumption 2.1(ii). We then performed the normalized placebo experiment 500 times to test whether the effects of the two alerts were significant. It can be seen from Figure 6 that the estimated treatment effects $\widehat{\tau}_t$ (black lines) were situated in the middle of the distribution of normalized placebo $z_t^{(k)}$ (grey lines) before the start of the two alerts and began to deviate from the distribution of $z_t^{(k)}$ after the alerts. The p-values for the orange and red alerts are shown in Tables S.2 and S.3 of the SM, which implies that the orange and red alerts had the most significant effects in the next 6-15 and 3-17 hours, respectively, since the start of the alert.

9 Conclusion

We propose a DSC method for the semiparametric time-varying additive autoregression outcome model, which allows for micro-level data with non-linear time-varying confounders, multiple treated units and spatial correlations in the data. The use of a semiparametric additive model provides much flexibility in both modeling and estimation processes, which potentially allows the approach to be applied to evaluate treatment effects in broader situations than the linear DSC. The proposed semiparametric DSC is also suitable for scenarios where different treated units are assigned to the treatment at different times.

Acknowledgements

We thank the two reviewers for their constructive comments and suggestions which have led to improvements in the presentation of this work. Our research was supported by National Natural Science Foundation of China grants 72501165, 12292983, 92358303, and Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China JYB2025XDXM801. We also thank for the technical support of the National Large Scientific and Technological Infrastructure "Earth System Numerical Simulation Facility" (https://cstr.cn/31134.02.EL). The work was partially conducted while the first author was a postdoctoral fellow at Peking University (supported by the China Postdoctoral Science Foundation 2023M730090).

Supporting information

We provide more details about the SPBK estimation and the hypothesis testing, additional results for simulation and case study, and technical details in the Supplementary Material.

References

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

Abadie A, Diamond A, Hainmueller J. Comparative politics and the synthetic control method. American Journal of Political Science, 2015, 59: 495–510.

Abadie A, Gardeazabal J. The economic costs of conflict: A case study of the Basque country. American Economic Review, 2003, 93: 113–132.

Abadie A, L'Hour J. A penalized synthetic control estimator for disaggregated data. Journal of the American Statistical Association, 2021, 116: 1817–1834.

Angrist J D, Pischke J-S. Mostly harmless econometrics: An empiricist's companion. Princeton University Press, 2009.

Athey S, Bayati M, Doudchenko N, et al. Matrix completion methods for causal panel data models. Journal of the American Statistical Association, 2021, 116: 1716–1730.

Athey S, Imbens G W. The state of applied econometrics: Causality and policy evaluation. Journal of Economic Perspectives, 2017, 31: 3–32.

Bai J S, Ng S. Matrix completion, counterfactuals, and factor analysis of missing data. Journal of the American Statistical Association, 2021, 116: 1746–1763.

Ben-Michael E, Feller A, Rothstein J. The augmented synthetic control method. Journal of the American Statistical Association, 2021, 116: 1789–1803.

Ben-Michael E, Feller A, Rothstein J. Synthetic controls with staggered adoption. Journal of the Royal Statistical Society Series B, 2021, 84: 351–381.

Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics, 2001, 29: 1165–1188.

Bonander C, Humphreys D, Degli Esposti M. Synthetic control methods for the evaluation of single-unit interventions in epidemiology: A tutorial. American Journal of Epidemiology, 2021, 190: 2700–2711.

Bruhn C A W, Hetterich S, Schuck-Paim C, et al. Estimating the population-level impact of vaccines using synthetic controls. Proceedings of the National Academy of Sciences, 2017, 114: 1524–1529.

Cavallo E, Galiani S, Noy I, et al. Catastrophic natural disasters and economic growth. The Review of Economics and Statistics, 2013, 95: 1549–1561.

Chen S X, Van Keilegom I. A review on empirical likelihood methods for regression. TEST, 2009, 18: 415–447.

De Boor C. A Practical Guide to Splines. New York: Springer, 1978.

Ding P, Li F. A Bracketing Relationship between Difference-in-Differences and Lagged-Dependent-Variable Adjustment. Political Analysis, 2019, 27: 605–615.

Doudchenko N, Imbens G W. Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Technical report, National Bureau of Economic Research, 2016.

Ferman B, Pinto C. Synthetic controls with imperfect pretreatment fit. Quantitative Economics, 2021, 12: 1197–1221.

Ghosh S, Chaudhuri S, Gangopadhyay U. Maximum likelihood estimation under constraints: Singularities and random critical points. IEEE Transactions on Information Theory, 2023, 69: 7976–7997.

Härdle W, Mammen E. Comparing nonparametric versus parametric regression fits. The Annals of Statistics, 1993, 21: 1926–1947.

Imbens G W, Wooldridge J M. Recent developments in the econometrics of program evaluation. Journal of Economic Literature, 2009, 47: 5–86.

Leadbetter M R, Lindgren G, Rootzén H. Extreme Value Theory and Strength of Materials. New York: Springer, 1983, 267–277.

Liang X, Zou T, Guo B, et al. Assessing Beijing's pm2.5 pollution: severity, weather impact, APEC and winter heating. Proceedings of the Royal Society A, 2015, 471: 20150257.

Lin J J, Gamalo-Siebers M, Tiwari R. Propensity score matched augmented controls in randomized clinical trials: A case study. Pharmaceutical Statistics, 2018, 17: 629–647.

Owen A B. Empirical likelihood ratio confidence intervals for a single functional. Biometrika, 1988, 75: 237–249.

Owen A B. Empirical likelihood ratio confidence regions. The Annals of Statistics, 1990, 18: 90–120.

Owen A B. Empirical Likelihood. New York: Chapman and Hall/CRC, 2001.

Qin J, Lawless J. Empirical likelihood and general estimating equations. The Annals of Statistics, 1994, 22: 300–325.

Roth J, Sant'Anna P H, Bilinski A, et al. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. Journal of Econometrics, 2023, 235: 2218–2244.

Rubin D B. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 1974, 66: 688–701.

Ryan A M. Well-balanced or too matchy–matchy? The controversy over matching in difference-in-differences. Health Services Research, 2018, 53: 4106–4110.

Shi C, Sridhar D, Misra V, et al. On the assumptions of synthetic control methods. In: Proceedings of the 25th International Conference on Artificial Intelligence and Statistics, 2022, 7163–7175.

Viviano D, Bradic J. Synthetic learner: Model-free inference on treatments over time. Journal of Econometrics, 2023, 234: 691–713.

Wang L, Yang L J. Spline-backfitted kernel smoothing of nonlinear additive autoregression model. The Annals of Statistics, 2007, 35: 2474–2503.

Zheng X Y, Chen S X. Dynamic synthetic control method for evaluating treatment effects in auto-regressive processes. Journal of the Royal Statistical Society Series B, 2024, 86: 155–176.