True structure change, spurious treatment effect? A novel approach to disentangle treatment effects from structure changes

Hao, Shiming

7 July 2021

True structure change, spurious treatment effect? A novel approach to disentangle treatment effects from structure changes

Shiming Hao [a, b,*]

[a] *School of Economics & Trade, Hunan University, 109 Fenglin Road, Changsha, 410082, Hunan, PR China*

[b] *Academy of social-economic development, 2 Lushan South Road, 410012, Hunan, PR China*

Abstract: This paper develops a new flexible approach to disentangle treatment effects from structure changes. It is shown that ignoring prior structure changes or endogenous regime switches in causal inferences will lead to false positive or false negative treatment effects estimations. A difference in difference in difference strategy and a novel approach based on Automatically Auxiliary Regressions (AARs) are designed to separately identify and estimate treatment effects, structure changes effects and endogenous regime switch effects. The new approach has several desirable features. First, it does not need instrument variables to handle endogeneities and it is easy to implement with hardly any technical barriers to the empirical researchers; second, it can be extended to isolate one treatment from other treatments when the outcome is the working of a series of treatments; third, it outperforms other popular competitors in small sample simulations and the biases caused by endogeneities vanish with sample size. The new method is illustrated then in a comparative study of supporting direct destruction theory on the impacts of Hanshin-Awaji earthquake and Schumpeterian creative destruction theory on the impacts of Wenchuan earthquake.

Key words: structure changes; treatment effects; latent variable; endogeneity; regime switch model; social interactions

1. Introduction

Is the estimated treatment effect you get really the true treatment effect you want to get? “Of cause”, you may argue, “my model satisfies the parallel assumption (Callaway & Sant'Anna, 2021; Sun & Abraham, 2021), conditional independent assumption (Huber & Melly, 2015; Machado, 2017) and (quasi-)exogenous conditions among others (White, 2006; Imbens & Rubin, 2015),[1] my data is impeccable, my results are convincing after detailed robustness tests (Leamer, 1983; Franks et al., 2019; Cinelli C. & Hazlett, 2020), and any other competitive hypothesis has been excluded, therefore the treatment effect estimated must be consistent and irrefutable”.

☆ We sincerely thank Cheng Hsiao, Hiroshi Fujiki, Hansheng Wan and participants at the 2020 China Meeting of The Econometric Society, 2019 Asian Meeting of the Econometric Society, 2020 International Symposium on Econometrics for constructive helps and comments on this paper. Hao acknowledges the funds provided by National Key Social Science Foundation Grant (15ZDB176) and China Law Society Grant (No. CLS (2018) Y11).

* Corresponding author at: School of Economics & Trade, Hunan University, 109 Fenglin Road, Changsha, 410082, Hunan, China. Tel.: +8615211196096.
E-mail address: haoshiming@hnu.edu.cn (S. Hao).

[1] Under exogenous conditions, three commonly made restrictions of the treatment assignment mechanism are individualistic, probabilistic and unconfoundedness (Imbens & Rubin, 2015).Under an individualistic assignment mechanism, the combination of a probabilistic and unconfoundedness has been referred to both as strong unconfoundedness and strong ignorability (Stuart, 2010).

Beyond all doubt, this kind of research paradigm and routine has come to be the golden rules and precious precepts since the "credibility revolution" swept through economic studies for more than the past half century (see, *inter alia*, Keynes, 1939, 1940; Tinbergen, 1940; Haavelmo, 1944; Hendry, 1980; Black, 1982; Leamer, 1983; Pratt & Schlaifer, 1984; Hackman, 2001; Angrist & Pischke, 2009). [1] Nevertheless, we will show in this paper that two crucial factors have been neglected in the literatures both empirically and methodologically: prior structure changes and endogenous regime switches. If prior structure changes and endogenous regime switches are not controlled, causal inference fails down.

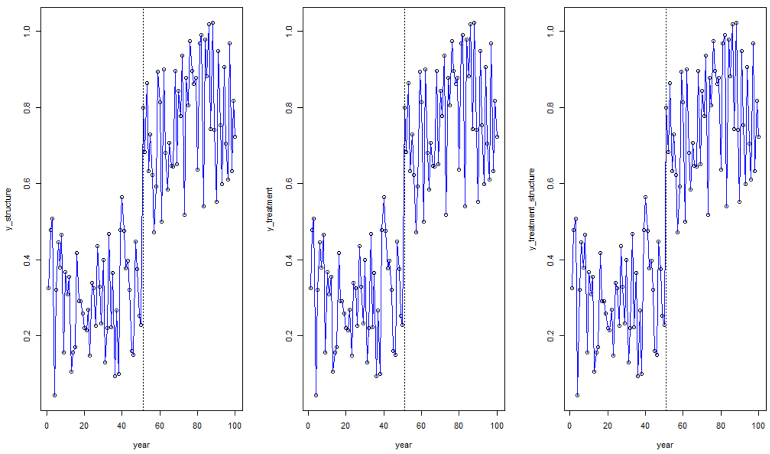
To see this, let us consider a simulated case adapted from a real empirical topic: the impact of electrification on modernization (Dinkelman, 2011; Alexopoulos & Cohen, 2016; Michaels et al., 2012; Lewis & Severnini, 2020). Suppose $\boldsymbol{y_{t,j}}$ represents the share of employment in industry j of a specific region at year t, the structure change or regime switch of the employment structure $\boldsymbol{y_{t,j}}$ is denoted as $\boldsymbol{s_{t,j}}$, which is totally driven by an unobservable latent variable $\boldsymbol{w_{t,j}}$, $\boldsymbol{s_{t,j}} = 1$ if $t \geq t_{s0,j}$. The mechanism of this kind of structure change is straightforward (Kim, 2004, 2009; Chang et al., 2017), for example $\boldsymbol{w_{t,j}}$ could be some latent innate endowments, such as industrial agglomeration densities, labor mobilities and other natural endowments, driving the changes of industrial employment structures (Moroney, 1975; Waring & Burgess, 2011). We are interested in evaluating the effect of promotion of electrification $\boldsymbol{D_{t,j}}$ on employments $\boldsymbol{y_{t,j}}$, where $\boldsymbol{D_{t,j}} = 1$ if $t \geq t_{D0,j}$. Therefrom, we could write down three data generating processes (DGPs):

$$\boldsymbol{y_{t,j,1}} = \left(\boldsymbol{w_{t,j}} + \boldsymbol{s_{t,j}}\beta_{j,1}\right)\eta_{j,1} + \boldsymbol{x_{t,j}}\alpha_{j,1} + \boldsymbol{\omega_{t,j}}, \boldsymbol{w_{t,j}} \perp \boldsymbol{x_{t,j}} \tag{1}$$

$$\boldsymbol{y_{t,j,2}} = \boldsymbol{w_{t,j}}\eta_{j,2} + \boldsymbol{D_{t,j}}\xi_{j,2} + \boldsymbol{x_{t,j}}\alpha_{j,2} + \boldsymbol{\omega_{t,j}}, \boldsymbol{w_{t,j}} \perp \left(\boldsymbol{D_{t,j}}, \boldsymbol{x_{t,j}}\right), \tag{2}$$

$$\boldsymbol{y_{t,j,3}} = \left(\boldsymbol{w_{t,j}} + \boldsymbol{s_{t,j}}\beta_{j,3}\right)\eta_{j,3} + \boldsymbol{D_{t,j}}\xi_{j,3} + \boldsymbol{x_{t,j}}\alpha_{j,3} + \boldsymbol{\omega_{t,j}}, \tag{3}$$

where $\boldsymbol{x_{t,j}}$ are covariates, $\boldsymbol{w'_{t,j}}$ and $\boldsymbol{x_{t,j}}$ are drawn from $\mathcal{U}(0,1)$, errors $\boldsymbol{\omega_{t,j}}$ are drawn from $\mathcal{N}(0,0.01)$, and $\boldsymbol{w_t} = sort(\boldsymbol{w'_t})$ for $t = 1,2,\dots,100$. DGP-(1) is a standard structure change model with regime switch (Chang et al., 2017), DGP-(2) is a standard treatment evaluation model where we assume $\boldsymbol{y_{t,j,2}} \perp \boldsymbol{D_{t,j}} | \boldsymbol{x_{t,j}}$, DGP-(3) is a combination of these two DGPs. If we assume $\beta_{j,1} = \beta_{j,2} = 1.2$, $\beta_{j,3} = 0.72$, $\eta_{j,1} = \eta_{j,2} = \eta_{j,3} = 0.25$, $\xi_{j,2} = 0.3$, $\xi_{j,3} = 0.12$, $\alpha_j = 0.5$ and $t_{s0,j} = t_{D0,j} = 50$, then these three DGPs' outcomes correspond to the following left, middle and right graph respectively:



[1] The main core of this revolution lies in the call for attentions to research designs under causal inference other than statistical exhaustions under relevance mining. To identify causality from correlations, researches need to argue whether the above assumptions and conditions are satisfied in their empirical works (Angrist & Pischke, 2009; Imbens & Rubin, 2015).

Figure 1. Who is who: structure change effect (left), treatment effect (middle), or both (right)?

As we shall see from Figure 1, these three outcomes are the same $\boldsymbol{y}_{t,j,1} = \boldsymbol{y}_{t,j,2} = \boldsymbol{y}_{t,j,3}$ although they are generated from three different GDPs. The question now is that if we observe one of these figures for the collected time-series $\boldsymbol{y}_{t,j}$, say the share of employment in industry j of a specific region, which DGP should we adopt? The left (1), the middle (2) or the right (3)? In this instance, if the true DGP is (3) with $\xi_{j,3} = 0.12$ but we adopt model (2) directly to evaluate the impacts of electrification, we will get over-estimated treatment effect $\xi_{j,2} = 0.3$. On the contrary, if the true DGP is (1), but we adopt model (2) directly, then we will get a spurious treatment effect $\xi_{j,2} = 0.3$, and the true effect is due to a structure change $\beta_{j,1}\eta_{j,1} = 1.2 \cdot 0.25 = 0.3$, here electrification has no impacts on industrial employments at all. This illustration shows us that ignoring prior structure changes and regime switches will fail down causal inferences in economics especially in regional policy evaluations (Alberto et al., 2010, 2015; Hsiao et al., 2012; Gobillon & Magnac, 2016; Xu, 2017).

Although it seems obvious, what beyond our expectation is that almost all empirical works in observation studies ignored this problem coincidentally. And there are also rare methodological approaches to deal with this issue in spite of the huge, increasing, updating literature on treatment effects and structure changes. Isolating the impact of one factor from other factors can be tricky and remains scattered in the literature (Fujiki & Hsiao, 2015; Lopez & Gutman, 2017). The most correlated paper to ours is Fujiki & Hsiao (2015), which is also the first paper noticing this issue. Through a backcasting technic, they propose a panel approach based on the well-known HCW method (Hsiao et al., 2012) to disentangle the effects of multiple treatments, they then find that the economic recessions after the Hanshin-Awaji earthquake are due to structure changes instead of the quake.

Other references that seem to be closest to this paper are multiple treatments (see, *inter alia*, Heckman et al., 2016) and multiple structure changes with endogenous regressors (see, *inter alia*, Hall, et al., 2012). However as shown in DGP-(3), this paper's setting is totally different from these two directions,[1] and it stands on its own feet for the following highlights: (1) different from multiple treatments, the mechanism of the structure change defined in DGP-(3) is totally different from treatments because $\boldsymbol{s}_t$ in (3) is driven by latent $\boldsymbol{w}_t$; and different from multiple structure changes, the mechanism of the treatment defined in DGP-(3) is totally different from structure changes because $\boldsymbol{D}_t$ in (3) is driven by $\boldsymbol{x}_t$;[2] (2) this paper distinguishes three sets of concepts: structure change, structure change effect and endogenous regime switch effect, while there is no strict distinction in the existing literature which will easily lead to misleading conclusions such as *false positive* or *false negative* mistakes;[3] (3) this paper is the first to simultaneously consider endogenous regime switches and endogenous treatments with unknown error distributions and unobservable latent variable, while the existing Bayesian methods reply on prior-known information (Kim, 2004, 2009; Chang et al., 2017).

This paper fills the research gap in disentangling treatment effects from structure changes or

[1] One may suspect that the setting and the issue proposed in this paper is actually a problem of multiple treatments or multiple structure changes (both $\boldsymbol{s}_t$ and $\boldsymbol{D}_t$ can be regarded as treatments or structure changes), hence weakens the innovation and potential value of this paper.

[2] In the multiple treatments literature, researchers usually assume that treatments are determined by confounders; while in multiple structure change literature, structure changes are determined by latent factors.

[3] *False positive* means that the true treatment effect is zero, but we get nonzero treatment effect estimations (maybe ATE, ATT or MTE); by contrast, *false negative* means that the true treatment effect is nonzero, but we get zero treatment effect estimations.

other treatment effects through a novel method. As far as we know, this is the second paper in this respect. Compared with Fujiki & Hsiao (2015)'s first try, the new method proposed can handle endogenous regime switches and endogenous treatments without IVs or other exogenous shocks, and is much more robust to the selection of control units. Simulations show that the new method outperforms the first try, especially in handling the endogenous problems caused by unobservable latent variable and omitted confounders.

This paper proceeds in the following way. Section 2 presents the DGP for endogenous regime switches and treatments in details, and illustrates it in three empirical cases. Section 3 introduces a new difference in difference strategy to identify the parameters. Section 4 proposes a novel estimation approach and establishes the estimators' asymptotic behaviors. Section 5 carries out small sample Monte Carlo studies and section 6 illustrates the new method through a comparative empirical study on the impacts of the earthquakes occurred in Hanshin-Awaji, Japan and Wenchuan, China. Section 7 concludes.

2. Models with endogenous structural changes and treatment effects

In this section, this paper introduces a new approach to model a social-economic outcome with structure changes and treatment effects. We will show that this new framework allows us to disentangling structural change effects from treatment effects. To show the issue that we want to reveal, instead of exhausting the technical complexities, we consider a simple time series setting, while the framework of this paper can be extended to panel data.

2.1. Endogenous structural changes and treatment effects

We decompose model (3) with endogenous regime switch and treatment effect into two layers, corresponding to two nested Data Generating Processes (DGPs)[1]. For the first layer, we study the interested economic indicators' latent growth pattern with a structural change and endogenous regime switch, where we assume the social-economic outcome's latent growth is driven by a latent factor

$$\mathbf{y}_{\boldsymbol{lp},\boldsymbol{t}} = \boldsymbol{w_t}\eta + \boldsymbol{s_t}\eta\beta + \boldsymbol{\nu_t}, \tag{4}$$

$$\boldsymbol{s_t} = s(\boldsymbol{w_t}) = 1 \cdot \mathbb{I}\{\boldsymbol{w_t} \geq \tau_s\} + 0 \cdot \mathbb{I}\{\boldsymbol{w_t} < \tau_s\}, \tag{5}$$

$\mathbf{y}_{\boldsymbol{lp},\boldsymbol{t}} = (y_{lp,1}, \dots, y_{lp,T})'$ is the latent growth part for some observed social-economic outcome that we are interested in but totally unobservable, subscript lp represents for the latent part with a structural change, $\boldsymbol{w_t} = (w_1, \dots, w_T)'$ denotes the latent factor which is also not observable to econometricians, η captures the impacts of the latent factor on the potential economic growth, s_t denotes the structure change where $\boldsymbol{s_t} = 0$ for $1 \leq t < t_{s0}$ and $\boldsymbol{s_t} = 1$ for $t_{s0} \leq t < T$, so the structure change take places at $t = t_{s0}$. $\boldsymbol{\nu_t}$ denotes an i.i.d. exogenous random shock whose density follows an unknown but symmetry distribution $\boldsymbol{\nu_t} =_{i.i.d.} \mathfrak{F}(0, \sigma_\nu^2)$, where we assume that $\mathbb{E}(\boldsymbol{\nu_t}) = 0$ and $\sigma_\nu^2 < \infty$. We assume $\mathbf{y}_{\boldsymbol{lp},\boldsymbol{t}}$ is driven by $(\boldsymbol{w_t}, \boldsymbol{s_t})$ and model (4) is correctly specified, $t = 1,2, \dots t_{s0}, \dots, T$. The latent variable $\boldsymbol{w_t}$ can describe macro-dynamic factors in fiscal studies, unobservable individuals' heterogeneousness in microeconometrics or path dependence and self-enforcement described in New Institutional theory.

[1] We call it as nested for the reason that we allow one layer exert influences on the other but not vice versa, hence our framework is distinguished from simultaneousness.

We assume that the structure change $\boldsymbol{s_t}$ is driven by a latent variable $\boldsymbol{w_t}$, hence the DGP for the structure change (5) we considered in this paper is a traditional Markov Switching Model where τ_s denotes the threshold, $\mathbb{I}\{\cdot\}$ is the indicator function (see Kim, 2004; Chang, 2017 for examples). Model (2) can be then rewritten as

$$\mathbf{y}_{lp,t} = \widetilde{\boldsymbol{w}}_t\eta + \boldsymbol{\nu}_t = (\boldsymbol{w}_t + \boldsymbol{s}_t\beta)\eta + \boldsymbol{\nu}_t, \tag{6}$$

where $\widetilde{\boldsymbol{w}}_t = (\boldsymbol{w}_t + \boldsymbol{s}_t\beta)$ measures the new latent growth part with a regime switch, β is the magnitude of structure change. Note that $(\mathbf{y}_{lp,t}, \boldsymbol{w}_t)$ is totally unobservable, hence we regard model (4-6) as a latent growth model for some observed social-economic outcome $\mathbf{y}_t$, in other words $\mathbf{y}_{lp,t}$ is the latent part of $\mathbf{y}_t$. $\mathbf{y}_{lp,t}$ can be then called the potential growth part, and $\mathbf{y}_t$ is the observed growth part.

For the second layer, we study the observed growth pattern with a treatment, where we assume the treatment is driven by some other exogenous indicators

$$\mathbf{y}_t = \mathbf{y}_{lp,t} + \boldsymbol{D}_t\xi + \boldsymbol{x}_t\alpha + \boldsymbol{\epsilon}_t, \tag{7}$$

$$\boldsymbol{D}_t = D(\boldsymbol{x}_t) = 1 \cdot \mathbb{I}\{\boldsymbol{x}_t \geq \tau_D\} + 0 \cdot \mathbb{I}\{\boldsymbol{x}_t < \tau_D\}, \tag{8}$$

where $\mathbf{y}_t = (y_1, \dots, y_T)'$ is the observed social-economic outcome's growth, $\mathbf{y}_{lp,t}$ is its latent part defined in (4), $\boldsymbol{D}_t$ denotes the treatment where $\boldsymbol{D}_t = 0$ for $1 \leq t < t_{D0}$ and $\boldsymbol{D}_t = 1$ for $t_{D0} \leq t < T$, so the treatment take places at $t = t_{D0}$, ξ is the treatment effect, and $\boldsymbol{x}_t = (x_{1t}, \dots, x_{pt})'$ denotes other P-dimensional confounders driving the observed economic growth as well as the treatment variable with $\alpha = (\alpha_1, \dots, \alpha_q)'$. We do not allow high-dimensional covariates in this paper, hence usually $P \ll T$. $\boldsymbol{\epsilon}_t$ is an i.i.d. random exogenous shock and we assume $\boldsymbol{\epsilon}_t =_{i.i.d.} \mathcal{N}(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon^2 < \infty$. We assume $\mathbf{y}_t$ is fully driven by $(\boldsymbol{w}_t, \boldsymbol{s}_t, \boldsymbol{x}_t, \boldsymbol{D}_t)$ and models (7-8) are correctly specified, $t = 1,2, \dots t_{s0}, \dots, t_{D0}, \dots, T$. Note that we do not require the treatment must be exogenous for the methods we developed in this paper, hence $\boldsymbol{D}_t$ could be a natural exogenous shock as well as an endogenous social-economic policy or intervention. However $\boldsymbol{s}_t$ is endogenous.

Combining layer one (4) and layer two (7), we get the total GDP for endogenous structure change and treatment effect

$$\mathbf{y}_t = (\boldsymbol{w}_t + s(\boldsymbol{w}_t)\beta)\eta + D(\boldsymbol{x}_t)\xi + \boldsymbol{x}_t\alpha + \boldsymbol{\epsilon}_t + \boldsymbol{\nu}_t, \tag{9}$$

where $s(\boldsymbol{w}_t)$ and $D(\boldsymbol{x}_t)$ are defined in (5) and (8) respectively, we assume $(\boldsymbol{\nu}_t, \boldsymbol{\epsilon}_t)$ is jointly i.i.d. distributed with

$$\begin{pmatrix} \boldsymbol{\nu}_t \\ \boldsymbol{\epsilon}_t \end{pmatrix} =_{i.i.d.} \mathfrak{G}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\nu^2 & \rho \\ \rho & \sigma_\epsilon^2 \end{pmatrix} \right),$$

$\mathfrak{G}$ is a Lebesgue measurable distribution defined on $\mathbb{R}^2$ with marginal distributions $\mathfrak{F}$ and $\mathcal{N}$, $\rho = Cov(\boldsymbol{\nu}_t, \boldsymbol{\epsilon}_t)$. β is the magnitude of structure change, η measures the endogenous regime switch effects describing the persistent impacts of the regime switch caused by structure change, $\beta\eta$ measures the structure change effects, ξ is the treatment effect. Note that we not only distinguish treatment effect from structure change effect, but also structure change effect from endogenous regime switch effect. The structure change effect $\beta\eta$ describes the direct effect of structure change on the outcome, while the endogenous regime switch effect η of $\widetilde{\boldsymbol{w}}_t$ describes the indirect effect of latent growth on the outcome through structure change. Usually $\eta \neq \beta\eta$ in empirical applications. $\beta\eta + \xi$ is the total effect we observed for the treatment and structure change. Note that we either allow $t_{s0} < t_{D0}$ or $t_{D0} < t_{s0}$ for the methods we developed, but for convenience we consider $t_{s0} < t_{D0}$ hence $t = 1,2, \dots, t_{s0}, \dots, t_{D0}, \dots, T$ for

(9) throughout this paper. Also note that we allow multiple structure changes and treatments, but for convenience we only consider univariate structure change and treatment in this paper. Special care should be paid to notice that we do not allow $t_{s0} = t_{D0}$ in this paper, hence the treatment and structure change cannot take place at a same time.

For the GDP (9) with a structure change and treatment effect, $(\mathbf{y}_t, \boldsymbol{s}_t, \boldsymbol{D}_t, \boldsymbol{x}_t)$ are observable to experimenters while the latent growth factor $\boldsymbol{w}_t$ is unobservable, what we are interested in are identification and estimation of the structural parameters $\theta = (w_1, \dots, w_T, \beta, \eta, \beta\eta, \xi, \alpha, \sigma_{\nu+\epsilon}, \rho) \subset \mathbb{R}^{T+7}$. Due to the unobservability of the latent growth part $\boldsymbol{w}_t$, we cannot use OLS directly to estimate model (9).[1] Apart from this, the methodology developed in this paper does not require $\boldsymbol{x}_t$ to be fully observed, hence the treatment $\boldsymbol{D}_t$ somehow could also be endogenous, we are then interested in consistently estimating $\theta' = (\beta\eta, \xi)$. Under this scenario, IVs are available methods to estimate (9) but good IVs are extremely difficult to find, MLE-type methods also fail here for we don't know the exact distribution of $\mathfrak{G}$ therein and the large sample performances of MLE with structure changes are not well understood. Bayesian methods of MCMC or EM to deal with latent variables also face the problem of prior specification, high-dimensional computation burden and inference difficulties (Li & Yu, 2012). As far as we know, our model is of potential interests to many social-economic researchers, but there are no suitable methods to deal with this problem and disentangling treatment effects from structural change effects remains a less explored (or even forgotten) area in econometrician's backyard garden.

2.2. Empirical illustrations

To show the issues we want to appeal, three real empirical cases appearing from the area of macro- and micro-economic studies are then illustrated.

- **Case One (regional policy evaluation):** Suppose now we are interested in studying the treatment effects of some macro-policy interventions on economic growth (measured by Gross Domestic Product, GDP), such as the 2008 economic stimulus package of China to fight against the global economic crisis, the general model (9) is set as

$$GDP_t = (economic\ fundamentals_t + the\ regime\ of\ the\ economy_t \cdot \beta) \cdot \eta + covariates_t \cdot \alpha + economic\ stimulus\ package_t \cdot \xi + \boldsymbol{\epsilon}_t + \boldsymbol{\nu}_t,$$
$$t = t_1 \dots, t_{s0}, \dots, t_{D0}, \dots, T, \tag{10}$$

where economic fundamentals are some latent factors that drive the development of the economy and determine the regimes of the economy (Chang et. al, 2017), covariates are other economic indicators needed to be controlled for, which will exert influences on the economy such as investment in fixed assets, foreign trades, foreign direct investments, domestic consumptions to name a few in the economic growth literature (Becker et al., 2010). There is a structure change at $t = t_{s0}$ and the treatment takes place at $t_{D0} = 2008$. What we are interested in is the consistent estimation and inference of the treatment effect ξ. However, if the researcher ignores the regime switch of the economy, say the DGP is misspecified as

$$GDP_t = covariates_t \cdot \alpha + economic\ stimulus\ package_t \cdot \xi + \boldsymbol{u}_t. \tag{11}$$

Then identification and estimation of ξ fail here whatever methods are adopted including case

[1] For the reason that $\boldsymbol{s}_t$ is endogenous, we call $\widetilde{\boldsymbol{w}}_t$ in (6) an endogenous regime switch.

(event) study, IVs, synthetic control methods or difference in difference even if the Conditional Independent Assumption (CIA) $\mathrm{GDP_t} \perp \text{economic stimulus package}_\mathrm{t}|\mathrm{covariates_t}$ is fulfilled. Furthermore, the most important problem we want to attract your attention is the failure of identification in (11). As we shall see

$$\begin{aligned}&\mathbb{E}(GDP_t|\mathbf{D}_t = 1, \mathbf{x}_t) - \mathbb{E}(GDP_t|\mathbf{D}_t = 0, \mathbf{x}_t)\\ &\qquad = \underbrace{\mathbb{E}(GDP_t^1 - GDP_t^0|\mathbf{D}_t = 1, \mathbf{x}_t)}_{ATT=\xi} + \underbrace{\mathbb{E}(GDP_t^1 - GDP_t^0|\mathbf{D}_t - \Delta_t = 1, \mathbf{x}_t)}_{\beta\eta}\end{aligned}$$

where $\mathbf{D}_t$ denotes the economic stimulus package, $\mathbf{s}_t$ denotes the structure change, $\mathbf{x}_t$ denotes covariates and $\Delta_t = \mathbf{D}_t - \mathbf{s}_t$. As long as $\beta\eta > 0$ (or $\beta\eta < 0$), the identification strategy of traditional methods will over- (or under-) estimate the treatment effect ξ. What's more, we will detect *false positive* treatment effect while the true effect is actually a structure change effect ($\beta\eta$) and the true treatment effect is zero $\xi = 0$, or *false negative* treatment effect while the true treatment effect and the structural change effect offset each other ($\beta\eta + \xi = 0$), and the true treatment effect is not zero $\xi \neq 0$. This case warns us that ignoring structure change factors or endogenous regime switches in evaluating policy interventions will lead us to biased conclusions and wrong policy implications.

- **Case Two (under-estimated job training effects):** Evaluating the effects of job training programs lies in the top topics of labor economics, where most of the researches mainly concern the problem of self-selection bias or Ashenfelter's dip. However, we will show in this paper that there is another bias omitted in the literatures. Suppose the general empirical input-output model of job training is set as

$$\begin{aligned}\log(wages_{it}) = (&personal\ ability_{it} + job\ position\ promotions_{it} \cdot \beta) \cdot \eta + covariates_{it} \cdot \alpha\\ &+ job\ training_{it} \cdot \xi + \boldsymbol{\iota_i} + \boldsymbol{\mu_t} + \boldsymbol{\epsilon_{it}} + \boldsymbol{\nu_{it}}, \qquad (12)\end{aligned}$$

where covariates may include education level, family background among others according to the literature, we suppose $\log(wages_{it}) \perp job\ training_{it}|covariates_{it}$ and all the right hand side variables in (12) are independent of $(\boldsymbol{\iota_i}, \boldsymbol{\mu_t}, \boldsymbol{\epsilon_{it}}, \boldsymbol{\nu_{it}})$, $i = 1, \dots, N$. It's easy to see that wages, personal ability and job promotion are directly influenced by job trainings, while job promotion is also directly influenced by personal ability. Hence, different from (10) where $t_{s0} < t_{D0}$, job promotion happens after job trainings, hence $t = t_1 \dots, t_{D0}, \dots, t_{s0}, \dots, T$. What we are interested in is the causal effects of job training on personal wages' change

$$\frac{\partial \boldsymbol{y_{it}}}{\partial \boldsymbol{D_{it}}} = \frac{\partial \boldsymbol{w_{it}}}{\partial \boldsymbol{D_{it}}}\eta + \frac{\partial \boldsymbol{s_{it}}}{\partial \boldsymbol{w_{it}}}\frac{\partial \boldsymbol{w_{it}}}{\partial \boldsymbol{D_{it}}}\eta\beta + \xi = \eta\left(\frac{\partial \boldsymbol{w_{it}}}{\partial \boldsymbol{D_{it}}} + \frac{\partial \boldsymbol{s_{it}}}{\partial \boldsymbol{D_{it}}}\beta\right) + \xi \qquad (13)$$

where $\boldsymbol{y_{it}}$ is $\log(\mathrm{wages_{it}})$, $\boldsymbol{w_{it}}$ represents personal ability which is totally unobservable hence latent, $\boldsymbol{s_{it}}$ is job promotion, the treatment $\boldsymbol{D_{it}}$ is job training. Of particular note is that the treatment effect we are interested in now is not ξ but (13) where $\boldsymbol{w_{it}}$ and $\boldsymbol{s_{it}}$ is not independent of $\boldsymbol{D_{it}}$. Suppose there exists a job promotion after job training, then we have $\partial \boldsymbol{D_{it}} = \partial \boldsymbol{s_{it}} = 1$. If we further assume that there is a unit ability improvement after job training $\boldsymbol{\partial w_{it}} = 1$ and $sign(\eta) = sign(\beta) = sign(\xi) > 0$ w.l.o.g., then the treatment effect turns out to be $\partial \boldsymbol{y_{it}}/\partial \boldsymbol{D_{it}} = \eta + \eta\beta + \xi$. Compared with ξ if the latent part is ignored, we can now see that ignoring personal ability's endogenous switch will under-estimate the training effect, where the under-estimated part $\eta + \eta\beta$ captures the indirect effects of job training on wages through unobservable personal ability improvements.

Under this scenario, one can verify that equation (12) is equivalent to

$$\log(wages_{it}) = covariates_{it} \cdot \alpha + job\ trainings_{it} \cdot \xi + 1/2 \cdot job\ trainings_{it} \cdot personal\ ability_{it} \cdot \phi + 1/2 \cdot job\ trainings_{it} \cdot job\ promotion_{it} \cdot \varphi + \boldsymbol{\iota_i} + \boldsymbol{\mu_t} + \boldsymbol{u_{it}}, \quad (14)$$

with $\phi = \eta$ and $\varphi = \eta\beta$ capturing the indirect treatment effect through personal ability improvement and job promotion.[1] From this point of view, we suspect that many labor literature under-estimated job training effects.

- **Case Three (optimal controls with social interactions):** We will show that the distinguish between structure changes, endogenous regime switches and treatment effects has its own unique meanings rooted in the microeconomic theories, ignoring structural factors will not only cause biased empirical estimations but also false theoretical conclusions.

Suppose a firm is facing a choice of different business models while the local government is facing a choice of whether or not carrying out a promotion of clean energy technology. Firms' decision sets are denoted as $\boldsymbol{s_{cit}} \in \{0,1\}$ where $\boldsymbol{s_{cit}} = 1$ if firm i in state c choses to turn its business model into a new one at $t = t_{s_0ci}$, otherwise $\boldsymbol{s_{cit}} = 0$; similarly, local governments' decisions are denoted as $\boldsymbol{D_{cit}} \in \{0,1\}$ where $\boldsymbol{D_{cit}} = 1$ if the new technology is carried out at $t = t_{D_0ci}$, otherwise $\boldsymbol{D_{cit}} = 0$. Note that the promotion of new technologies by the local governments can be seen as a treatment to the firms, and we allow heterogeneous responses so that t_{s_0ci} and t_{D_0ci} can vary among different states c and firms i. Given the public information set $\mathfrak{V}_{sD} = (\boldsymbol{w_{cit}}, \boldsymbol{x_{cit}})$, in which $\boldsymbol{w_{cit}}$ represents the firms-level observable information driving the choice of $\boldsymbol{s_{cit}}$, and $\boldsymbol{x_{cit}}$ captures local governments' information driving $\mathbf{D_{cit}}$, we let $\mathfrak{W}_{sD} = (\boldsymbol{\epsilon_{cit}}, \boldsymbol{v_{cit}})$ denote some unobservable private information to the econometricians, of which each component is only observable to the local governments or firms themselves respectively. If we rewrite the decision set (5) and (8) as $\boldsymbol{s_{cit}} = a_{ci} \cdot \mathbb{I}\{\boldsymbol{w_{cit}} \geq \tau_{si}\} + b_{ci} \cdot \mathbb{I}\{\boldsymbol{w_{cit}} < \tau_{si}\}$ and $\boldsymbol{D_{cit}} = c_{ci} \cdot \mathbb{I}\{\boldsymbol{x_{cit}} \geq \tau_{Di}\} + d_{ci} \cdot \mathbb{I}\{\boldsymbol{x_{cit}} < \tau_{Di}\}$ respectively for some real values a_{ci},b_{ci},c_{ci},$d_{ci} \in \{0,1\}$, $i = 1,2,\dots,N_c$, $c = 1,2,\dots,C$, and $t_{ci} = 1,\dots,t_{s_0ci},\dots,t_{D_0ci},\dots,T$, then given the total information set $(\mathfrak{V}, \mathfrak{W})$, the problem we are facing now is the optimal choice of $\Theta_{ci} = \{a_{ci}, b_{ci}, c_{ci}, d_{ci}, \mathrm{t}_{s_0ci}, \mathrm{t}_{D_0ci}\} \subset \mathbb{R}^6$ for firm i and the corresponding local goverment. Note that in this game, we require local governments' actions are taken after firms, so $\boldsymbol{s_{cit}}$ is observable to local governments while $\boldsymbol{D_{cit}}$ is not observable to firms. Suppose there exist a smooth function $\mathrm{G}_l^{sD}(a_{ci}, b_{ci}, c_{ci}, d_{ci})$ satisfying

$$\dot{\mathbf{s}}_{\boldsymbol{cit}} = \begin{cases} \mathrm{G}_s^{00}(a_{ci}, b_{ci}, c_{ci}, d_{ci}), 1 \leq t_{ci} < \mathrm{t}_{s_0ci} \\ \mathrm{G}_s^{10}(a_{ci}, b_{ci}, c_{ci}, d_{ci}), \mathrm{t}_{s_0ci} \leq t_{ci} < \mathrm{t}_{D_0ci} \\ \mathrm{G}_s^{11}(a_{ci}, b_{ci}, c_{ci}, d_{ci}), \mathrm{t}_{D_0ci} \leq t_{ci} < \mathrm{T} \end{cases}, \dot{\mathbf{D}}_{\boldsymbol{cit}} = \begin{cases} \mathrm{G}_D^{00}(a_{ci}, b_{ci}, c_{ci}, d_{ci}), 1 \leq t_{ci} < \mathrm{t}_{s_0ci} \\ \mathrm{G}_D^{10}(a_{ci}, b_{ci}, c_{ci}, d_{ci}), \mathrm{t}_{s_0ci} \leq t_{ci} < \mathrm{t}_{D_0ci} \\ \mathrm{G}_D^{11}(a_{ci}, b_{ci}, c_{ci}, d_{ci}), \mathrm{t}_{D_0ci} \leq t_{ci} < \mathrm{T} \end{cases},$$

where the state function $\mathrm{G}_l^{sD}(a_{ci}, b_{ci}, c_{ci}, d_{ci})$ are specified to describe 4 states on which the decision set $l = s, D$ can lay depending on the occurrence of s-regimes and D-regimes, $\dot{\mathbf{s}}_{\boldsymbol{cit}}, \dot{\mathbf{D}}_{\boldsymbol{cit}}$ describes the motion of the decision state variables $\boldsymbol{s_{cit}}$, $\boldsymbol{D_{cit}}$ respectively on the time interval. The optimization problem turns out to be

[1] One should be cautious that η and $\eta\beta$ now capture treatment effect instead of structure change effect or endogenous switch effect as in (10) shown in case one. We will show in the next section of this paper that if $\boldsymbol{s_t}$ is not independent of $\boldsymbol{D_t}$ or $\boldsymbol{w_t}$ is not independent of $\boldsymbol{x_t}$, then we cannot disentangle treatment from structure change through identification. Under this scenario, the treatment effect and structure change effect mix with each other, so that η, $\eta\beta$ capture the indirect treatment effect though the latent variable $\boldsymbol{w_t}$ and structure change $\boldsymbol{s_t}$ (the structure change is indirectly caused by treatment through the latent variable).

$$\underset{\Theta\subset\mathbb{R}^{6\sum_{c=1}^{C}N_c}}{\operatorname{argmax}}\ \mathrm{V}(\Theta) = \underset{\Theta\subset\mathbb{R}^{6\sum_{c=1}^{C}N_c}}{\operatorname{argmax}} \left(\mathbb{E}\left(\int_{1}^{\mathrm{t}_{s_0}} (\mathcal{H}_1 - \lambda_S^{00}\mathrm{G}_S^{00} - \lambda_D^{00}\mathrm{G}_D^{00})\, dt + \int_{\mathrm{t}_{s_0 ci}}^{\mathrm{t}_{D_0}} (\mathcal{H}_2 - \lambda_S^{10}\mathrm{G}_S^{10} - \lambda_D^{10}\mathrm{G}_D^{10})\, dt + \int_{\mathrm{t}_{D_0}}^{\mathrm{T}} (\mathcal{H}_3 - \lambda_S^{11}\mathrm{G}_S^{11} - \lambda_D^{11}\mathrm{G}_D^{11})\, dt \right)\right) \quad (15)$$

with the Hamiltonian $\mathcal{H}_{\Bbbk} = \mathcal{F}(a_{ci}, b_{ci}, c_{ci}, d_{ci})e^{-\delta t} + \lambda_S^{sD}\mathrm{G}_S^{sD} + \lambda_D^{sD}\mathrm{G}_D^{sD}, k = 1,2,3$ and a well-defined smooth object function $\mathcal{F}(a_{ci}, b_{ci}, c_{ci}, d_{ci})$ subjected to $\dot{\mathbf{s}}_{\boldsymbol{cit}}$ and $\dot{\mathbf{D}}_{\boldsymbol{cit}}$, λ_l^{sD} is the co-state variable associated with the decision state l when the firms and local governments are in regime $(\boldsymbol{s_{cit}}, \boldsymbol{D_{cit}})$, the expectation operator is taken over all c and i. Note that $6\sum_{c=1}^{C} N_c$ can be very large, so this is a high-dimensional optimization problem. Under some regular conditions of Boucekkine et al. (2013) and Seidl (2019), for each c and i, the solution of (15) Θ_{ci}^* satisfies a series of first-order equations corresponding to a *Matching Condition*

$$\mathbf{y}_{\boldsymbol{cit_{ci}}} = \left(\boldsymbol{w_{cit_{ci}}} + s\left(\boldsymbol{w_{cit_{ci}}}\right)\beta_{ci}\right)\eta_{ci} + D\left(\boldsymbol{x_{cit_{ci}}}\right)\xi_{ci} + \boldsymbol{x_{cit_{ci}}}\alpha_{ci} + \boldsymbol{\epsilon_{cit_{ci}}} + \boldsymbol{\nu_{cit_{ci}}} \quad (16)$$

for a well-defined object function $\mathcal{F}(a_{ci}, b_{ci}, c_{ci}, d_{ci})$, where $\mathrm{y}_{cit_{ci}}$ is the total social welfare, $\left(\boldsymbol{w_{cit_{ci}}} + s\left(\boldsymbol{w_{cit_{ci}}}\right)\beta_{ci}\right)\eta_{ci}$ is firm's welfare and $D\left(\boldsymbol{x_{cit_{ci}}}\right)\xi_{ci} + \boldsymbol{x_{cit_{ci}}}\alpha_{ci}$ is local government's welfare, $\beta_{ci}\eta_{ci}$ captures heterogeneous effects of firms' decisions on social welfare while ξ_{ci} captures heterogeneous treatment effects of the promotion of clean energy technologies for different c and i. In this perspective, if the social interactions and games between firms and local governments are neglected from the optimizations (15), the *matching condition* will not convergent to (16), the solution of (15) will not convergent to Θ_{ci}^*, and the overall solution will not convergent to the optimal Θ^*. It follows that neglecting the confrontation and adjustment of enterprises ahead of time to the expected behavior of the governments will lead to unexpected social policy effects, this can be the reason why some of the social interventions lose their effects or cause opposite effects.

Suppose a firm's decision s_{cit} is also affected by his peers' decisions $\boldsymbol{s_{cjt}}$, $j \in \mathcal{P}_{ci}$ where $\mathcal{P}_{ci}$ is the reference group for firm i in state c (Manski, 2013), then given the incomplete information set $(\mathfrak{V}_{ss'}, \mathfrak{W}_{ss'})$ where $\mathfrak{V}_{ss'} = \left(\boldsymbol{x_{cit}}, \boldsymbol{x_{cjt}}\right)$ and $\mathfrak{W}_{ss'} = \left(\boldsymbol{\nu_{cit}}, \boldsymbol{\nu_{cjt}}\right)$, the utility function for firm i in this Bayesian-Nash game under social interactions can be specified as

$$\mathcal{U}_{ci}\left(\mathbf{s}_{\boldsymbol{cit}}, \mathbf{s}_{\boldsymbol{cjt}}\right) = \left(\boldsymbol{x_{cit}}\alpha_{ci} + \delta \sum_{j\in\mathcal{P}_{ci}} \alpha_{ci,j}\boldsymbol{x_{cjt_{cj}}}\right)\mathbf{s}_{\boldsymbol{cit}} - \frac{1}{2}\mathbf{s}_{\boldsymbol{cit}}^2 - \frac{\phi}{2}\left(\mathbf{s}_{\boldsymbol{cit}} - \sum_{j\in\mathcal{P}_{ci}} \gamma_{ci,j}\mathbf{s}_{\boldsymbol{cjt}}\right)^2 + \mathbf{s}_{\boldsymbol{cit}}\boldsymbol{D_{cit}}\xi_{ci}. \quad (17)$$

The specific meaning of (17) is defined in Blume et.al (2015). Note that δ and ϕ capture peer effects, and we allow firms' utility not only influenced by his peers' actions but also local governments' treatments. The solution of the first-order condition for (17) is

$$\mathbb{E}(\boldsymbol{s_{cit}}|\boldsymbol{x}) = \frac{\phi}{1+\phi}\sum_{j\in\mathcal{P}_{ci}} \gamma_{ci,j}\mathbb{E}\left(\mathbf{s}_{\boldsymbol{cjt}}|\boldsymbol{x}\right) + \frac{\delta}{1+\phi}\sum_{j\in\mathcal{P}_{ci}} \alpha_{ci,j}\boldsymbol{x_{cjt_{cj}}} + \frac{\alpha_{ci}}{1+\phi}\boldsymbol{x_{cit_{ci}}} + \frac{\xi_{ci}}{1+\phi}\mathbb{E}(\boldsymbol{D_{cit}}|\boldsymbol{x}). \quad (18)$$

If we assume the promotion of clean energy technology by the local government will only take effects on firm i in state c other than i's peers $\mathcal{P}_{ci}$, and only firm i in state c is facing the choice of different business models, then we have $\mathbf{y}_{\boldsymbol{cjt_{cj}}} = \alpha_{cj}\boldsymbol{x_{cjt_{cj}}} + \boldsymbol{\nu_{cjt_{cj}}}$ with $\mathbf{s}_{\boldsymbol{cjt}} = 0$ and

$\mathbf{D}_{\boldsymbol{cjt}} = 0$ for all $j \in \mathcal{P}_{ci}$. Therefore from (16), we can get

$$\boldsymbol{s_{cit}} = \frac{1}{\beta_{ci}\eta_{ci}}\mathbf{y}_{\boldsymbol{cit_{ci}}} - \frac{1}{\beta_{ci}}\boldsymbol{w}_{\boldsymbol{cit_{ci}}} - \frac{\alpha_{ci}}{\beta_{ci}\eta_{ci}}\boldsymbol{x}_{\boldsymbol{cit_{ci}}} - \frac{\xi_{ci}}{\beta_{ci}\eta_{ci}}\boldsymbol{D}_{\boldsymbol{cit}} - \frac{1}{\beta_{ci}\eta_{ci}}\left(\boldsymbol{\epsilon}_{\boldsymbol{cit_{ci}}} + \boldsymbol{\nu}_{\boldsymbol{cit_{ci}}}\right) \quad (19)$$

for firm i, and

$$\mathbf{s}_{\boldsymbol{cjt}} = \mathbf{y}_{\boldsymbol{cjt_{cj}}} - \alpha_{cj}\boldsymbol{x}_{\boldsymbol{cjt_{cj}}} - \boldsymbol{\nu}_{\boldsymbol{cjt_{cj}}}, \mathbb{E}\left(\boldsymbol{\nu}_{\boldsymbol{cjt_{cj}}}\right) = 0 \quad (20)$$

for i's peers $j \in \mathcal{P}_{ci}$. Substitute (20) into (18), we get the general social equilibrium equation

$$\mathbb{E}(\boldsymbol{s_{cit}}|\boldsymbol{x}) = \frac{\phi}{(1+\phi)(n-1)}\sum_{j\in\mathcal{P}_{ci}}\mathbb{E}\left(\mathbf{y}_{\boldsymbol{cjt_{cj}}}|\boldsymbol{x}\right) - \frac{1}{(1+\phi)(n-1)}(\phi-\delta)\left(\sum_{j\in\mathcal{P}_{ci}}\alpha_{ci,j}\boldsymbol{x}_{\boldsymbol{cjt_{cj}}}\right)$$

$$+\frac{\alpha_{ci}}{1+\phi}\boldsymbol{x_{cit}} + \frac{\xi_{ci}}{1+\phi}\mathbb{E}(\boldsymbol{D_{cit}}|\boldsymbol{x}). \quad (21)$$

Simultaneously consider equation (17) and (19), if $\phi - \delta = 0$ we then get

$$\boldsymbol{w}_{\boldsymbol{cit_{ci}}} = \frac{1}{\eta_{ci}}\mathbf{y}_{\boldsymbol{cit_{ci}}} + \frac{1+\beta_{ci}\eta_{ci}}{\eta_{ci}(n-1)}\sum_{j\in\mathcal{P}_{ci}}\mathbb{E}\left(\mathbf{y}_{\boldsymbol{cjt_{cj}}}|\boldsymbol{x}\right), \quad (22)$$

where $1+\phi = -\beta_{ci}\eta_{ci}$, and we set $\gamma_{ci,j} = 1/n-1$ according to the literature (e.g.: Manski, 2013). The *matching condition* with social interactions then turns out to be

$$\frac{-1-\beta_{ci}\eta_{ci}}{(n-1)}\sum_{j\in\mathcal{P}_{ci}}\mathbb{E}\left(\mathbf{y}_{\boldsymbol{cjt_{cj}}}|\boldsymbol{x}\right) = \boldsymbol{s}_{\boldsymbol{cit_{ci}}}\beta_{ci}\eta_{ci} + \boldsymbol{D}_{\boldsymbol{cit_{ci}}}\xi_{ci} + \boldsymbol{x}_{\boldsymbol{cit_{ci}}}\alpha_{ci} + \boldsymbol{\epsilon}_{\boldsymbol{cit_{ci}}} + \boldsymbol{\nu}_{\boldsymbol{cit_{ci}}}. \quad (23)$$

As we can see here that the latent variable $\boldsymbol{w}_{\boldsymbol{cit_{ci}}}$ for the *matching condition* (16) under social interactions is a compound function of firms' welfares $\mathbf{y}_{\boldsymbol{cit_{ci}}}$and their peers' average welfares $1/(n-1)\sum_{j\in\mathcal{P}_{ci}}\mathbb{E}\left(\mathbf{y}_{\boldsymbol{cjt_{cj}}}|\boldsymbol{x}\right)$. Parameter $\beta_{ci}\eta_{ci}$ here in (23) captures some sort of endogenous peer effects of peers' average welfares on firms' welfares as well as heterogeneous effects of firms' decisions on total social welfare.

3. Identification

3.1. Identification of endogenous structural changes and treatment effects

As discussed before in the introduction part, most of the empirical studies are particularly interested in the following specifications

$GDP1$: $\mathbf{y}_{\boldsymbol{t},\mathbf{1}} = (w(\boldsymbol{D_t}) + s(\boldsymbol{w_t})\beta_1)\eta_1 + D(\boldsymbol{x_t})\xi_1 + \boldsymbol{x_t}\alpha_1 + \boldsymbol{\epsilon}_{\boldsymbol{t},\mathbf{1}} + \boldsymbol{\nu}_{\boldsymbol{t},\mathbf{1}},$ (24)

$GDP2$: $\mathbf{y}_{\boldsymbol{t},\mathbf{2}} = (\boldsymbol{w_t} + s(\boldsymbol{w_t})\beta_2)\eta_2 + D(\boldsymbol{x_t})\xi_2 + \boldsymbol{x_t}\alpha_2 + \boldsymbol{\epsilon}_{\boldsymbol{t},\mathbf{2}} + \boldsymbol{\nu}_{\boldsymbol{t},\mathbf{2}}, \boldsymbol{w_t} \perp D(\boldsymbol{x_t}),$ (25)

$GDP3$: $\mathbf{y}_{\boldsymbol{t},\mathbf{3}} = (\boldsymbol{w_t} + s(\boldsymbol{w_t})\beta_3)\eta_3 + \boldsymbol{x_t}\alpha_3 + \boldsymbol{\nu}_{\boldsymbol{t},\mathbf{3}},$ (26)

$GDP4$: $\mathbf{y}_{\boldsymbol{t},\mathbf{4}} = D(\boldsymbol{x_t})\xi_4 + \boldsymbol{x_t}\alpha_4 + \boldsymbol{\epsilon}_{\boldsymbol{t},\mathbf{4}}.$ (27)

corresponding to the GDP with treatment effect for (27), GDP with structural change effect and endogenous regime switch effect for (26), GDP with independent structural change effect and treatment effect for (25) and GDP with dependent structural change effect and treatment effect for (24). These are four different but widely used model specifications in empirical studies where we usually assume $\beta_k\eta_k \neq \beta_{k'}\eta_{k'}$, $\xi_k \neq \xi_{k'}$ and $\alpha_k \neq \alpha_{k'}$for $k, k' \in \{1,2,3,4\}$, and each of the models is correctly specified. The difference between model (24) and (25) is that the latent variable $\boldsymbol{w_t}$ is influenced by treatment $\boldsymbol{D_t}$ in GDP1 (hence $\boldsymbol{w_t}$ is a function of $\boldsymbol{D_t}$) while $\boldsymbol{w_t}$

is independent of $\boldsymbol{D_t}$ in GDP2, which further implies that the structure change is caused by treatment in (24) while the structural change effect has no relationship with treatment effect in (25). GDP1 is common in economic studies for example the economy's structure changes (or transitions) are caused by an earthquake which could be regarded as an exogenous treatment (Okuyama, 2015).

The Direct Acyclic Graphs (DAGs) for these four DGPs are shown in Figure 2, corresponding to the following adjacency matrices

$$\mathcal{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \mathcal{B} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathcal{C} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \mathcal{D} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

where $\boldsymbol{\mathcal{L}_{mn}} = 1$ if there is a direct link between nodes $m, n \in \{\mathbf{D}_t, \mathbf{w}_t, \mathbf{s}_t, \mathbf{x}_t, \mathbf{y}_t\}$, otherwise $\boldsymbol{\mathcal{L}_{mn}} = 0$ for $\boldsymbol{\mathcal{L}} = \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{C}}, \boldsymbol{\mathcal{D}}$. We impose the following regularity conditions:

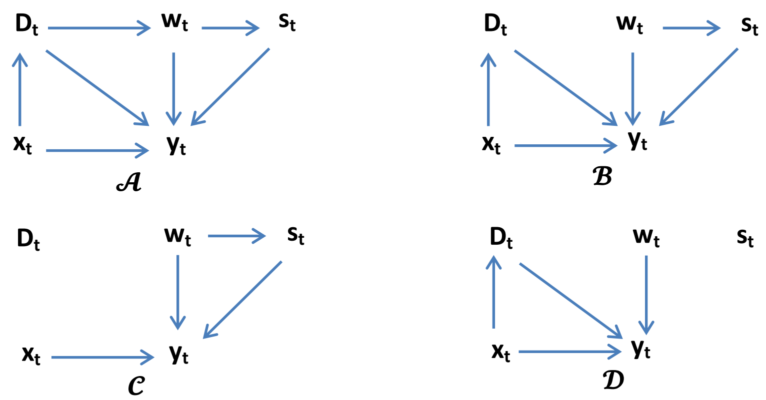


Figure 2 DAGs for endogenous structure change and treatment effect

Assumption 1. $\mathbf{y}_t \perp D(\boldsymbol{x_t})|\boldsymbol{x_t}$ for the GDPs 1, 2, 4 and $\mathbf{y}_t \perp s(\boldsymbol{w_t})|\boldsymbol{w_t}$ for the GDPs 2-3.

Assumption 2. $\boldsymbol{\epsilon_{kt}}$ are i.i.d. with $\boldsymbol{\epsilon_{kt}} =_d \mathcal{N}(0, \sigma_{k\epsilon}^2)$, and $\boldsymbol{\nu_{kt}}$ are i.i.d. with $\boldsymbol{\nu_{kt}} =_d \mathfrak{F}_k(0, \sigma_{k\nu}^2)$ where $\mathfrak{F}_k$ is a Lebesgue measurable distribution, $k = 1,2,3,4$, $t = 1, \dots, t_{s0}, \dots, t_{D0}, \dots, T$; and $\{\boldsymbol{\epsilon_{k1}}, \boldsymbol{\epsilon_{k2}}, \dots, \boldsymbol{\epsilon_{k\bar{t}}}\} \perp \{\boldsymbol{\epsilon_{k\bar{t}+1}}, \boldsymbol{\epsilon_{k\bar{t}+2}}, \dots, \boldsymbol{\epsilon_{kT}}\}$ for $2 \le \bar{t} \le T-2$. Particularly, $\|\boldsymbol{\epsilon_{2t}}\|_\infty + \|\boldsymbol{\epsilon_{2t}}\|_{-\infty} = 0$ and $\|\boldsymbol{\nu_{2t}}\|_\infty + \|\boldsymbol{\nu_{2t}}\|_{-\infty} = 0$ on a given finite support.

Assumption 3. $r(\boldsymbol{X_1}) = 3 + p$, $r(\boldsymbol{X_2}) = 3 + p$, $r(\boldsymbol{X_3}) = 2 + p$ and $r(\boldsymbol{X_4}) = 1 + p$, where $\boldsymbol{X_1} = (w(\boldsymbol{D_t}), s(\boldsymbol{w_t}), D(\boldsymbol{x_t}), \boldsymbol{x_t})_{T\times(3+p)}$ for DGP1, $\boldsymbol{X_2} = (\boldsymbol{w_t}, s(\boldsymbol{w_t}), D(\boldsymbol{x_t}), \boldsymbol{x_t})_{T\times(3+p)}$ for DGP2, $\boldsymbol{X_3} = (\boldsymbol{w_t}, s(\boldsymbol{w_t}), \boldsymbol{x_t})_{T\times(2+p)}$ for DGP3 and $\boldsymbol{X_4} = (D(\boldsymbol{x_t}), \boldsymbol{x_t})_{T\times(1+p)}$ for DGP4.

Assumption 4. If we let $\boldsymbol{\mathcal{L}_{,n}}$ denote the n's column of the matrix $\boldsymbol{\mathcal{L}_{T\times N}}$, then there is a pseudo-subspace spanned by $\boldsymbol{\mathcal{L}_{,n}}$: $\mathscr{k}_1\boldsymbol{\mathcal{L}_{,1}} + \mathscr{k}_2\boldsymbol{\mathcal{L}_{,2}} + \dots + \mathscr{k}_N\boldsymbol{\mathcal{L}_{,N}}$ for all nonzero numbers $\mathscr{k}_n \in \mathbb{R}\backslash\{0\}$, we denote this pseudo-subspace as $\mathcal{M}^-(\boldsymbol{\mathcal{L}})$.[1]

Assumption 1 is the CIA condition widely used in causal inference framework, we also

[1] Our notation of the pseudo-subspace is defined in accordance with generalized inverse.

require that the structure change s_t is independent of outcomes y_t given the latent variable w_t such that we are able to identify the structural change effects $\beta\eta$ as well as the endogenous regime switch effects η, a stronger assumption is $\mathbb{E}(\boldsymbol{y_t w_t}) = \mathbb{E}(\boldsymbol{y_t})\mathbb{E}(\boldsymbol{w_t})$ but is not required here. Note that $\boldsymbol{y_t} \perp s(\boldsymbol{w_t})|\boldsymbol{w_t}$ is no longer satisfied for the GDP 1, but $\boldsymbol{y_t} \perp s(\boldsymbol{w_t})|(\boldsymbol{w_t}, \boldsymbol{D_t})$ is satisfied therein (Dawid, 1979). Assumption 3 requires that all the models (24-27) are estimable, and we rule out the situation that $t_{s0} = t_{D0}$, we do not allow treatment and structure change take place at a same time point.[1] The difference between the usual subspace and pseudo-subspace in Assumption 4 is that we do not allow $\mathcal{k}_n = 0$, so one can see from the adjacency matrices that $\mathcal{M}^-(\boldsymbol{\mathcal{B}}) = \mathcal{M}^-(\boldsymbol{\mathcal{C}} + \boldsymbol{\mathcal{D}})$ while $\mathcal{M}^-(\boldsymbol{\mathcal{A}}) \neq \mathcal{M}^-(\boldsymbol{\mathcal{C}} + \boldsymbol{\mathcal{D}})$. The definition of pseudo-subspace will tell us under what conditions the structure change effect can be disentangled from treatment effect, which is shown in the following proposition,

Proposition 1. If the Assumptions 1, 3 and 4 are satisfied, then the structure change effect $\beta\eta$, endogenous regime switch effect η and treatment effect ξ can be separately identified from each other under DGP2 with adjacency matrix $\boldsymbol{\mathcal{B}}$ but not under DGP1 with adjacency matrix $\boldsymbol{\mathcal{A}}$.

Remark 1. Under the Assumptions 1, 3 and 4, structure change effects $\beta\eta$, endogenous regime switch effect η and treatment effect ξ can be identified separately and distinguished from each other as long as $\mathbf{w_t} \perp \boldsymbol{D_t}$ and $t_{s0} \neq t_{D0}$.

Remark 2. A less weak condition for identification is $\mathbf{w_t} \perp \mathbf{D_t}|\mathbf{x_t}$.

The intuition behind Proposition 1 is straightforward, if the structure change is caused by treatment as shown in DGP1 with adjacency matrix $\boldsymbol{\mathcal{A}}$, structural change effect, endogenous switch effect and treatment effect will then mix with each other, making it's impossible to distinguish between these three effects. For example, "structural changes of the economy struck by a natural disaster may occur due to the initial destructions and disruptions caused by the event and to the recovery and reconstruction activities, where the structural changes will result in new human capital accumulation and technology replacement" (Horwich, 2000; Noy, 2009; Okuyama, 2015; etc.). At this point, it's difficult to disentangle structure change effect $\beta\eta$ from treatment effect ξ, and endogenous regime switch effect η from treatment effect ξ, but we can get a total mixing effect $\beta\eta + \xi$. Our Proposition implies that empirical researchers should take the differences between structure change effect $\beta\eta$, endogenous switching effect η and treatment effect ξ seriously, one should be clear which effect to be identified and estimated in their empirical settings. If they neglect this, it would be easy to make *false positive* or *false negative* mistake.

3.2. Who is who, difference in difference in difference

As shown in Proposition 1, model (22) is unidentifiable, hence the specification we are interested in this paper is

$$\boldsymbol{y_t} = (\boldsymbol{w_t} + \boldsymbol{s_t}\beta)\eta + \boldsymbol{D_t}\xi + \boldsymbol{x_t}\alpha + \boldsymbol{\epsilon_t} + \boldsymbol{v_t}, \boldsymbol{w_t} \perp \boldsymbol{D_t}, t = 1, \dots, t_{s0}, \dots, t_{D0}, \dots, T. \quad (28)$$

Note that model (28) is a parameterization of the semi-parametric model (25), we require model (28) is correctly specified in empirical studies and the Assumptions 1-3 are also suitable for this model. The questions we are interested in now are the identifications of the structure change

[1] In fact if $t_{s0} = t_{D0}$, structure change and treatment in $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ would be perfectly collinear, we cannot distinguish treatment from structure change.

effect $\beta\eta$, endogenous switch effect η and treatment effect ξ. The problem of identifications in this case can be seen as a pseudo causal inference problem on networks, herein the structure change $\boldsymbol{s_t}$ and treatment $\boldsymbol{D_t}$ can be seen as two nodes on the network but we do not allow them interference with each other.[1] Following the notations of Hudgens & Halloran (2008) and Forastiere et al. (2021), we let $\mathbf{y}_t^{(s,D)}$ denote the outcome under different realizations of $\boldsymbol{s}, \boldsymbol{D} \in \{1,0\}$, so we observe $\mathbf{y}_t^{(0,0)}$ for $1 \le \boldsymbol{t} < t_{s0}$, $\mathbf{y}_t^{(1,0)}$ for $t_{s0} \le \boldsymbol{t} < t_{D0}$ and $\mathbf{y}_t^{(1,1)}$ for $t_{D0} \le \boldsymbol{t} < T$. The outcome can be then rewritten as

$$\mathbf{y}_t = \mathbf{y}_t^{(0,0)} \cdot \mathbb{I}\{1 \le \boldsymbol{t} < t_{s0}\} + \mathbf{y}_t^{(1,0)} \cdot \mathbb{I}\{t_{s0} \le \boldsymbol{t} < t_{D0}\} + \mathbf{y}_t^{(1,1)} \cdot \mathbb{I}\{t_{D0} \le \boldsymbol{t} < T\} \qquad (29)$$

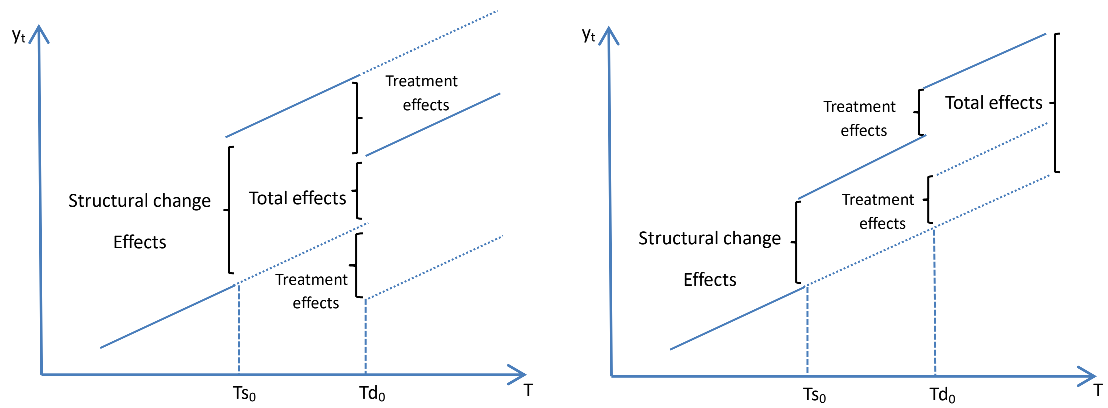
for any $t \in \{1, \dots, T\}$. $\mathbf{y}_t^{(0,0)}$ for $t_{s0} \le \boldsymbol{t} < t_{D0}$ then denotes the counterfactual outcome if there is no structure change; $\mathbf{y}_t^{(0,0)}$ for $t_{D0} \le \boldsymbol{t} < t_T$ denotes the counterfactual outcome if there is no structure change and treatment; $\mathbf{y}_t^{(0,1)}$ for $t_{D0} \le \boldsymbol{t} < t_T$ denotes the counterfactual outcome if there is no structure change but treatment (Splawa-Neyman et al., 1923; Rubin, 1978; Holland, 1986). Under this scenario, we will show that $(\beta\eta, \eta, \xi)$ can be separately identified through a Difference in Difference (DID) analogous strategy under corresponding *Parallel Trend Assumption*.

Assumption 5. There exist $i \in \mathbb{Z} = \{1,2,3, \dots\}$ such that

$$\mathbb{E}\left(\Delta \boldsymbol{y}_t^{(0,0)} | \boldsymbol{s_t}, \boldsymbol{D_t}\right)_{1 \le t < t_{s0}} = \mathbb{E}\left(\Delta \boldsymbol{y}_t^{(0,0)} | \boldsymbol{s_t}, \boldsymbol{D_t}\right)_{t_{s0} \le t < t_{D0}} = \mathbb{E}\left(\Delta \boldsymbol{y}_t^{(0,0)} | \boldsymbol{s_t}, \boldsymbol{D_t}\right)_{t_{D0} \le t \le T},$$

where $\Delta \boldsymbol{y}_t^{(0,0)} = \boldsymbol{y}_t^{(0,0)} - \boldsymbol{y}_i^{(0,0)}$ and the expectation operator is taken over t.[2]

Assumption 5 is analogous to the *Parallel Trend Assumption* widely used in DID settings, it implies that all the break effects detected in the observed $\mathbf{y}_t$ can only be attributed to structure change $\boldsymbol{s_t}$, treatment $\boldsymbol{D_t}$ or both. If there is no structure change or treatment, then we would not detect any break effect for $\mathbf{y}_t^{(0,0)}$, $1 \le t < T$. The only difference between Assumption 5 as shown in Figure 2 and the traditional *Parallel Trend Assumption* is that our strategy is a before-after design while the traditional *Parallel Trend Assumption* is a potential counterfactual framework. So the new strategy compares the states before and after the changes while the traditional approach compares the states with and without the changes. The following theorem shows that under Assumption 5, these two approaches equivalent to each other in our settings.



[1] From this perspective, the weak condition implied by Proposition 1 can be seen as a SUTVA where we require $\boldsymbol{s_t}$ be independent of $\boldsymbol{w_t}$.

[2] Expectation operator is taken over t means that $\mathbb{E}(\boldsymbol{x})_{t_0 \le t \le t_1} = 1/(t_1 - t_0 + 1) \sum_{t=t_0}^{t_1} \boldsymbol{x_t}$ for $\mathrm{t} = \{1, \dots, t_0, \dots, t_1, \dots, T\}$.

Figure 3. A DID in DID identification strategy

Theorem 1. If the Assumptions 1-5 are satisfied, then we have

$$\beta\eta + \xi = \mathbb{E}\left(\boldsymbol{y}_{\boldsymbol{t}}^{(\mathbf{1,1})} | \boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0} \le t \le T} - \mathbb{E}\left(\boldsymbol{y}_{\boldsymbol{t}}^{(\mathbf{0,0})} | \boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{s0} \le t < t_{D0}}, \tag{30}$$

$$\xi = \mathbb{E}\left(\boldsymbol{y}_{\boldsymbol{t}}^{(\mathbf{1,1})} | \boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0} \le t \le T} - \mathbb{E}\left(\boldsymbol{y}_{\boldsymbol{t}}^{(\mathbf{1,0})} | \boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{s0} \le t < t_{D0}}, \tag{31}$$

$$\beta\eta = \mathbb{E}\left(\boldsymbol{y}_{\boldsymbol{t}}^{(\mathbf{1,0})} | \boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{s0} \le t < t_{D0}} - \mathbb{E}\left(\boldsymbol{y}_{\boldsymbol{t}}^{(\mathbf{0,0})} | \boldsymbol{s}_t, \boldsymbol{D}_t\right)_{1 \le t < t_{s0}}, \tag{32}$$

where the expectation operator is taken over t.

Interestingly, Theorem 1 indicates that under Assumption 5, Rubin's counterfactual inference framework is equivalent to a before-after event study design in our settings (28), where $\boldsymbol{y}_{\boldsymbol{t}}^{(\mathbf{0,0})}$ in (30), $\boldsymbol{y}_{\boldsymbol{t}}^{(\mathbf{1,0})}$ in (31) and $\boldsymbol{y}_{\boldsymbol{t}}^{(\mathbf{0,0})}$ in (32) are potential outcomes for $t_{s0} \le t < t_{D0}$, $t_{s0} \le t < t_{D0}$ and $1 \le t < t_{s0}$ respectively. In this perspective, the structure change effect $\beta\eta$, treatment effect ξ and total effect $\beta\eta + \xi$ can be separately identified and consistently estimated as long as the counterfactuals $\boldsymbol{y}_{\boldsymbol{t}}^{(\mathbf{0,0})}$ for $t_{s0} \le t < t_{D0}$, $\boldsymbol{y}_{\boldsymbol{t}}^{(\mathbf{1,0})}$ for $t_{s0} \le t < t_{D0}$ and $\boldsymbol{y}_{\boldsymbol{t}}^{(\mathbf{0,0})}$ for $1 \le t < t_{s0}$ can be consistently estimated.

4. A flexible estimation approach

In real empirical studies, the method of IVs may fail here to estimate the parameters in (28) for suitable IVs are difficult to find; the method of DID, propensity scores matching, synthetic controls among others may also fail here for the latent variable is unobservable and suitable control units are not always available.[1] In this regard, we propose a new method and a 3-steps estimation procedure to estimate the parameters $(\beta, \eta, \xi, \alpha, \rho, \sigma_{v+\epsilon}, w_1, \dots, w_T) \subset \mathbb{R}^{5+P+T}$, where the latent variable $\boldsymbol{w} = (w_1, \dots, w_T)'$ could be treated as unknown parameters w.l.o.g.

4.1. A three-step estimation approach

Step 1: estimation of $\beta\eta$ through Automatically Auxiliary Regressions.

In the first step, we will show that the structure change effect $\beta\eta$ could be consistently estimated through a new proposed called Automatically Auxiliary Regressions (hereafter AARs). Recall that the social-economic outcome and structure change in (28), $\boldsymbol{y} = (y_{t=1}, y_{t=2}, \dots, y_{t=t_{D0}-1})'$ and $\boldsymbol{s} = (s_{t=1}, s_{t=2}, \dots, s_{t=t_{D0}-1})'$, could be treated as functions of time: $\mathbf{y}(\boldsymbol{t})$ and $\mathbf{s}(\boldsymbol{t})$. Hence by the Wasserstein approximation theorem, we consider polynomial approximations of $\mathbf{y}(\boldsymbol{t})$ and $\mathbf{s}(\boldsymbol{t})$ of order q with respect to t

$$\begin{cases} \boldsymbol{y}^{s=1}(\boldsymbol{t}) = \frac{1}{q} a_{y,q}^{s=1} \boldsymbol{t}^{\boldsymbol{q}} + \dots + \frac{1}{2} a_{y,2}^{s=1} \boldsymbol{t}^{\mathbf{2}} + a_{y,1}^{s=1} \boldsymbol{t} + a_{y,0}^{s=1} + \boldsymbol{\varepsilon}_{\boldsymbol{y}}^{s=1} \equiv \beta_{y,q}^{s=1} \boldsymbol{t}^{\boldsymbol{q}} + \dots + \beta_{y,2}^{s=1} \boldsymbol{t}^{\mathbf{2}} + \beta_{y,1}^{s=1} \boldsymbol{t} + \beta_{y,0}^{s=1} + \boldsymbol{\varepsilon}_{\boldsymbol{y}}^{s=1}(\boldsymbol{t}) \\ \boldsymbol{s}^{s=1}(\boldsymbol{t}) = \frac{1}{q} a_{s,q}^{s=1} \boldsymbol{t}^{\boldsymbol{q}} + \dots + \frac{1}{2} a_{s,2}^{s=1} \boldsymbol{t}^{\mathbf{2}} + a_{s,1}^{s=1} \boldsymbol{t} + a_{s,0}^{s=1} + \boldsymbol{\varepsilon}_{\boldsymbol{s}}^{s=1} \equiv \beta_{s,q}^{s=1} \boldsymbol{t}^{\boldsymbol{q}} + \dots + \beta_{s,2}^{s=1} \boldsymbol{t}^{\mathbf{2}} + \beta_{s,1}^{s=1} \boldsymbol{t} + \beta_{s,0}^{s=1} + \boldsymbol{\varepsilon}_{\boldsymbol{s}}^{s=1}(\boldsymbol{t}) \end{cases}, (3$$

[1] Except for the method of IVs, all causal inference methods in observation studies need to find control units of good qualities, which are scarce or even impossible in empirical studies such as evaluating the effects of some one-cuts-fit-all social policies.

where $\beta_{\ell,q}^{s=1} \equiv a_{\ell,q}^{s=1}/q$ captures the weight of the q's polynomial, $\beta_{\ell,0}^{s=1} \equiv a_{\ell,0}^{s=1}$ is the intercept, ε_ℓ follows some unknown distribution, $\ell = y, s$, $\boldsymbol{t} = 1, \ldots, t_{s0}, \ldots, t_{D0-1}$ and $q \leq t_{D0-1}$.[1] The superscript $s = 1$ denotes the first step, where $\mathbf{y^{s=1}(t)} = \mathbf{y(t)}$ in our first step. Taking derivatives of $\mathbf{y}(\boldsymbol{t})$, $\mathbf{s}(\boldsymbol{t})$ with respect to t, we get

$$\begin{cases} \mathbf{y}'_{\boldsymbol{s=1}}(\boldsymbol{t}) = q\beta_{y,q}^{s=1}\boldsymbol{t^{q-1}} + \cdots + 2\beta_{y,2}^{s=1}\boldsymbol{t} + \beta_{y,1}^{s=1} + \boldsymbol{\varepsilon_y^{s=1}}'(\boldsymbol{t}) \\ \mathbf{s}'_{\boldsymbol{s=1}}(\boldsymbol{t}) = q\beta_{s,q}^{s=1}\boldsymbol{t^{q-1}} + \cdots + 2\beta_{s,2}^{s=1}\boldsymbol{t} + \beta_{s,1}^{s=1} + \boldsymbol{\varepsilon_s^{s=1}}'(\boldsymbol{t}) \end{cases}. \tag{34}$$

The question we are interested in now is how we can model the dynamics between the economic outcome $\mathbf{y}(\boldsymbol{t})$ and the structure change $\mathbf{s}(\boldsymbol{t})$. Following the literature of symbolic computations (e.g. Alonso et al., 1997; Alonso et al., 2007; Gutierrez & Urroz, 2020), it is easy to verify that there exists a function $\mathcal{K}(\cdot) \in \mathcal{L}^2(\mathbb{R})$ such that the automatically auxiliary dynamics between $\mathbf{y}(\boldsymbol{t})$ and $\mathbf{s}(\boldsymbol{t})$ is satisfied with

$$\mathcal{K}\left(\mathbf{y}'_{\boldsymbol{s=1}}(\boldsymbol{t}), \mathbf{s}'_{\boldsymbol{s=1}}(\boldsymbol{t})\right) = a_{s,2}^{s=1}\mathbf{y}'_{\boldsymbol{s=1}}(\boldsymbol{t}) - a_{y,2}^{s=1}\mathbf{s}'(\boldsymbol{t}) + a_{y,2}^{s=1}a_{s,1}^{s=1} - a_{s,2}^{s=1}a_{y,1}^{s=1} = 0, \tag{35}$$

where the polynomials are chosen as $q = 2$.[2] Estimate equation (33) by OLS to get the estimators $\hat{a}_{s,2}^{s=1}$, $\hat{a}_{s,1}^{s=1}$, $\hat{a}_{y,2}^{s=1}$ and $\hat{a}_{y,1}^{s=1}$,[3] input all the estimators into (35) we then get

$$\hat{\mathbf{y}}'_{\boldsymbol{s=1}}(\boldsymbol{t}) = \frac{\hat{a}_{y,2}^{s=1}\mathbf{s}'_{\boldsymbol{s=1}}(\boldsymbol{t}) + \hat{a}_{s,2}^{s=1}\hat{a}_{y,1}^{s=1} - \hat{a}_{y,2}^{s=1}\hat{a}_{s,1}^{s=1}}{\hat{a}_{s,2}^{s=1}}. \tag{36}$$

Further consider the following auxiliary regression

$$\mathbf{y}_{\boldsymbol{t}}^{\boldsymbol{s=1}} = \delta_0^{s=1} + \delta_1^{s=1}\hat{\mathbf{y}}'_{\boldsymbol{s=1}}(\boldsymbol{t}) + \boldsymbol{\varpi}_{\boldsymbol{t}}^{\boldsymbol{s=1}}, t = 1, \ldots, t_{s0}, \ldots, t_{D0-1}, \tag{37}$$

where we impose no extra restrictions on $\boldsymbol{\varpi}_{\boldsymbol{t}}^{\boldsymbol{s=1}}$. Run OLS to estimate (37) again, we will then get $\hat{\mathbf{y}}_{\boldsymbol{t}}^{\boldsymbol{s=1}} = \hat{\delta}_0^{s=1} + \hat{\delta}_1^{s=1}\hat{\mathbf{y}}'_{\boldsymbol{s=1}}(t)$.

Under the following assumptions, we will prove that there exists a consistent estimator for the structure change effect defined in (32) through the AARs procedure (33-37):

Assumption 6. $\boldsymbol{\varepsilon_{y,t}^{s=1}}$ is i.i.d. with $\boldsymbol{\varepsilon_{y,t}^{s=1}} =_{i.i.d.} \mathfrak{S}_{s=1}\left(0, \sigma_{\boldsymbol{\varepsilon,s=1}}^2\right)$, where $\mathfrak{S}$ is a Lebesgue measurable distribution and $\sigma_{\boldsymbol{\varepsilon,s=1}} < \infty$, $t = 1, \ldots, t_{s0}, \ldots, t_{D0-1}$.

Assumption 7. $\eta\boldsymbol{\Delta_{s=1}}(\boldsymbol{w_t}) + \sum_p \alpha_p\boldsymbol{\Delta_{s=1}}\left(\boldsymbol{x_{t,p}}\right) = 0$, where

$$\boldsymbol{\Delta_{s=1}}(\boldsymbol{w_t}) \equiv \mathbb{E}\left(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta}\odot\boldsymbol{w_t}\right)_{s=1, t_{s0}\leq t\leq t_{D0-1}} - \mathbb{E}\left(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta}\odot\boldsymbol{w_t}\right)_{s=1, 1\leq t< t_{s0}},$$

$$\boldsymbol{\Delta_{s=1}}\left(\boldsymbol{x_{t,p}}\right) \equiv \mathbb{E}\left(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta}\odot\boldsymbol{x_{t,p}}\right)_{s=1, t_{s0}\leq t\leq t_{D0-1}} - \mathbb{E}\left(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta}\odot\boldsymbol{x_{t,p}}\right)_{s=1, 1\leq t< t_{s0}},$$

for $p = 1, \ldots, P$, $\odot$ denotes the Hadamard product, and

$$\boldsymbol{\delta_T} \equiv \begin{pmatrix} 2\cdot 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 2\cdot t_{D0-1} & 1 & 0 \end{pmatrix}_{t_{D0-1}\times 3}, \qquad \boldsymbol{\delta_\beta} \equiv \begin{pmatrix} \beta_{y2} \\ \beta_{y1} \\ \beta_{y0} \end{pmatrix}.$$

Lemma 1. Suppose that there exist constants $0 < c_w, c_{x_p} < min(1/(t_{s0} - 1), 1/(t_{D0} - t_{s0}))$ such that $\boldsymbol{\delta_t}\boldsymbol{\delta_\beta}\odot\boldsymbol{w_t} = O((t_{D0} - t_{s0})^{c_w})$, $\boldsymbol{\delta_t}\boldsymbol{\delta_\beta}\odot\boldsymbol{x_{t,p}} = O((t_{D0} - t_{s0})^{c_{x_p}})$ for $t_{s0} \leq t \leq t_{D0-1}$ and $\boldsymbol{\delta_t}\boldsymbol{\delta_\beta}\odot\boldsymbol{w_t} = O((t_{s0} - 1)^{c_w})$, $\boldsymbol{\delta_t}\boldsymbol{\delta_\beta}\odot\boldsymbol{x_{t,p}} = O((t_{s0} - 1)^{c_{x_p}})$ for $1 \leq t \leq t_{s0-1}$, $p = 1, \ldots, P$. Then as $t_{s0} \to \infty$ and $t_{D0} \to \infty$, we have $\left[(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta})'\boldsymbol{\delta_T}\boldsymbol{\delta_\beta}\right]^{-1}(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta})'\boldsymbol{\ell_t}\mathbb{E}\left(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta}\right)_{s=1,t} =$

[1] "Automatically" means that without any model specifications, we can approximate the structure change and treatment through a q-order polynomial series as shown in (33), then the dynamics of the change rates between these two variables can be shown in a continuous function shown in (35); "Auxiliary Regression" means that we can model the observed social-economic outcome and its change rate through a reduced-form model shown in (37).

[2] The choice of q will not influence the consistency of our estimator.

[3] Any other estimation methods can also be adopted here, not limited to OLS. We consider OLS for convenience.

$o_p(1)$ for $t_{s0} \leq t \leq t_{D0-1}$ and $\left[(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right]^{-1}(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\ell_t}\mathbb{E}(\boldsymbol{\delta_T\delta_\beta})_{s=1,t} = o_p(1)$ for $1 \leq t \leq t_{s0-1}$, $\boldsymbol{\ell_t} \in \{\boldsymbol{w_t}, \boldsymbol{x_{t,1}}, \dots, \boldsymbol{x_{t,P}}\}$, the expectation operator is taken over t.

Theorem 2. Under the identification condition (32), as $t_{s0} \to \infty$ and $t_{D0} \to \infty$ we have

$$\widehat{\beta\eta}_{AARs} = \mathbb{E}(\hat{\boldsymbol{y}}_{\boldsymbol{t}}^{\boldsymbol{s=1}})_{t_{s0} \leq t \leq t_{D0-1}} - \mathbb{E}(\hat{\boldsymbol{y}}_{\boldsymbol{t}}^{\boldsymbol{s=1}})_{1 \leq t \leq t_{s0-1}} \longrightarrow_p \beta\eta$$

for model (28), the expectation operator is taken over t. If the Assumptions 2, 5, 6 and 7 hold true, we then have $\mathbb{E}(\widehat{\beta\eta}_{AARs}) = \beta\eta$.

Remark 3. The idea of AARs is a bit like Indirect Inference methods in structural econometrics, where both methods draw the idea of auxiliary regressions (Li, 2010). As we see here we do not need IVs or control groups to consistently estimate the structure change effect even if there exist latent variable $\boldsymbol{w_t}$ and endogenous structure change $\boldsymbol{s_t}$. The most critical condition for AARs to hold in small samples is Assumption 7, which requires that there are no structure changes or treatment effects in the latent variable $\boldsymbol{w_t}$ and the covariates $\boldsymbol{x_t}$, hence the detected effect $\widehat{\beta\eta}_{AARs}$ can only be due to the structure change $\boldsymbol{s_t}$ instead of other variations. Note that Assumption 7 is easily satisfied if $\boldsymbol{\delta_T\delta_\beta w_t}$ and $\boldsymbol{\delta_T\delta_\beta x_{t,p}}$ are martingale difference sequences (MDS), or $\boldsymbol{w_t}$ and $\boldsymbol{x_{tp}}$ are stationary for $t = 1, \dots, t_{s0}, \dots, t_{D0-1}$. If Assumption 7 is not satisfied in empirical studies, the AARs' estimator will be biased in small samples. However the good news is that AARs' estimator is still consistent in large samples. In summary, AARs' estimator will be unbiased and consistent if Assumption 7 is satisfied, otherwise it will be biased although consistent.[1] We will give the empirical researchers some advices on testing Assumption 7 in the following step 2.

Step 2: estimations of $\beta\eta + \xi$ and ξ through AARs.

As shown in Theorem 1, the treatment effect ξ could be consistently estimated as long as the structural change effect $\beta\eta$ and the total effect $\beta\eta + \xi$ could be consistently estimated. Consider the polynomials approximations similar to (34)

$$\begin{cases} \mathbf{y}'_{\boldsymbol{s=2}}(\boldsymbol{t}) = q\beta_{y,q}^{s=2}\boldsymbol{t^{q-1}} + \cdots + 2\beta_{y,2}^{s=2}\boldsymbol{t} + \beta_{y,1}^{s=2} + \boldsymbol{\varepsilon}_{\boldsymbol{y}}^{\boldsymbol{s=2}\prime}(\boldsymbol{t}) \\ \mathbf{D}'_{\boldsymbol{s=2}}(\boldsymbol{t}) = q\beta_{\boldsymbol{D},q}^{s=2}\boldsymbol{t^{q-1}} + \cdots + 2\beta_{\boldsymbol{D},2}^{s=2}\boldsymbol{t} + \beta_{\boldsymbol{D},1}^{s=2} + \boldsymbol{\varepsilon}_{\boldsymbol{D}}^{\boldsymbol{s=2}\prime}(\boldsymbol{t}) \end{cases}. \tag{38}$$

where $q = 2$, $\beta_{\ell,q}^{s=2} \equiv a_{\ell,q}^{s=2}/q$ and $\beta_{\ell,0}^{s=2} \equiv a_{\ell,0}^{s=2}$, note that

$$\hat{\mathbf{y}}^{\boldsymbol{s=2}}(\boldsymbol{t}) = \mathbf{y}(\boldsymbol{t}) \cdot \mathbb{I}\{1 \leq \boldsymbol{t} < t_{s0}\} + (\mathbf{y}(\boldsymbol{t}) - \widehat{\beta\eta}_{\mathrm{AARs}}) \cdot \mathbb{I}\{t_{s0} \leq \boldsymbol{t} < t_{D0}\} + \mathbf{y}(\boldsymbol{t}) \cdot \mathbb{I}\{t_{D0} \leq \boldsymbol{t} < T\}.$$

Obviously, we have

$$\hat{\mathbf{y}}^{\boldsymbol{s=2}}(\boldsymbol{t}) \longrightarrow_{\mathrm{p}} \mathbf{y}_{\boldsymbol{t}}^{(\mathbf{0,0})} \cdot \mathbb{I}\{1 \leq \boldsymbol{t} < t_{s0}\} + \mathbf{y}_{\boldsymbol{t}}^{(\mathbf{0,0})} \cdot \mathbb{I}\{t_{s0} \leq \boldsymbol{t} < t_{D0}\} + \mathbf{y}_{\boldsymbol{t}}^{(\mathbf{1,1})} \cdot \mathbb{I}\{t_{D0} \leq \boldsymbol{t} < T\}$$

in the light of Theorem 2, where $\mathbf{y}_{\boldsymbol{t}}^{(\mathbf{0,0})}$ denote the counterfactual outcomes without structure change for $1 \leq t < t_{D0}$. Consider the following auxiliary regression

$$\mathbf{y}_{\boldsymbol{t}}^{\boldsymbol{s=2}} = \delta_0^{s=2} + \delta_1^{s=2}\hat{\mathbf{y}}'_{\boldsymbol{s=2}}(\boldsymbol{t}) + \boldsymbol{\varpi}_{\boldsymbol{t}}^{\boldsymbol{s=2}}, \tag{39}$$

where we impose no extra restrictions on $\boldsymbol{\varpi}_{\boldsymbol{t}}^{\boldsymbol{s=2}}$. Similar to (35-36), we can get

$$\hat{\mathbf{y}}'_{\boldsymbol{s=2}}(\boldsymbol{t}) = \frac{\hat{a}_{y,2}^{s=2}\mathbf{D}'_{\boldsymbol{s=2}}(\boldsymbol{t}) + \hat{a}_{\boldsymbol{D},2}^{s=2}\hat{a}_{y,1}^{s=2} - \hat{a}_{y,2}^{s=2}\hat{a}_{\boldsymbol{D},1}^{s=2}}{\hat{a}_{\boldsymbol{D},2}^{s=2}} \tag{40}$$

for $t = 1, \dots, t_{s0}, \dots, t_{D0-1}, \dots, T$. Regress (39) by OLS and get $\hat{\mathbf{y}}_{\boldsymbol{t}}^{\boldsymbol{s=2}} = \hat{\delta}_0^{s=2} + \hat{\delta}_1^{s=2}\hat{\mathbf{y}}'_{\boldsymbol{s=2}}(\boldsymbol{t})$.

[1] Assumption 7 guarantees unbiasness while Lemma 1 implies consistency.

Assumption 8. $\eta\boldsymbol{\Delta}_{s=2}(\boldsymbol{w_t}) + \sum_p \alpha_p \boldsymbol{\Delta}_{s=2}(\boldsymbol{x_{t,p}}) = 0$, where

$$\boldsymbol{\Delta}_{s=2}(\boldsymbol{w_t}) \equiv \mathbb{E}(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta}\odot\boldsymbol{w_t})_{s=2,t_{D0}\le t\le T} - \mathbb{E}(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta}\odot\boldsymbol{w_t})_{s=2,1\le t<t_{D0}},$$

$$\boldsymbol{\Delta}_{s=2}(\boldsymbol{x_{t,p}}) \equiv \mathbb{E}(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta}\odot\boldsymbol{x_{t,p}})_{s=2,t_{D0}\le t\le T} - \mathbb{E}(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta}\odot\boldsymbol{x_{t,p}})_{s=2,1\le t<t_{D0}},$$

for $p = 1, \dots, P$, and

$$\boldsymbol{\delta_T} \equiv \begin{pmatrix} 2\cdot 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 2\cdot T & 1 & 0 \end{pmatrix}_{T\times 3}, \qquad \boldsymbol{\delta_\beta} \equiv \begin{pmatrix} \beta_{y2} \\ \beta_{y1} \\ \beta_{y0} \end{pmatrix},$$

$\boldsymbol{\delta_\beta}$ is defined in (38).

Assumption 9. $\boldsymbol{\varepsilon_{y,t}^{s=2}}$ is i.i.d. with $\boldsymbol{\varepsilon_{y,t}^{s=2}} =_{i.i.d.} \mathfrak{S}_{s=2}(0, \sigma^2_{\boldsymbol{\varepsilon,s=2}})$, where $\mathfrak{S}$ is a Lebesgue measurable distribution and $\sigma_{\boldsymbol{\varepsilon.s=2}} < \infty$, $t = 1, \dots, t_{s0}, \dots, t_{D0-1}$.

Lemma 2. If there exist constants $0 < \tilde{c}_w, \tilde{c}_{x_p} < min(1/(t_{D0}-1), 1/(T-t_{D0}+1))$ such that $\boldsymbol{\delta_t}\boldsymbol{\delta_\beta}\odot\boldsymbol{w_t} = O((T-t_{D0}+1)^{\tilde{c}_w})$, $\boldsymbol{\delta_t}\boldsymbol{\delta_\beta}\odot\boldsymbol{x_{t,p}} = O((T-t_{D0}+1)^{\tilde{c}_{xp}})$ for $t_{D0} \le t \le T$ and $\boldsymbol{\delta_t}\boldsymbol{\delta_\beta}\odot\boldsymbol{w_t} = O((t_{D0}-1)^{\tilde{c}_w})$, $\boldsymbol{\delta_t}\boldsymbol{\delta_\beta}\odot\boldsymbol{x_{t,p}} = O((t_{D0}-1)^{\tilde{c}_{xp}})$ for $1 \le t \le t_{D0-1}$, $p = 1, \dots, P$. Then as $t_{D0} \to \infty$ and $T \to \infty$, we have $[(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta})'\boldsymbol{\delta_T}\boldsymbol{\delta_\beta}]^{-1}(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta})'\boldsymbol{\ell_t}\mathbb{E}(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta})_{s=2,t} = o_p(1)$ for $t_{D0} \le t \le T$ and $[(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta})'\boldsymbol{\delta_T}\boldsymbol{\delta_\beta}]^{-1}(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta})'\boldsymbol{\ell_t}\mathbb{E}(\boldsymbol{\delta_T}\boldsymbol{\delta_\beta})_{s=2,t} = o_p(1)$ for $1 \le t \le t_{D0-1}$, $\boldsymbol{\ell_t} \in \{\boldsymbol{w_t}, \boldsymbol{x_{t,1}}, \dots, \boldsymbol{x_{t,P}}\}$, the expectation operator is taken over t.

Corollary 1. Under the identification conditions (30-31), as $t_{D0} \to \infty$ and $T \to \infty$ we have

$$\widehat{\beta\eta+\xi}_{AARs} = \mathbb{E}(\boldsymbol{\hat{y}_t^{s=2}})_{t_{D0}\le t\le T} - \mathbb{E}(\boldsymbol{\hat{y}_t^{s=2}})_{1\le t\le t_{D0-1}} \longrightarrow_p \beta\eta + \xi,$$

and

$$\hat{\xi}_{AARs} = \widehat{\beta\eta+\xi}_{AARs} - \widehat{\beta\eta}_{AARs} \longrightarrow_p \xi$$

for model (28), the expectation operator is taken over t. If the Assumptions 1-9 hold true, we then have $\mathbb{E}(\widehat{\beta\eta+\xi}_{AARs}) = \beta\eta + \xi$ and $\mathbb{E}(\hat{\xi}_{AARs}) = \xi$.

As shown in Theorem 2 and Corollary 1, the most critical conditions for the unbiasness are Assumptions 7-8, which require that there are no structure and treatment breaks in the latent variable $\boldsymbol{w_t}$ and other covariates $\boldsymbol{x_{t,p}}$.[1] Hence the first step to adopt AARs in small samples is to check whether these conditions are satisfied, we provide a rule of thumb to test these Assumptions. Note that from (28), we have $\boldsymbol{u_{t,s=1}} \equiv \boldsymbol{w_t}\eta + \boldsymbol{x_t}\alpha + \boldsymbol{\epsilon_t} + \boldsymbol{\nu_t} = \mathbf{y_t} - \boldsymbol{s_t}\beta\eta$ for the time series $t = 1, \dots, t_{s0}, \dots, t_{D0-1}$, from which we can get $\boldsymbol{\hat{u}_{t,s=1}} = \mathbf{y_t} - \boldsymbol{s_t}\widehat{\beta\eta}_{AARs}$. Hence, under Assumption 2, testing Assumption 7 turns into testing $\mathcal{H}_0\colon \mathbb{E}(\boldsymbol{\hat{u}_t})_{s=1,1\le t<t_{s0}} - \mathbb{E}(\boldsymbol{\hat{u}_t})_{s=1,t_{s0}\le t\le t_{D0-1}} \neq 0$, $\mathcal{H}_1\colon \mathbb{E}(\boldsymbol{\hat{u}_t})_{s=1,1\le t<t_{s0}} - \mathbb{E}(\boldsymbol{\hat{u}_t})_{s=1,t_{s0}\le t\le t_{D0-1}} = 0$. Similarly, we can get $\boldsymbol{\hat{u}_{t,s=2}} = \mathbf{y_t} - \boldsymbol{s_t}\widehat{\beta\eta}_{AARs} - \boldsymbol{D_t}\hat{\xi}_{AARs}$ where $\boldsymbol{u_{t,s=2}} \equiv \boldsymbol{w_t}\eta + \boldsymbol{x_t}\alpha + \boldsymbol{\epsilon_t} + \boldsymbol{\nu_t} = \mathbf{y_t} - \boldsymbol{s_t}\beta\eta - \boldsymbol{D_t}\xi$ for $t = 1, \dots, T$. Hence testing Assumption 8 turns into testing

[1] Similar to Regression Discontinuity Design (RD), these assumptions turn into the requirement of smoothness and continuousness of $\boldsymbol{w_t}$ and $\boldsymbol{x_{t,p}}$ near t_{s0} and t_{D0} when estimating local structure change effect and local treatment effect.

$\mathcal{H}_0$: $\mathbb{E}(\widehat{\boldsymbol{u}}_t)_{s=2,1\le t<t_{D0}} - \mathbb{E}(\widehat{\boldsymbol{u}}_t)_{s=2,t_{D0}\le t\le T} \neq 0$, $\mathcal{H}_1$: $\mathbb{E}(\widehat{\boldsymbol{u}}_t)_{s=2,1\le t<t_{D0}} - \mathbb{E}(\widehat{\boldsymbol{u}}_t)_{s=2,t_{D0}\le t\le T} = 0$. If $\mathcal{H}_0$ are rejected by inference, then AARs could be adopted.

Step 3: estimations of α, η, β and $\mathbb{E}(\boldsymbol{w_t})$.

From Theorem 2 and Corollary 1 , we have

$$\mathbf{y_t} - \boldsymbol{s_t}\widehat{\beta\eta}_{AARs} - \boldsymbol{D_t}\hat{\hat{\xi}}_{AARs} = \boldsymbol{w_t}\eta + \boldsymbol{x_t}\alpha + \boldsymbol{\epsilon_t} + \boldsymbol{\nu_t} + o_p(1). \tag{41}$$

It is easy to verify that, as $T \to \infty$, $\hat{\alpha}_{AARs} = (\boldsymbol{x_t}'\boldsymbol{x_t})^{-1}\boldsymbol{x_t}'(\mathbf{y_t} - \boldsymbol{s_t}\widehat{\beta\eta}_{AARs} - \boldsymbol{D_t}\hat{\hat{\xi}}_{AARs}) \longrightarrow_p \alpha$ under some standard regular conditions such as $\boldsymbol{x_t}'\boldsymbol{x_t}$ is of full rank implied by Assumption 3, and $\boldsymbol{w_t}$ is independent of $\boldsymbol{D_t}$ implied by model (28) and Proposition 1.

From (41), as $T \to \infty$ we then have

$$\mathbf{y_t} - \boldsymbol{s_t}\widehat{\beta\eta}_{AARs} - \boldsymbol{D_t}\hat{\hat{\xi}}_{AARs} - \boldsymbol{x_t}\hat{\alpha}_{AARs} = \boldsymbol{w_t}\eta + \boldsymbol{\epsilon_t} + \boldsymbol{\nu_t} + o_p(1), \tag{42}$$

the endogenous switch effect can be then estimated,

$$\hat{\eta}_{AARs} = \left(\frac{\|\boldsymbol{w_t}\|_{\infty} + \|\boldsymbol{w_t}\|_{-\infty}}{\left\|\mathbf{y_t} - \boldsymbol{s_t}\widehat{\beta\eta}_{AARs} - \boldsymbol{D_t}\hat{\hat{\xi}}_{AARs} - \boldsymbol{x_t}\hat{\alpha}_{AARs}\right\|_{\infty} + \left\|\mathbf{y_t} - \boldsymbol{s_t}\widehat{\beta\eta}_{AARs} - \boldsymbol{D_t}\hat{\hat{\xi}}_{AARs} - \boldsymbol{x_t}\hat{\alpha}_{AARs}\right\|_{-\infty}}\right)^{-1},$$

where $\|\boldsymbol{w_t}\|_{\infty}$ denotes the $\mathcal{L}_{\infty}$ norm while $\|\boldsymbol{w_t}\|_{-\infty}$ denotes the $\mathcal{L}_{-\infty}$ norm. Therefrom, under Assumptions 1-9 we can get $\hat{\eta}_{AARs} \longrightarrow_p \eta$, $\hat{\beta}_{AARs} = \widehat{\beta\eta}_{AARs}/\hat{\eta}_{AARs} \longrightarrow_p \beta$ and $\widehat{\mathbb{E}(\boldsymbol{w_t})}_{AARs} = 1/\hat{\eta}_{AARs}\,(\mathbf{y_t} - \boldsymbol{s_t}\widehat{\beta\eta}_{AARs} - \boldsymbol{D_t}\hat{\hat{\xi}}_{AARs} - \boldsymbol{x_t}\hat{\alpha}_{AARs}) \longrightarrow_p \mathbb{E}(\boldsymbol{w_t})$ as $T \to \infty$.

If $\|\boldsymbol{w_t}\|_{\infty}$ and $\|\boldsymbol{w_t}\|_{-\infty}$ are known priors, our estimation mission completes here. But the problem we are facing in most of the empirical studies is that we know nothing about the latent variable $\boldsymbol{w_t}$, hence $\|\boldsymbol{w_t}\|_{\infty}$ and $\|\boldsymbol{w_t}\|_{-\infty}$ should be estimated. The good news is that they can be estimated through social interaction models as shown in section 2.2.

Let $\mathbf{y_{j,t}}$ denote the outcomes for $\mathbf{y_t}$'s neighbors, $\mathcal{P} = \{1,2,\dots,n\}$ is the reference group of size n, then $j \in \mathcal{P}$ represents $\mathbf{y_t}$'s peers and $j \in \mathbb{Z}\backslash\mathcal{P}$ represents $\mathbf{y_t}$'s other neighbors except for the peers, $\mathbb{Z} = \{1,2,\dots\}$ represents total neighbors set and is of size N. We may denote $\mathbf{y_{i,t}} \equiv \mathbf{y_t}$ w.l.o.g. Recall the Near-epoch Dependency (NED) condition widely used in time-series/spatial literatures (see e.g. Jenish & Prucha, 2012),

Assumption 10. Assume there exist $d_1 < \infty$, $d_2 < \infty$ and $\mathcal{s} > 0$ such that

$$\inf_{1\le t\le T} \left|\boldsymbol{y_t} - \mathbb{E}(\boldsymbol{y_t}|\boldsymbol{\mathcal{F}}(\mathcal{s}))\right| \le d_1 \cdot \varphi(N), \tag{43}$$

$$\sup_{1\le t\le T} \left|\boldsymbol{y_t} - \mathbb{E}(\boldsymbol{y_t}|\boldsymbol{\mathcal{F}}(\mathcal{s}))\right| \le d_2 \cdot \varphi(N) \tag{44}$$

for $\varphi(N) \ge 0$ with $lim_{N\to\infty}\varphi(N) = 0$, and $\boldsymbol{\mathcal{F}}(\mathcal{s}) = \sigma\{\boldsymbol{y_{j,t}}: \varrho(i,j) < \mathcal{s}\}$ be the sigma-field generated by the random variables $\mathbf{y_{j,t}}$ located in the $\mathcal{s}$ - neighborhood of location $\boldsymbol{y_{i,t}}$. Usually $\varrho(i,j) = \min_{j\in\mathbb{Z}}|\boldsymbol{y_{i,t}} - \boldsymbol{y_{j,t}}|$, we may assume $\mathcal{s} = N^c$ for some $c > 0$. Then $\boldsymbol{y_{i,t}}$ is said to be uniformly $\mathcal{L}_1(d)$ –inf-NED and $\mathcal{L}_1(d)$ – sup-NED on $\boldsymbol{y_{j,t}}$ for (43-44) respectively over t.

When the distance between $\boldsymbol{y_{i,t}}$ and $\boldsymbol{y_{j,t}}$ tends to be infinite ($\mathcal{s} = \infty$), $\boldsymbol{y_{i,t}}$'s neighbor $\boldsymbol{y_{j,t}}$ will have no prediction power on $\boldsymbol{y_{i,t}}$, which implies that the network is sparse or asymptotically sparse (Graham & de Paula, 2020). Consequently, there are three methods to estimate $\|\boldsymbol{w_t}\|_{\infty}$ and $\|\boldsymbol{w_t}\|_{-\infty}$ under two different scenarios.

- Situation One: common latent variable.

Suppose the economies on the network share a same latent variable such that the economies' latent growths $\left(\mathbf{y}_{lp,i,t}, \mathbf{y}_{lp,j,t}\right)$ are all driven by $\boldsymbol{w_t}$, meanwhile suppose the structure change and treatment only occur in $\boldsymbol{y_{i,t}}$ and all neighbors $j \in \{1,2,\ldots\}$ are ordered from small to large by distance $\varrho(i,j)$.[1] Then the GDP for the neighbors can be set as

$$\boldsymbol{y_{j,t}} = \boldsymbol{w_t}\eta_j + \boldsymbol{x_{j,t}}\alpha_j + \boldsymbol{v_j}, \tag{45}$$

where $\boldsymbol{v_j} =_{i.i.d.} \mathfrak{F}_k\left(0, \sigma_j^2\right)$ for some Lebsgue measurable distribution $\mathfrak{F}_k$, $j = 1,2,\ldots,N$.

Assumption 11. (sparsity and unique social equilibrium) $n \ll N$ and $\frac{1}{N}\sum_{j=1}^{N}\eta_j = 1$, $\eta_j = \gamma_j$ for $j \in \mathbb{Z}$, η_j is defined in (45) and γ_j is defined in (48).

Assumption 11 requires that, on the one hand, the size of the reference group $\mathcal{P}$ is far smaller than that of the neighbors group $\mathbb{Z}$, which implies that the number of friends is limited on the network; on the other hand, the influential weights of the latent variable on neighbors' latent growths form a convex set $\{\eta_1, \ldots, \eta_N\}$, so there is a unique social equilibrium on the network. We can then get

$$\phi_{en}\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{y_{j,t}} = \phi_{en}\boldsymbol{w_t} + \phi_{en}\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{x_{j,t}}\alpha_j + \phi_{en}\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{v_j} \tag{46}$$

from (45) and Assumption 11, the coefficient $|\phi_{en}| < 1$ captures endogenous peer effect of $1/n\sum_{j\in\mathcal{P}}\boldsymbol{y_{j,t}}$ on $\boldsymbol{y_t}$ by the classical social interaction model (Graham & de Paula, 2020). As $N \to \infty$, by the Law of Large Numbers we can get

$$\inf_{1\le t\le T}\phi_{en}\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{y_{j,t}} = \inf_{1\le t\le T}\phi_{en}\boldsymbol{w_t} + o_p(1) \quad and \quad \sup_{1\le t\le T}\phi_{en}\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{y_{j,t}} = \sup_{1\le t\le T}\phi_{en}\boldsymbol{w_t} + o_p(1)$$

where $\phi_{en}\sum_{j\in\mathcal{P}}\alpha_j$ captures some sort of exogenous peer effect of $1/n\sum_{j\in\mathcal{P}}\boldsymbol{x_{j,t}}$ on $\boldsymbol{y_t}$, $(\phi_{en}/N)\alpha_N = o_p(1)$ and $1/N\sum_{j\in\mathbb{Z}}\phi_{en}\boldsymbol{x_{j,t}}\alpha_j = o_p(1)$ according to Assumption 10. The intuition behind this is straightforward, note that we have $\boldsymbol{y_t} \leftarrow_{\phi_{en}/N} \boldsymbol{y_{j,t}} \leftarrow_{\alpha_j} \boldsymbol{x_{j,t}}$ where $\leftarrow_{\phi_{en}/N}$ denotes $\boldsymbol{y_{j,t}}$'s impact on $\boldsymbol{y_t}$ is ϕ_{en}/N. Hence if $\phi_{en}/N = 0$, we will get $(\phi_{en}/N)\alpha_j = 0$. Under this scenario, $\|\boldsymbol{w_t}\|_\infty$ and $\|\boldsymbol{w_t}\|_{-\infty}$ could be estimated by the largest and smallest value of the neighbors' mean outcome over $t = 1,\ldots,T$ on the network.

Apart from this, consider $\boldsymbol{w_t}$ is omitted from (45): $\mathbf{y}_{j,t} = \boldsymbol{x_{j,t}}\alpha_j + \boldsymbol{\mathcal{v}_j}$, direct OLS estimation will lead to

$$\widehat{\boldsymbol{\mathcal{v}}}_{\boldsymbol{j}} = \boldsymbol{y_{j,t}} - \boldsymbol{x_{j,t}}\hat{\alpha}_j \to_p \boldsymbol{w_t}\eta_j + \boldsymbol{v_j}. \tag{47}$$

as $T \to \infty$. By Assumption 11 again, as $N \to \infty$ we can get

$$\inf_{1\le t\le T}\frac{1}{N}\sum_{j=1}^{N}\widehat{\boldsymbol{\mathcal{v}}}_{\boldsymbol{j}} = \inf_{1\le t\le T}\boldsymbol{w_t} + o_p(1) \quad and \quad \sup_{1\le t\le T}\frac{1}{N}\sum_{j=1}^{N}\widehat{\boldsymbol{\mathcal{v}}}_{\boldsymbol{j}} = \sup_{1\le t\le T}\boldsymbol{w_t} + o_p(1),$$

which implies that $\|\boldsymbol{w_t}\|_\infty$ and $\|\boldsymbol{w_t}\|_{-\infty}$could also be estimated by the largest and smallest value of the estimated residuals of (47) over $t = 1,\ldots,T$.

- Situation Two: heterogeneous latent variables.

Different from Situation One, it would be more plausible that the latent variable be different among $\mathbf{y}_{i,t}$'s neighbors $\boldsymbol{y_{j,t}}$ on the economies' network, so each economy is driven by its unique

[1] As we can see here, $\mathbf{y}_{i,t}$'s neighbors $\mathbf{y}_{j,t}$ could be treated as control units in the causal inference literature, $\mathbf{y}_{i,t}$ is the treated unit.

latent factor $\boldsymbol{w_{t,j}}$. In this case, consider

$$\boldsymbol{w_t} = \frac{\gamma_1}{N}\boldsymbol{w_{j=1,t}} + \frac{\gamma_2}{N}\boldsymbol{w_{j=2,t}} + \cdots + \frac{\gamma_2}{N}\boldsymbol{w_{j=N,t}} \tag{48}$$

for some $\gamma_j \in \mathbb{R}$ w.l.o.g., and

$$\boldsymbol{y_{j,t}} = \boldsymbol{w_{j,t}}\eta_j + \boldsymbol{x_{j,t}}\alpha_j + \boldsymbol{\pi_j}, \tag{49}$$

where $\boldsymbol{\pi_j}$ i.i.d. to some symmetry distribution with zero means and finite variance. Note that $\gamma_j/\eta_j = 1$ by Assumption 11, hence we can get

$$\phi_{en}\eta\boldsymbol{w_t} = \phi_{en}\eta\frac{1}{N}\sum_{j=1}^{N}\gamma_j\boldsymbol{y_{j,t}} - \phi_{en}\eta\frac{1}{N}\sum_{j=1}^{N}\gamma_j\alpha_j\boldsymbol{x_{j,t}} - \phi_{en}\eta\frac{1}{N}\sum_{j=1}^{N}\gamma_j\boldsymbol{\pi_j} \tag{50}$$

from (48) and (49), where the coefficient η captures endogenous regime switch effect shown in (28). Therefore by Assumption 10, as $N \to \infty$ we can get $(\phi_{en}/N)\eta\gamma_N\alpha_N = o_p(1)$, $\phi_{en}/N\sum_{j\in\mathbb{Z}}\eta\gamma_j\alpha_j\boldsymbol{x_{j,t}} = o_p(1)$ similar to (46), and

$$\inf_{1\le t\le T}\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{y_{j,t}} = \inf_{1\le t\le T}\boldsymbol{w_t} + o_p(1) \quad and \quad \sup_{1\le t\le T}\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{y_{j,t}} = \sup_{1\le t\le T}\boldsymbol{w_t} + o_p(1),$$

which is equivalent to Situation One under Assumptions 10-11.

4.2. Asymptotic behaviors

The asymptotic distributions of the estimators are established in the following theorem:

Theorem 3. Suppose the Assumptions 1-9 hold true, $T_0/T_1 \to c_1 \in (0,\tau)$, $T_0/T_2 \to c_2 \in (0,\tau)$ and $T_1/T_2 \to c_3 \in (0,\tau)$ with $\tau < \infty$ where $T_0 = t_{s0} - 1$, $T_1 = t_{D0} - t_{s0}$ and $T_2 = T - t_{D0} + 1$. Then for the structure change effect, as $T_0 \to \infty$ and $T_1 \to \infty$ we have

$$\begin{aligned}\sqrt{T_0}\left(\widehat{\beta\eta}_{AARs} - \beta\eta\right) \\ \rightsquigarrow \mathcal{N}\Bigg(0,\frac{T_0 Var(\boldsymbol{\epsilon_t})_{t_{s0}\le t\le t_{D0-1}}}{T_1^2}\boldsymbol{\mathcal{S}_{t_{s0}\le t\le t_{D0-1}}}\boldsymbol{\mathcal{Q}_{t_{s0}\le t\le t_{D0-1}}}\boldsymbol{\mathcal{S}'_{t_{s0}\le t\le t_{D0-1}}} \\ + \frac{T_0 Var(\boldsymbol{\epsilon_t})_{1\le t\le t_{s0-1}}}{T_0^2}\boldsymbol{\mathcal{S}_{1\le t\le t_{s0-1}}}\boldsymbol{\mathcal{Q}_{1\le t\le t_{s0-1}}}\boldsymbol{\mathcal{S}'_{1\le t\le t_{s0-1}}}\Bigg),\end{aligned}$$

where $\boldsymbol{\mathcal{S}_{t_{s0}\le t\le t_{D0-1}}} = \left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}\boldsymbol{\delta_T\delta_\beta I_{T_0+T_1}}$, $\boldsymbol{\mathcal{Q}_{t_{s0}\le t\le t_{D0-1}}} = \mathbb{E}\left(\boldsymbol{\delta_T\delta_\beta}\odot\boldsymbol{\delta_T\delta_\beta}\right)_{t_{s0}\le t\le t_{D0-1}}$ and $\boldsymbol{\mathcal{S}_{1\le t\le t_{s0-1}}} = \left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}\boldsymbol{\delta_T\delta_\beta I_{T_0}}$, $\boldsymbol{\mathcal{Q}_{1\le t\le t_{s0-1}}} = \mathbb{E}\left(\boldsymbol{\delta_T\delta_\beta}\odot\boldsymbol{\delta_T\delta_\beta}\right)_{1\le t\le t_{s0-1}}$, $\boldsymbol{\delta_T\delta_\beta}$ is defined in Theorem 2. For the total effect, as $T = T_0 + T_1 + T_2 \to \infty$ we have

$$\begin{aligned}\sqrt{T_0+T_1}\left(\widehat{\beta\eta+\xi}_{AARs} - (\beta\eta+\xi)\right) \\ \rightsquigarrow \mathcal{N}\Bigg(0,\frac{(T_0+T_1)Var(\boldsymbol{\epsilon_t})_{t_{D0}\le t\le T}}{T_2^2}\boldsymbol{\mathcal{S}_{t_{D0}\le t\le T}}\boldsymbol{\mathcal{Q}_{t_{D0}\le t\le T}}\boldsymbol{\mathcal{S}'_{t_{D0}\le t\le T}} \\ + \frac{(T_0+T_1)Var(\boldsymbol{\epsilon_t})_{1\le t\le t_{D0-1}}}{(T_0+T_1)^2}\boldsymbol{\mathcal{S}_{1\le t\le t_{D0-1}}}\boldsymbol{\mathcal{Q}_{1\le t\le t_{D0-1}}}\boldsymbol{\mathcal{S}'_{1\le t\le t_{D0-1}}}\Bigg)\end{aligned}$$

with $\boldsymbol{\mathcal{S}_{t_{D0}\le t\le T}} = \left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}\boldsymbol{\delta_T\delta_\beta I_T}$, $\boldsymbol{\mathcal{Q}_{t_{D0}\le t\le T}} = \mathbb{E}\left(\boldsymbol{\delta_T\delta_\beta}\odot\boldsymbol{\delta_T\delta_\beta}\right)_{t_{s0}\le t\le t_{D0-1}}$ and $\boldsymbol{\mathcal{S}_{1\le t\le t_{D0-1}}} = \left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}\boldsymbol{\delta_T\delta_\beta I_{T_1+T_2}}$, $\boldsymbol{\mathcal{Q}_{1\le t\le t_{D0-1}}} = \mathbb{E}\left(\boldsymbol{\delta_T\delta_\beta}\odot\boldsymbol{\delta_T\delta_\beta}\right)_{1\le t\le t_{D0-1}}$, $\boldsymbol{\delta_T\delta_\beta}$ is defined in Corollary 1. For the treatment effect, as $T_0 \to \infty$ and $T_1 \to \infty$ we have

$$\sqrt{T_0}(\hat{\xi}_{AARs} - \xi) \rightsquigarrow \mathcal{N}\left(0, \frac{T_0 Var(\boldsymbol{\epsilon_t})_{t_{D0}\le t\le T}}{T_2^2} \boldsymbol{\mathcal{S}_{t_{D0}\le t\le T} \mathcal{Q}_{t_{D0}\le t\le T} \mathcal{S}'_{t_{D0}\le t\le T}} + \frac{T_0 Var(\boldsymbol{\epsilon_t})_{1\le t\le t_{D0-1}}}{(T_0+T_1)^2} \boldsymbol{\mathcal{S}_{1\le t\le t_{D0-1}} \mathcal{Q}_{1\le t\le t_{D0-1}} \mathcal{S}'_{1\le t\le t_{D0-1}}}\right)$$

with $\boldsymbol{\mathcal{S}_{t_{D0}\le t\le T}} = \left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}\boldsymbol{\delta_T\delta_\beta I_T}$, $\boldsymbol{\mathcal{Q}_{t_{D0}\le t\le T}} = \mathbb{E}\left(\boldsymbol{\delta_T\delta_\beta}\odot\boldsymbol{\delta_T\delta_\beta}\right)_{t_{s0}\le t\le t_{D0-1}}$ and $\boldsymbol{\mathcal{S}_{1\le t\le t_{D0-1}}} = \left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}\boldsymbol{\delta_T\delta_\beta I_{T_1+T_2}}$, $\boldsymbol{\mathcal{Q}_{1\le t\le t_{D0-1}}} = \mathbb{E}\left(\boldsymbol{\delta_T\delta_\beta}\odot\boldsymbol{\delta_T\delta_\beta}\right)_{1\le t\le t_{D0-1}}$, $\boldsymbol{\delta_T\delta_\beta}$ is defined in Corollary 1. For the individual effect, as $T \to \infty$ we have

$$\sqrt{T}(\hat{\alpha}_{AARs} - \alpha_{AARs}) \rightsquigarrow \mathcal{N}(0, T\sigma_\alpha^2(\boldsymbol{x_t}'\boldsymbol{x_t})^{-1})$$

where $\sigma_\alpha^2 = Var(\mathbf{y_t} - \boldsymbol{s_t}\beta\eta_{AARs} - \boldsymbol{D_t}\xi_{AARs})_{1\le t\le T}$, the expectation and variance operator are taken over t.

Note that the convergence speeds of the AARs estimators are different from each other, among which the convergence speed of the total effect is the fastest. In empirical studies, we usually require the pre-structural period T_0 to be large enough. The asymptotic behavior of $\hat{\beta}_{AARs} = \widehat{\beta\eta}_{AARs}/\hat{\eta}_{AARs}$ is nontrivial, and is established in e.g.: Hinkley (1969) and Nadarajah (2006). Note that the first and high-order moments of $\hat{\beta}_{AARs}$ do not exist, it is beyond this paper's scope to carry out inference for $\hat{\beta}_{AARs}$.

5. Monte Carlo Simulations

We consider DGP

$$\begin{cases} \mathbf{y}_{\boldsymbol{lp,t}} = \boldsymbol{w_t}\eta + \boldsymbol{s_t}\eta\beta + \boldsymbol{\nu_t} \\ \mathbf{y_t} = \mathbf{y}_{\boldsymbol{lp,t}} + \boldsymbol{D_t}\xi + \boldsymbol{x_t}\alpha + \boldsymbol{\epsilon_t}' \end{cases} \tag{51}$$

with $\beta = 2$, $\eta = 1.25$, $\xi = -1.7$ and $\alpha = 0.5$ respectively for sample size $T = 100, 250, 500$, where $t_{s0} = 35$, $t_{D0} = 70$ for $T = 100$, $t_{s0} = 87$, $t_{D0} = 175$ for $T = 250$ and $t_{s0} = 175$, $t_{D0} = 350$ for $T = 500$ according to the convergence condition shown in Theorem 3. Hence $T_0/T \longrightarrow 0.35$, $T_1/T \longrightarrow 0.36$ and $T_3/T \longrightarrow 0.3$ for $T = 100, 250, 500$ respectively in our simulations. we set $\boldsymbol{w_t'} =_d \mathcal{U}(0,1)$ and $\boldsymbol{w_t} = sort(\boldsymbol{w_t'})$, $sort(\cdot)$ denotes the sort function, $\|\boldsymbol{w_t}\|_\infty = max_t\boldsymbol{w_t}$, $\|\boldsymbol{w_t}\|_{-\infty} = min_t\boldsymbol{w_t}$; $\boldsymbol{x_t} =_d \mathcal{U}(0,1)$, $\boldsymbol{\nu_t} =_d \mathcal{N}(0,0.01)$ and $\boldsymbol{\epsilon_t} =_d \mathcal{N}(0,0.01)$. The control units are generated as

$$\mathbf{y}_{\boldsymbol{t,j}} = \boldsymbol{w_t}\eta + \boldsymbol{x_t}\alpha + \boldsymbol{\epsilon_t} + \boldsymbol{\nu_{t,j}} \tag{52}$$

for $j = 1, \dots, 20$ with $\boldsymbol{\nu_{t,j}} =_d \mathcal{N}(0,0.01)$.

In the first case, we consider $\boldsymbol{x_t}$ is observable hence controlled in our model. To test Assumption 7, we consider a stronger inference method than the one suggested in section 4.1: $\mathcal{H}_0: \not{f}(\hat{\boldsymbol{u}}_{\boldsymbol{t}}^{\boldsymbol{s=1}})_{1\le t<t_{s0}} - \not{f}(\hat{\boldsymbol{u}}_{\boldsymbol{t}}^{\boldsymbol{s=1}})_{t_{s0}\le t\le t_{D0-1}} \neq 0$, $\mathcal{H}_1: \not{f}(\hat{\boldsymbol{u}}_{\boldsymbol{t}}^{\boldsymbol{s=1}})_{1\le t<t_{s0}} - \not{f}(\hat{\boldsymbol{u}}_{\boldsymbol{t}}^{\boldsymbol{s=1}})_{t_{s0}\le t\le t_{D0-1}} = 0$ where $\hat{\boldsymbol{u}}_{\boldsymbol{t}}^{\boldsymbol{s=1}} = \mathbf{y_t} - \boldsymbol{s_t}\widehat{\beta\eta}_{AARs}$ and $\not{f}(\cdot)$ is the empirical distribution. Similarly, to test Assumption 8, we consider $\mathcal{H}_0: \not{f}(\hat{\boldsymbol{u}}_{\boldsymbol{t}}^{\boldsymbol{s=2}})_{1\le t<t_{s0}} - \not{f}(\hat{\boldsymbol{u}}_{\boldsymbol{t}}^{\boldsymbol{s=2}})_{t_{D0}\le t\le T} \neq 0$, $\mathcal{H}_1: \not{f}(\hat{\boldsymbol{u}}_{\boldsymbol{t}}^{\boldsymbol{s=2}})_{1\le t<t_{s0}} - \not{f}(\hat{\boldsymbol{u}}_{\boldsymbol{t}}^{\boldsymbol{s=2}})_{t_{D0}\le t\le T} = 0$ where $\hat{\boldsymbol{u}}_{\boldsymbol{t}}^{\boldsymbol{s=2}} = \mathbf{y_t} - \boldsymbol{s_t}\widehat{\beta\eta}_{AARs} - \boldsymbol{D_t}\hat{\xi}_{AARs}$. If $\mathcal{H}_0$ are rejected, then Assumptions 7-8 hold true. We adopt Wilcoxon rank sum method to test the Null hypothesizes. Several competitive estimators including OLS and Generalized Synthetic Control Method (Gsynth)

are compared with the new method.[1] The results are shown below in Table 1 and Figure 4 (a).

In the second case, we suppose the confounder $\boldsymbol{x_t}$ is unobservable to the empirical researchers hence omitted in model (51), consequently the treatment $\boldsymbol{D_t}$ and the structure change $\boldsymbol{s_t}$ are both endogenous. We are interested in estimating $(\eta\beta, \xi)$, the results are shown in Table 2 and Figure 4 (b).

Table 1

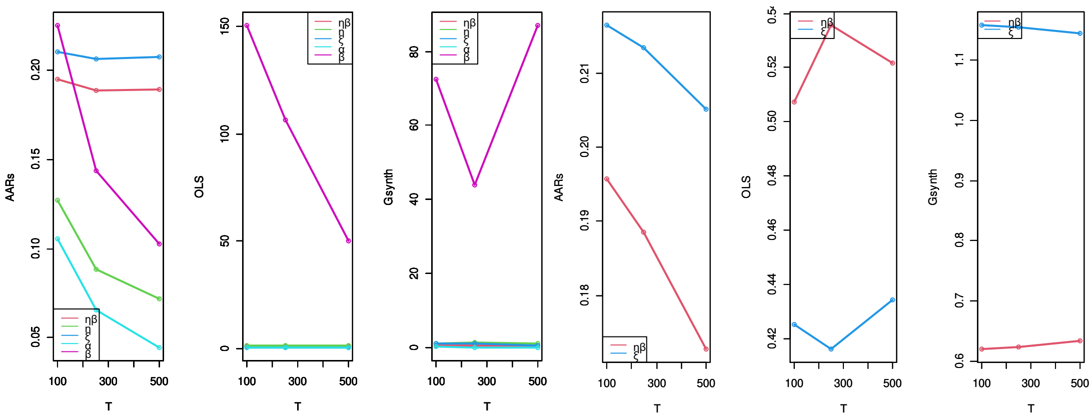
Small sample performances of several estimators ($\boldsymbol{x_t}$ is observable and controlled).

Biases		$\lvert\beta\eta - \widehat{\beta\eta}\rvert$	$\lvert\beta - \hat{\beta}\rvert$	$\lvert\eta - \hat{\eta}\rvert$	$\lvert\xi - \hat{\xi}\rvert$	$\lVert w_t - \hat{w}_t\rVert$	$\lvert\alpha - \hat{\alpha}\rvert$
	AARs	**0.195**	**0.225**	**0.127**	**0.210**	**8.366**	**0.105**
T=100	OLS	0.512	150.453	1.166	0.420	574.398	0.187
	Gsynth	0.616	72.527	1.267	1.136	4822.992	0.260
	AARs	**0.189**	**0.144**	**0.089**	**0.207**	**18.758**	**0.066**
T=250	OLS	0.513	106.349	1.167	0.425	1028.038	0.180
	Gsynth	0.643	43.855	1.294	1.162	6963.779	0.004
	AARs	**0.190**	**0.103**	**0.072**	**0.208**	**35.532**	**0.044**
T=500	OLS	0.515	50.112	1.163	0.424	1001.816	0.181
	Gsynth	0.636	86.979	1.229	0.687	27020.25	0.152

Table 2

Small sample performances of several estimators ($\boldsymbol{x_t}$ is unobservable and omitted).

Biases		$\lvert\beta\eta - \widehat{\beta\eta}\rvert$	$\lvert\xi - \hat{\xi}\rvert$
	AARs	**0.196**	**0.216**
T=100	OLS	0.507	0.425
	Gsynth	0.620	1.158
	AARs	**0.189**	**0.213**
T=250	OLS	0.536	0.416
	Gsynth	0.624	1.155
	AARs	**0.173**	**0.205**
T=500	OLS	0.522	0.434
	Gsynth	0.634	1.145



[1] Gsynth is a generalization of DID and Synthetic Control Method (SCM) (Alberto & Gardeazabal, 2003; Alberto et. al, 2010; Xu, 2017).

(a) (b)

Figure 4. Biases of three estimators in the first case (a) and second case (b).

From Table 1 and Table 2, it is shown that AARs dominates the other two estimators: (1) AARs owns good small sample performances and the biases of AARs are the smallest; (2) the biases of AARs are decreasing with sample size proving the consistencies of the estimators, while OLS and Gsynth are not consistent due to model misspecification caused by the unobservable and uncontrollable latent variable $\boldsymbol{w_t}$; (3) AARs can be adopted to estimate the structure change effect and treatment effect even though there exist confounders not controlled in the model, wherein both $\boldsymbol{s_t}$ and $\boldsymbol{D_t}$ are endogenous.[1]

6. A tale of two cities: the shocks of earthquakes

Will natural disasters such as earthquakes inevitably lead to economic recessions? Economists hold different views on this topic, although the literature is diverse, all viewpoints can be summed up into two theories: direct destruction theory and Schumpeterian creative destruction theory. For the former, researchers hold the view that natural disasters will destroy economic constructions and labor aggregations directly, move economies away from their steady-state levels of objectives and result in economic recessions, see Barro & Lee (1993); Raddatz (2005); Noy (2009) and Cavallo et al. (2013) for examples; while the latter holds that natural disasters update capital stocks and encourage the adoptions of new technologies, which will lead to improved Total Factor Productivity (TFP) and growth of GDP, see Schumpeter (1942), Caballero (1994), Skidmore & Toya (2002), Shabnam (2014) and references therein for examples.

Which of these two theories is correct? There is no final conclusion. To throw new lights on this topic, we turn back to the Wenchuan and Kobe earthquakes. By the proposed AARs method, we find that the reason for the conflicting conclusions on the impacts of earthquakes lies in the neglecting of business cycle effects captured by structure changes and endogenous regime switches. The great Hanshin-Awaji earthquake of Richter scale 7.3 took place on January 17, 1995 in Kobe city, Japan. This strong earthquake caused extremely serious damages to Kobe, the main city in the Hanshin economic zone of Japan. According to the statistics, more than 6500 people died in the earthquake stricken area (more than 4000 people were killed by smashing and suffocation, accounting for more than 90% of the deaths), about 27000 people were injured, and 108000 buildings were destroyed with nearly 300000 homeless victims (Horwich, 2000). Meanwhile, the Wenchuan earthquake of Richter scale 8.0 took place on May 12, 2008 in Wenchuan county, China. The quake lead to 69227 people killed, 17923 people missing, 374643 people injured, and 1993.03 million people lost their homes. The total affected population reached 46.256 million. As of September 2008, the direct economic loss caused by the quake was 845.14 billion Yuan (RMB). The Wenchuan earthquake is the most destructive, the most extensive and the heaviest disaster loss since the founding of the People's Republic of China.[1]

To evaluate the impact of the earthquake on economic development, we consider the following equation:

$$\mathbf{y_t} = (\boldsymbol{w_t} + \boldsymbol{s_t}\beta)\eta + \boldsymbol{D_t}\xi + \boldsymbol{x_t}\alpha + \boldsymbol{\epsilon_t} + \boldsymbol{\nu_t}, t = t_1, \dots, t_{s0}, \dots, t_{D0}, \dots, t_T, \quad (53)$$

[1] This can be a good news to empirical researchers who are always limited by good quality IVs to handle endogeneities.

[1] For more details about this earthquake, we refer to Ng et al. (2015).

where $\mathbf{y}_t$ denotes $log(GDP)$, $\boldsymbol{D_t} = 1 \cdot \mathbb{I}\{\boldsymbol{t} \geq 1995\} + 0 \cdot \mathbb{I}\{\boldsymbol{t} \leq 1994\}$ with $t_{D0} = 1995$, $t_1 = 1955$, $t_2 = 1956$,..., $t_T = 2009$ for Kobe and $\boldsymbol{D_t} = 1 \cdot \mathbb{I}\{\boldsymbol{t} \geq 2008\} + 0 \cdot \mathbb{I}\{\boldsymbol{t} \leq 2007\}$ with $t_{D0} = 2008$, $t_1 = 1978$, $t_2 = 1980$, $t_3 = 1985$, $t_4 = 1990$, $t_5 = 1995$, $t_6 = 2000$, $t_7 = 2001$, $t_8 = 2002$,..., $t_T = 2018$ for Wenchuan.[2] To adopt the AARs method, we use the Huberized CUSUM test implemented in the R package "robcp" to estimate the structure change points, and then get $\boldsymbol{s_t} = 1 \cdot \mathbb{I}\{\boldsymbol{t} \geq 1980\} + 0 \cdot \mathbb{I}\{\boldsymbol{t} \leq 1979\}$ with $\hat{t}_{s0} = 1980$ for Kobe; and $\boldsymbol{s_t} = 1 \cdot \mathbb{I}\{\boldsymbol{t} \geq 1983\} + 0 \cdot \mathbb{I}\{\boldsymbol{t} \leq 1982\}$ with $\hat{t}_{s0} = 1983$ for Wenchuan. We include other neighbor-cities' $log(GDPs)$ on the economic network as covariates $\boldsymbol{x_t}$ for Kobe and Wenchuan to implement our AARs and OLS estimations (Ng et al., 2015),[3] meanwhile we also treat these neighbor-cities as control units free of earthquakes to implement the Gsynth estimation. 46 cities are then selected as controls for Kobe which are shown in Fujiki & Hsiao (2015), and 20 cities selected for Wenchuan.[4] The data collection for Kobe is described in Fujiki & Hsiao (2015) while the data for Wenchuan are collected from *Statistical yearbook of Sichuan Province*. Therefrom, we get $\|\boldsymbol{w_t}\|_{\infty} = 3.9985$, $\|\boldsymbol{w_t}\|_{-\infty} = 0.4809$ for Kobe and $\|\boldsymbol{w_t}\|_{\infty} = 10.6383$, $\|\boldsymbol{w_t}\|_{-\infty} = 5.5295$ for Wenchuan by step 3 in section 4.1.

6.1. Long-run effects

We compare the long-run effects of the earthquakes with structure changes and without structure changes through OLS, Gsynth and AARs, which are shown in Table 3 and 4 respectively:

Table 3

A tale of two cities: the long-run effects of earthquakes on economic developments with prior structure changes (Hanshin-Awaji, Japan & Wenchuan, China).

	Hanshin-Awaji (Japan)			**Wen-chuan (China)**		
	AARs	OLS	Gsynth	AARs	OLS	Gsynth
Treatment effect (ξ)	-0.147 (0.205)	0.506 (0.438)	-0.007 (0.052)	0.739*** (0.307)	1.130 (1.560)	-0.024 (0.055)
Structural change effect ($\beta\eta$)	1.698*** (0.180)	3.216*** (0.310)	-0.006 (0.052)	1.708*** (0.249)	8.833*** (1.187)	-0.397*** (0.041)
Endogenous regime switch effect (η)	0.752*** (0.017)	0.444*** (0.043)	-0.0003*** (9.425e-05)	0.887*** (0.001)	0.443*** (0.033)	-0.003*** (3.290e-06)
Structure change	2.259	7.249	19.887	1.926	19.938	142.251

[2] Limited by the availabilities, these are the most comprehensive data we can collect.
[3] The covariates $\boldsymbol{x_t}$ are independent of treatment $\boldsymbol{D_t}$ for the earthquake can be seen as an exogenous shock.
[4] The selection of control units are based on two criterions: (1) we need the controls not influenced by earthquakes at t_{D0} as well as any other structure changes at t_{s0} and (2) the economic development levels are similar to that of the treated unit. We adopt the method of Fujiki & Hsiao (2015) to select control units for Wenchuan, 20 cities and districts selected are Chengdu, Zigong, Panzhihua, Luzhou, Deyan, Mianyan, Guangyuan, Suining, Neijiang, Leshan, Nanchong, Meishan, Yibin, Guangan, Dazhou, Yaan, Bazhong, Ziyan, Ganzi Tibetan Autonomous Prefecture and Liangshan Yi Autonomous Prefecture.

(β)						
Threshold (τ_s)	1.135	1.923	-2563.491	7.503	15.022	-2386.022
Statistics-1 (P. value)	204 (0.926)	\	\	18 (0.036)	\	\
Statistics-2 (P. value)	201 (0.990)	\	\	30 (0.235)	\	\
Controls	√	√	√	√	√	√
Adj. R^2	0.990	0.999	0.998	0.980	0.999	0.997
T_0	25	25	25	5	5	5
T_1	15	15	15	8	8	8
T_2	15	15	15	11	11	11
T	55	55	55	24	24	24

Standard errors in parentheses. *<0.05, **<0.01, ***<0.001.

Table 4

A tale of two cities? the long-run effects of earthquakes on economic developments with neglected structure changes (Hanshin-Awaji, Japan & Wenchuan, China).

	Hanshin-Awaji (Japan)			**Wen-chuan (China)**		
	AARs	OLS	Gsynth	AARs	OLS	Gsynth
Treatment effect (ξ)	1.804*** (0.156)	3.723*** (0.535)	-0.289*** (0.055)	2.020*** (0.198)	9.963*** (1.857)	-0.421*** (0.036)
Statistics (P. value)	372 (0.179)	\	\	93 (0.228)	\	\
Structure changes ($\beta\eta, \beta, \eta$)	×	×	×	×	×	×
Controls	√	√	√	√	√	√
Adj. R^2	0.991	0.999	0.998	0.845	0.999	0.998
T_0	40	40	40	13	13	13
T_1	40	40	40	13	13	13
T_2	15	15	15	11	11	11
T	55	55	55	24	24	24

Standard errors in parentheses. *<0.05, **<0.01, ***<0.001.

We find that: (1) even if the structure changes are considered in all methods, OLS and Gsynth still get unreasonable estimation results for the treatment effect and structural change effect. More specifically, OLS over-estimates the treatment effect ξ, structural change effect $\beta\eta$, structure change β, threshold level τ_s, and under-estimates the endogenous regime switch effect η; while Gsynth gets nearly *false negative* estimations for the treatment effect, structural change, endogenous regime switch effect, threshold level and over-estimates the structure change for both Hanshin-Awaji earthquake and Wenchuan earthquake; (2) AARs' result supports the Schumpeterian creative destruction theory for Wenchuan earthquake, which implies that technological improvements and new governmental investments after the earthquake increased log(GDP) by 0.7394; but does not support the direct destruction theory and the Schumpeterian creative destruction theory for Hanshin-Awaji, which indicates that the earthquake has no long-run effects on Kobe's economic developments; (3) most of all, neglecting structure changes over-estimates the treatment effects of both earthquakes.

6.2. Short-run effects

But why the Hanshin-Awaji earthquake has no long-run effects on economic development while the Wenchuan earthquake shows long-run positive effects? Will earthquakes really not lead to economic recessions? To find these out, we turn to estimate the short-run effects of the earthquakes. We estimate the dynamic treatment effects and endogenous regime switch effects with $t_T = 1995 + i$ for Kobe, where $i = 1,2,3,4,5,6,7,8,9,10,11,12,13,14$ respectively; and $t_T = 2008 + i$ for Wenchuan, where $i = 1,2,3,4,5,6,7,8,9,10$ respectively. The estimation results are shown in Figure 5 and Table 5:

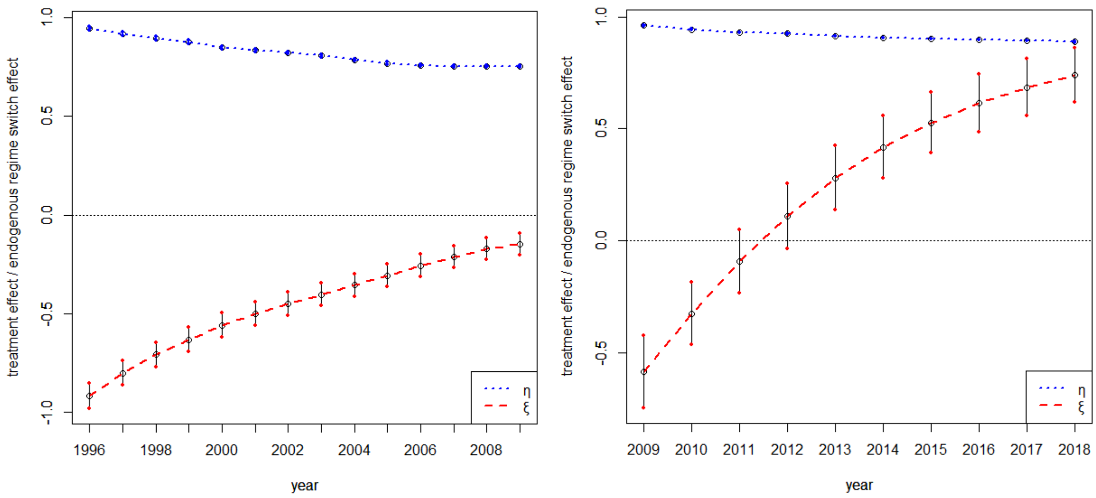


Figure 5. Short-run effects of earthquakes: Hanshin-Awaji, Japan (left) & Wenchuan, China (right).

Table 5

A tale of two cities: the short-run effects of earthquakes on economic developments with prior structure changes (Hanshin-Awaji, Japan & Wenchuan, China).

	Hanshin-Awaji (Japan)		**Wen-chuan (China)**	
Forward step since quake	Treatment effect (ξ)	Endogenous regime switch effect (η)	Treatment effect (ξ)	Endogenous regime switch effect (η)
$i = 1$	-0.915***	0.946***	-0.585*	0.961***
	(0.211)	(0.027)	(0.319)	(0.002)
$i = 2$	-0.780***	0.918***	-0.326	0.944***
	(0.207)	(0.028)	(0.284)	(0.002)
$i = 3$	-0.708***	0.896***	-0.094	0.930***
	(0.213)	(0.027)	(0.296)	(0.001)
$i = 4$	-0.630***	0.877***	0.108	0.924***
	(0.215)	(0.027)	(0.315)	(0.001)

$i = 5$	-0.557***	0.848***	0.280	0.914***
	(0.212)	(0.025)	(0.320)	(0.001)
$i = 6$	-0.500**	0.834***	0.417	0.904***
	(0.213)	(0.025)	(0.319)	(0.001)
$i = 7$	-0.448**	0.822***	0.527*	0.901***
	(0.212)	(0.024)	(0.316)	(0.001)
$i = 8$	-0.402**	0.807***	0.615**	0.897***
	(0.211)	(0.023)	(0.313)	(0.001)
$i = 9$	-0.355**	0.785***	0.685***	0.893***
	(0.209)	(0.021)	(0.310)	(0.001)
$i = 10$	-0.306*	0.768***	0.739***	0.887***
	(0.207)	(0.020)	(0.307)	(0.001)
$i = 11$	-0.256	0.758***	\	\
	(0.207)	(0.019)		
$i = 12$	-0.211	0.752***	\	\
	(0.206)	(0.018)		
$i = 13$	-0.171	0.752***	\	\
	(0.205)	(0.018)		
$i = 14$	-0.147	0.752***	\	\
	(0.205)	(0.017)		

Standard errors in parentheses. *<0.05, **<0.01, ***<0.001.

Interestingly, we find that: (1) both earthquakes show negative short-run impacts on economic development, supporting the direct destruction theory in short-runs after earthquakes, while the short-run effects last 10 years for Kobe but only 1 year for Wenchuan; (2) this adverse short-run effect diminished for Kobe in the long-run while turns into beneficial long-run effect for Wenchuan. The reason why Wenchuan turns into new growth in the long-run lies in the fact that strong measures have been taken timely and continuously by the central government of China and the local government of Wenchuan after quake.

These two findings reveal us that earthquakes will exert adverse effects on economic development in short-runs, so the direct destruction theory holds true in short-runs, while this effect will diminish or turn into positive effects in the long-run depending on whether the government has taken remedial measures. If positive measures are taken, the development of the economy will recover soon and even turn into new growths, as described by the Schumpeterian creative destruction theory; otherwise, it recovers slowly.

Put the pieces together, through the proposed AARs method, we conclude that one of the reasons for the disunity literature lies in the neglecting of prior structure changes and endogenous regime switches when evaluating earthquake shocks on economic developments.

7. Concluding Remarks

Be cautious of prior structure changes and endogenous regime switches when you are carrying out a regional program evaluation! As shown in this paper, neglecting prior structure changes and endogenous regime switches will lead to over-estimated, under-estimated or even *false positively estimated* or *false negatively estimated* treatment effects, resulting in server misleading research conclusions and wrong policy implications. Unfortunately, what is worrying is that almost all published empirical studies ignored this point, which is exactly what this paper wants to attract your attention.[1] The good news is that a new method, called AARs, is proposed in this paper to deal with this issue. Through an automatically auxiliary dynamics, the AARs is able to disentangling structure change effects from treatment effects, and the parameters can be consistently estimated though a flexible 3-step estimation procedure. This new approach has several clear advantages: first of all, we allow endogenous structure changes with an unobservable latent variable and endogenous treatments with totally unobservable (or partially observable) confounders. Mostly important, we do not need IVs or any other exogenous shocks to help us achieve identification; second, we allow multiple structure changes and multiple treatments; third, the new method is highly flexible and easy to implement, there are nearly no technical barriers for empirical researchers.

Instead of sophisticated and exhausted technical explorations, the main purpose of this paper is to present the problem we want to call for appearing in current empirical studies through a simple model. Although it is simple, the basic idea and the baseline specification can be extended to handle complex situations, among which particular interests are: (1) smooth structure changes. The endogenous regime switch considered in this paper is designed as an abrupt structure break, but it is more reasonable to “allow the structure change to take a period of times to take effects” (Chen & Hong, 2012), disentangling smooth structure transitions from treatments would be attractive; (2) time-varying structure change effects, endogenous regime switch effects and treatment effects. The model considered in this paper assumes that all these effects remain the same over time, but it is more realistic and meaningful to take time into consideration in modeling the dynamics of structure changes and policy transitions; (3) more general specifications: nonparametric or semi-parametric settings. It would be quite attractive to consider nonparametric nested systems wherein the structure changes and treatments are determined in much more flexible forms of the thresholds. Efforts on these directions are undergoing.

References

Abadie A., Diamond A., Hainmueller J. 2010. Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program. Journal of the American Statistical Association. 105 (490), 493–505.

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

Alexopoulos M, Cohen J. 2016. The medium is the measure: Technical change and employment, 1909—1949. Review of economics and statistics. 98, 792–810.

Alonso, C., Gutierrez, J., Recio, T. 1997. A note on separated factors of separated

[1] For example, Abadie et. al (2015) finds that the negative impact of the reunification of Germany after World War II on the economy continued until 2003. However through the proposed AARs method, we find that, this kind of influence only lasted until 1998. The reason why there seems no effect in the short run after the reunification is that the continuous growth of structure change effects neutralized the treatment effects. That is to say, the reunification of Germany had a negative impact on the economy in the short term, but this effect was neutralized by the inertia of economic growth before unification. Abadie et. al (2015) ignores the structure change effects, which leads to the over-estimation of the influence of German unification on economic growth by 1.051374 units (after taking log).

polynomials. Journal of Pure and Applied Algebra. 121(3), 217-222.
Alonso, C., Gutierrez, J., Recio, T. 2007. A rational function decomposition algorithm by near-separated polynomials. Journal of Symbolic Computation. 19(6), 527-544.
Angrist J.D., Pischke J.S. 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.
Becker, J., Fuest, C. 2010. Taxing foreign profits with international mergers and acquisitions. International Economic Review. 51(1), 171-186.
Black F. 1982. The Trouble with Econometric Models. Financial Analyst Journal. 38, 29—37.
Blume L.E., Brock W.A., Durlauf S.N., Jayaraman R. 2015. Linear Social Interactions Models. Journal of Political Economy. 123(2): 444-496.
Boucekkine, R., Pommeret, A., Prieur F. 2013. Optimal regime switching and threshold effects. Journal of Economic Dynamics & Control. 37(12), 2979-2997.
Caballero, R.J., Hammour, M.L. 1994. The cleansing effect of recessions. American Economic Review. 84(5), 1350–1368.
Callaway B., Sant'Anna P.H.C. 2021. Difference-in-Differences with multiple time periods. Journal of Econometrics (forthcoming).
Cavallo E., Galiani S., Noy I. et al. 2013. Catastrophic natural disasters and economic growth. Review of Economics & Statistics. 95(5), 1549-1561.
Chang Y., Choi Y., Park J.Y. 2017. A new approach to model regime switch. Journal of Econometrics. 196, 127-143
Chen, B., Hong, Y. 2012. Testing for smooth structural changes in time series models via nonparametric regression. Econometrica. 80(3), 1157-1183.
Cinelli C., Hazlett C. 2020. Making Sense of Sensitivity: Extending Omitted Variable Bias. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 82(1), 39–67.
Dawid, A. P. 1979. Conditional independence in statistical theory. Jour. Roy. Statist. Soc. B. 41(1), 1-31.
Dinkelman T. 2011. The effects of rural electrification on employment: New evidence from South Africa. American Economic Review. 101, 3078– 3108.
Forastiere L., Airoldi E.M., Mealli F. 2021. Identification and estimation of treatment and interference effects in observational studies on networks. Journal of the American Statistical Association (forthcoming).
Franks A., D’Amour A., Feller A. 2019. Flexible sensitivity analysis for observational studies without observable implications. Journal of the American Statistical Association, 114(528), 1574-1596.
Fujiki H., Hsiao C. 2015. Disentangling the effects of multiple treatments—measuring the net economic impact of the 1995 great Hanshin-Awaji earthquake. Journal of Econometrics. 186, 66-73.
Gobillon L., Magnac T. 2016. Regional policy evaluation: Interactive fixed effects and synthetic controls. Review of Economics and Statistics. 98 (3), 535–551.
Graham B., de Paula A. 2020. The Econometric Analysis of Network Data. Academic Press/Elsevier.
Gutierrez, J., Urroz J.J. 2020. On some classes of irreducible polynomials. Journal of Symbolic Computation. 105, 64-70.
Haavelmo T. 1944. The Probability Approach in Econometrics. Econometrica. 12, 1—115.

Hall, A.R., Han, S., Boldea, O. 2012. Inference regarding multiple structural changes in linear models with endogenous regressors. Journal of Econometrics. 170(2), 281-302.

Heckman J.J. 2001. Econometrics and Empirical Economics. Journal of Econometrics. 100, 3—5.

Heckman, J. J., Humphries, J. E., Veramendi, G. 2016. Dynamic treatment effects. Journal of Econometrics. 191(2), 276-292.

Hendry D.F. 1980. Econometrics—Alchemy or Science? Economica. 47, 387—406.

Hinkley, D.V. 1969. On the ratio of two correlated normal random variables. Biometrika. 56(3),635-639.

Holland, P.W. 1986. Statistics and causal inference. Journal of the American statistical Association 81, 945– 960.

Horwich G. 2000. Economic lessons of the Kobe earthquake. Economic Development & Cultural Change. 48(3), 521-542.

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

Huber, M., Melly, B. 2015. A test of the conditional independence assumption in sample selection models. Journal of Applied Econometrics, 30, 1144-1168.

Hudgens M.G., Halloran M.E. 2008. Toward causal inference with interference. Journal of the American Statistical Association. 103, 832-842.

Imbens G.W., Rubin D.B. 2015. Causal inference in statistics, social, and biomedical sciences. Cambridge University Press.

Jenish, N., Prucha, I.R. 2012. On spatial processes and asymptotic inference under near-epoch dependence. Journal of Econometrics, 170(1), 178-190.

Keynes J.M. 1939. Professor Tinbergen's Method. Economic Journal. 49, 558—568.

Keynes J.M. 1940. Comment. Economic Journal. 50, 154—156.

Kim C. J. 2004. Markov-switching models with endogenous explanatory variables. J. Econometrics 122, 127–136.

Kim C. J. 2009. Markov-switching models with endogenous explanatory variables II: A Two-Step MLE procedure. J. Econometrics 148, 46–55.

Leamer E. 1983. Let's Take the Con Out of Econometrics. American Economic Review. 73, 31—43.

Lewis, J. Severnini, E. 2020. Short- and long-run impacts of rural electrification: evidence from the historical rollout of the U.S. power grid. Journal of Development Economics. 143, 1-19.

Lopez M.J., Gutman R. 2017. Estimation of causal effects with multiple treatments: a review and new ideas. Statistical Science. 32(3), 432-454.

Machado, C. 2017. Unobserved selection heterogeneity and the gender wage gap. Journal of Applied Econometrics. 32(7), 1348-1366.

Manski, C. F. 2013. Identification of treatment response with social interactions. Econometrics Journal. 16(1), S1-S23.

Michaels G., Rauch F., Redding S.J. 2012. Urbanization and structural transformation. The Quarterly Journal of Economics. 127, 535–586

Moroney, J. R. 1975. Natural resource endowments and comparative labor costs: a hybrid model of comparative advantage. Journal of Regional Science. 15(2), 139-151.

Nadarajah S. 2006. On the ratio for some elliptically symmetric distributions. Journal of Multivariate Analysis. 97(2), 342-358.

Ng D., Gan, L., Hernandez, M. A. 2015. Do natural disasters cause an excessive fear of heights? Evidence from the Wenchuan earthquake. Journal of Urban Economics. 90, 79-89.

Noy, I. 2009. The macroeconomic consequences of disasters. Journal of Development Economics. 88(2), 221-231.

Okuyama, Y. 2015. Regional economic impacts of terrorist attacks, natural disasters and metropolitan policies. Journal of regional science. 56(2), 366-368.

Pratt J.W., Schlaifer R. 1984. On the Nature and Discovery of Structure. Journal of the American Statistical Association. 79, 9—21

Raddatz, C. 2005. Are external shocks responsible for the instability of output in low income countries? Journal of Development Economics. 84(1), 155-187.

Rubin D.B. 1978. Bayesian inference for causal effects: the role of randomization. Annals of Statistics. 6(1), 34-58.

Schumpeter, J.A. 1942. Capitalism, socialism and democracy. New York: Harper.

Seidl, A. 2019. Zeno points in optimal control models with endogenous regime switching. Journal of Economic Dynamics and Control. 100, 353-368.

Shabnam, N. 2014. Natural Disasters and Economic Growth: A Review. Int. J. Disaster Risk Sci. 5,157–163.

Skidmore, M., Toya, H. 2002. Do natural disasters promote long-run growth? Economic Inquiry. 40(4), 664–687.

Splawa-Neyman, J., Dabrowska, D. Speed, T. 1990. On the application of probability theory to agricultural experiments. Statistical Science. 5, 465–472.

Stuart, E.A. 2010. Matching methods for causal inference: a review and a look forward. Statistical Science. 2*5*(1), 1-21.

Sun, L., Abraham, S. 2021. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics (forthcoming).

Tinbergen J. 1940. On a Method of Statistical Business-Cycle Research: A Reply. Economic Journal. 50, 141—154.

Tong, L. 2010. Indirect inference in structural econometric models. Journal of Econometrics, 15*7*(1), 120-128.

Waring P. Burgess J. 2011. Continuity and change in the Australian minimum wage setting system: the legacy of the commission. Journal of Industrial Relations. 53(5), 681-697.

White, H. 2006. Time-series estimation of the effects of natural experiments. Journal of Econometrics. 135(1-2), 527-566.

Xu Y.Q. 2017. Generalized synthetic control method: Causal inference with interactive fixed effects models. Political Analysis. 25 (1), 57–76.

Yong, L., Yu, J. 2011. Bayesian hypothesis testing in latent variable models. Journal of Econometrics. 1*66*(2), 237-246.

Online Appendix

All proofs of the Propositions and Theorems in the main text are collected in this Online Appendix.

Proof of Proposition 1. Under Assumption 1, the adjacency matrices for the DGPs (24-27) are shown as $\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{B}}$, $\boldsymbol{\mathcal{C}}$, and $\boldsymbol{\mathcal{D}}$. By Assumptions 3-4, if we define $\boldsymbol{X} = (\boldsymbol{w_t}, s(\boldsymbol{w_t}), D(\boldsymbol{x_t}), \boldsymbol{x_t})$ and

$$\boldsymbol{E} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{E_1} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{E_2} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

then it is easy to see that the parameters $\theta_2 = (\beta_2, \eta_2, \beta_2\eta_2, \xi_2, \alpha_2)$ can be identified in DGP (25) if and only if $\boldsymbol{y_{t,2}} \subset \mathcal{M}^-(\boldsymbol{X}(\boldsymbol{\mathcal{B}} + \boldsymbol{E})')$, the parameters $\theta_3 = (\beta_3, \eta_3, \beta_3\eta_3, \alpha_3)$ can be identified in DGP (26) if and only if $\boldsymbol{y_{t,3}} \subset \mathcal{M}^-(\boldsymbol{X}(\boldsymbol{\mathcal{C}} + \boldsymbol{E_1})')$ and the parameters $\theta_4 = (\xi_4, \alpha_4)$ can be identified in DGP (27) if and only if $\boldsymbol{y_{t,4}} \subset \mathcal{M}^-(\boldsymbol{X}(\boldsymbol{\mathcal{D}} + \boldsymbol{E_2})')$. We then have

$$\begin{aligned} \boldsymbol{y_{t,1}} &\subset \mathcal{M}^-(\mathbf{X}(\boldsymbol{\mathcal{B}} + \boldsymbol{E})') \\ &= \mathcal{M}^-(\mathbf{X}\boldsymbol{\mathcal{B}}') + \mathcal{M}^-(\mathbf{X}\boldsymbol{E}') \\ &= \mathcal{M}^-(\mathbf{X}(\boldsymbol{\mathcal{C}} + \boldsymbol{\mathcal{D}})') + \mathcal{M}^-(\mathbf{X}(\boldsymbol{E_1} + \boldsymbol{E_2})') \\ &= \mathcal{M}^-(\mathbf{X}(\boldsymbol{\mathcal{C}} + \boldsymbol{E_1})') + \mathcal{M}^-(\mathbf{X}(\boldsymbol{\mathcal{D}} + \boldsymbol{E_2})') \end{aligned}$$

by Assumption 4 again. Note that $\boldsymbol{y_{t,1}} \subset \mathcal{M}^-(\mathbf{X}(\boldsymbol{\mathcal{C}} + \boldsymbol{E_1})') + \mathcal{M}^-(\mathbf{X}(\boldsymbol{\mathcal{D}} + \boldsymbol{E_2})')$ implies $\boldsymbol{y_{t,1}} \subset (\boldsymbol{y_{t,2}} + \boldsymbol{y_{t,3}})$, which shows that the structure change β_2, structure change effect $\beta_2\eta_2$, endogenous regime switch effect η_2 and treatment effect ξ_2 in GDP (25) can be identified as long as the structure change β_3, structure change effect $\beta_3\eta_3$, endogenous regime switch effect η_3 in DGP (26) and the treatment effect ξ_4 in DGP (27) can be identified. On the contrary, one can verify that $\boldsymbol{y_{t,1}} \subset (\boldsymbol{y_{t,2}} + \boldsymbol{y_{t,3}})$ will not imply $\boldsymbol{y_{t,2}} \subset \mathcal{M}^-(\boldsymbol{X}(\boldsymbol{\mathcal{A}} + \boldsymbol{E})')$, which confirms that the structural change effects $\beta\eta$, endogenous regime switch effects η and treatment effects ξ can be identified separately under DGP (25) with adjacency matrix $\boldsymbol{\mathcal{B}}$ but not under DGP (24) with adjacency matrix $\boldsymbol{\mathcal{A}}$.

Proof of Theorem 1. Under model (28) and Assumption 5, there exist $i \in \mathbb{Z} = \{1,2,3,\dots\}$ such that

$$\begin{aligned} \beta\eta + \xi &= \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(\mathbf{1,1})} | \boldsymbol{s_t} = 1, \boldsymbol{D_t} = 1\right) - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(\mathbf{0,0})} | \boldsymbol{s_t} = 0, \boldsymbol{D_t} = 0\right) \\ &= \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(\mathbf{1,1})} | \boldsymbol{s_t} = 1, \boldsymbol{D_t} = 1\right) - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(\mathbf{0,0})} | \boldsymbol{s_t} = 1, \boldsymbol{D_t} = 1\right) + \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(\mathbf{0,0})} | \boldsymbol{s_t} = 1, \boldsymbol{D_t} = 1\right) \\ &\quad - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(\mathbf{0,0})} | \boldsymbol{s_t} = 0, \boldsymbol{D_t} = 0\right) \\ &= \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(\mathbf{1,1})} | \boldsymbol{s_t}, \boldsymbol{D_t}\right)_{T \geq t \geq t_{D0}} - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(\mathbf{0,0})} | \boldsymbol{s_t}, \boldsymbol{D_t}\right)_{T \geq t \geq t_{D0}} + \mathbb{E}\left(\Delta\mathbf{y}_{\boldsymbol{t}}^{(\mathbf{0,0})} | \boldsymbol{s_t}, \boldsymbol{D_t}\right)_{T \geq t \geq t_{D0}} \\ &\quad - \mathbb{E}\left(\Delta\mathbf{y}_{\boldsymbol{t}}^{(\mathbf{0,0})} | \boldsymbol{s_t}, \boldsymbol{D_t}\right)_{t_{D0} > t \geq t_{s0}} + \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t-i}}^{(\mathbf{0,0})} | \boldsymbol{s_t}, \boldsymbol{D_t}\right)_{T \geq t \geq t_{D0}} \\ &\quad - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t-i}}^{(\mathbf{0,0})} | \boldsymbol{s_t}, \boldsymbol{D_t}\right)_{t_{D0} > t \geq t_{s0}} \\ &= \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(\mathbf{1,1})} | \boldsymbol{s_t}, \boldsymbol{D_t}\right)_{T \geq t \geq t_{D0}} - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t-i}}^{(\mathbf{0,0})} | \boldsymbol{s_t}, \boldsymbol{D_t}\right)_{t_{D0} > t \geq t_{s0}} - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(\mathbf{0,0})} - \mathbf{y}_{\boldsymbol{t-i}}^{(\mathbf{0,0})} | \boldsymbol{s_t}, \boldsymbol{D_t}\right)_{T \geq t \geq t_{D0}} \\ &\quad + 0 \end{aligned}$$

$$= \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(1,1)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{T\geq t\geq t_{D0}} - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t-i}}^{(0,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0}>t\geq t_{s0}}$$

$$- \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(0,0)} - \mathbf{y}_{\boldsymbol{t-i}}^{(0,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0}>t\geq t_{s0}}$$

$$= \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(1,1)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{T\geq t\geq t_{D0}} - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(0,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0}>t\geq t_{s0}}$$

for the total effect, similarly we can get

$$\beta\eta = \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(1,0)}|\boldsymbol{s}_t = 1, \boldsymbol{D}_t = 0\right) - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(0,0)}|\boldsymbol{s}_t = 0, \boldsymbol{D}_t = 0\right)$$

$$= \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(1,0)}|\boldsymbol{s}_t = 1, \boldsymbol{D}_t = 0\right) - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(0,0)}|\boldsymbol{s}_t = 1, \boldsymbol{D}_t = 0\right) + \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(0,0)}|\boldsymbol{s}_t = 1, \boldsymbol{D}_t = 0\right)$$

$$- \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(0,0)}|\boldsymbol{s}_t = 0, \boldsymbol{D}_t = 0\right)$$

$$= \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(1,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0}>t\geq t_{s0}} - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(0,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0}>t\geq t_{s0}} + \mathbb{E}\left(\Delta\mathbf{y}_{\boldsymbol{t}}^{(0,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0}>t\geq t_{s0}}$$

$$- \mathbb{E}\left(\Delta\mathbf{y}_{\boldsymbol{t}}^{(0,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{s0}\geq t\geq 1} + \mathbb{E}\left(\Delta\mathbf{y}_{\boldsymbol{t-i}}^{(0,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0}>t\geq t_{s0}}$$

$$- \mathbb{E}\left(\Delta\mathbf{y}_{\boldsymbol{t-i}}^{(0,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{s0}\geq t\geq 1}$$

$$= \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(1,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0}>t\geq t_{s0}} - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t-i}}^{(0,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{s0}>t\geq 1} - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(0,0)} - \mathbf{y}_{\boldsymbol{t-i}}^{(0,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0}>t\geq t_{s0}}$$

$$= \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(1,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0}>t\geq t_{s0}} - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t-i}}^{(0,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{s0}>t\geq 1} - \mathbb{E}\left(\Delta\mathbf{y}_{\boldsymbol{t-i}}^{(0,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{s0}\geq t\geq 1}$$

$$= \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(1,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0}>t\geq t_{s0}} - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(0,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{s0}>t\geq 1}$$

for the structure change effect, and

$$\xi = (\beta\eta + \xi) - \beta\eta$$

$$= \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(1,1)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{T\geq t\geq t_{D0}} - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(0,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0}>t\geq t_{s0}} - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(1,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0}>t\geq t_{s0}}$$

$$+ \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(0,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{s0}>t\geq 1}$$

$$= \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(1,1)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{T\geq t\geq t_{D0}} - \mathbb{E}\left(\mathbf{y}_{\boldsymbol{t}}^{(1,0)}|\boldsymbol{s}_t, \boldsymbol{D}_t\right)_{t_{D0}>t\geq t_{s0}}$$

for the treatment effect by Assumption 5.

Proof of Lemma 1. If we denote $t_0 = t_{s0} - 1$ and $t_1 = t_{D0} - t_{s0}$, then it is trivial to see that

$$\left[(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right]^{-1}(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{w_t}\mathbb{E}(\boldsymbol{\delta_T\delta_\beta})_{s=1,t_{s0}\leq t\leq t_{D0-1}}$$

$$= \frac{1}{t_1}\frac{\sum_{t=t_0+1}^{t_0+t_1}\left(2\boldsymbol{t}\beta_{y2} + \beta_{y1}\right)}{\sum_{t=t_0+1}^{t_0+t_1}\left(2\boldsymbol{t}\beta_{y2} + \beta_{y1}\right)^2}\sum_{t=t_0+1}^{t_0+t_1}\left(2\boldsymbol{t}\beta_{y2} + \beta_{y1}\right)\boldsymbol{w_t}$$

$$= \frac{1}{t_1}\frac{2\beta_{y2}\frac{1}{2}(t_0+1)(t_0+t_1) + t_1\beta_{y1}}{4\beta_{y2}^2\mathbb{Q} + t_1\beta_{y2}^2 + 2\beta_{y1}\beta_{y2}(t_0+t_1)(t_0+1)}\sum_{t=t_0+1}^{t_0+t_1}\left(2t\beta_{y2} + \beta_{y1}\right)\boldsymbol{w_t},$$

where $\mathbb{Q}=\left(\frac{1}{3}((t_0+t_1)^3-(t_0+1)^3)+\frac{1}{2}((t_0+t_1)^2+(t_0+1)^2)+\frac{1}{6}\left((t_0+t_1)-(t_0+1)\right)\right)$.

After simplification, we can get

$$\left[(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right]^{-1}(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{w_t}\mathbb{E}(\boldsymbol{\delta_T\delta_\beta})_{s=1,t_{s0}\le t\le t_{D0-1}}$$

$$=\frac{1}{(t_0+1)t_1}\frac{\beta_{y2}+\frac{t_1}{(t_0+1)(t_0+t_1)}\beta_{y1}}{4\beta_{y2}^2\widetilde{\mathbb{Q}}+\frac{t_1}{(t_0+1)^2(t_0+t_1)}\beta_{y2}^2+2\beta_{y1}\beta_{y2}\frac{1}{t_0+1}}\sum_{t=t_0+1}^{t_0+t_1}(2t\beta_{y2}+\beta_{y1})\,\boldsymbol{w_t}$$

where $\widetilde{\mathbb{Q}}=\left(\frac{1}{3}\left(\left(\frac{t_0+t_1}{t_0+1}\right)^2-\frac{t_0+1}{t_0+t_1}\right)+\frac{1}{2}\left(\frac{t_0+t_1}{(t_0+1)^2}+\frac{1}{t_0+t_1}\right)+\frac{1}{6}\frac{1}{(t_0+1)^2}\right)$. By the conditions of Lemma 1, as $t_{D0}\to\infty$ we have $\left[(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right]^{-1}(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{w_t}\mathbb{E}(\boldsymbol{\delta_T\delta_\beta})_{s=1,t_{s0}\le t\le t_{D0-1}}=o_p(1)$, and similarly as $t_{s0}\to\infty$ we can get $\left[(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right]^{-1}(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{w_t}\mathbb{E}(\boldsymbol{\delta_T\delta_\beta})_{s=1,1\le t\le t_{s0-1}}=o_p(1)$. In the same way, it is easy to show that $\left[(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right]^{-1}(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{x_{t,p}}\mathbb{E}(\boldsymbol{\delta_T\delta_\beta})_{s=1,t_{s0}\le t\le t_{D0-1}}=o_p(1)$ and $\left[(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right]^{-1}(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{x_{t,p}}\mathbb{E}(\boldsymbol{\delta_T\delta_\beta})_{s=1,1\le t\le t_{s0-1}}=o_p(1)$ for $p=1,\dots,P$.

Proof of Theorem 2. (i) We proof consistency. As shown in the main text, when the polynomial order is set as $q=2$, by Assumption 7, equation (34) can be rewritten as

$$\mathbf{y}'_{t,s=1}=\boldsymbol{\delta_T\delta_\beta}+\boldsymbol{\varepsilon}'_{t,y}, \tag{A.1}$$

where $\boldsymbol{\delta_\beta}$ denotes the parameter matrix. Note that if Assumption 6 holds true, we can get

$$\boldsymbol{\varepsilon}'_{t,y}=_{i.i.d.}\mathcal{N}\left(0,\sigma_\varepsilon^2\cdot\boldsymbol{\varepsilon}'_y(\mathbf{0})\right) \tag{A.2}$$

by the delta method. Estimate equation (A.1) through OLS, by (A.2) and the law of large numbers, we then have

$$\hat{\mathbf{y}}'_{t,s=1}=\boldsymbol{\delta_T}(\boldsymbol{\delta}'_T\boldsymbol{\delta_T})^{-1}\boldsymbol{\delta}'_T\mathbf{y}'_{t,s=1}=\boldsymbol{\delta_T}(\boldsymbol{\delta}'_T\boldsymbol{\delta_T})^{-1}\boldsymbol{\delta}'_T(\boldsymbol{\delta_T\delta_\beta}+\boldsymbol{\varepsilon}'_{t,y})=\boldsymbol{\delta_T\delta_\beta}+o_p(1). \tag{A.3}$$

Input (A.3) into (37), under Assumption 2, we can get

$$\begin{aligned}
\hat{\mathbf{y}}_t^{s=1}&=\left(\hat{\mathbf{y}}'_{t,s=1}{}'\hat{\mathbf{y}}'_{t,s=1}\right)^{-1}\hat{\mathbf{y}}'_{t,s=1}{}'\mathbf{y}_t^{s=1}\hat{\mathbf{y}}'_{t,s=1}\\
&=\left(\hat{\mathbf{y}}'_{t,s=1}{}'\hat{\mathbf{y}}'_{t,s=1}\right)^{-1}\hat{\mathbf{y}}'_{t,s=1}{}'\left((\boldsymbol{w_t}+\boldsymbol{s_t}\beta)\eta+\boldsymbol{x_t}\alpha+\boldsymbol{\epsilon_t}+\boldsymbol{\nu_t}\right)\hat{\mathbf{y}}'_{t,s=1}\\
&=\left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{w_t}\eta\boldsymbol{\delta_T\delta_\beta}+\left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{s_t}\beta\eta\boldsymbol{\delta_T\delta_\beta}\\
&\quad+\left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{x_t}\alpha\boldsymbol{\delta_T\delta_\beta}\\
&\quad+\left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}(\boldsymbol{\delta_T\delta_\beta})'(\boldsymbol{\epsilon_t}+\boldsymbol{\nu_t})\boldsymbol{\delta_T\delta_\beta}+o_p(1)\\
&=\left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{w_t}\eta\boldsymbol{\delta_T\delta_\beta}+\left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{s_t}\beta\eta\boldsymbol{\delta_T\delta_\beta}\\
&\quad+\left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}(\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{x_t}\alpha\boldsymbol{\delta_T\delta_\beta}+o_p(1)+o_p(1/(t_{D0}-1))
\end{aligned}$$

for $t = 1, \ldots, t_{s0}, \ldots, t_{D0} - 1$, $s = 1$ denotes first step. By Assumption 5 and Lemma 1, as $t_{D0} \to \infty$ we finally get

$$\widehat{\beta\eta}_{AARs} = \mathbb{E}\left(\hat{\mathbf{y}}_{\boldsymbol{t}}^{s=1}\right)_{t_{s0} \le t \le t_{D0}-1} - \mathbb{E}\left(\hat{\mathbf{y}}_{\boldsymbol{t}}^{s=1}\right)_{1 \le t \le t_{s0}-1}$$

$$= \beta\eta + \eta\Delta_{s=1}(\boldsymbol{w_t}) + \sum_{p=1}^{P} \alpha_p \Delta_{s=1}\left(\boldsymbol{x_{t,p}}\right) + o_p(1)$$

$$= \beta\eta + o_p(1)$$

for the structure change effect, the expectation operator is taken over t.

(ii) We proof unbiasness. Rewrite (28) as

$$\boldsymbol{y_t} = (\boldsymbol{w_t} + \boldsymbol{s_t}\beta)\eta + \boldsymbol{D_t}\xi + \boldsymbol{x_t}\alpha + \boldsymbol{\epsilon_t} + \boldsymbol{\nu_t}, \boldsymbol{w_t} \perp \boldsymbol{D_t}, t = 1, \ldots, t_{s0}, \ldots, t_{D0}, \ldots, T. \qquad (\mathrm{A}.4)$$

where we define $\boldsymbol{y_t} = diag(y_1, \ldots, y_T)$, $\boldsymbol{w_t} = diag(w_1, \ldots, w_T)$, $\boldsymbol{s_t} = diag(s_1, \ldots, s_T)$, $\boldsymbol{D_t} = diag(D_1, \ldots, D_T)$, $\boldsymbol{x_t} = diag(x_1, \ldots, x_T)$, $\boldsymbol{\epsilon_t} = diag(\epsilon_1, \ldots, \epsilon_T)$ and $\boldsymbol{\nu_t} = diag(\nu_1, \ldots, \nu_T)$ w.l.o.g. By Assumption 7, we then have

$$\mathbb{E}\left(\hat{\mathbf{y}}_{\boldsymbol{t}}^{s=1}\right) = \left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}(\boldsymbol{\delta_T\delta_\beta})'\mathbb{E}\left(\boldsymbol{w_t}\eta + \sum_{p=1}^{P}\boldsymbol{x_{t,p}}\alpha_p + \boldsymbol{s_t}\beta\eta\right)\boldsymbol{\delta_T\delta_\beta}$$

$$= \left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}(\boldsymbol{\delta_T\delta_\beta})'\mathbb{E}(\boldsymbol{s_t}\beta\eta)\boldsymbol{\delta_T\delta_\beta}$$

hence,

$$\mathbb{E}\left(\widehat{\beta\eta}_{AARs}\right) = \mathbb{E}\left(\mathbb{E}\left(\hat{\mathbf{y}}_{\boldsymbol{t}}^{s=1}\right)_{t_{s0} \le t \le t_{D0}-1} - \mathbb{E}\left(\hat{\mathbf{y}}_{\boldsymbol{t}}^{s=1}\right)_{1 \le t \le t_{s0}-1}\right)$$

$$= \hat{\mathbf{y}}^{\boldsymbol{s=1}}_{t_{s0} \le t \le t_{D0}-1} - \hat{\mathbf{y}}^{\boldsymbol{s=1}}_{1 \le t \le t_{s0}-1}$$

$$= \left((\boldsymbol{\delta_{t\in\mathcal{T}_1}\delta_\beta})'\boldsymbol{\delta_{t\in\mathcal{T}_1}\delta_\beta}\right)^{-1}\left(\boldsymbol{\delta_{t\in\mathcal{T}_1}\delta_\beta}\right)'\mathbf{1}\beta\eta\boldsymbol{\delta_{t\in\mathcal{T}_1}\delta_\beta}$$

$$- \left((\boldsymbol{\delta_{t\in\mathcal{T}_2}\delta_\beta})'\boldsymbol{\delta_{t\in\mathcal{T}_2}\delta_\beta}\right)^{-1}\left(\boldsymbol{\delta_{t\in\mathcal{T}_2}\delta_\beta}\right)'\mathbf{0}\beta\eta\boldsymbol{\delta_{t\in\mathcal{T}_2}\delta_\beta} = \beta\eta,$$

where $\boldsymbol{\mathcal{T}_1} = \{t_{s0}, \ldots, t_{D0} - 1\}$, $\boldsymbol{\mathcal{T}_2} = \{1, \ldots, t_{s0} - 1\}$, $\mathbf{1} = diag(1, \ldots, 1)_{(t_{D0}-t_{s0})\times(t_{D0}-t_{s0})}$ and $\mathbf{0}_{t_{D0}-t_{s0}} = diag(0, \ldots, 0)_{(t_{s0}-1)\times(t_{s0}-1)}$.

Proof of Lemma 2. The proof is similar to Lemma 1, hence omitted here.

Proof of Corollary 1. The proof is similar to Theorem 2, hence omitted here.

Proof of Theorem 3. From the proof of Theorem 2, by the Lindeberg-Levy CLT theorem, as $T_0 \to \infty$ and $T_1 \to \infty$ we have

$$\sqrt{T_0}\left(\widehat{\beta\eta}_{AARs} - \beta\eta\right)$$

$$= \sqrt{T_0}\left(\frac{1}{T_1}\sum_{t=T_0+1}^{T_0+T_1}\left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}(\boldsymbol{\delta_T\delta_\beta})'(\boldsymbol{\epsilon_t} + \boldsymbol{\nu_t})\boldsymbol{\delta_T\delta_\beta}\right.$$

$$\left. - \frac{1}{T_0}\sum_{t=1}^{T_0}\left((\boldsymbol{\delta_T\delta_\beta})'\boldsymbol{\delta_T\delta_\beta}\right)^{-1}(\boldsymbol{\delta_T\delta_\beta})'(\boldsymbol{\epsilon_t} + \boldsymbol{\nu_t})\boldsymbol{\delta_T\delta_\beta}\right) + o_p(1) \rightsquigarrow \mathcal{N}\left(0, \sigma^2_{\beta\eta}\right).$$

Note that $\boldsymbol{\delta_T\delta_\beta} = \left(2\beta_{y2} + \beta_{y1}, 4\beta_{y2} + \beta_{y1}, \ldots, 2(t_{D0} - 1)\beta_{y2} + \beta_{y1}\right)'$, hence by Assumption 2, we can get

$$Var\left(\sqrt{T_0}\left(\widehat{\beta\eta}_{AARs}-\beta\eta\right)\right)$$

$$=\frac{T_0 Var(\boldsymbol{\epsilon_t}+\boldsymbol{\nu_t})_{t_{s0}\le t\le t_{D0}-1}}{T_1^2}\frac{\left(\sum_{t=T_0+1}^{T_0+T_1}\left(2\boldsymbol{t}\beta_{y2}+\beta_{y1}\right)\right)^2}{\sum_{t=T_0+1}^{T_0+T_1}\left(2\boldsymbol{t}\beta_{y2}+\beta_{y1}\right)^2}$$

$$+\frac{T_0 Var(\boldsymbol{\epsilon_t}+\boldsymbol{\nu_t})_{1\le t\le t_{s0}-1}}{T_0^2}\frac{\left(\sum_{t=1}^{T_0}\left(2\boldsymbol{t}\beta_{y2}+\beta_{y1}\right)\right)^2}{\sum_{t=1}^{T_0}\left(2\boldsymbol{t}\beta_{y2}+\beta_{y1}\right)^2},$$

which is exactly the one shown in matrix form in Theorem 3, the variance operator is taken over t. We can adopt AARs estimators to estimate the error terms $\boldsymbol{\epsilon_t}+\boldsymbol{\nu_t}$: $(\hat{\boldsymbol{\epsilon}}_{\boldsymbol{t}}+\hat{\boldsymbol{\nu}}_{\boldsymbol{t}})_{AARs}=\boldsymbol{y_t}-\boldsymbol{w_t}\hat{\eta}_{AARs}-\boldsymbol{s_t}\widehat{\beta\eta}_{AARs}-\boldsymbol{D_t}\hat{\hat{\xi}}_{AARs}-\boldsymbol{x_t}\hat{\alpha}_{AARs}$, $t=1,\dots,t_{s0},\dots,t_{D0}-1$. Note that under Theorem 2 and Corollary 1, we have $(\hat{\boldsymbol{\epsilon}}_{\boldsymbol{t}}+\hat{\boldsymbol{\nu}}_{\boldsymbol{t}})_{AARs}\rightarrow_p \boldsymbol{\epsilon_t}+\boldsymbol{\nu_t}$ as $T_0\to\infty$, $T_1\to\infty$. The asymptotic behaviors of the estimators $\widehat{\beta\eta+\xi}_{AARs}$, $\hat{\hat{\xi}}_{AARs}$ and $\hat{\alpha}_{AARs}$ can be established in the same way, hence omitted here. The conditions $T_0/T_1\to c_1\in(0,\tau)$, $T_0/T_2\to c_2\in(0,\tau)$ and $T_1/T_2\to c_3\in(0,\tau)$ with $\tau<\infty$ are required to ensure that the asymptotic variances of these distributions will not vanish with sample size.