

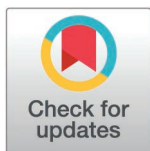
RESEARCH ARTICLE

Estimating endogenous treatments effects under long-range dependency without untreated controls

Shiming Hao *

Law School, Yunnan University, Chenggong, Kunming, Yunnan, P. R. China

* haoshiming@ynu.edu.cn



OPEN ACCESS

Citation: Hao S (2026) Estimating endogenous treatments effects under long-range dependency without untreated controls. PLoS One 21(6): e0347847. <https://doi.org/10.1371/journal.pone.0347847>

Editor: Abhik Ghosh, Indian Statistical Institute, INDIA

Received: December 22, 2025

Accepted: April 8, 2026

Published: June 3, 2026

Copyright: © 2026 Shiming Hao. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The R package "endoLTE", which supports the findings of this study, is openly available on GitHub at: <https://github.com/codescollection/endoLTE>. The package can be installed following the step-by-step instructions provided on the repository's homepage and includes the implementation of the estimator, simulation scripts, and empirical

Abstract

The identification and estimation of social policy effects through time-series natural experiments is fundamental in modern econometrics. However, challenges come from the heterogeneities caused by staggered treatment adoptions and the endogeneities caused by omitted variables. In this paper, we propose a novel method to identify and estimate staggeringly adopted endogenous treatments effects with several treatments when there is no available (suitable) untreated unit and instrument variable. First, we propose a conditional mean symmetry condition by projecting the potential outcomes onto a sub-linear space spanned by the proposed common proximal variable. Under this condition, we can rule out confounding biases. Second, a proposed weak index restriction constructed by Bernstein expansions satisfying conditional mean independence property enables us to consistently estimate multiple heterogeneous treatments effects, and the proposed estimators are robust to weak common proximal variable. We show that the asymptotic distribution of the step-wise estimator is a fractional Brownian motion process with long range dependency. Third, we propose a bootstrap procedure to circumvent the inference difficulty brought by time-series dependencies. Monte Carlo simulations show that our proposed estimator and inference framework work well in small samples, and our contributions are further illustrated by an empirical example with unilateral divorce law reforms.

Introduction

This paper considers the identification and estimation of the following treatments effects panel model:

$$y_{it} = S_{it}\beta_{it} + W_{it}\eta_{it} + D_{it}\xi_{it} + X_{it}\alpha_{it} + Z_{it}\nu_{it} + \mu_i + \mu_t + \epsilon_{it},$$

$$\text{cov}(\lambda_{it}, \epsilon_{it}) \neq 0, \quad \lambda_{it} \in \{S_{it}, D_{it}\}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T, \quad (1)$$

datasets. The data underlying this study are publicly accessible at the following link: <https://github.com/codescollection/endoLTE/tree/main/data>. The code to replicate our simulation results is accessible at: <https://github.com/codescollection/endoLTE/tree/main/replication%20codes>. The package is distributed under the MIT License. Please cite this version if you use the code in your own research.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

y_{it} is a scale response variable for unit i at time t , i.e., some observed social-economic outcome; S_{it} is a treatment taking place at time t_{Si} and driven by the confounder W_{it} , D_{it} is another treatment taking place at time t_{Di} and driven by the confounder X_{it} , e.g., the treatments (S_{it} , D_{it}) may be some economic stimulate policies promoted by the governments while the confounders (W_{it} , X_{it}) may be some social-economic factors driving the growth of y_{it} and the stimulation of the policies (S_{it} , D_{it}) [1–4]. We assume that W_{it} and X_{it} are all of dimension one. Z_{it} are d_z dimensional exogeneous control variables which are independent of the treatments (S_{it} , D_{it}) and the model error ϵ_{it} . μ_i and μ_t are individual fixed effect and time fixed effect. The parameters of interest are the heterogeneous treatment effects β_{it} and ξ_{it} . We consider three main challenges that commonly arise in empirical studies:

1. **The endogeneity problem.** We assume that the confounders (W_{it} , X_{it}) are unobservable. Meanwhile, we relax the Gauss-Markov Theorem to allow the treatments to be correlated with the error term: $\text{cov}(\lambda_{it}, D_{it}) \neq 0$ for $\lambda_{it} \in \{S_{it}, D_{it}\}$, which means that even if the confounders are observed and included in model (1), the model is still endogeneous. In empirical studies, Instrument Variables (IVs) approaches are considered to be an effective way to handle the problem under this scenario [5], however good quality IVs are hard to be found [6–8]. Hence, we directly assume in this paper that no suitable IVs are available for empirical analysis.
2. **The heterogeneity problem.** On the one hand, as shown in model (1), all parameters are time-varying and unit-varying (two-way heterogeneous), which implies that treating the treatments effects as constant will lead to false estimations of ATE and ATT [9–11]. On the other hand, we further assume that the treatments (S_{it} , D_{it}) are staggeringly adopted, i.e., there exist at least two $i \neq i' \in \{1, 2, \dots, N\}$ such that $t_{Si} \neq t_{Si'}$ and $t_{Di} \neq t_{Di'}$. It is well-known that regional policy evaluation approaches such as Two-Way Fixed Effect Estimations (TWFE), Synthetic Control Methods (SCM) and Difference in Difference (DID) among others will lead to false estimations under this scenario [12–19], e.g., the contamination bias [20].
3. **The absence of untreated units.** When the social-economic policies (S_{it} , D_{it}) are one-size-fit-all, e.g., the government's policies are carried out nationwide, no untreated units are available for identification and estimation of the treatments effects. Regional policy evaluation approaches such as TWFE, SCM and DID among others are no longer applicable under this scenario. Empirical researches tend to find or construct untreated units from historical data or cross-national data (e.g., [21]), however the validities are suspectable. To say the least, even if the policies are not one-size-fit-all, we are stilling facing the problem of selecting “good-quality” untreated units [22]. In fact, adopting unsuitable untreated units will lead to biased estimations [23,24].

Although these three problems are common in empirical studies, the methodological literature handling these three dilemmas simultaneously as far as we know is in straitened circumstances. Most recent discussions have paid attentions to the heterogeneous problem [20,25], however if untreated units are absent, the proposed

methods are not applicable. Even if untreated units are present, estimating model (1) by the proposed methods with the endogeneity problem will also lead to biased estimations. How to identify and estimate heterogeneous treatments effects with endogeneity and staggeringly adopted treatments in untreated units and IVs absent time-series natural experiments remains an unsolved problem.

To overcome these difficulties, a novel method is proposed. We propose a Conditional Mean Symmetry condition by projecting the counterfactual outcome onto the sub-linear space spanned by the proposed Common Proximal variable, wherein the Common Proximal variable is of high correlation with the unobservable confounders and independent of the error term. The Conditional Mean Symmetry condition is, on the one hand, shown to be equivalent to the Weak Index Restrictions used in the identification of semiparametric models [26,27]; on the other hand it is also equivalent to the matching condition used in observation studies which rules out confounding biases (see, e.g., [28–31]). Our framework is built on Professor Halbert White and colleague's pioneering works [32–34], and we further illustrate the relationships between Granger (G-) causality, Sims causality and structural causality in untreated units absent time series natural experiments [35], as well as how we can disentangle one treatment effect from another through a forward point-wise estimation approach based on the proposed Conditional Mean Symmetry condition.

We contribute to the recent emerging literature on heterogeneous treatments effects in event studies with variations in treatments timings in two main fronts. First, we allow non-stationary potential outcomes. Stationarity is usually required in time-series natural experiments to guarantee valid inference [10,23,36,37], however social-economic outcomes are always non-stationary, e.g., trend stationary processes. Causal inference with non-stationary outcomes is somehow nettlesome. In view of this, our proposed bootstrap inference framework allows the potential outcomes to be non-stationary processes, e.g., the potential outcomes are trend-stationary processes with long-range dependency, where the deterministic trend captures the long-term growth and the stationary component follows a long memory model. As far as we are concerned, this is the first time we consider non-stationarity with long range dependency in time-series natural experiments.

Second, the Conditional Mean Symmetry identification assumption proposed in this paper presents an innovative identification paradigm that extends causal inference to empirical settings—such as national policy evaluations—where the DID framework is not applicable due to the absence of a valid comparison group. If there is no untreated unit, [32–34,38,39] and [40] have considered the possibility of using only treated units to estimate causal effects. However, they did not consider scenarios involving multiple endogenous treatments and long-range dependence, both of which are common in real-world data. Our approach fills this gap.

The rest of the paper is organized as shown in the roadmap Fig 1. The second part introduces the identification assumptions and describes how we can identify multiple treatments effects without untreated units under the Neyman-Rubin counterfactual framework. The estimation procedures and inference method for a single (univariate) time series model are gathered in the third part, and the fourth part extends the proposed methods to handle panel (multivariate) settings. The fifth part reports the results of the numerical examples and the final part concludes. All the proofs and additional analysis are collected in the Online Supplementary Materials S1 File-S5 File. To facilitate the use of our method, we provide an open-source R package endoLTE at <https://github.com/codescollection/endoLTE> that implements all estimators and inference procedures proposed in this paper.

Identification

We will start our identification and estimation approach by a single (univariate) time-series model with homogeneous treatments effects. Then, we will show how the proposed approach can be extended to the heterogeneous settings and staggered panel settings, as shown in model (1).

Global identification

Consider the following 4 specifications:

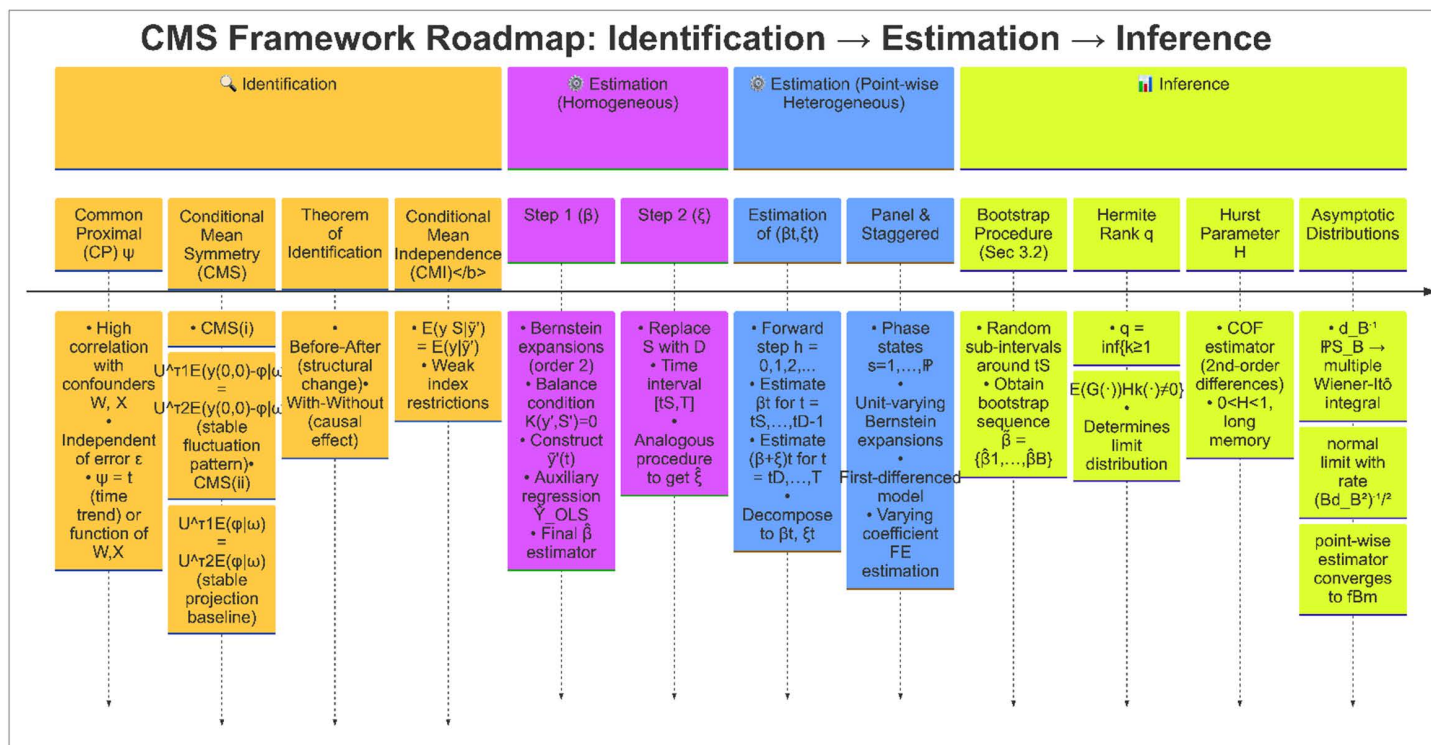


Fig 1. Roadmap of identification, estimation and inference.

<https://doi.org/10.1371/journal.pone.0347847.g001>

$$\text{DGP1: } y_t = S_t \beta_1 + W_t \eta_1 + D_t \xi_1 + X_t \alpha_1 + Z_t \nu_1 + \epsilon_{t,1}, \quad W_t \leftrightarrow D_t, \quad (2)$$

$$\text{DGP2: } y_t = S_t \beta_2 + W_t \eta_2 + D_t \xi_2 + X_t \alpha_2 + Z_t \nu_2 + \epsilon_{t,2}, \quad W_t \perp D_t, \quad (3)$$

$$\text{DGP3: } y_t = S_t \beta_3 + W_t \eta_3 + Z_t \nu_3 + \epsilon_{t,3}, \quad (4)$$

$$\text{DGP4: } y_t = D_t \xi_4 + X_t \alpha_4 + \epsilon_{t,4}, \quad (5)$$

(S_t, D_t) are treatments with the confounders (W_t, X_t) , Z_t are exogenous control variables. DGP1 assumes that the confounder W_t is correlated with the treatment D_t , i.e., $W_t \leftrightarrow D_t$, while DGP2 assumes that W_t is independent of D_t . DGP3 and DGP4 respectively describes a single treatment effect model.

The Directed Acyclic Graphs (DAGs) for DGPs (2)–(5) are shown in Fig 2, corresponding to the adjacency matrices A, B, C, D .

Definition 1 (Pseudo-subspace). If we let $\mathcal{L}_{*,n}$ denote the n -th column of the matrix $\mathcal{L}_{T \times N}$, then there exists a pseudo-subspace spanned by $\mathcal{L}_{*,n}: k_1 \mathcal{L}_{*,1} + k_2 \mathcal{L}_{*,2} + \dots + k_N \mathcal{L}_{*,N}$ for all nonzero constants $k_n \in \mathbb{R} \setminus \{0\}$, $n = 1, 2, \dots, N$. We denote this pseudo-subspace as $\mathcal{M}^{-1}(\mathcal{L})$.

Assumption 1 (Rank condition) In models (2)–(5), for the random vectors $W = (W_1, W_2, \dots, W_T)'$, $S = (S_1, S_2, \dots, S_T)'$, $D = (D_1, D_2, \dots, D_T)'$, $X = (X_1, X_2, \dots, X_T)'$, and the matrix $Z = (Z_1, Z_2, \dots, Z_T)'$, denote $\mathbf{x}_1 = (W, S, D, X, Z)$ for DGP1,

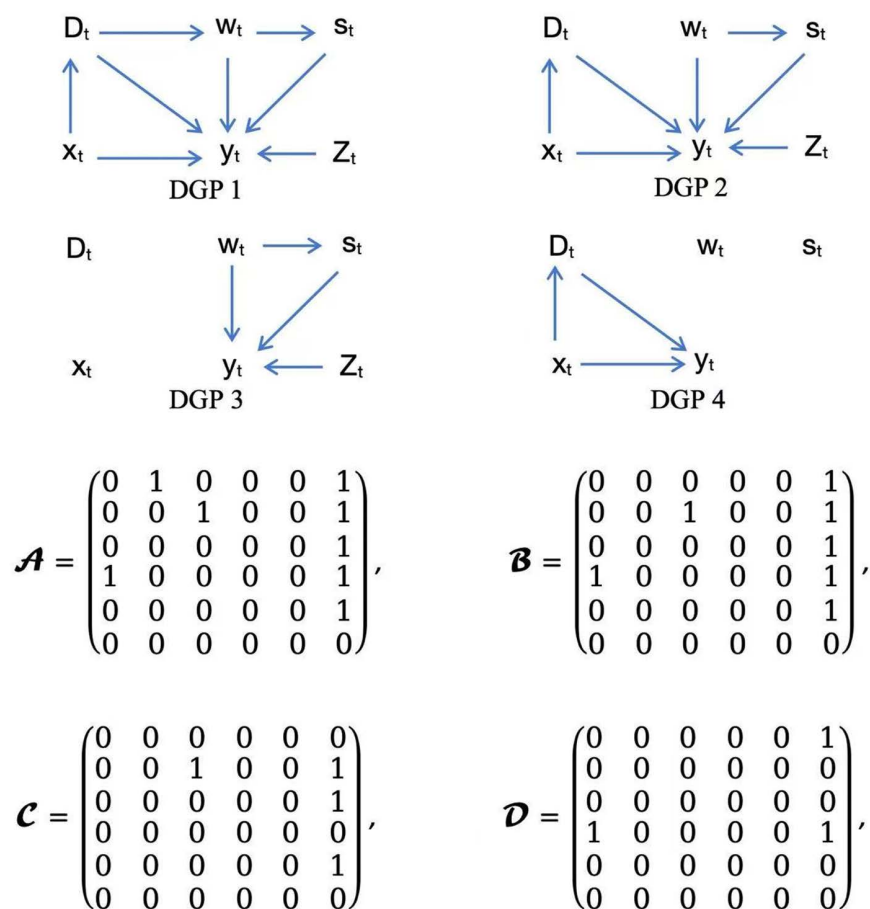


Fig 2. DAGs for the treatments effects models (2-5) (DGPs 1-4) with corresponding adjacency matrices.

<https://doi.org/10.1371/journal.pone.0347847.g002>

$\mathbf{x}_2 = (W, S, D, X, Z)$ for DGP2, $\mathbf{x}_3 = (W, S, Z)$ for DGP3 and $\mathbf{x}_4 = (X, D)$ for DGP4. $\mathbf{x}_1$ and $\mathbf{x}_2$ are matrices of dimension $T \times (4 + d_z)$, $\mathbf{x}_3$ is of dimension $T \times (2 + d_z)$ and $\mathbf{x}_4$ is of dimension $T \times 2$. Then we have $\text{rank}(\mathbf{x}_1' \mathbf{x}_1) = 4 + d_z$, $\text{rank}(\mathbf{x}_2' \mathbf{x}_2) = 4 + d_z$, $\text{rank}(\mathbf{x}_3' \mathbf{x}_3) = 2 + d_z$ and $\text{rank}(\mathbf{x}_4' \mathbf{x}_4) = 2$.

The difference between the usual subspace and the pseudo-subspace defined in Definition 1 is that we do not allow $k_n = 0$, so one can see from the adjacency matrices that $\mathcal{M}^{-1}(\mathcal{B}) = \mathcal{M}^{-1}(\mathcal{C} + \mathcal{D})$ while $\mathcal{M}^{-1}(\mathcal{A}) \neq \mathcal{M}^{-1}(\mathcal{C} + \mathcal{D})$. Assumption 1 requires $t_S \neq t_D$, we do not allow treatments take place at the same time, e.g., the governments' policies are carried out at the same time. In fact if $t_S = t_D$, treatments (S_t, D_t) would be mixed with each other, making it impossible to distinguish one from another.

Proposition 1. If Assumption 1 is satisfied, then the treatment effect β and the treatment effect ξ can be separately identified from each other under DGP2 with the adjacency matrix $\mathcal{B}$ but not under DGP1 with the adjacency matrix $\mathcal{A}$.

Why and how it is possible to identify treatment effect in untreated units absent time-series natural experiments

As shown in Proposition 1, model (2) is unidentifiable, hence only the following treatments effects model (6) is identifiable. We now start with this univariate time-series model with homogeneous coefficients, and show how the estimation and inference procedure could be extended to handle heterogeneous panel settings latter.

$$y_t = W_t\eta + S_t\beta + D_t\xi + X_t\alpha + Z_t\gamma + \varepsilon_t, \quad \text{cov}(\lambda_t, \varepsilon_t) \neq 0, \quad t = 1, 2, \dots, t_S, \dots, t_D, \dots, T, \quad (6)$$

where $\lambda_t \in \{S_t, D_t\}$, $W_t \perp D_t$, Z_t are exogenous control variables independent of λ_t and ε_t . The parameters of interest are (β, ξ) , but (D_t, S_t) are endogenous due to the unobservability of (W_t, X_t) and $\text{cov}(\lambda_t, \varepsilon_t) \neq 0$. Instead of finding IVs for the endogenous treatment variables D_t and S_t , we propose to find (or construct) a Common Proximal Variable ψ_t satisfying the following Assumptions 2–3 and Definition 2. On this basis, for model (6), we consider the Bernstein expansions of the random variables $y_t \equiv y(t)$, $S_t \equiv S(t)$ and $D_t \equiv D(t)$ on the Common Proximal Variable ψ_t of order 2 [41]:

$$y(t) = \sum_{k=0}^2 \frac{2!}{k!(2-k)!} \psi_t^k (1-\psi_t)^{2-k} f_t\left(\frac{k}{2}\right) = \rho_{y,2} \psi_t^2 + \rho_{y,1} \psi_t + \rho_{y,0}, \quad (7)$$

$$\lambda(t) = \sum_{k=0}^2 \frac{2!}{k!(2-k)!} \psi_t^k (1-\psi_t)^{2-k} f_t\left(\frac{k}{2}\right) = \rho_{\lambda,2} \psi_t^2 + \rho_{\lambda,1} \psi_t + \rho_{\lambda,0}, \quad \lambda \in \{S, D\}, \quad (8)$$

where $f_\lambda(0) = 0$, $f_\lambda(\frac{1}{2}) = \frac{1}{2} \rho_{\lambda,1}$, $f_\lambda(1) = \rho_{\lambda,2} + \rho_{\lambda,1}$, $\rho_{y,2}, \rho_{y,1}, \rho_{\lambda,2}, \rho_{\lambda,1} \in \mathbb{R} \setminus \{0\}$ are called Bernstein coefficients and the expansions (7)–(8) we get are called Bernstein polynomials of $y(t)$, $\lambda(t)$ on ψ_t , where $\lambda(t) \in \{S(t), D(t)\}$. Note that the first order derivative of $y(t)$ and $\lambda(t)$ with respect to ψ_t are

$$y'(t) = 2\rho_{y,2}\psi_t + \rho_{y,1}, \quad \lambda'(t) = 2\rho_{\lambda,2}\psi_t + \rho_{\lambda,1}.$$

Assumption 2 (i) The random variables W and ψ are of high correlation, X and ψ are also of high correlation, e.g., ψ is a proximal variable for W and X . (ii) ψ is independent of ε , and $E(\varepsilon) = 0$

Assumption 3 There exists a continuous function $g \in C^p$, $p \geq 1$, such that: (i) $W_t = ag(y'(t)) + b + u_t$, $X_t = cg(y'(t)) + d + w_t$, $a, b, c, d \in \mathbb{R} \setminus \{0\}$, C^p is the Lebesgue space $C^p(\Omega_g, \mathcal{F}_g, \mu_g)$ with $\|g\|_p = (\int |g|^p d\mu_g)^{1/p}$ defined on Ω_g , u_t, w_t are martingale processes independent of ψ_t and $E(u) = 0$, $E(w) = 0$. (ii) For the Bernstein expansions, (W, X) and $g(y')$ are of high correlations respectively, where $g(y') = (g(y'(1)), g(y'(2)), \dots, g(y'(T)))$. (iii) The composite error $u_t + w_t + \varepsilon_t$ is independent of λ_t, ψ_t and Z_t , where $\lambda_t \in \{S_t, D_t\}$.

Definition 2 (Common Proximal Variable). For model (6), a random variable ψ_t satisfying Assumptions 2–3 is called a Common Proximal Variable (CP).

Remark 1 (Doob decomposition theorem). For most of the social-economic outcomes y_t , if y_t is changing with time, i.e., y_t is a function of t , as we defined in (7); or if there are theoretical and empirical evidences supporting that W_t, X_t and Z_t are changing with time, we would then take $\psi_t = t$ satisfying Assumptions 2–3 without loss of generality. Particularly, if both $(W_t, \mathcal{F}_{W_t})_{t \in \tau}$ and $(X_t, \mathcal{F}_{X_t})_{t \in \tau}$ are sub-martingales for $\tau = \{1, 2, \dots, T\}$, and we have

$$W_t = ag(\psi_t) + b + u_t, \quad X_t = cg(\psi_t) + d + w_t \quad \text{for } \psi_t = t, \quad (9)$$

where $g(\cdot) \in \mathcal{L}^p$ is a continuous and non-decreasing function, the coefficients are real non-zero numbers $a, b, c, d \in \mathbb{R} \setminus \{0\}$, $E(u) = E(w) = 0$, then (9) is exactly the famous Doob-Meyer decomposition: the sub-martingales W_t and X_t can be decomposed into a non-decreasing process $g(\psi_t)$ plus martingales u_t and w_t respectively by Assumption 3(i).

Following [42] and [43], we let $y_t(S, D)$ denote the potential outcome under different realizations of $S, D \in \{1, 0\}$, so we observe $y_t(0, 0)$ for $1 \leq t < t_S$, $y_t(1, 0)$ for $t_S \leq t < t_D$ and $y_t(1, 1)$ for $t_D \leq t < T$. Then y_t in (6) can be rewritten as

$$y_t = y_t(0, 0) \cdot \mathbb{I}\{1 \leq t < t_S\} + y_t(1, 0) \cdot \mathbb{I}\{t_S \leq t < t_D\} + y_t(1, 1) \cdot \mathbb{I}\{t_D \leq t < T\} \quad (10)$$

for any $t \in \{1, \dots, T\}$. $y_t(0, 0)$ for $t_S \leq t < t_D$ then denotes the counterfactual outcome if there is no treatment S_t ; $y_t(0, 0)$ for $t_D \leq t < t_T$ denotes the counterfactual outcome if there are no treatments (S_t, D_t) ; $y_t(0, 1)$ for $t_D \leq t < t_T$ denotes the counterfactual outcome if there is no treatment S_t [44,45].

Definition 3 (Sub-sample expectation operator on the Hilbert space). Let $\mathcal{H} \subset \mathcal{L}^p(\Omega_m, \mathcal{F}_m, \mu_m)$ be the Hilbert space generated by the random process $\{m_t\}_{t \in \tau}$, $\mathcal{H}$ contains all the functions $\sum_{t \in \ell} c_t m_t$ for $\ell \subset \tau = \{1, 2, \dots, T\}$ and $c_t \in \mathcal{R}$. Define a family of operators U^ℓ on $\mathcal{H}$ as $U^\ell E(m) = \frac{1}{|\ell|} \sum_{t \in \ell} m_t$, $|\ell|$ is the size of the set ℓ . Then $U^\ell E(\cdot)$ is called the sub-interval expectation operator on the sub-space $\ell \subset \tau = \{1, 2, \dots, T\}$.

Definition 4 (Mean symmetry and conditional mean symmetry). For any random variable $x \in \mathcal{L}^p(\Omega_x, \mathcal{F}_x, \mu_x)$, if $U^{\tau_1} E(x) = U^{\tau_2} E(x)$ for $\tau_1, \tau_2 \subset \tau = \{1, 2, \dots, T\}$ where $\tau_1 \cap \tau_2 = \emptyset$, $\tau_1 \cup \tau_2 = \tau$, then we say that x satisfies mean symmetry. The conditional mean symmetry is then defined by projecting x onto another random variable $y \in \mathcal{L}^p(\Omega_y, \mathcal{F}_y, \mu_y)$ such that $U^{\tau_1} E(x | y) = U^{\tau_2} E(x | y)$.

Assumption 4 (Long range dependency) In model (6), $\text{cov}(\varepsilon_{t+n}, \varepsilon_t) \sim H(2H-1)n^{-2(1-H)}$ as $n \rightarrow \infty$ for $t = 1, 2, \dots, T$ and $T \rightarrow \infty$, the parameter $H \in (0, 1)$ is referred to as the Hurst parameter, ε_t is stationary. W_t , X_t and Z_t are stationary or trend stationary processes, hence y_t is a stationary or trend stationary process with long range dependency.

Assumption 5 (Conditional Mean Symmetry, CMS) There exist series of random sequences $\phi = \{\phi_1, \phi_2, \dots, \phi_T\}$ and $\omega = \{\omega_1, \omega_2, \dots, \omega_T\}$ such that: (i) conditional parallel: $U^{\tau_1} E(y(0, 0) - \phi | \omega) = U^{\tau_2} E(y(0, 0) - \phi | \omega)$, where $y(0, 0) = \{y_1(0, 0), y_2(0, 0), \dots, y_T(0, 0)\}$ is the counterfactual outcome sequence; and (ii) conditional mean symmetry: $U^{\tau_1} E(\phi | \omega) = U^{\tau_2} E(\phi | \omega)$, where $\tau_1 = \{1, 2, \dots, t_\lambda - 1\}$, $\tau_2 = \{t_\lambda, t_\lambda + 1, \dots, T\}$, $\lambda \in \{S, D\}$.

If we let $\chi_t = y_t(0, 0) - \phi_t$ be a martingale process with respect to the σ -filtration $\sigma(y_r(0, 0) - \phi_r, y_{r-1}(0, 0) - \phi_{r-1}, \dots, y_1(0, 0) - \phi_1)$, $r \leq t$, then $\{\chi_t\}_{t \in \tau}$ satisfies Assumption 5(i). This is because $E(\chi_t | \sigma(\chi_r, \dots, \chi_1)) = \chi_r$ and $E(\chi_t) = E(\chi_r)$, hence we can get Assumption 5(i) with $\omega = (y_r(0, 0) - \phi_r, y_{r-1}(0, 0) - \phi_{r-1}, \dots, y_1(0, 0) - \phi_1)'$. Under this scenario, Assumption 5(ii) is also satisfied. To see this, suppose that there exists an affine function $h \in \mathcal{C}^p(\Omega_h, \mathcal{F}_h, \mu_h)$ such that $E(\phi | \omega) = h(\omega)$. Note that if ω_t is a stationary process, by the von-Neumann Ergodic Theorem (see [46,47]), define $\phi^- = \{\phi_{t+k}\}_{t \in \ell}$ and $\omega^- = \{\omega_{t+k}\}_{t \in \ell}$, we can get

$$\lim_{|\ell| \rightarrow \infty} U^\ell E(\phi | \omega) = \lim_{|\ell| \rightarrow \infty} U^\ell E(\phi^- | \omega^-) = \rho \in \mathcal{H}, \text{ for any } k \in \mathbb{Z},$$

which exactly implies Assumption 5(ii) with $\ell = \{1, 2, \dots, t_\lambda - 1\}$ and $k = t_\lambda - 1$. Assumption 5 indicates that χ_t is an ergodic but non-mixing process on the Hilbert space [48]. This implies that the random sequence $\phi = (\phi_1, \phi_2, \dots, \phi_T)'$ could be chosen by satisfying the condition that $(\chi_t, \mathcal{F}_{\chi_t})_{t \in \tau}$ is a martingale process with $\chi_t = y_t(0, 0) - \phi_t$, e.g., the random sequences ϕ could be the untreated units as shown in Fig 2 (grey solid lines). This remind us that, in time-series natural experiments, one of the criterions of judging whether the selected untreated units are “good enough” is that the difference between the counterfactual outcome and the untreated unit should be a martingale process over the whole time interval.

For time series, structure change (break) is usually defined as a *before-after change*, i.e., there is a spontaneous trend, slope, structure or regime change before and after some point of time; while treatment effect is defined as a *with-without change*, i.e., the growth path of the observed y_t with D_t (or S_t), ceteris paribus, would be totally different without D_t (or S_t). Hence treatment effect is defined in a counterfactual way, while structure change or other distribution shifts effect is not [28–30,49,50]. This is also the main difference between Granger Causality, Sims Causality and the Neyman-Rubin counterfactual framework (e.g., [35,38,51,52]). Fig 3 illustrates the differences through hypothetical DGPs. It can be seen that without further restrictions (assumptions), structure change has no direct relationship with treatment effect. A (non-)zero treatment effect does not necessarily imply a (non-)zero structure change, and vice versa. Empirical researchers may hold the intuition that non-zero treatment effect implies non-zero structure change, as illustrated in Fig 3c. However, this is not always the case, see Fig 3b and (d) for examples.

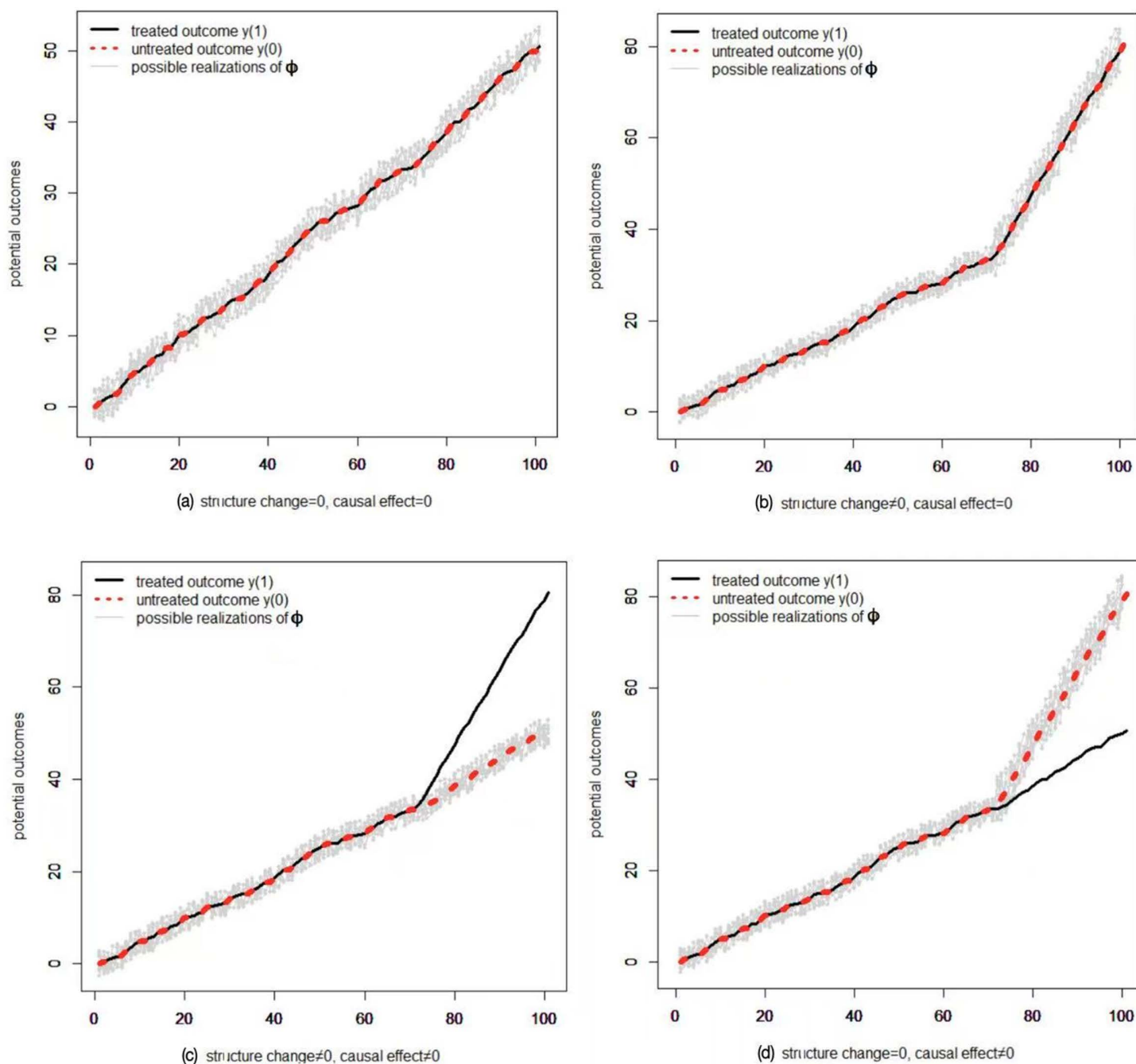


Fig 3. Comparisons between structure change and treatment effect through hypothetical examples (a-d).

<https://doi.org/10.1371/journal.pone.0347847.g003>

To make the interpretations and credibility of our assumptions clear and transparent, an intuitive illustration of Assumption 5 is further given in Fig 4 and 5, it can be seen that the random sequences $y_t(0, 0)$ and ϕ_t will not satisfy Assumption 5 (the green dots do not equal to each other in Fig 4) until we project them on to the space spanned by (the red dots equal to each other), where we take $\omega_t = \psi_t$. Note that in this paper we allow the observed outcome to be non-stationary (e.g.,

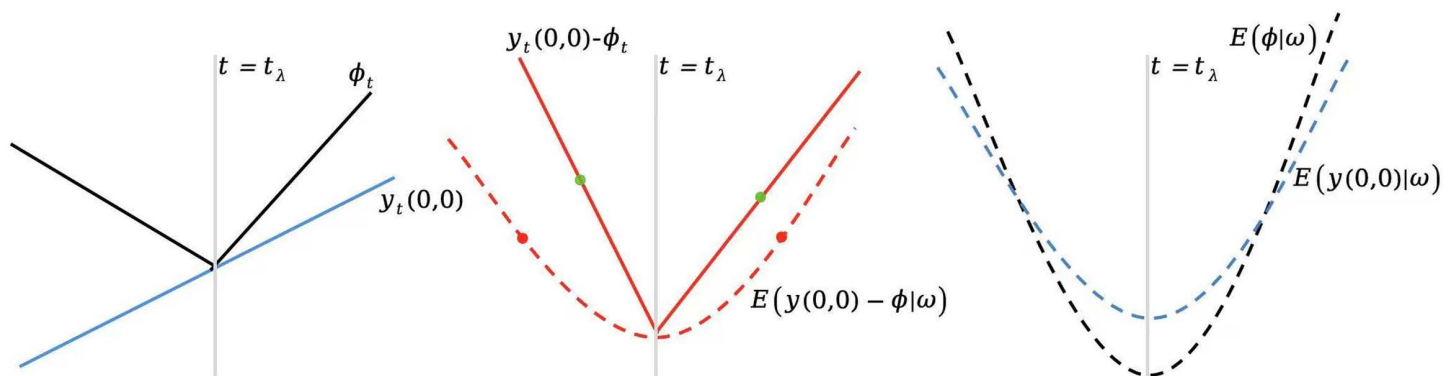


Fig 4. Hypothetical illustration example of Conditional Mean Symmetry (Assumption 5). the middle panel shows Assumption 5 (i), the right panel shows Assumption 5 (i) and (ii). The blue solid line shows the original (before-projection) counterfactual outcome $y_t(0,0)$, the black solid line shows the original (before-projection) random sequence ϕ_t , while the blue dashed line shows the after-projection counterfactual outcome $E(\phi(0,0) | \omega)$ and the black dashed line shows the after-projection sequence $E(\phi | \omega)$. The red solid line shows the difference of before-projections $y_t(0,0) - \phi_t$, while the red dashed line shows the difference of after-projections $E(y(0,0) - \phi | \omega)$. The green dots show the sub-sample means of before-projections: $U^{\tau_1} E(y(0,0) - \phi)$ and $U^{\tau_2} E(y(0,0) - \phi)$, while the red dots show the sub-sample means of after-projections: $U^{\tau_1} E(y(0,0) - \phi | \omega)$ and $U^{\tau_2} E(y(0,0) - \phi | \omega)$, $\tau_1 = \{1, 2, \dots, t_\lambda - 1\}$, $\tau_2 = \{t_\lambda, t_\lambda + 1, \dots, T\}$, $\lambda \in \{S, D\}$. The vertical line shows the time when the treatment takes place.

<https://doi.org/10.1371/journal.pone.0347847.g004>

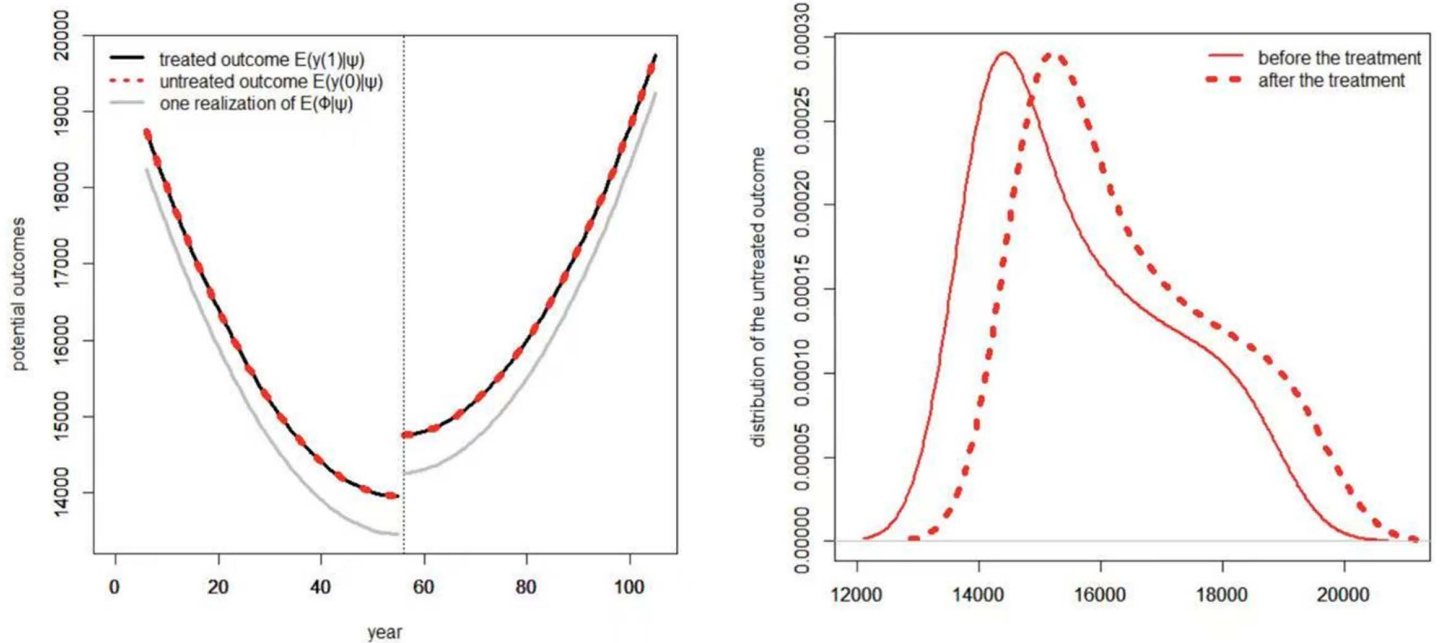
trend stationary process), while the literatures usually require (conditional) stationary for identification and inference in time-series natural experiments, see, e.g., [53,54].

Remark 2 (Matching condition) Assumption 5 rules out confounding biases caused by (W_t, X_t) and guarantees what identified are pure treatments effects. To see this, consider the DGPs (a-d) shown in Fig 3, the after-projections of the potential outcomes are shown in Fig 5 respectively. It can be seen that the DGP (b) or (d) does not satisfy Assumption 5(ii) (Conditional Mean Symmetry fails down due to the fact that there exists a structure change), and neglecting the distribution shifts shown in the right panel of Fig 5(a) will lead to false positive identification results, i.e., the identified effect stems from structural shifts, which has been misleadingly interpreted as evidence of a treatment effect; or false negative mistake, just as shown in DGP (b) or (d) of Fig 3 which satisfies Assumption 5(i) but not (ii).

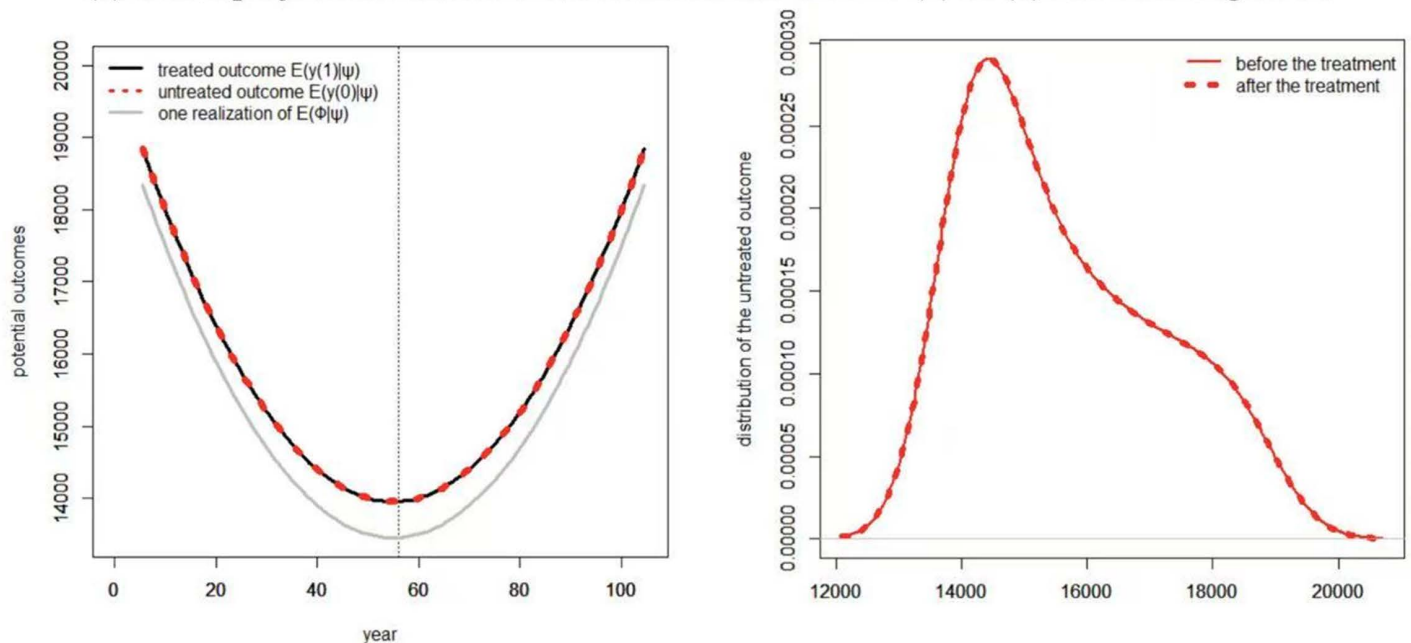
Assumption 5 implicitly implies that the treatment effect can be identified if and only if the distribution of the untreated outcome (after-projection) for the pre-treatment period shares a common support with that of the treatment period (Fig 3a and (c) shown in Fig 5b). As stressed by [28] and [32], reliable estimates obtained when the support for the treated regime coincides with or is a subset of that for the non-treated regime, and not necessarily otherwise. This is the core Matching Condition used in observational studies (e.g., [28,29]). Assumption 5 then permits us to avoid confounding the effect of treatment with shifts in the distribution of other causes.

Remark 3. Recall the general panel data model $y_{it} = g(B_{it}, \alpha_i, \epsilon_{it})$, the time-homogeneity condition assumes that $\epsilon_{it} | B_{it}, \alpha_i \rightarrow_d \epsilon_{is} | B_{is}, \alpha_i$ for all $t, s \in \{1, 2, \dots, T\}$, where B_{it} is the explanatory variable and $\rightarrow_d$ denotes distributional equivalence, the disturbance ϵ_{it} is then conditional strict stationary [55,56]. When $B_{it} \in \{0, 1\}$, the time-homogeneity condition, under our scenario (6), is equivalent to the assumption that the conditional distribution of the counterfactual untreated outcome $y_t(0,0)$ given (X_t, W_t) does not change before and after the treatment, as shown in the right panel of Fig 5(b). However, Assumption 5 does not necessarily imply the time-homogeneity condition, so Assumption 5 is less restrictive. Hence, with the distribution of factors other than D_t (or S_t) not varying over time, the changes in D_t (or S_t) over time can help identify the ceteris paribus effect of D_t (or S_t) on y_t this is exactly why and how we can achieve causal inference in time-series natural experiment through Assumption 5.

Assumption 6 (SUTVA and no anticipation) We assume stable unit treatment value (SUTVA) and no anticipation effect exist.



(a) After-projection outcomes and distributions of DGP (b) or (d) shown in Figure 2.



(b) After-projection outcomes and distributions of DGP (a) or (c) shown in Figure 2.

Fig 5. After-projection outcomes (left panel), and distributions of DGPs (b) and (c) shown in Fig 3 (right panel). We take $\omega_t = \psi_t$. For the left panel, the dark solid line shows the after-projections of the treated outcome $E\{Y(1)|\omega\}$, the red dashed line shows the after-projections of the untreated outcome $E\{Y(0)|\omega\}$, and the grey solid line shows one possible realization of $E\{\phi|\omega\}$. For the right panel, the red solid line shows the distribution of the untreated outcome (after-projection) for the pre-treatment period $t = 1, 2, \dots, t_S - 1$, while the red dashed line shows the distribution of the untreated outcome (after-projection) for the treatment period $t = t_S, t_S + 1, \dots, T$. The vertical line shows the time when the treatment takes place. For expositional convenience, we only consider model (6) with single treatment S_t .

<https://doi.org/10.1371/journal.pone.0347847.g005>

Theorem 1 (Identifications). For model (6) with $\tau_4 = \{t_D, t_D + 1, \dots, T\}$, $\tau_3 = \{1, 2, \dots, t_D - 1\}$, $\tau_2 = \{t_S, t_S + 1, \dots, t_D - 1\}$ and $\tau_1 = \{1, 2, \dots, t_S - 1\}$: (i) Before-After (structure change effect): if Assumption 5(i) is satisfied, then we have

$$\beta + \xi = U^{\tau_4} E(y(1, 1) | S, D, \omega) - U^{\tau_3} E(y(0, 0) | S, D, \omega),$$

$$\xi = U^{\tau_4} E(y(1, 1) | S, D, \omega) - U^{\tau_2} E(y(1, 0) | S, D, \omega),$$

$$\beta = U^{\tau_2} E(y(1, 0) | S, D, \omega) - U^{\tau_1} E(y(0, 0) | S, D, \omega).$$

(ii) With-Without (treatment effect): if Assumptions 5(i)-(ii) are both satisfied, we will then get

$$\beta + \xi = U^{\tau_4} E(y(1, 1) - y(0, 0) | S, D, \omega),$$

$$\xi = U^{\tau_4} E(y(1, 1) - y(1, 0) | S, D, \omega),$$

$$\beta = U^{\tau_2} E(y(1, 0) - y(0, 0) | S, D, \omega).$$

The Before-After part (i) identifies the structure change effect, while the With-Without part (ii) identifies the treatment effect.

We are now able to give a formal distinction between treatment effect and structure change effect:

Definition 5 (Causal field and multiverse). For model (6), the random processes triplets (y_t, D_t, X_t) is called a causal field where $E(y|D, X) = E(y|X)$ and $E(yD) \neq E(y)E(D)$; while the random process triplets (y_t, D_t, W_t) is called a structure change field where $E(y|D, W) \neq E(y|D)$ and $E(yD) \neq E(y)E(D)$. Similarly, the triplet (y_t, S_t, W_t) is a causal field and the triplet (y_t, S_t, X_t) is a structure change field. For the pairs (y_t, S_t) and (y_t, D_t) without further information, we cannot tell whether they are causal fields or structure change fields. The quadruplets (y_t, D_t, S_t, X_t) is called a time-bowl (the quadruplets (y_t, D_t, S_t, W_t) is also a time-bowl), the fields under all time-bowls constitute a multiverse:

$$\{(y_t, D_t, X_t), (y_t, D_t, W_t), (y_t, S_t, X_t), (y_t, S_t, W_t)\}.$$

The interesting *time-bowl* and *multiverse* concepts are borrowed from the 2023 DC movie The Flash. The *time-bowl* corresponds to a specific universe and the *multiverse* is constructed by different specific universes, the Flash Man (the protagonist of the movie) can enter any *time-bowl* from the *multiverse* by his superpower. I'm inspired by these originalities, we econometricians are just like the Flash Man whose superpower are different model assumptions under which we can disentangle one *time-bowl* (e.g., causal field) from another *time-bowl* (e.g., structure change field).

It can be seen from Theorem 1 and Definition 5 that the parameters of interest (ξ, β) in model (6) have different meanings under different assumptions. The before-after part under Assumption 5(i) identifies a structure change effect, while the with-without part under Assumptions 5(i)-(ii) identifies a causal effect. This finding implies that the parallel assumption alone in DiD settings is not sufficient enough to disentangle treatment effect from structure change effect. The presence of a non-zero estimated effect under the parallel trends assumption does not necessarily imply a causal treatment effect. An alternative explanation is that the treated units were undergoing a spontaneous structural change, while the untreated units were not. Under Assumption 5, a (non-)zero treatment effect implies a (non-)zero structure change as shown in

Fig 3(a) and (c), but not vice versa as shown in Fig 3(b) and (d). In observation studies, the triplet (y_t, D_t, X_t) or (y_t, S_t, W_t) describes an “economic story”, which ascribes a causal meaning to the empirical findings. Without such a “story” (the relationship between y_t and D_t (S_t) given X_t (W_t)), we can hardly claim that what we finally get is actually a causal relationship.

The plausibility of the CMS assumption

To justify the Conditional Mean Symmetry (CMS) assumption and demonstrate its plausibility, we illustrate its logic step-by-step using a specific economic example: the impact of a national minimum wage policy on employment. Suppose a national minimum wage policy (the treatment variable D_t) was enacted on January 1, 2000. We want to evaluate the causal effect of this policy on the youth employment rate (y_t). This is a classic time-series natural experiment with no untreated units, as the entire country is affected.

The challenging we are facing now is that the employment rate is influenced not only by the policy but also by numerous unobservable confounding factors, such as: (i) Macroeconomic cycles: Employment is high during economic booms and low during recessions. (ii) Industrial structure adjustments: An increase in the service sector’s share and a decrease in manufacturing also affect overall employment. (iii) Demographic changes: Changes in the proportion of the youth population. These factors simultaneously influence whether the policy might be considered (as economic pressures form the backdrop for policy enactment) and the employment outcome.

Step 1: Finding the “Common Proximal Variable (CP)” ψ_t . According to Remark 1, under specific economic scenarios, we can take $\psi_t = t$ (the time trend) as the CP variable. The rationale for this choice is that the confounding factors (economic cycles, industrial structure, demographics) all change systematically over time. By Assumptions 2–3, we can use a function of time (e.g., $g(\cdot)$) to roughly capture these trends, i.e., we assume that the macroeconomic condition W_t can be decomposed as $W_t = ag(\psi_t) + b + u_t$, and the industrial structure X_t can be decomposed as $X_t = cg(\psi_t) + d + w_t$. Here, $g(\psi_t)$ is a smoothly varying function of time, while u_t and w_t are short-term random fluctuations around this trend and are independent of the error term ε_t .

Step 2: Applying CMS(i) (Conditional Parallelism). In our example, CMS(i) implies that $y_t(0)$ is the employment rate “if there were no minimum wage policy”. ϕ_t is the “shadow variable” used for adjustment. It can be understood as the part of the outcome highly correlated with the CP variable (ψ_t). We can simplify this to the part of the employment rate explained by the time trend. ω_t here can be constructed from $y'(t)$.

The economic meaning of CMS(i) is: After removing the part explained by the time trend (and the confounding factors associated with it) (ϕ_t), the behavior (e.g., its expected value) of the remaining “pure random shock” part ($y_t(0) - \phi_t$) is the same before (τ_1) and after (τ_2) the policy implementation. In layman’s terms: Without the policy, employment would fluctuate due to economic cycles. CMS(i) assumes that the “pattern” of these fluctuations is the same before and after the policy. For example, during an economic upturn, employment would be $x\%$ above the trend line; during a downturn, it would be $y\%$ below the trend line. This “pattern of deviation from the trend” is stable.

Step 3: Applying CMS(ii) (Mean Symmetry). In our example, ϕ_t can be roughly understood as the “potential employment level driven by the long-term trend”. CMS(ii) assumes that: The employment level determined by the long-term trend itself did not undergo a structural break around the time of the policy. That is, we cannot say that the “long-term trend line” itself had a discontinuity in its slope or intercept right after January 1, 2000, compared to before. The trend line is continuous and smooth across the policy implementation point.

In this specific example, the plausibility of the CMS assumption depends on judgment of the economic context:

- **Plausibility of CMS(i):** The assumption that the “pattern of deviation of employment from its long-term trend” is stable is a common hypothesis in macroeconomics. Business cycles (boom-recession-recovery) do exhibit a certain regularity and cyclicity [2]. As long as there are no other major structural changes (e.g., war, technological revolution), this pattern of fluctuation around the trend is likely to be similar.

- *Plausibility of CMS(ii)*: The assumption that the “long-term trend itself” did not have a structural break around the policy implementation is precisely what we need to verify. The purpose of the minimum wage policy is to change the employment level, so it could very well cause a change in the intercept or slope of the trend line. CMS(ii) essentially states that the “trend” part captured by ϕ_t does not include the effect of the policy. The entire policy effect is intended to be captured by the parameter we ultimately estimate, ξ .

The CMS assumption is a stronger assumption than the parallel trends assumption used in DiD settings in that it attempts to substitute cross-unit comparability with within-unit stability over time; however, it is also more flexible because it does not require the existence of a unit unaffected by the treatment.

Empirical manifestations of CMS violations

Understanding how CMS assumptions would be violated in real-world data is crucial for applied researchers. Below, I describe what empirical violations would look like for each component of the CMS assumption, using concrete economic examples.

Violations of CMS(i):

- *Pattern 1: Time-varying volatility in residuals*. Imagine plotting the residuals $Y_t - \hat{Y}_t$ (where $\hat{Y}_t$ is the projection onto the CP variable) over time. A violation of CMS(i) would show: For the pre-treatment period, residuals fluctuate within a stable range (e.g., $\pm 1\%$); while for the post-treatment period, the pattern of fluctuations changes fundamentally—perhaps they become more volatile, or the auto-correlation structure changes. In the minimum wage example, suppose before 2000, employment deviations from trend were mild and short-lived (2–3 quarters). After 2000, these deviations become persistent and last 5–7 years. This suggests the underlying economic dynamics have changed, violating the “same fluctuation pattern” assumption.
- *Pattern 2: Structural break in the relationship with the CP*. Another violation occurs if the relationship between the counterfactual untreated outcome and the CP variable changes over time. This would appear as: For the pre-treatment period, $E(Y(0, 0)|\omega)$ follows a function $K(\omega)$; however for the post-treatment period, the function shifts to $P(\omega) \neq K(\omega)$. Consider the divorce law reform example from the later empirical analysis part of the paper. If the relationship between divorce rates and time (the CP variable) fundamentally changed after the 1997 Asian financial crisis—not just a level shift but a change in how divorce responds to economic conditions—this would violate CMS(i). The “adjustment” ϕ_t that worked before the crisis would no longer be appropriate after.

Violations of CMS(ii):

- *Pattern 1: Level shifts in the projected component*. The ϕ_t series shows a discrete jump exactly at the treatment time that cannot be explained by the treatment itself. In the minimum wage case, suppose ϕ_t captures the “long-term trend” in employment driven by demographics. If a new demographic survey methodology was introduced in 2000 that redefined how youth population is measured, the trend line would shift mechanically—violating CMS(ii) even if the minimum wage policy had no effect.
- *Pattern 2: Slope changes in the projected component*. The rate of change of ϕ_t shifts at the treatment boundary. In the divorce law example, suppose the CP variable is time t , capturing the natural upward trend in divorce rates due to secular social changes. If the 1997 financial crisis permanently accelerated the pace of social change (making people more individualistic faster), then the slope of ϕ_t would increase after 1997. This would violate CMS(ii) because the ϕ_t used to adjust pre-treatment data would not be appropriate post-treatment.
- *Pattern 3: Correlation between ϕ_t and the treatment indicator*. ϕ_t becomes correlated with D_t or S_t after controlling for ω_t . This would appear as: In a regression of ϕ_t on D_t and ω_t , the coefficient on D_t is statistically significant. This indicates that the “pure trend” component is actually contaminated by the treatment itself.

In summary, the violations of CMS(i) come from: units learn from the treatment and respond differently to subsequent shocks; the treatment fundamentally alters how the economy/society responds to other factors; the set of units changes over time (in panel settings); and treatment effects spill over to affect the untreated state. The violations of CMS(ii) would come from: other policy changes or shocks coinciding with the treatment; the treatment itself alters the long-term trends it was supposed to be independent of; and selecting a proximal variable that is itself affected by the treatment.

Diagnose of the CMS assumption

When untreated units (control units) are present, the random sequence ϕ_t could be selected as the untreated units, and the CMS assumption could be tested by the empirical process-based methods proposed by [57], and apply them to our constructed residual term $y_t(0, 0) - \phi_t$ on the pre-treatment period $[1, t_{s^*}]$ where we assume $t_{s^*} < t_s < t_D$. Truncate the sample to the period $\tau_1 = [1, t_0 - 1]$ and $\tau_2 = [t_0, t_s^*]$ (the “faked pre- and post-treatment” interval), if the estimated pseudo effect $\Delta_{-}(i) = U^{\tau_1} E(y(0, 0) - \phi | \omega) - U^{\tau_2} E(y(0, 0) - \phi | \omega)$ or $\Delta_{-}(ii) = U^{\tau_1} E(\phi | \omega) - U^{\tau_2} E(\phi | \omega)$ is significantly nonzero, then CMS is not satisfied. However, when untreated units (control units) are absent, CMS could not be tested directly. Two approaches can be adapted to construct possible indirect tests for CMS:

1. Placebo tests using a “fake treatment time.” We know no actual treatment occurred on the “faked pre- and post-treatment” interval τ_1 and τ_2 , so the true treatment effect should be zero. Apply the paper’s two-step estimation procedure to this hypothetical interval to estimate the “treatment effect.” If we obtain estimates that are statistically significantly different from zero, this provides strong evidence that the CMS assumption (especially CMS(i)) might be violated. Because even in the absence of treatment, our method falsely identifies an “effect,” suggesting the projection mechanism used to construct the counterfactual is biased.
2. Testing implications of Conditional Mean Independence (CMI). Theorem 3 of the paper states that the proposed two-step estimation procedure implies conditional mean independence: $E(YS|\hat{Y}) = E(Y|\hat{Y})$. This means that, conditional on $\hat{Y}_t$, the treatment variable S_t should be unrelated to potential outcomes. While potential outcomes are unobservable, we can test whether S_t is related to some functions of $\hat{Y}_t$. In the sample including the true pre-treatment period (τ_1) and the post-treatment period (τ_2), we can test whether $E(S_t | \hat{Y}_t)$ is constant. A non-constant relationship would indicate a violation of CMI and thus CMS.

Choice of the CP variable.

- (i) ψ_t should be a sufficient statistics of the confounders or should be of high-correlation with the confounders. This guarantees that the CP variable is not weak.
- (ii) ψ_t should be independent of ε_t , which guarantees that the CP variable is valid.

These two requirements imply that for trend stationary social-economic outcome y_t , the CP could be selected as (i) time t ; or (ii) common factor f_t extracted from $Z_{it} = \mu_t + \lambda_i f_t + e_{it}$, where Z_{it} are multiple macroeconomic variables (such as GDP, consumption, and investment across regions) correlated with y_t but independent of the policy, f_t then is a good CP variable as a measure of the “economic cycle”.

Researchers should select CP variable by integrating economic theory with the context of the research question. [Table 1](#) provides some concrete examples.

Time ($\psi_t = t$) is a natural and convenient CP because many confounders trend over time. However, there are several scenarios where time fails to satisfy the CP assumptions: (i) Treatment-induced structural break in the trend: If the treatment itself alters the long-term trajectory of the confounders (or the outcome’s relationship with time), then using a simple time trend violates CMS(ii); (ii) Non-linear or non-smooth confounding: If confounders evolve in a way that cannot be captured by a smooth function of time (e.g., abrupt shifts due to wars, natural disasters, or technological breakthroughs), a simple time trend will not proxy them well. The residuals u_t and w_t would then be correlated with time, violating the martingale difference

Table 1. Candidate Common Proximal variables for typical empirical settings.

Treatment/Outcome	Unobserved Confounders	Candidate CP variable	Rationale
The effect of Minimum Wage Policy on youth employment rate	Macroeconomic conditions (business cycle), industrial structure, demographic trends	Time trend or GDP growth of trading partners	Time captures long-run trends; external GDP is unaffected by domestic policy but correlated with domestic cycles
The effect of education reform on students' test scores	Teaching quality, family background, school resources	Regional GDP per capita (if exogenous to education policy)	GDP reflects economic development that influences both education inputs and student performance
The effect of environmental regulation on air quality	Industrial activity, weather patterns, technology adoption	Industrial production index of major trading partners	External demand drives local industrial activity but may be exogenous to local environmental policy
The effect of health-care policy on life expectancy	Nutrition, lifestyle, sanitation	Caloric intake per capita (if largely driven by income trends)	Nutrition is a fundamental determinant of health and evolves with economic growth, but may be correlated with policy if policy affects income

The table provides examples of candidate Common Proximal variables for various treatment-outcome pairs, along with the unobservable confounders they aim to capture and the rationale for their choice.

<https://doi.org/10.1371/journal.pone.0347847.t001>

assumption; (iii) Multiple confounders with different trends: When there are several confounders with different time profiles (e.g., one increasing linearly, another following a U-shape), a single time trend cannot simultaneously capture all. The CP would be misspecified; (iv) Seasonality or high-frequency fluctuations: In quarterly or monthly data, time may not capture seasonal patterns or business cycle frequencies that act as confounders. In such cases, a more sophisticated CP (e.g., a business cycle indicator) is needed; (v) When the treatment timing coincides with another aggregate shock: If the treatment date coincides with a major event (e.g., a financial crisis), the time trend alone cannot distinguish between the treatment effect and the crisis effect. The crisis would appear as a structural break in the trend, violating CMS(i) and (ii).

Multiple CPs are admissible, and researchers can attempt to select the “best” one—or a combination of them. However, within the specific CMS framework of this paper, the theory is developed for a univariate CP ψ_t . If multiple CPs are available, the researcher faces a model selection problem: choosing which variable (or which combination) best satisfies the CMS assumption. This can be left for future research.

Estimation by weak index restrictions and inference with long range dependency

Substitute (9) into model (6) and directly estimate model (6) will not be correct due to the fact that $\text{cov}(\lambda_t, \varepsilon_t) \neq 0$, $\lambda_t \in \{S_t, D_t\}$. In this section, we propose a novel method to deal with this issue without IVs.

A two-step estimation approach by weak index restrictions. Step 3.1. (Estimation of β) Based on the Bernstein expansions (7)–(8), we will get

$$y'(t) = 2\rho_{y,2}\psi_t + \rho_{y,1}, \quad S'(t) = 2\rho_{S,2}\psi_t + \rho_{S,1} \quad \text{for } t = 1, 2, \dots, t_S, \dots, t_D - 1.$$

By symbolic computations, see [58,59], there exists a continuous function $K(\cdot, \cdot) \in C^2(\mathbb{R})$ such that the relationship between $y'(t)$ and $S'(t)$ satisfies:

$$K(y'(t), S'(t)) = a_{S,2}y'(t) - a_{y,2}S'(t) + a_{y,2}a_{S,1} - a_{S,2}a_{y,1} = 0 \quad \text{with } a_{\sigma,b} = b\rho_{\sigma,b},$$

$\rho_{\sigma,b}$ is the Bernstein coefficient defined in (7)–(8) for $b \in \{1, 2\}$ and $\sigma \in \{y, S\}$. Estimate (7)–(8) by OLS to get the estimators $\hat{a}_{\sigma,b} = b\hat{\rho}_{\sigma,b}$ respectively for b and σ , and input all the estimators into the equation $K(y'(t), S'(t))$, we will get

$$\hat{y}'(t) = (\hat{a}_{y,2}S'(t) + \hat{a}_{S,2}\hat{a}_{y,1} - \hat{a}_{y,2}\hat{a}_{S,1})/\hat{a}_{S,2}.$$

Then further consider the following auxiliary regression:

$$Y = XB + W, \quad (11)$$

$$\hat{Y}_{OLS} = X\hat{B}_{OLS} = X(X'X)^{-1}X'Y, \quad (12)$$

$$X = (I \hat{y}')_{(t_D-1) \times 2}, \quad B = (\delta_0 \delta_1)', \quad (13)$$

$$I = (1, 1, \dots, 1)'_{(t_D-1) \times 1}, \quad (14)$$

$$Y = (y_1, y_2, \dots, y_{t_D-1})', \quad (15)$$

$$\hat{y}' = (\hat{y}'(1), \hat{y}'(2), \dots, \hat{y}'(t_D - 1))', \quad (16)$$

W is an unobserved random shock. Like Indirect Inference [60], we do not require $\hat{B}_{OLS}$ to be consistent for B in the auxiliary regression (11). Substitute (6) into $\hat{Y}_{OLS}$, we will then get

$$\begin{aligned} U^{\tau_2}E(\hat{Y}_{OLS}) - U^{\tau_1}E(\hat{Y}_{OLS}) &= (U^{\tau_2}E(\hat{y}) - U^{\tau_1}E(\hat{y})) \left(\beta \frac{U^{\tau}E(S\hat{y}) - U^{\tau}E(S)U^{\tau}E(\hat{y})}{U^{\tau}E(\hat{y}\hat{y}) - U^{\tau}E(\hat{y})U^{\tau}E(\hat{y})} \right. \\ &\quad + \eta \frac{U^{\tau}E(W\hat{y}) - U^{\tau}E(W)U^{\tau}E(\hat{y})}{U^{\tau}E(\hat{y}\hat{y}) - U^{\tau}E(\hat{y})U^{\tau}E(\hat{y})} + \alpha \frac{U^{\tau}E(X\hat{y}) - U^{\tau}E(X)U^{\tau}E(\hat{y})}{U^{\tau}E(\hat{y}\hat{y}) - U^{\tau}E(\hat{y})U^{\tau}E(\hat{y})} \\ &\quad \left. + \gamma \frac{U^{\tau}E(Z\hat{y}) - U^{\tau}E(Z)U^{\tau}E(\hat{y})}{U^{\tau}E(\hat{y}\hat{y}) - U^{\tau}E(\hat{y})U^{\tau}E(\hat{y})} + \frac{U^{\tau}E(\varepsilon\hat{y}) - U^{\tau}E(\varepsilon)U^{\tau}E(\hat{y})}{U^{\tau}E(\hat{y}\hat{y}) - U^{\tau}E(\hat{y})U^{\tau}E(\hat{y})} \right) \\ &\equiv \beta \mathcal{A}_1 + \eta \mathcal{A}_2 + \alpha \mathcal{A}_3 + \gamma \mathcal{A}_4 + \mathcal{A}_5 = \beta \mathcal{A}_1 + \eta \mathcal{A}_2 + \alpha \mathcal{A}_3 + \gamma \mathcal{A}_4 \end{aligned}$$

for $\tau_1 = \{1, 2, \dots, t_S - 1\}$, $\tau_2 = \{t_S, \dots, t_D - 1\}$ and $\tau = \{1, 2, \dots, t_D - 1\}$, the last equality holds true ($\mathcal{A}_5 = 0$) by (6) and Assumption 2 (ii). Notice that by Assumption 3, model (6) could be rewritten as

$$y_t = (a\eta + c\alpha)g(y'(t)) + S_t\beta + D_t\xi + Z_t\gamma + e + u_t + w_t + \varepsilon_t, \quad \text{cov}(\lambda_t, \varepsilon_t) \neq 0, \quad (17)$$

where $e = b\eta + d\alpha$, $\lambda_t \in \{S_t, D_t\}$, $t = 1, 2, \dots, t_D - 1$, thereby we can get

$$U^{\tau_2}E(\hat{Y}_{OLS}) - U^{\tau_1}E(\hat{Y}_{OLS}) = \beta \mathcal{A}_1 + (a\eta + c\alpha)(U^{\tau_2}E(\hat{y}) - U^{\tau_1}E(\hat{y})) + \gamma \mathcal{A}_4.$$

Estimate (17) by semiparametric methods, under Assumption 2 (ii) we will get $(a\eta + c\alpha) = (a\eta + c\alpha) + o_p(1)$ and $\hat{\gamma} = \gamma + o_p(1)$ as $t_D - 1 \rightarrow \infty$, see, e.g., [61]. Substitute these estimators into $U^{\tau_2}E(\hat{Y}_{OLS}) - U^{\tau_1}E(\hat{Y}_{OLS})$, we will finally get

$$\hat{\beta} = \frac{U^{\tau_2}E(\hat{Y}_{OLS}) - U^{\tau_1}E(\hat{Y}_{OLS}) - (a\eta + c\alpha)(U^{\tau_2}E(\hat{y}) - U^{\tau_1}E(\hat{y})) - \hat{\gamma} \mathcal{A}_4}{\mathcal{A}_1}.$$

Assumption 6 As $t_D - 1 \rightarrow \infty$, we have (i) $\frac{t_D - t_S}{t_D - 1} = o_p(1)$;

(ii) $\sum_{t=t_S}^{t_D-1} Z_t < \infty$ and $\sum_{t=t_S}^{t_D-1} y_t < \infty$;

(iii) $U^T E(y^2) < \infty$, $U^T E(Z^2) < \infty$, $U^T E(yZ) < \infty$, $U^T E(y) < \infty$ and $U^T E(Z) < \infty$;

(iv) $\sum_{t=1}^{t_D-1} S_t(u_t + w_t + \varepsilon_t) < \infty$, $\tau = \{1, 2, \dots, t_D - 1\}$.

Theorem 2. (UNBIASEDNESS AND CONSISTENCY) Under the Assumptions 1–5 and Assumption 6, $E(\hat{\beta}) = \beta$, and $\hat{\beta} = \beta + o_p(1)$.

STEP 3.2. (ESTIMATION OF ξ) The estimation approach is similar to Step 3.1 except for replacing the time interval $[1, t_D - 1]$ in (11) with $[t_S, T]$, and replacing $S'(t)$ with $D'(t)$. The estimator we finally get is

$$\hat{\xi} = \frac{U^{\tau_4} E(\hat{Y}_{OLS}) - U^{\tau_1} E(\hat{Y}_{OLS}) - \hat{\beta} \tilde{\mathcal{A}}_1 - (\widehat{a\eta + c\alpha})(U^{\tau_4} E(\hat{Y}) - U^{\tau_3} E(\hat{Y})) - \hat{\gamma} \tilde{\mathcal{A}}_3}{\tilde{\mathcal{A}}_4},$$

where $\tau_3 = \{t_S, t_S + 1, \dots, t_D - 1\}$, $\tau_4 = \{t_D, t_D + 1, \dots, T\}$ and $\tau = \{t_S, t_S + 1, \dots, T\}$, $\hat{Y}_{OLS} = X\hat{B}_{OLS} = X(X'X)^{-1}X'Y$ with $X = (I\hat{Y})_{(T-t_S+1) \times 2}$, $B = (\delta_0 \delta_1)'$, $I = (1, 1, \dots, 1)_{(T-t_S+1) \times 1}$, $Y = (y_{t_S}, y_{t_S+1}, \dots, y_T)'$, and $\hat{y} = (\hat{y}'(t_S), \hat{y}'(t_S + 1), \dots, \hat{y}'(T))'$. The coefficients are

$$\tilde{\mathcal{A}}_1 = (U^{\tau_4} E(\hat{Y}) - U^{\tau_1} E(\hat{Y})) \frac{U^{\tau} E(S\hat{Y}) - U^{\tau} E(S)U^{\tau} E(\hat{Y})}{U^{\tau} E(\hat{Y}\hat{Y}) - U^{\tau} E(\hat{Y})U^{\tau} E(\hat{Y})},$$

$$\tilde{\mathcal{A}}_3 = (U^{\tau_4} E(\hat{Y}) - U^{\tau_3} E(\hat{Y})) \frac{U^{\tau} E(Z\hat{Y}) - U^{\tau} E(Z)U^{\tau} E(\hat{Y})}{U^{\tau} E(\hat{Y}\hat{Y}) - U^{\tau} E(\hat{Y})U^{\tau} E(\hat{Y})},$$

$$\tilde{\mathcal{A}}_4 = (U^{\tau_4} E(\hat{Y}) - U^{\tau_3} E(\hat{Y})) \frac{U^{\tau} E(D\hat{Y}) - U^{\tau} E(D)U^{\tau} E(\hat{Y})}{U^{\tau} E(\hat{Y}\hat{Y}) - U^{\tau} E(\hat{Y})U^{\tau} E(\hat{Y})}.$$

Under the Assumptions 1–5, by Theorem 1 and similar to Theorem 2, we can get $E(\hat{\xi}) = \xi$, and $\hat{\xi} = \xi + o_p(1)$ as long as conditions similar to these in Assumption 6 are satisfied, i.e., (i) $\frac{T - t_D}{t_D - t_S} = o_p(1)$;

(ii) $\sum_{t=t_D}^T Z_t < \infty$ and $\sum_{t=t_D}^T y_t < \infty$;

(iii) $U^T E(y^2) < \infty$, $U^T E(Z^2) < \infty$, $U^T E(yZ) < \infty$, $U^T E(y) < \infty$ and $U^T E(Z) < \infty$;

(iv) $\sum_{t=t_S}^T D_t(u_t + w_t + \varepsilon_t) < \infty$, $\tau = \{t_S, t_S + 1, \dots, T\}$.

The following statements clarifies why and how the above proposed estimation approach satisfies Assumption 5 and guarantees Theorem 1:

Theorem 3. (CONDITIONAL MEAN INDEPENDENCE, CMI) For the 2-steps estimation approach, we have $E(\mathcal{Y}S|\hat{Y}) = E(\mathcal{Y}|\hat{Y})$, where $\mathcal{Y}_t = (y_t(1, 0), y_t(0, 0))$ for Step 3.1; and $E(\mathcal{Y}D|\hat{Y}) = E(\mathcal{Y}|\hat{Y})$, where $\mathcal{Y} = (y_t(1, 1), y_t(0, 0))$ for Step 3.2.

Interestingly, Theorem 3 indicates that given $\hat{Y}(t)$, which is constructed by the CP ψ_t , the treatments (S_t, D_t) are nearly randomly assigned, which is a weak version of the *unconfoundedness condition* $(y_t(1, 1), y_t(0, 0)) \perp \lambda_t | \hat{Y}(t)$ for $\lambda_t \in \{S_t, D_t\}$.

Remark 4. (WEAK INDEX RESTRICTIONS) Theorem 3 is natural in the sense that the CP $\psi_t = h(W_t, X_t)$ is an index function of the confounders (W_t, X_t) by Assumptions 2–3. The conditions $E(\pi|X) = E(\pi|\psi)$ and $E(\pi|W) = E(\pi|\psi)$, in which

$\pi = u + w + \varepsilon$ is the error term defined in model (17), are the well known weak or mean index restrictions widely used in semiparametric identifications ([62–64]). Under weak index restrictions, the treatments effects can be uniquely identified and consistently estimated even if untreated units and IVs are not available.

Remark 5. (CMI IMPLIES CMS) To see how CMI implies CMS, for expositional purposes, we consider model (6) with $\beta = 0$ (Remark 5 is still valid for the case where $\beta \neq 0$. But for convenience, we only consider the single treatment situation), and we assume that $t_D = T/2$. For Assumption 5 (ii), we take $\omega_t = \psi_t$. Because $\hat{y}'(t)$ is a continuous and measurable function of ψ_t , hence $\omega_t = \hat{y}'(t)$ also satisfies Assumption 5 (ii). For Assumption 5 (i), note that

$$\hat{y}'(t) = (\hat{a}_{y,2}S'(t) + \hat{a}_{S,2}\hat{a}_{y,1} - \hat{a}_{y,2}\hat{a}_{S,1})/\hat{a}_{S,2} \equiv \alpha(t) \cdot t + \beta(t)$$

for some $\alpha(t), \beta(t) \in C^2(\mathbb{R})$, hence one can verify that $\hat{y}'(t \in \tau_1) = a \cdot \hat{y}'(t \in \tau_2) + b$ for $a=1$ and $b = 2\rho_{y,2} \cdot |\tau_2|$, where $\hat{y}'(t \in \tau_1) = (\hat{y}'(1), \hat{y}'(2), \dots, \hat{y}'(t_D - 1))'$ and $\hat{y}'(t \in \tau_2) = (\hat{y}'(t_D), \hat{y}'(t_D + 1), \dots, \hat{y}'(T))'$. By the proof of Theorem 3, we can get $y_t(0, 0) = a(t) \cdot \hat{y}'(t) + b(t) + \varepsilon_t$ for some $a(t), b(t) \in C^2(\mathbb{R})$, and $E(\varepsilon) = 0$. Substitute $y_t(0, 0)$ into Assumption 2 (i), we will then get

$$U^j E(y(0, 0) | \hat{y}') = a(t \in \tau_j) \cdot \hat{y}'(t \in \tau_j) + b(t \in \tau_j), \quad j = 1, 2.$$

After simplifications, by Assumption 5 (i),

$$\begin{aligned} & U^{\tau_2} E(y(0, 0) | \hat{y}') - U^{\tau_1} E(y(0, 0) | \hat{y}') \\ &= \{a \cdot a(t \in \tau_2) - a(t \in \tau_1)\} \cdot \hat{y}'(t \in \tau_1) \\ &+ \{b \cdot a(t \in \tau_2) + b(t \in \tau_2) - b(t \in \tau_1)\} = 0 \end{aligned}$$

if and only if the following balance conditions are satisfied:

$$a \cdot a(t \in \tau_2) - a(t \in \tau_1) = 0 \quad \text{and} \quad b \cdot a(t \in \tau_2) + b(t \in \tau_2) - b(t \in \tau_1) = 0.$$

Simultaneously considering Theorem 3 and the above balance conditions, CMI implies CMS if and only if

$$\begin{aligned} & a(t \in \tau_2) - a(t \in \tau_1) = 0, \\ & 2\rho_{y,2} \cdot |\tau_2| \cdot a(t \in \tau_2) + b(t \in \tau_2) - b(t \in \tau_1) = 0, \\ & U^{\tau_2} E(y(1, 1)) - a(t \in \tau_2) \cdot U^{\tau_2} E(\hat{y}') - b(t \in \tau_2) - c(t \in \tau_2) \cdot U^{\tau_2} E(D) = 0, \\ & U^{\tau_1} E(y(1, 1)) - a(t \in \tau_1) \cdot U^{\tau_1} E(\hat{y}') - b(t \in \tau_1) - c(t \in \tau_1) \cdot U^{\tau_1} E(D) = 0, \end{aligned}$$

where $U^{\tau_2} E(D) = 1$ and $U^{\tau_1} E(D) = 0$; $c(t \in \tau_2)$ and $c(t \in \tau_1)$ are defined in the proof of Theorem 3 (see the online Supplementary Material [S2 File](#)), $\tau_1 = \{1, 2, \dots, t_D - 1\}$, $\tau_2 = \{t_D, t_D + 1, \dots, T\}$. The above simultaneous functions are equivalent to

$$\begin{pmatrix} 1 & 0 & \Sigma & \Omega \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 \\ -\Xi & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a(t \in \tau_2) \\ b(t \in \tau_1) \\ b(t \in \tau_2) \\ c(t \in \tau_2) \end{pmatrix} = \begin{pmatrix} \Pi \\ \Phi \\ \Psi \\ \Gamma \end{pmatrix}, \quad \text{which we denote as } X\theta = \Delta,$$

where

$$\begin{aligned}\Xi &= 1/\{2\beta_{y,2} \cdot |\tau_2| - (U^{\tau_2} E(\hat{Y}) - U^{\tau_1} E(\hat{Y}))\}, \\ \Omega &= U^{\tau_2} E(\hat{Y}), \\ \Phi &= U^{\tau_1} E(Y(1, 1)), \\ \Pi &= -\{U^{\tau_2} E(Y(1, 1)) - U^{\tau_1} E(Y(1, 1))\}/\{2\beta_{y,2} \cdot |\tau_2| - (U^{\tau_2} E(\hat{Y}) - U^{\tau_1} E(\hat{Y}))\}, \\ \Psi &= U^{\tau_1} E(\hat{Y}), \\ \Sigma &= 2\beta_{y,2} \cdot |\tau_2|, \\ \Gamma &= U^{\tau_2} E(Y(1, 1)).\end{aligned}$$

Under this scenario one can verify that, by some elementary but tedious computations, $\text{rank}(X'X) = 4$, hence θ can be uniquely point identified and estimated by $(X'X)^{-1}X'\Delta$. This implies that CMI is compatible with CMS, there is an equivalent relationship between CMI and CMS.

Inference with long range dependency

We consider a bootstrap procedure. For the estimation Step 3.1, consider divide the whole time interval $[1, 2, \dots, t_S, \dots, t_D - 1]$ into sub-intervals $I_b = [t_{s,b}, t_{s,b} + 1, \dots, t_s, \dots, t_{e,b}]$ around t_s randomly, where $t_{s,b} \in [1, t_S - 1]$ and $t_{e,b} \in [t_S, t_D - 1]$ for $b = 1, 2, \dots, B$. On each interval, we can get $\hat{\beta}_b$ by Step 3.1. Thereby, by Theorem 2, the bootstrap sequence $\tilde{B} = \{\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_B\}$ should be an i.i.d. stationary process with $E(\hat{\beta}_b) = \mu_\beta$. The parameter of interest is $\sigma_\beta^2 = \text{Var}(\hat{\beta}_b)$. By [65] and [66], the limit behavior of the partial sum $\mathbb{P}S_B = \sum_{b=1}^B (M_b - \sigma_\beta^2)$ depends on the Hermite rank of the function G , where $M_b = (\hat{\beta}_b - \mu_\beta)^2$ and $\hat{\beta}_b = G(y_{t \in I_b}, S_{t \in I_b}, D_{t \in I_b}, Z_{t \in I_b}, \psi_{t \in I_b})$. Let $H_k(\beta) = (-1)^k \exp(\frac{1}{2}\beta^2) d^k \exp(-\frac{1}{2}\beta^2)/d\beta^k$, $\beta \in \mathbb{R}$, $k \geq 0$ denote the k -th Hermite polynomial. The Hermite rank q of G is then defined as $q = \inf\{k \geq 1 : E(G(\mathcal{R}))H_k(\mathcal{R}) \neq 0\}$, $\mathcal{R} = (y, S, D, Z, \psi)$. Suppose that $\text{cov}(\hat{\beta}_b, \hat{\beta}_{b'}) = (b' - b)^{-H} \mathcal{L}(b' - b)$ for $0 < H < q^{-1}$ and $\mathcal{L} : (0, \infty) \rightarrow \mathbb{R}$ is a slowly varying function at infinity [67], i.e., $\lim_{t \rightarrow \infty} \mathcal{L}(\tau k)/\mathcal{L}(k) = 1$ for all $\tau > 0$. Set $d_B = B^{2-qH} \mathcal{L}^q(B)^{1/2}$, as $B \rightarrow \infty$ we have

$$\begin{aligned}d_B^{-1} \mathbb{P}S_B &\rightsquigarrow_d \Theta^{-q/2} c_q \int_{\mathbb{R}^q} \frac{\exp(i(B_1 + \dots + B_q)) - 1}{i(B_1 + \dots + B_q)} \\ &\quad \cdot \prod_{k=1}^q |B_k|^{(H-1)/2} dW(B_1) \dots dW(B_q)\end{aligned}\tag{18}$$

where $\Theta = 2\Gamma(H) \cos(\frac{1}{2}H\pi)$, $c_q = E(G(\mathcal{R})H_q(\mathcal{R}))/q!$. $d_B^{-1} \mathbb{P}S_B$ converges in distribution to a multiple Wiener-Itô integral with respect to the random spectral measure W of Gaussian white-noise processes [46,66,68].

For $q=1$, $d_B^{-1} \mathbb{P}S_B$ has a normal limit distribution with mean zero and variance $2c_q/(1-H)(2-H)$; and for $q \geq 2$, the limit distribution is non-normal. Herein, in this paper we chose to reconstruct the normalizing factor d_B so as to get a more straightforward limit distribution. As noted by [68,69], the wrong choice of d_B will make the distribution of the statistics $d_B^{-1} \mathbb{P}S_B$ converge to a degenerate limit because $d_B^{-1} \mathbb{P}S_B$ tends to zero so fast. An appropriate construction could be $\mathbb{H}_B = (Bd_B^2)^{-1/2}(\mathbb{P}S_B - B \cdot \sigma_{\beta,\eta}^2)$, then we will get

THEOREM 4. (Asymptotic of the bootstrap estimator) if $q \geq 1$ and $B^{-1} = o(1)$, then

$$\sup_{x \in \mathbb{R}} |P_B(\mathbb{H}_B \leq x) - \Phi(x/\sigma_q)| = o(1),$$

where $\sigma_q^2 = 2(2-qH)^{-1}(1-qH)^{-1}c_q^2q!$, Φ is the cumulative distribution function of the normal distribution with mean zero and variance 1, P_B is the probability measure defined on the bootstrap sequence. Furthermore,

$$\sup_{x \in \mathbb{R}} \left| P_B(\mathbb{H}_B \leq x) - \Phi(d_B^{-1} \mathbb{P} S_B \leq x) \right| = o(1)$$

if and only if $q = 1$.

Usually $E(\beta^2) \neq 0$ for model (6), hence $E(G(\mathcal{R}))H_{k=1}(\mathcal{R}) \neq 0$ for most of the DGPs, thereby the Hermite rank can be taken as $\hat{q} = 1$. On this basis, we propose to estimate the Hurst parameter H by using the change-of-frequency (COF) estimator based on the second-order difference of $\tilde{B}$ ([70,71]):

$$\hat{H} = \frac{1}{2} \log_2 \left(\frac{\sum_{i=1}^{B-4} (\tilde{B}_{i+4} - 2\tilde{B}_{i+2} + \tilde{B}_i)^2}{\sum_{i=1}^{B-2} (\tilde{B}_{i+2} - 2\tilde{B}_{i+1} + \tilde{B}_i)^2} \right),$$

where $\log_2(\cdot)$ is the base-2 logarithm. The critical values and confidence interval can be then constructed correspondingly by [72]. The proposed bootstrap procedure can be then applied similarly to the estimation Step 3.2, i.e., we divide the whole time interval $[1, 2, \dots, T]$ into sub-intervals around t_d randomly, then repeat Step 3.1 and Step 3.2 to get the bootstrap sequence $\tilde{B} = \{\hat{\xi}_1, \hat{\xi}_2, \dots, \hat{\xi}_B\}$, the inference framework thereby follows. Be cautious that, our bootstrap inference framework for $\hat{\xi}$ accounts for the estimation uncertainty brought by the first estimation Step 3.1, because $\hat{\xi}$ is a smooth function of $\hat{\beta}$ as shown in Step 3.2.

Heterogeneous treatments effects and panel settings with staggered treatments adoptions

A forward point-wise estimation approach of the heterogeneous treatments effects

Consider now the following one-way heterogeneous treatments effects model (the coefficients are time-varying):

$$y_t = W_t \eta_t + S_t \beta_t + D_t \xi_t + X_t \alpha_t + Z_t \gamma_t + \nu_t + \varepsilon_t, \quad t = 1, 2, \dots, t_S, \dots, t_D, \dots, T \quad (19)$$

where $\text{cov}(\lambda_t, \varepsilon_t) \neq 0$, $\lambda_t \in \{S_t, D_t\}$. All variables and parameters follow previous definitions. The parameters of interests are β_t for $t = t_S, \dots, t_D, \dots, T$ and ξ_t for $t = t_D, \dots, T$.

STEP 4.1. (Estimation of β_t for $t = t_S, \dots, t_D - 1$) We estimate the following model respectively for the forward point-wise step $h = 0, 1, 2, \dots, t_D - t_S - 1$:

$$y_t - \hat{\beta}_t^- \cdot I\{t \leq t_S\} - \hat{\beta}_t^+ \cdot I\{t > t_S + h\} = W_t \eta_t + S_t \beta_t + X_t \alpha_t + Z_t \gamma_t + \nu_t + \varepsilon_t, \quad (20)$$

where $t = 1, 2, \dots, t_S - 1 + h$, $\hat{\beta}_t^- = 0$ and $\hat{\beta}_t^+ = \hat{\beta}_{t-h}$. Estimate model (20) by Step 3.1 for $t_D - t_S - 1$ times, we will get point-wise estimators $\hat{\beta}_{t_S}, \hat{\beta}_{t_S+1}, \hat{\beta}_{t_S+2}, \dots, \hat{\beta}_{t_D-1}$ respectively for the point-wise step h .

STEP 4.2. (Estimation of $\beta_t + \xi_t$ for $t = t_D, \dots, T$) Further consider the following model respectively for the forward point-wise step $h = 0, 1, 2, \dots, T - t_D$:

$$\tilde{y}_t = W_t \eta_t + SD_t(\beta + \xi)_t + X_t \alpha_t + Z_t \gamma_t + \nu_t + \varepsilon_t, \quad t = 1, 2, \dots, t_S, \dots, t_D + h \quad (21)$$

where

$$\tilde{y}_t = y_t \cdot I\{1 \leq t \leq t_S - 1\} + (y_t - \hat{\beta}_t \cdot I\{t_S \leq t \leq t_D - 1\}) + y_t \cdot I\{t = t_D\} + (y_t - (\widehat{\beta + \xi})_t^+) \cdot I\{t > t_D + h\},$$

$SD_t \equiv S_t D_t$ is the total treatment indicator, $(\beta + \xi)_t \equiv \beta_t + \xi_t$ is the total treatment effect, $(\widehat{\beta + \xi})_t^+ = (\widehat{\beta + \xi})_{t-h}$. In accordance with Step 4.1, estimate model (21) by Step 3.1 for $T - t_D$ times, we will get point-wise estimators $(\widehat{\beta + \xi})_{t_D}, (\widehat{\beta + \xi})_{t_D+1}, \dots, (\widehat{\beta + \xi})_T$ for the point-wise step h respectively.

STEP 4.3. (Estimation of β_t and ξ_t for $t = t_D, \dots, T$) Combining Step 4.1 and 4.2 to get $\hat{\phi} \equiv (\hat{\beta}_{t_S}, \hat{\beta}_{t_S+1}, \dots, \hat{\beta}_{t_D-1}, (\widehat{\beta + \xi})_{t_D}, (\widehat{\beta + \xi})_{t_D+1}, \dots, (\widehat{\beta + \xi})_T)$, consider the following forward one step-wise regression model

$$\hat{\phi} - \hat{\theta}_t^- \cdot I\{t \leq t_D\} - \hat{\theta}_t^+ \cdot I\{t > t_D + h\} = D_t \theta_t + \sigma_t \quad \text{for } t = t_S, t_S + 1, \dots, t_D - 1 + h,$$

where $\hat{\theta}_t^- = 0$, $\hat{\theta}_t^+ = \hat{\theta}_{t-h}$, $h = 0, 1, 2, \dots, T - t_D$. Estimate this model by Step 3.1 for $T - t_D$ times, we will finally get the estimators

$$\hat{\xi}_{t_D} = \hat{\theta}_{t_D}, \quad \hat{\xi}_{t_D+1} = \hat{\theta}_{t_D+1}, \quad \hat{\xi}_{t_D+2} = \hat{\theta}_{t_D+2}, \quad \dots, \quad \hat{\xi}_T = \hat{\theta}_T$$

and

$$\hat{\beta}_{t_D} = (\widehat{\beta + \xi})_{t_D} - \hat{\xi}_{t_D}, \quad \hat{\beta}_{t_D+1} = (\widehat{\beta + \xi})_{t_D+1} - \hat{\xi}_{t_D+1}, \quad \dots, \quad \hat{\beta}_T = (\widehat{\beta + \xi})_T - \hat{\xi}_T$$

respectively for the point-wise step h .

Let ζ_t , $t \in \mathbb{Z}$ be *i.i.d.* random variables and $\mathcal{F}_t = (\dots, \zeta_{t-1}, \zeta_t)$ be the corresponding sigma-filtration. For model (19) with $\mathcal{X}_t = (g(\mathcal{Y}'(t)), S_t, Z_t)$, and π_t defined in Remark 4, we assume that $\mathcal{X}_t = \mathbb{G}(t/T, \mathcal{F}_t)$, $\pi_t = \mathbb{H}(t/T, \mathcal{F}_t)$, $\mathcal{X}_t \pi_t = \mathbb{GH}(t/T, \mathcal{F}_t)$ and $\mathcal{Y}'(t) \pi_t = \mathbb{JH}(t/T, \mathcal{F}_t)$, $\mathbb{G}$ and $\mathbb{H}$ are measurable functions such that $\mathbb{G}(t/T, \mathcal{F}_t)$ and $\mathbb{H}(t/T, \mathcal{F}_t)$ are well defined for each t and $E(\pi|y', Z, S) = 0$ (see [73–75] for further definitions). The variance-covariance matrix of these random processes can be denoted as:

$$\mathcal{M}_{\mathcal{X}}(t) = E(\mathbb{G}\mathbb{G}'), \quad \mathcal{M}_{\pi}(t) = E(\mathbb{H}\mathbb{H}'), \quad \mathcal{M}_{\mathcal{X}\pi}(t) = E(\mathbb{G}\mathbb{H}\mathbb{G}\mathbb{H}'), \quad \mathcal{M}_{\mathcal{Y}'\pi}(t) = E(\mathbb{J}\mathbb{H}\mathbb{J}\mathbb{H}').$$

We now construct the asymptotic behavior of the point-wise estimator $\hat{\beta}(\tau)$, the asymptotic behavior of $\hat{\xi}(\tau)$ can be similarly constructed, hence omitted here for brevity.

Theorem 5. (Asymptotics of the point-wise estimator) Assume that the functions $\mathbb{G}$, $\mathbb{H}$, $\mathbb{GH}$, $\mathbb{JH}$ are stochastically Lipschitz continuous processes, and the smallest eigenvalues of the matrices $\mathcal{M}_{\mathcal{X}}(t)$, $\mathcal{M}_{\pi}(t)$, $\mathcal{M}_{\mathcal{X}\pi}(t)$ and $\mathcal{M}_{\mathcal{Y}'\pi}(t)$ are bounded away from zero. Define $\hat{\beta}(\tau) = \hat{\beta}_{t_S-1+\tau}$ respectively for $\tau = 1, 2, \dots, T - t_S + 1$, then we have

$$\begin{aligned} (t_S + \tau - 1)^H (\hat{\beta}(\tau) - \beta(\tau) - \text{Bias}(\tau)) &\rightsquigarrow_d \\ &\frac{1}{\bar{A}_1} \text{Co} \left(\int_0^{r_\tau T} \mathcal{M}_{\mathcal{X}}(s) ds \right)^{-1} (\mathbb{Q}(r_\tau T) - \mathbb{Q}(0)) \\ &- \frac{1}{\bar{A}_2} (\tilde{\mathbb{Q}}(r_\tau T) - \tilde{\mathbb{Q}}(0)) - \frac{1}{\bar{A}_3} (\overline{\mathbb{Q}}(r_\tau T) - \overline{\mathbb{Q}}(0)), \end{aligned}$$

and

$$\text{Bias}(\tau) = \frac{1}{\bar{A}_1} \text{Co} \left(\frac{1}{r_\tau T} \sum_{s=1}^{r_\tau T} X'_s X_s \right)^{-1} \left(\frac{1}{r_\tau T} \sum_{s=1}^{r_\tau T} X'_s (\hat{B}(\tau = s) - B(\tau = s)) \right) = o_p(c),$$

where

$$Co = (U^{T_2}E(\hat{y}) - U^{T_1}E(\hat{y}), 0, \check{A}_4, 0), \quad c = \frac{(t_S + \tau - 1)^H}{(t_S + \tau)}, \quad r_\tau = \frac{t_S + \tau}{T},$$

$H \in (0, 1)$ is the Hurst parameter,

$$\begin{aligned} \hat{B}(\tau = 1) &= (\underbrace{0, 0, \dots, 0}_{t_S}, 0)', \\ \hat{B}(\tau = 2) &= (\underbrace{0, 0, \dots, 0}_{t_S}, \hat{\beta}(\tau = 1), 0)', \\ \hat{B}(\tau = 3) &= (\underbrace{0, 0, \dots, 0}_{t_S}, \hat{\beta}(\tau = 1), \hat{\beta}(\tau = 2), 0)', \\ &\vdots \\ \hat{B}(\tau = T - t_S + 1) &= (\underbrace{0, 0, \dots, 0}_{t_S}, \hat{\beta}(\tau = 1), \hat{\beta}(\tau = 2), \dots, \hat{\beta}(\tau = T - t_S))', \end{aligned}$$

and

$$\begin{aligned} \check{A}_1 &= (U^{T_2}E(\hat{y}) - U^{T_1}E(\hat{y})) \frac{U^T E(S\hat{y}) - U^T E(S)U^T E(\hat{y})}{U^T E(\hat{y}\hat{y}) - U^T E(\hat{y})U^T E(\hat{y})}, \\ \check{A}_2 &= \frac{1}{U^T E(S\hat{y}) - U^T E(S)U^T E(\hat{y})}, \\ \check{A}_4 &= (U^{T_2}E(\hat{y}) - U^{T_1}E(\hat{y})) \frac{U^T E(Z\hat{y}) - U^T E(Z)U^T E(\hat{y})}{U^T E(\hat{y}\hat{y}) - U^T E(\hat{y})U^T E(\hat{y})}, \\ \check{A}_3 &= \frac{U^T E(\hat{y})}{U^T E(S\hat{y}) - U^T E(S)U^T E(\hat{y})}. \end{aligned}$$

for $\mathcal{T}_1 = \{1, 2, \dots, t_S - 1\}$, $\mathcal{T}_2 = \{t_S, t_S + 1, \dots, t_S + \tau\}$ and $\mathcal{T} = \{1, 2, \dots, t_S + \tau\}$. $\mathbb{Q}(\tau)$, $\tilde{\mathbb{Q}}(\tau)$, $\overline{\mathbb{Q}}(\tau)$ are Gaussian processes with covariance

$$E(\mathbb{Q}(\tau_1)\mathbb{Q}(\tau_2)') = \int_0^{\min(\tau_1, \tau_2)} \mathcal{M}_{\mathcal{X}\pi}(s) ds, \quad E(\tilde{\mathbb{Q}}(\tau_1)\tilde{\mathbb{Q}}(\tau_2)') = \int_0^{\min(\tau_1, \tau_2)} \mathcal{M}_{\mathcal{Y}'\pi}(s) ds,$$

and

$$E(\overline{\mathbb{Q}}(\tau_1)\overline{\mathbb{Q}}(\tau_2)') = \int_0^{\min(\tau_1, \tau_2)} \mathcal{M}_\pi(s) ds.$$

Theorem 5 states that the asymptotic distribution of the step-wise estimator is a fractional Brownian motion (fBm) process; when $H=0.5$ the asymptotic distribution turns out to be the usual Brownian motion. Inference for the step-wise estimator is as the same as the proposed bootstrap procedure.

Panel settings with staggered treatments adoptions and two-way heterogeneities

We now extend model (19) to allow for two-way heterogeneities. Consider the setting with N units and each unit i receives one (several) policy(s) or treatment(s), the start times and end times of the treatments adoptions may, but

need not, vary by unit. For the balanced staggered treatments adoptions, the whole time interval $[1, 2, \dots, T]$ can be divided into $\mathbb{P}$ sub-intervals: $[t_{s=1}, t_{s=1} + 1, \dots, t_{s=2} - 1]$, $[t_{s=2}, t_{s=2} + 1, \dots, t_{s=3} - 1]$, $[t_{s=3}, t_{s=3} + 1, \dots, t_{s=4} - 1]$, $\dots$, $[t_{s=\mathbb{P}-1}, t_{s=\mathbb{P}-1} + 1, \dots, t_{s=\mathbb{P}} - 1]$, $[t_{s=\mathbb{P}}, t_{s=\mathbb{P}} + 1, \dots, T]$ where $s = 1, 2, \dots, \mathbb{P}$ denotes the s -th sub-interval, t_s is the initial time of the s -th sub-interval; and $t_{s+1} - t_s = \ell_s$, ℓ_s is the length of the s -th sub-interval.

We assume that, on the s -th sub-interval, each unit i receives a treatment $d_{i,s} = (d_{i,t=t_s}, d_{i,t=t_{s+1}}, \dots, d_{i,t=t_{s+1}-1})$, where $d_{i,t=t_s} = \dots = d_{i,t=t_{s+1}-1} \in \{0, 1\}$. Then $d_{*,s} = (d_{i=1,s}, d_{i=2,s}, \dots, d_{i=N,s})'_{N \times \ell_s}$ denotes the treatments indicator for all units on the s -th sub-interval. Correspondingly, $\zeta_{*,s} = (\zeta_{i=1,s}, \zeta_{i=2,s}, \dots, \zeta_{i=N,s})'_{N \times \ell_s}$ denotes the two-way heterogeneous treatments effects for all units on the s -th sub-interval with $\zeta_{i,s} = (\zeta_{i,t=t_s}, \zeta_{i,t=t_{s+1}}, \dots, \zeta_{i,t=t_{s+1}-1})$, $\zeta_{i,t}$ represents the treatment effect for unit i at time t . A graphical and intuitive illustration of these definitions is given in the Online Supplementary Material [S3 File](#).

Definition 6 (Phase states and phase transitions). The sub-intervals $s = 1, 2, \dots, \mathbb{P}$ are called phase states, the switch from phase state s to $s+1$ is called a phase transition. A phase state corresponds to a treatment state, the switch from one treatment state $d_{i,s}$ to another $d_{i,s+1}$ for the same unit i is called a heterogeneous phase transition if and only if $d_{i,s} \neq d_{i,s+1}$, otherwise it is a homogeneous phase transition.

The DGP of panel settings with staggered treatments adoptions we consider now is

$$Y_{*,*} \equiv \bigoplus_{s=1}^{\mathbb{P}} \left\{ \mathbb{D}_{*,s} \circ \zeta_{*,s} + X_{*,s} \circ \alpha_{*,s} + Z_{*,s} \circ \nu_{*,s} \right\} + \mu_{*,*} + \nu_{*,*} + \epsilon_{*,*}, \quad (22)$$

where $\bigoplus$ denotes the horizontal stacking of matrices, i.e., $C_{N,(T_1+T_2)} = A_{N,T_1} \bigoplus B_{N,T_2}$ representing the stacking of matrices A and B , $\circ$ denotes the Hadamard product, $Y_{*,*} \equiv (Y_{*,s=1}, Y_{*,s=2}, \dots, Y_{*,s=\mathbb{P}})'_{N \times T}$ is the outcomes of all units on the whole time interval, $Y_{*,s} = (Y_{i=1,s}, Y_{i=2,s}, \dots, Y_{i=N,s})'_{N \times \ell_s}$ is the outcomes of all units on the s -th sub-interval, and $Y_{i,s} = (Y_{i,t=t_s}, Y_{i,t=t_{s+1}}, \dots, Y_{i,t=t_{s+1}-1})'_{1 \times \ell_s}$ is the outcome of unit i on the s -th sub-interval, $T = \sum_s \ell_s$ is the total length of the sub-intervals; $X_{*,s} = (X_{i=1,s}, X_{i=2,s}, \dots, X_{i=N,s})'_{N \times \ell_s}$ are unobservable (or omitted) confounders driving $\mathbb{D}_{*,s}$ and $Y_{*,s}$ for $s = 1, 2, \dots, \mathbb{P}$ with $X_{i,s} = (X_{i,t=t_s}, X_{i,t=t_{s+1}}, \dots, X_{i,t=t_{s+1}-1})'_{1 \times \ell_s}$ representing the confounder for unit i on the s -th sub-interval, the coefficients are $\alpha_{*,s} = (\alpha_{i=1,s}, \alpha_{i=2,s}, \dots, \alpha_{i=N,s})'_{N \times \ell_s}$ with $\alpha_{i,s} = (\alpha_{i,t=t_s}, \alpha_{i,t=t_{s+1}}, \dots, \alpha_{i,t=t_{s+1}-1})'_{1 \times \ell_s}$; $Z_{*,s}$ are observed exogenous control variables, which is similarly structured as $X_{*,s}$, the structure of $\nu_{*,s}$ is similar to $\alpha_{*,s}$; $\mu_{*,*}$ is unit fixed effect and $\nu_{*,*}$ is time fixed effect. We assume that: $\text{cov}(\mathbb{D}_{*,s=j}, \epsilon_{i,t}) \neq 0$, and $\text{cov}(\epsilon_{i,t+n}, \epsilon_{i,t}) \sim H(2H-1)n^{-2(1-H)}$ as $n \rightarrow \infty$, $H \in (0, 1)$ is the Hurst parameter, $i' \neq i \in \{1, 2, \dots, N\}$, $t = 1, 2, \dots, T$. Note that Model (1) is a special case of model (22) with $\mathbb{P} = 3$. The parameter of interest are the two-way heterogeneous treatments effects $\zeta_{*,s}$ on every phase state $s = 1, 2, \dots, \mathbb{P}$.

The non-matrix form of (22) with forward point-wise step h is:

$$Y_{i,t} - \tilde{\zeta}_{i,t}^- \cdot \mathbb{I}(t \leq t_{i,k}) - \tilde{\zeta}_{i,t}^+ \cdot \mathbb{I}(t \geq t_{i,k} + h) = d_{i,t} \zeta_{i,t} + X_{i,t} \alpha_{i,t} + Z_{i,t} \gamma_{i,t} + \mu_i + \nu_t + \epsilon_{i,t}, \quad t = 1, 2, \dots, t_{i,k} - 1 + h \quad \text{for } h = 1, 2, \dots, T - t_{i,k}, \quad (23)$$

$t_{i,k}$ is the first time unit i receives treatment, $k \in \{2, 3, \dots, \mathbb{P}\}$; $\tilde{\zeta}_t^- = 0$ and $\tilde{\zeta}_t^+ \equiv \hat{\zeta}_{t-h}$. For a given h , consider the step-wise first-order difference of model (23):

$$\Delta Y_{i,t} - \tilde{\zeta}_{i,t}^- \cdot \mathbb{I}(t \leq t_{i,k}) - \tilde{\zeta}_{i,t}^+ \cdot \mathbb{I}(t \geq t_{i,k} + h) = \Delta d_{i,t} \zeta_{i,t} + \Delta X_{i,t} \alpha_{i,t} + \Delta Z_{i,t} \gamma_{i,t} + \Delta \nu_t + \Delta \epsilon_{i,t},$$

where $\Delta \sigma_{i,t} = \sigma_{i,t} - \frac{1}{t_{i,k}+h} \sum_{t=1}^{t_{i,k}+h} \sigma_{i,t}$, $\sigma_{i,t} \in \{y_{i,t}, d_{i,t}, X_{i,t}, \nu_t, \epsilon_{i,t}\}$. Suppose that $\sum_{t=1}^{t_{i,k}+h} d_{i,t} < \infty$ for all i , as $t_{i,k} + h \rightarrow \infty$ we will get

$$\widetilde{\Delta Y}_{i,t} = d_{i,t} \zeta_{i,t} + \Delta X_{i,t} \alpha_{i,t} + \Delta Z_{i,t} \gamma_{i,t} + \Delta \nu_t + \Delta \epsilon_{i,t} + o_p(1) \quad (24)$$

with $\widetilde{\Delta Y}_{i,t} = \Delta Y_{i,t} - \tilde{\zeta}_{i,t}^- \cdot \mathbb{I}(t \leq t_{i,k}) - \tilde{\zeta}_{i,t}^+ \cdot \mathbb{I}(t \geq t_{i,k} + h)$.

Assumption 7 For the Proximal Variables $\psi_{i,t}$: (i) the random variables $\Delta X_{i,t}$ and $\Delta \psi_{i,t}$ are of high correlations, $\Delta \nu_t$ and $\Delta \psi_{i,t}$ are also of high correlation; (ii) $\Delta \psi_{i,t}$ is independent of $\Delta \epsilon_{i,t}$; (iii) there exists a continuous function $g \in C^p$, $p \geq 1$, such that:

$$\Delta X_{i,t} = \tilde{a} g(\tilde{\Delta} y'(i, t)) + \tilde{b} + \tilde{u}_{i,t}, \quad \Delta \nu_t = \tilde{c} g(\tilde{\Delta} y'(i, t)) + \tilde{d} + \tilde{w}_{i,t},$$

with $\tilde{\Delta} y'(i, t) = y'(i, t) - \frac{1}{t_{i,k}+h} \sum_{t=1}^{t_{i,k}+h} y'(i, t)$, $y'(i, t)$ is defined similarly to model (7), $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d} \in \mathbb{R} \setminus \{0\}$.

In line with models (7), (8), (11) and (17), model (24) are then turned into varying coefficients panel forms. Estimate this model by varying coefficient fixed effects methods (see, e.g., [76–78]), and by the estimation Steps 3.1–3.2, Steps 4.1–4.3 and Assumption 7, we will get consistent estimators $\hat{\zeta}_{i,t}$ respectively for $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, T$ by applying Theorem 2 and Theorem 5. The bootstrap inference method proposed can be herein applied. See the Online Supplementary Material [S3 File](#) for details.

Numerical examples

Monte Carlo simulations

Data Generating Process. To evaluate the performances of our method, we generate potential outcomes from a varying coefficients linear model with staggered treatments:

$$y_{it} = W_{it}\eta_{it} + D_{it}\xi_{it} + Z_{it}\gamma_{it} + U_{it}, \quad t = 1, 2, \dots, T, \quad i = 1, 2, \dots, N, \quad (25)$$

where $D_{it} = [P_i]$, $X_{it} = [P_{i,1}]$, $W_{it} = [P_{i,2}] + W_{t\Delta}^H$, $Z_{it} = [P_{i,3}] + W_{t\Delta}^H$, $U_{it} = X_{it}\alpha_{it} + e_{it} + \mu_i + \nu_t$ and $P_{i,1} \sim N(0, \sigma_{P,i}^2)$, $P_i = I\{P_{i,1} \geq q_\tau(P_{i,1})\}$, $P_{i,2} \sim U(a_i, b_i)$, $P_{i,3} \sim U(c_i, d_i)$, $\alpha_{it} \sim U(e_i, f_i)$, $e_{it} \sim N(0, \sigma_{e,i}^2)$ where U denotes the uniform distribution, $q_\tau(\cdot)$ denotes the τ -th quantile; $\eta_{it} = 10.25 + i$, $\gamma_{it} = 0.1 \cdot i$, $\xi_{it} \sim U(g_i, h_i)$. $[A]$ denotes the order operator of the vector A , i.e., $[A] = (A_{[1]}, A_{[2]}, \dots, A_{[n]})$ where $A_{[1]} \leq A_{[2]} \leq \dots \leq A_{[n]}$ for $A \in \{P_i, P_{i,1}, P_{i,2}, P_{i,3}\}$; $W_{t\Delta}^H$ is a discrete-time fractional Brownian motion (fBm) with the Hurst parameter H :

$$W_{t\Delta}^H = \sigma^{-\kappa\Delta} W_{(t-1)\Delta}^H + (1 - \sigma^{-\kappa\Delta})\mu + \epsilon_t, \quad \epsilon_t = \rho \int_{(t-1)\Delta}^{t\Delta} \sigma^{-\kappa(i\Delta-s)} d\bar{B}_s^H,$$

$\Delta = L/T$ denotes the sampling interval, $\kappa \in \mathbb{R}^+$, $\rho \in \mathbb{R}^+$ and $\mu \in \mathbb{R}$ are given parameters, $\bar{B}_t^H$ denotes a zero-mean Gaussian process with the covariance function $\text{cov}(\bar{B}_t^H, \bar{B}_s^H) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H})$, $t, s \geq 0$ [79]. $W_{t\Delta}^H$ is the discrete-time representation of the continuous-time fractional Ornstein-Uhlenbeck process $dW_t = \kappa(\mu - W_t)dt + \rho d\bar{B}_t^H$ [80].

Model (25) is endogeneous because X_{it} is correlated with D_{it} , and we assume that U_{it} is unobservable. We take $\psi_t = t$ by Remark 1. The method based on fast Fourier transformation (FFT) to generate $W_{t\Delta}^H$ is provided in [81] and [79]. In this paper we consider $H=0.700$, $\kappa = 0.237$, $\mu = 2.417$, $\rho = 0.700$, $T=500$, $N=5$ and $\Delta = 1/2^8$, hence $L = 1.953125$. We set $\sigma_{P,i}^2 = 0.02 + i \cdot 0.02$, $a_i=0$, $b_i=1$, $c_i = -0.01 \cdot i$, $d_i = 0.01 \cdot i$, $e_i=14$, $f_i=16$, $g_i=-i$, $h_i=i$, $\tau = 0.5$ and $\sigma_{e,i}^2 = 0.01$ for model (25). The estimands we are interested in are

$$\theta_{ATT,1} = \frac{1}{NT_D} \sum_{i=1}^N \sum_{t=t_{i,d}}^T \xi_{1,it}, \quad \theta_{ATT,2} = \frac{1}{N} \sum_{j=1}^N \left(\frac{1}{T-t_{i,d}} \sum_{t=t_{i,d}}^T \xi_{2,it} \right),$$

where $\xi_{1,it}$ is the panel model estimand; $\xi_{2,it}$ is the mean group estimand which averages coefficients obtained from separate time-series regressions for each individual; $t_{i,d}$ is the time unit i receives treatment, $T_D = \sum_{i=1}^N (T - t_{i,d})$.

We also compare our method with the literature when untreated units are available. We generate 10 untreated units for each treated unit through model (25), i.e., the j -th untreated outcome for the treated unit i is generated as follows:

$$Y_{co,it,j} = W_{it}\eta_{it} + Z_{it}\gamma_{it} + U_{it} + U(-0.05j, 0.05j), \quad t = 1, 2, \dots, T, \quad j = 1, 2, \dots, 10,$$

which guarantees that $Y_{it} - Y_{co,it,j}$ is a martingale process for every i and j over the time interval $\{t = 1, \dots, T\}$.

Estimation Performances. The simulation results in terms of bias and mean squared error (MSE) are shown in Table 2. It would be interesting to see that the treatment effects could be consistently estimated by our methods. The biases of the proposed method ($\hat{\theta}_{ATT,1}$, $\hat{\theta}_{ATT,2}$) are generally the smallest among all competitors and the biases are decreasing with sample size. The other estimators do not perform well due to the heterogeneities or endogeneities.

We now examine the performances of the confidence interval (CI) for the proposed methods. It can be seen from Table 3 that the coverage probabilities (CP) are increasing with sample size, and generally perform good.

Weak and Invalid CP. To illustrate the consequences if the chosen proximal variable is weak or invalid, we carried out additional Monte Carlo simulations studies. (i) Weak CP is generated by $\psi_{t,weak} = \psi_t + N(0, \sigma)$, i.e., we introduce different levels of noise σ into the CP variable and examine the resulting effects on the estimation outcomes. (ii) Invalid CP is generated by $\psi_{t,invalid} = \psi_t + \rho e_{1t} + \rho e_{2t} + \dots + \rho e_{Nt}$, i.e., we introduce different levels of invalidity ρ into the CP variable and examine the resulting effects on the estimation outcomes. The results are shown in the following Tables 4–8.

Interestingly, we find that the bias in estimating the treatment effect is increasing in the invalidity of the CP variable, but not in its weakness. That is, our proposed methods remains largely unbiased even when the CP variable is weak (as shown in Table 4). To understand this phenomenon, a theoretical analysis proving that our estimators are robust to weak CP variable is provided in the Online Supplementary Material S5 File.

Table 2. Simulation results with long range dependency ($H=0.70$).

Estimators	T=100,N=5		T=300,N=5		T=500,N=5	
	Av. Bias	MSE	Av. Bias	MSE	Av. Bias	MSE
without untreated units						
$\hat{\theta}_{ATT,1}$	0.428	0.211	0.291	0.094	0.142	0.021
$\hat{\theta}_{ATT,2}$	0.585	0.429	0.361	0.165	0.140	0.018
TWFE	1.874	3.838	0.914	1.306	0.843	0.710
did_staggered (or)	0.552	0.341	0.494	0.260	0.598	0.371
did_staggered (dr)	0.552	0.341	0.494	0.260	0.598	0.371
did_staggered (ipw)	0.552	0.341	0.494	0.260	0.598	0.371
before_after	15.020	22.632	13.020	16.423	5.506	30.35
with untreated units						
synthdid	4.993	25.197	4.326	18.991	0.458	0.210
gsynth	4.787	22.929	4.131	17.073	0.458	0.210
did_staggered (or)	0.952	0.955	0.194	0.042	0.225	0.054
did_staggered (dr)	0.952	0.955	0.194	0.042	0.225	0.054
did_staggered (ipw)	0.952	0.955	0.194	0.042	0.225	0.054
did_staggered (sunab)	1.738	3.089	0.977	0.959	0.458	0.210

Notes: based on 1000 simulations; “did_staggered (or)”, “did_staggered (ipw)” and “did_staggered (dr)” represents for the outcome regression estimator, inverse probability weighting estimator and double robust estimator proposed in [18,82]; “synthdid” is the synthetic DiD estimator proposed in [83]; “gsynth” is the generalized synthetic control method proposed by [84], and “sunab” is the estimator proposed by. Before_after is the estimator simply compares the means of the treated outcomes before and after the treatments.

<https://doi.org/10.1371/journal.pone.0347847.t002>

Table 3. Empirical coverage probability of the 95% CI for the treatment effect.

Estimator	T = 100	T = 300	T = 500
N = 1			
$\theta_{ATT,1}$ (CP)	0.912	0.931	0.950
N = 5			
$\theta_{ATT,1}$ (CP)	0.848	0.951	0.856
$\theta_{ATT,2}$ (CP)	0.842	0.962	0.957

Notes: based on 1000 simulations. The coverage probability is estimated by $CP_{CI} = \frac{1}{NTD} \sum_{i=1}^N \sum_{t=t_{i,d}}^T \hat{\sigma}_{it}$, where $\hat{\sigma}_{it} = \mathbb{I}\{\xi_{it} \leq \xi_{it}^* \leq \xi_{it}^*\}$, $[\xi_{it}^*, \xi_{it}^*]$ is the estimated 95% confidence interval for the treatment effect ξ_{it} .

<https://doi.org/10.1371/journal.pone.0347847.t003>

Table 4. Simulation studies when the CP is weak or invalid (T=30, N=5; CP=t).

Weak CP	$\sigma = 15$	$\sigma = 45$	$\sigma = 150$	$\sigma = 550$
$\theta_{ATT,1}$	0.436	0.451	0.461	0.467
$\theta_{ATT,2}$	0.867	2.380	1.509	1.532

<https://doi.org/10.1371/journal.pone.0347847.t004>

Table 5. Simulation studies when the CP is weak or invalid (T=50, N=5; CP=CF treated).

Weak CP	$\sigma = 15$	$\sigma = 45$	$\sigma = 150$	$\sigma = 550$
$\theta_{ATT,1}$	1.707	1.837	1.787	1.824
$\theta_{ATT,2}$	1.658	2.042	1.690	1.825

<https://doi.org/10.1371/journal.pone.0347847.t005>

Table 6. Simulation studies when the CP is weak or invalid (T=50, N=5; CP=CF untreated).

Weak CP	$\sigma = 15$	$\sigma = 45$	$\sigma = 150$	$\sigma = 550$
$\theta_{ATT,1}$	1.782	1.764	1.712	1.782
$\theta_{ATT,2}$	2.242	1.807	1.681	1.810

<https://doi.org/10.1371/journal.pone.0347847.t006>

Table 7. Simulation studies when the CP is weak or invalid (T=50, N=5; CP=CF treated).

Invalid CP	$\rho = 1$	$\rho = 50$	$\rho = 100$	$\rho = 500$
$\theta_{ATT,1}$	0.375	0.448	0.454	0.456
$\theta_{ATT,2}$	0.645	0.576	0.564	0.547

<https://doi.org/10.1371/journal.pone.0347847.t007>

Table 8. Simulation studies when the CP is weak or invalid (T=50, N=5; CP=CF untreated).

Invalid CP	$\rho = 1$	$\rho = 50$	$\rho = 100$	$\rho = 500$
$\theta_{ATT,1}$	1.690	0.816	0.838	0.902
$\theta_{ATT,2}$	1.639	0.798	0.824	0.901

<https://doi.org/10.1371/journal.pone.0347847.t008>

Short panels, and simple serial correlation. In the paper, we employ the Change-of-Frequency (COF) estimator to estimate the Hurst parameter H . This estimator is based on the second differences of the bootstrap sequence, and the consistent convergence of the Hurst parameter does not depend on T , but rather on B . In practice, the Hurst parameter is only used to construct the normalization factor $\hat{d}_B$. As long as the bootstrap sample size B is sufficiently large, the

confidence intervals computed from short time sequences can still maintain the nominal coverage level. To see this, we have conducted simulation studies specifically designed for short time sequences.

From Table 9, we find that even for short panel data, the treatment effect estimator remains consistent. At the same time, we observe from Fig 6 that the bias of the Hurst parameter estimator decreases as the number of bootstrap replications increases, indicating that the proposed COF estimator is feasible. Hence, the estimators perform well even if the time series is short.

Apart from these, we carried out additional simulation studies to illustrate our methods' performances with simple serial correlation, i.e., the model's error term ϵ_{it} is generated by an AR(1) model with auto-correlation coefficient ρ .

The robustness of the proposed bootstrap inference framework to both short and long memory shown in Table 10 stems from its sub-interval resampling design, which non-parametrically captures the underlying dependence structure without imposing parametric assumptions. For short memory (e.g., AR(1) with exponentially decaying auto-correlations), the randomly partitioned sub-intervals I_b are asymptotically independent as the sample size T grows. Consequently, the bootstrap sequence $\tilde{B} = (\tilde{\beta}_1, \dots, \tilde{\beta}_B)$ becomes approximately *i.i.d.*, rendering conventional inference valid. For long-range dependence (LRD) where auto-correlations decay hyperbolically, the bootstrap sequence itself retains the long memory property, with its covariance structure satisfying $\text{Cov}(\tilde{\beta}_b, \tilde{\beta}_{b'}) \sim (b' - b)^{-H} \mathcal{L}(b' - b)$. Rather than ignoring this dependence, the proposed method directly characterizes the limiting distribution of the partial sum $\mathbb{PS}_B$. As established in Theorem 4, after appropriate normalization, the statistic $\mathbb{H}_B$ converges in distribution to a standard normal with the Hermite rank $q = 1$. This unified framework automatically adapts to the dependence strength: under short memory, the Hurst parameter H plays no distorting role, and under long memory, it is explicitly incorporated into the normalization factor d_B , ensuring valid coverage regardless of the true memory process.

Unilateral divorce law reforms and divorce growth

We illustrate our proposed methods by analyzing the heterogeneous treatment effects of no-fault divorce law reforms on divorce growth in China. Unilateral (or no-fault) divorce reform carried out worldwide allows either spouse to end a marriage, redistributing property rights and bargaining power relative to fault-based divorce regimes [85,86][85]. The Marriage Law of the People's Republic of China (promulgated in 2003 and implemented in 2004) stipulates that the condition for divorce is emotional breakdown rather than one party's fault, so both spouses can file a divorce lawsuit even if both spouses are not at fault (Before *The Marriage Law of the People's Republic of China (2004)*, the condition for one spouse to file a divorce lawsuit in China is that the other party is at fault, i.e., bigamy, cohabitation with others, domestic violence, maltreat or abandon. Many divorce lawsuits cannot be supported by the court because one spouse cannot provide sufficient evidence to prove the other's fault. For more details on Chinese divorce system). What's special for China is that

Table 9. Simulation study for short panels.

T, N	Estimator	Avg Bias	MSE	CP (95%)
$T=10, N=5$	$\theta_{ATT,1}$	1.630	2.576	0.548
	$\theta_{ATT,2}$	0.864	1.098	0.693
$T=20, N=5$	$\theta_{ATT,1}$	0.928	1.315	0.643
	$\theta_{ATT,2}$	0.943	1.268	0.814
$T=30, N=5$	$\theta_{ATT,1}$	0.478	0.149	0.815
	$\theta_{ATT,2}$	0.515	0.220	0.753
$T=50, N=5$	$\theta_{ATT,1}$	0.436	0.178	0.920
	$\theta_{ATT,2}$	0.647	0.288	0.910

<https://doi.org/10.1371/journal.pone.0347847.t009>

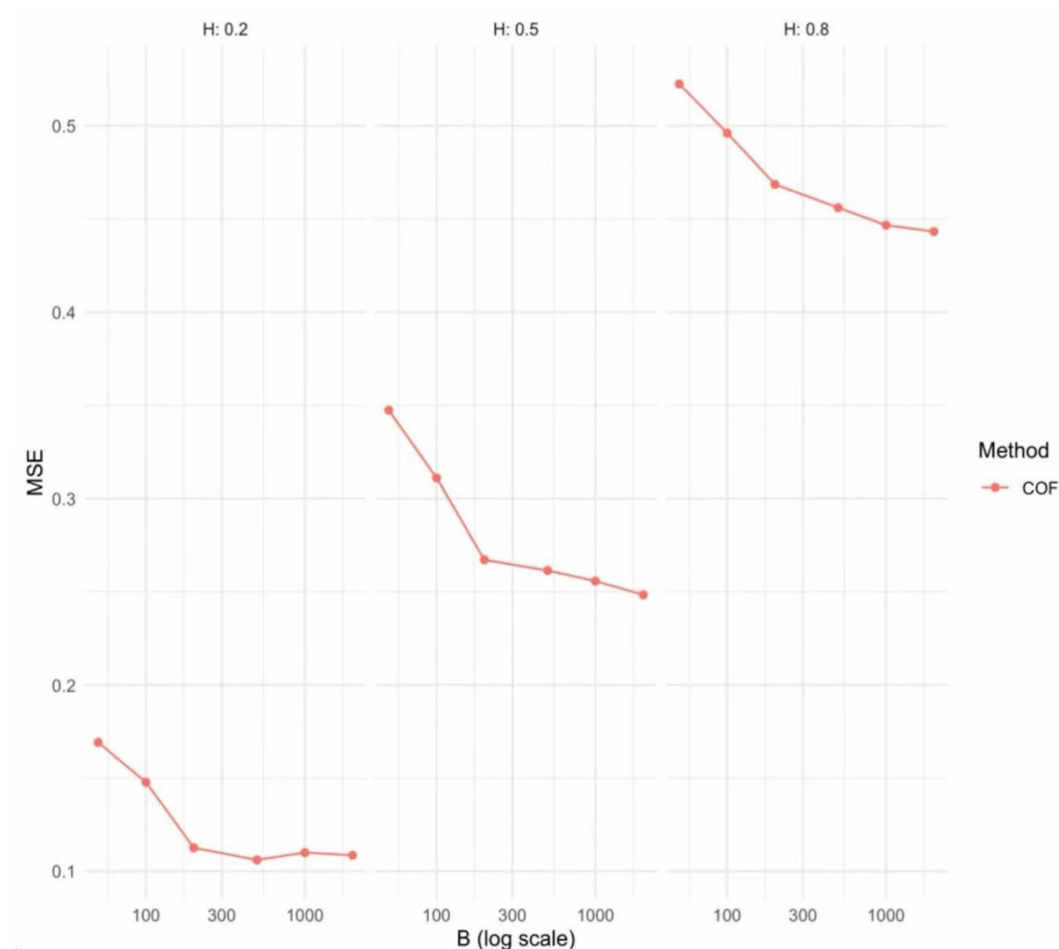


Fig 6. Simulation studies for the COF estimator. The figure shows bias of Hurst estimator vs. bootstrap replications with different level of long range dependency.

<https://doi.org/10.1371/journal.pone.0347847.g006>

Table 10. Simulation studies with simple serial correlation.

	Estimator	Avg Bias	MSE	CP (95%)
$\rho = 0.1$	$\theta_{ATT,1}$	0.947	1.304	0.782
	$\theta_{ATT,2}$	1.212	1.654	0.843
$\rho = 0.3$	$\theta_{ATT,1}$	0.875	1.302	0.814
	$\theta_{ATT,2}$	0.768	1.266	0.904
$\rho = 0.5$	$\theta_{ATT,1}$	0.938	1.276	0.788
	$\theta_{ATT,2}$	0.765	1.031	0.811

<https://doi.org/10.1371/journal.pone.0347847.t010>

the implement of the law is “one size fits all”, which means that all 34 provincial districts across the country are all treated units. Hence we are unable to collect untreated units for policy evaluation. The method proposed in this paper is therefore most suitable.

All 30 provinces' registered divorce data per year from 1990 to 2010 are collected from the National Bureau of Statistics (The data of Chongqing etc. are not provided by the official statistics. Please see the website of the National Bureau of Statistics for reasons: <https://data.stats.gov.cn/english/easyquery.htm?cn=E0103>). The divorces all show an upward growth trend, hence we take $\psi_{it} = t$ for all i ; we take Z_{it} to be the GDP per capita, because GDP per capita is found to be highly correlated with factors driving the growth of divorce [85]. D_{it} takes value 1 after the year 2003 and 0 otherwise.

As shown in Definition 5 and Theorem 1, if there exist structural change S_{it} in divorce before the implementation of the law D_{it} , treatment effect could be contaminated with structure change effect. Hence, structure change effect should be removed when evaluating the law enforcement effect. S_{it} is then detected by the Chow test through the R package “sctest” using data ranging from 1990 to 2002. The structure changes and treatments statuses are shown in Fig 7(a). Fig 7(b) further shows that staggering effect exist in structure changes [16].

The parameters we are interested in are the average treatment effect on the treated, the average structure change effect and the total average effect:

$$\theta_{ATT} = (1/|N_{tr}|T_D) \cdot \sum_{i \in N_{tr}} \sum_{t \in T_D} \xi_{it},$$

$$\theta_{ASC} = (1/|N_{tr}|T_S) \cdot \sum_{i \in N_{tr}} \sum_{t \in T_S} \beta_{it},$$

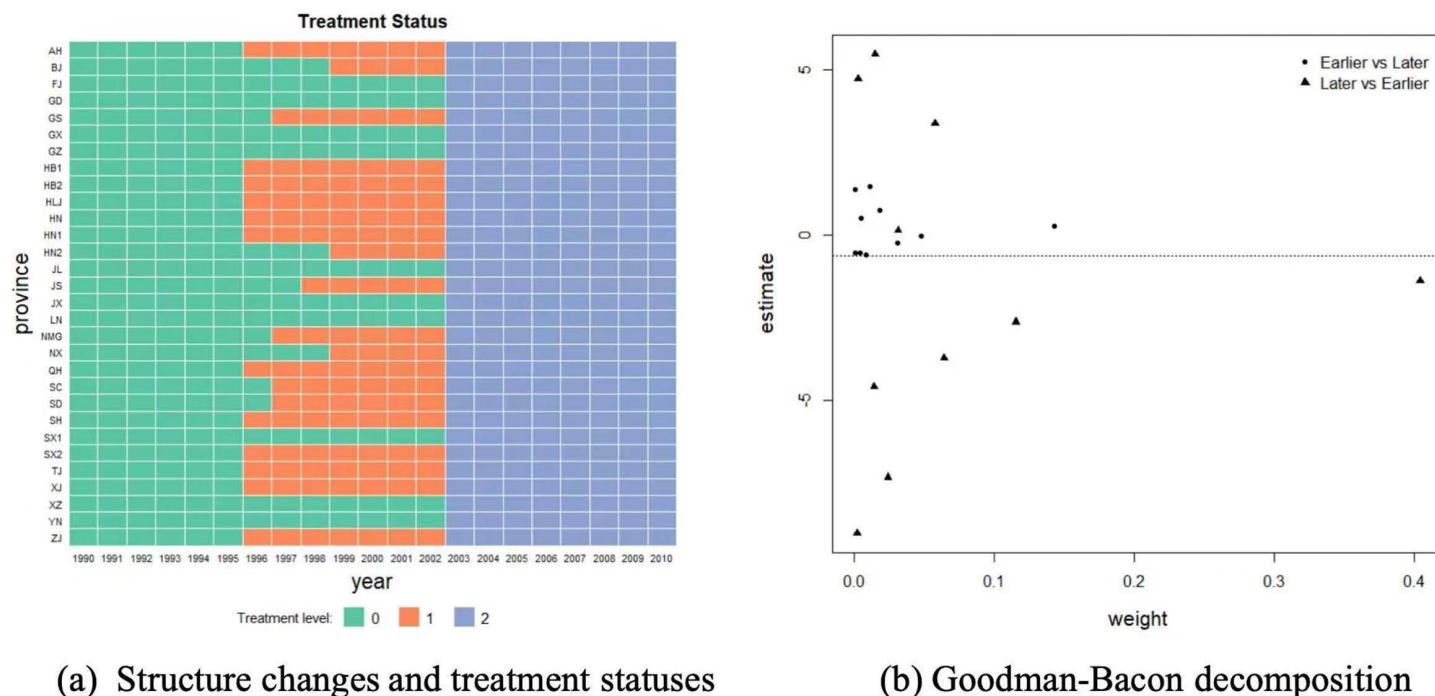


Fig 7. Structure changes and treatments statuses (left), and Goodman-Bacon decomposition (right). Treatment level 1 indicates the periods when structure changes commence and persist, level 2 indicates the periods when treatment starts and continues, as well as the periods when structure changes last after level 1.

<https://doi.org/10.1371/journal.pone.0347847.g007>

$$\theta_{TAE} = (1/|N_{tr}|T_{S0}) \cdot \sum_{i \in N_{tr}} \sum_{t \in T_S} (\beta_{it} + \xi_{it}),$$

in which $N_{tr} = \{1, 2, \dots, 30\}$, $T_D = \{2003, 2004, \dots, 2010\}$, $T_S = \bigcup_{i \in N_{tr}} \{T_{i,S}, T_{i,S} + 1, \dots, 2010\}$, and $T_{i,S} \in \{1996, 1997, 1998, 1999\}$ is the start time of the structure change for province i as shown in Fig 7(a). The estimation results are collected in Table 11.

It is shown that the structure changes around 1996–1999 lead to a decline of divorce (33240 couples on average), while the implements of unilateral divorce reform since 2003 lead to an increase of divorce (44780 couples on average). TWFE and OLS underestimates the treatment effects and structure change effects for neglecting the heterogeneities and endogeneities caused by two-way variability and potential confounders.

The 1997 Asian financial crisis exerted great negative economic impacts on China, many people lost their jobs or experienced a decrease in asset and income; meanwhile, a major flood occurred across China in 1998, resulting in exceeding 20 billion USD direct economic losses. Under these scenarios, couples will choose to overcome difficulties together, resulting in a decrease in divorce [85]. On the other hand, the unilateral divorce reform lowers the threshold and cost of divorce, making it easier to divorce, hence resulting in an increase of divorce. This finding is in accordance with the literature [86,87].

We also report the point-wise estimations of the ASCs and ATTs in Fig 8 (left: $ASC_t = (1/|N_{tr}|) \cdot \sum_{i \in N_{tr}} \beta_{it}$; right: $ATT_t = (1/|N_{tr}|) \cdot \sum_{i \in N_{tr}} \xi_{it}$). It can be seen that neglecting the treatment D_{it} (Fig 8(a)) or structure changes S_{it} (Fig 8(b)) will lead to contamination bias [20], and the bias will lead to wrong empirical conclusions. This warns us be aware of other potential distributional shifts if our focus is the effect of some specific social policy treatment.

Conclusion

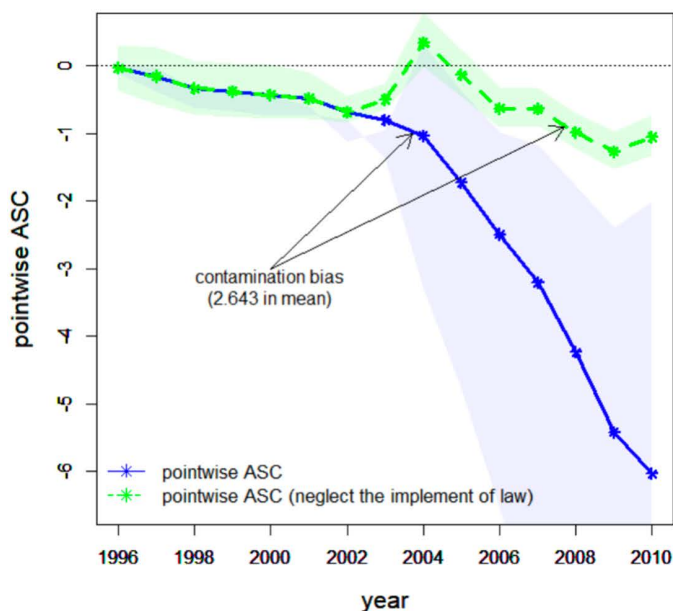
We propose a novel method to disentangle one treatment effect from another in time-series natural experiment settings with staggered and endogenous treatments adoptions, we allow the social-economic outcome to be long range dependent and no untreated units along with IVs are available. The Conditional Mean Symmetry condition constructed by the proposed Common Proximal Variable enables us to identify the treatments effects under the Neyman-Rubin counterfactual framework, while a two-step point-wise estimation approach allows us to consistently estimate the heterogeneous treatments effects. We show that the asymptotic distribution of the estimator is a fractional Brownian motion process

Table 11. Unilateral divorce law reform effect on divorce growth.

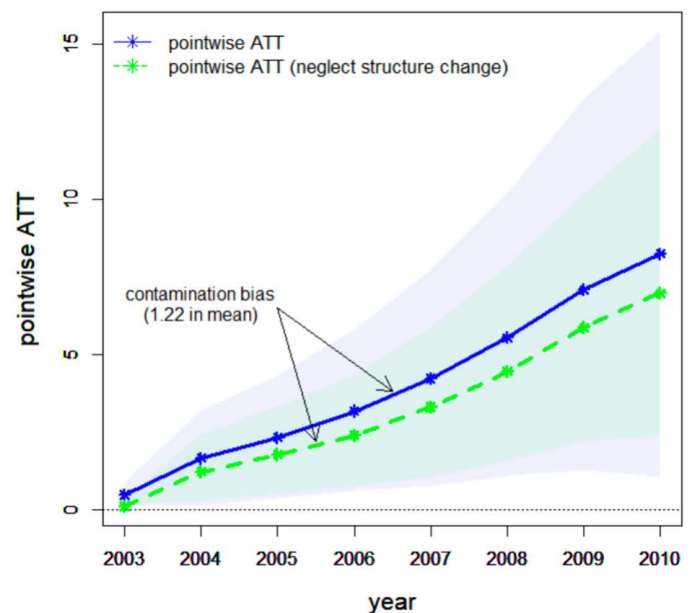
	new_1		new_2		TWFE		Pooling_ols	
θ_{ASC}	−3.324*** (0.008)	−1.131 (2.290)	−2.886*** (0.049)	−1.036*** (0.003)	−2.332*** (0.509)	−5.896*** (0.586)	0.706 (0.789)	−1.079 (0.782)
θ_{ATT}	4.478*** (0.010)	2.740*** (0.002)	4.679*** (0.087)	2.677*** (0.002)	2.163*** (0.366)	6.029*** (0.391)	2.788*** (0.609)	5.336*** (0.582)
θ_{TAE}	−0.681***	0.643	−0.134***	0.632***	−0.169	0.133	3.494***	4.257***
CP variable	✓	✓	✓	✓	×	×	×	×
Control variable	✓	×	✓	×	✓	×	✓	×
year fixed effects	✓	✓	✓	✓	✓	✓	×	×
province fixed effects	✓	✓	✓	✓	✓	✓	×	×

Notes: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. “new_1” and “new_2” refer to two different estimand θ_1 and θ_2 (see text).

<https://doi.org/10.1371/journal.pone.0347847.t011>



(a) Point-wise ASC



(b) point-wise ATT

Fig 8. Point-wise estimation of ASC and ATT.

<https://doi.org/10.1371/journal.pone.0347847.g008>

with long range dependency, thereby a bootstrap procedure is considered for inference. We provide a new identification and estimation framework for empirical researchers and decision-makers whose interest lies in evaluating social policies effects.

Supporting information

S1 File. Notation table.

(PDF)

S2 File. Main proofs of Proposition and Theorems.

(PDF)

S3 File. Detailed estimation procedure for panel model.

(PDF)

S4 File. Illustration of panel settings with staggered treatments adoptions.

(PDF)

S5 File. Impact of choosing a weak or invalid CP variable on estimation and inference.

(PDF)

Author contributions

Conceptualization: Shiming Hao.

Writing – original draft: Shiming Hao.

References

- Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet.* 2008;9(12):938–50. <https://doi.org/10.1038/nrg2482> PMID: [19015656](https://pubmed.ncbi.nlm.nih.gov/19015656/)
- Easterly W, Rebelo S. Fiscal policy and economic growth. *J Monet Econ.* 1993;32(3):417–58.
- Zagler M, Dürnecker G. Fiscal policy and economic growth. *J Econ Surv.* 2003;17(3):397–418.
- Ke X, Hsiao C. Economic impact of the most drastic lockdown during COVID-19 pandemic-The experience of Hubei, China. *J Appl Econ (Chichester Engl).* 2022;37(1):187–209. <https://doi.org/10.1002/jae.2871> PMID: [34518735](https://pubmed.ncbi.nlm.nih.gov/34518735/)
- Angrist J, Pischke J. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press. 2009.
- Bound J, Jaeger D, Baker R. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J Am Stat Assoc.* 1995;90(430):443–50.
- Young A. Consistency without inference: Instrumental variables in practical application. *Eur Econ Rev.* 2022;147:104–12.
- Cinelli C, Hazlett C. An omitted variable bias framework for sensitivity analysis of instrumental variables. *Biometrika.* 2025;112(2). <https://doi.org/10.1093/biomet/asaf004>
- Hahn J. On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects. *Econometrica.* 1998;66(2):315. <https://doi.org/10.2307/2998560>
- Chernozhukov V, Wuthrich K, Zhu Y. An exact and robust conformal inference method for counterfactual and synthetic controls. *J Am Stat Assoc.* 2021;116(536):1849–64.
- Roth J, Sant'Anna P, Bilinski A, Poe J. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *J Econom.* 2023;235(2):2218–44.
- Borusyak K, Jaravel X. Revisiting event study designs. In: 2017.
- Imai K, Kim I. When should we use unit fixed effects regression models for causal inference with longitudinal data?. *Am J Polit Sci.* 2019;63(2):467–90.
- de Chaisemartin C, D'Haultfoeulle X. Two-way fixed effects estimators with heterogeneous treatment effects. *Am Econ Rev.* 2020;110(9):2964–96.
- Sun L, Abraham S. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *J Econom.* 2021;225(2):175–99.
- Goodman-Bacon A. Difference-in-differences with variation in treatment timing. *J Econom.* 2021;225(2):254–77.
- Athey S, Imbens G. Design-based analysis in difference-in-differences settings with staggered adoption. *J Econom.* 2022;226(1):62–79.
- Callaway B, Sant'Anna P. Difference-in-differences with multiple time periods. *J Econom.* 2021;225(2):200–30.
- Roth J, Sant'Anna PHC. When Is Parallel Trends Sensitive to Functional Form?. *ECTA.* 2023;91(2):737–47. <https://doi.org/10.3982/ecta19402>
- de Chaisemartin C, D'Haultfoeulle X. Two-way fixed effects and differences-in-differences estimators with several treatments. *J Econom.* 2023;236(2):480–500.
- Qiu Y, Chen X, Shi W. Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. *J Popul Econ.* 2020;33(4):1127–72. <https://doi.org/10.1007/s00148-020-00778-2> PMID: [32395017](https://pubmed.ncbi.nlm.nih.gov/32395017/)
- Masini R, Medeiros MC. Counterfactual analysis and inference with nonstationary data. *J Bus Econ Stat.* 2022;40(1):227–39.
- Abadie A. Using synthetic controls: feasibility, data requirements, and methodological aspects. *J Econ Lit.* 2021;59(2):391–425.
- Shi Z, Huang J. Forward-selected panel data approach for program evaluation. *J Econom.* 2023;234:512–35.
- Arkhangelsky D, Imbens G. *Causal models for longitudinal and panel data.* CEMFI. 2023.
- Manski CF. Identification of binary response models. *J Am Stat Assoc.* 1988;83(403):729–38.
- Powell JL. Elsevier. 1994. p. 2443–521.
- Heckman J, Ichimura H, Todd P. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. *Rev Econ Stud.* 1997;64(4):605–54.
- Hirano K, Imbens GW. Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization. *Health Services & Outcomes Research Methodology.* 2001;2(3–4):259–78. <https://doi.org/10.1023/a:1020371312283>
- Hirano K, Imbens GW, Ridder G. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. *Econometrica.* 2003;71(4):1161–89. <https://doi.org/10.1111/1468-0262.00442>
- Masten M, Poirier A. Choosing exogeneity assumptions in potential outcome models. *Econom J.* 2023;26(3):327–49.
- White H. Time-series estimation of the effects of natural experiments. *J Econom.* 2006;135(1):527–66.
- White H, Lu X. Granger causality and dynamic structural systems. *J Financ Econom.* 2010;8(2):193–243.
- Lu X, Su L, White H. Granger Causality And Structural Causality In Cross-section And Panel Data. *Econom Theory.* 2016;33(2):263–91. <https://doi.org/10.1017/s0266466616000086>
- Lechner M. The relation of different concepts of causality used in time series and microeconometrics. *Econom Rev.* 2011;30(1):109–27.

36. Li K, Bell D. Estimation of average treatment effects with panel data: asymptotic theory and implementation. *J Econom*. 2017;197(1):65–75.
37. Abadie A, Cattaneo MD. Introduction to the special section on synthetic control methods. *J Am Stat Assoc*. 2021;116(536):1713–5.
38. Angrist JD, Kuersteiner G. Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score. *Rev Econ Stat*. 2011;93(3):725–47.
39. Angrist JD, Jordà O, Kuersteiner G. Semiparametric estimates of monetary policy effects: String theory revisited. *J Bus Econ Stat*. 2018;36(3):371–87.
40. Bojinov I, Shephard N. Time series experiments and causal estimands: Exact randomization tests and trading. *J Am Stat Assoc*. 2019;114(528):1665–82.
41. Koralov L, Sinai YG. *Theory of Probability and Random Processes*. 2 ed. Springer. 2007.
42. Hudgens MG, Halloran ME. Toward Causal Inference With Interference. *J Am Stat Assoc*. 2008;103(482):832–42. <https://doi.org/10.1198/016214508000000292> PMID: [19081744](https://pubmed.ncbi.nlm.nih.gov/19081744/)
43. Forastiere L, Airolidi EM, Mealli F. Identification and estimation of treatment and interference effects in observational studies on networks. *J Am Stat Assoc*. 2021;116(534):901–18.
44. Rubin DB. Bayesian Inference for Causal Effects: The Role of Randomization. *Ann Statist*. 1978;6(1). <https://doi.org/10.1214/aos/1176344064>
45. Holland P. Statistics and causal inference. *J Am Stat Assoc*. 1986;81(396):945–60.
46. Jacod J, Shiryaev AN. *Limit theorems for stochastic processes*. 2nd ed. Springer. 2003.
47. Revuz D, Yor M. *Continuous martingales and brownian motion*. Springer. 2013.
48. Samorodnitsky G. Long memory and self-similar processes. *Annales de la Faculté des sciences de Toulouse : Mathématiques*. 2009;15(1):107–23. <https://doi.org/10.5802/afst.1115>
49. Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*. 1974;66(5):688–701.
50. Heckman J, Ichimura H, Smith J, Todd P. Characterizing Selection Bias Using Experimental Data. *Econometrica*. 1998;66(5):1017. <https://doi.org/10.2307/2999630>
51. Granger CWJ. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*. 1969;37(3):424. <https://doi.org/10.2307/1912791>
52. Sims CA. Macroeconomics and Reality. *Econometrica*. 1980;48(1):1. <https://doi.org/10.2307/1912017>
53. Menchetti F, Bojinov I. Estimating the effectiveness of permanent price reductions for competing products using multivariate Bayesian structural time series models. *Ann Appl Stat*. 2022;16(1). <https://doi.org/10.1214/21-aos1498>
54. Menchetti F, Cipollini F, Mealli F. Combining counterfactual outcomes and ARIMA models for policy evaluation. *The Econometrics Journal*. 2022;26(1):1–24. <https://doi.org/10.1093/ectj/utac024>
55. Fernández-Val I, Freeman H, Weidner M. Low-rank approximations of nonseparable panel models. *Econom J*. 2021;24(2):C40–77.
56. Ishihara T. Identification and estimation of time-varying nonseparable panel data models without stayers. *J Econom*. 2019;215(1):184–208.
57. Delgado MA, Escanciano JC. Nonparametric tests for conditional symmetry in dynamic models. *J Econom*. 2007;141:652–82.
58. Alonso C, Gutierrez J, Recio T. A note on separated factors of separated polynomials. *Journal of Pure and Applied Algebra*. 1997;121(3):217–22. [https://doi.org/10.1016/s0022-4049\(96\)00131-4](https://doi.org/10.1016/s0022-4049(96)00131-4)
59. Alonso C, Gutierrez J, Recio T. A rational function decomposition algorithm by near-separated polynomials. *J Symb Comput*. 2007;19(6):527–44.
60. Li T. Indirect inference in structural econometric models. *Journal of Econometrics*. 2010;157(1):120–8. <https://doi.org/10.1016/j.jeconom.2009.10.027>
61. Robinson PM. Root-N-Consistent Semiparametric Regression. *Econometrica*. 1988;56(4):931. <https://doi.org/10.2307/1912705>
62. Manski CF. Identification of Binary Response Models. *Journal of the American Statistical Association*. 1988;83(403):729–38. <https://doi.org/10.1080/01621459.1988.10478655>
63. Chamberlain G. Efficiency Bounds for Semiparametric Regression. *Econometrica*. 1992;60(3):567. <https://doi.org/10.2307/2951584>
64. Powell JL. Estimation of semiparametric models. *Handbook of econometrics*. 1994. p. 2443–521.
65. Taqqu MS. Weak convergence to fractional brownian motion and to the rosenblatt process. *Z Wahrscheinlichkeitstheorie verw Gebiete*. 1975;31(4):287–302. <https://doi.org/10.1007/bf00532868>
66. Dobrushin RL, Major P. Non-central limit theorems for non-linear functional of Gaussian fields. *Z Wahrscheinlichkeitstheorie verw Gebiete*. 1979;50(1):27–52. <https://doi.org/10.1007/bf00535673>
67. Resnick S. On the foundations of multivariate heavy-tail analysis. *Journal of Applied Probability*. 2004;41(A):191–212. <https://doi.org/10.1239/jap/1082552199>
68. Lahiri S. *Resampling Methods for Dependent Data*. Springer. 2013.
69. Lahiri S. Refinements in asymptotic expansions for sums of weakly dependent random vectors. *Ann Probab*. 1993;21:791–9.

70. Lang G, Roueff F. Semi-parametric Estimation of the Hölder Exponent of a Stationary Gaussian Process with Minimax Rates. *Statistical Inference for Stochastic Processes*. 2001;4(3):283–306. <https://doi.org/10.1023/a:1012227325436>
71. Barndorff-Nielsen OE, Corcuera JM, Podolskij M. Limit Theorems for Functionals of Higher Order Differences of Brownian Semi-Stationary Processes. *Springer Proceedings in Mathematics & Statistics*. Springer Berlin Heidelberg. 2012. p. 69–96. https://doi.org/10.1007/978-3-642-33549-5_4
72. Cai Z, Juhl T. The distribution of rolling regression estimators. *J Econom*. 2022;235(2):1447–63.
73. Zhou Z, Wu W. Simultaneous inference of linear models with time varying coefficients. *J R Stat Soc Ser B*. 2010;72:513–31.
74. Wu W, Zhou Z. Gaussian approximations for non-stationary multiple time series. *Stat Sin*. 2011;21:1397–413.
75. Wu W, Zhou Z. Multiscale jump and volatility analysis for high-frequency financial data. *J Am Stat Assoc*. 2018;113(523):1215–27.
76. Li Q, Racine J. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press. 2006.
77. Su L, Wang X, Jin S. Sieve estimation of time-varying panel data models with latent structures. *J Bus Econ Stat*. 2019;37(3):334–49.
78. Su L, Wang X, Yang Z. Sieve estimation of panel data models with cross-section dependence. *J Econom*. 2023;235(2):1447–63.
79. Kubilius K, Mishura Y, Ralchenko K. *Parameter Estimation in Fractional Diffusion Models*. Springer. 2017.
80. Kroese D, Botev Z, Taimre T, Vaisman R. *Data science and machine learning: mathematical and statistical methods*. Chapman and Hall/CRC. 2019.
81. Wang X, Xiao W, Yu J. Modeling and forecasting realized volatility with the fractional Ornstein-Uhlenbeck process. *J Econom*. 2023;232(2):389–415.
82. Sant'Anna P, Zhao J. Doubly robust difference-in-differences estimators. *J Econom*. 2020;219(1):101–22.
83. Arkhangelsky D, Athey S, Hirshberg DA, Imbens GW, Wager S. Synthetic difference-in-differences. *Am Econ Rev*. 2021;111(12):4088–118.
84. Xu Y. Generalized synthetic control method: causal inference with interactive fixed effects models. *Polit Anal*. 2017;25(1):57–76.
85. Brien M, Lillard L, Stern S. Cohabitation, marriage, and divorce in a model of match quality. *Int Econ Rev*. 2006;47(2):451–94.
86. Stevenson B, Wolfers J. Bargaining in the shadow of the law: Divorce laws and family distress. *Q J Econ*. 2006;121(1):267–88.
87. Kim D, Oka T. Divorce law reforms and divorce rates in the USA: An interactive fixed-effects approach. *J Appl Econom*. 2014;29(2):231–45.