

Endogenous Kink threshold regression

Jianhan Zhang* Chaoyi Chen^{†‡} Yiguo Sun[§] Thanasis Stengos[¶]

August 26, 2022

Abstract

This paper considers an endogenous kink threshold regression model with an unknown threshold value in a time series as well as a panel data framework, where both the threshold variable and regressors are allowed to be endogenous. We construct our estimators from a control function approach and derive the consistency and asymptotic distribution of our proposed estimators. Monte Carlo simulations are used to assess the finite sample performance of our proposed estimators. Finally, we apply our model to analyze the effects of COVID-19 cases on the labor market of the US and Canada.

Keywords: Control function approach; COVID-19; Endogeneity; Kink regression model; Unemployment rate

JEL Codes: C24, C33, E24

*Department of Economics and Finance, University of Guelph, Guelph, Ontario, N1G 2W1, Canada; Email: jzhang56@uoguelph.ca.

[†]Magyar Nemzeti Bank (Central Bank of Hungary), Budapest, 1054, Hungary; Email: chenc@mnb.hu.

[‡]MNB Institute, John von Neumann University, Kecskemét, 6000, Hungary.

[§]Department of Economics and Finance, University of Guelph, Guelph, Ontario, N1G 2W1, Canada; Email: yisun@uoguelph.ca.

[¶]Department of Economics and Finance, University of Guelph, Guelph, Ontario, N1G 2W1, Canada; Email: tstengos@uoguelph.ca.

1 Introduction

Threshold regression (TR) model is popularly used to capture potential shifts in economic relationships; e.g., [Tong \(1990\)](#) and [Hansen \(2000\)](#). Notwithstanding, the conventional TR model requires the regression function is discontinuous at the true threshold level. But in many empirical applications, there is no reason to expect a discontinuous regression model. As a modification, [Chan and Tsay \(1998\)](#) propose a continuous threshold autoregressive model to allow for a piece-wise linear function of the threshold variable. Notably, this model allows the threshold regression to be continuous. Still, the slope has a discontinuity at the true threshold level and, thus, is widely regarded as a special case of the broad class of threshold autoregressive models. Extending [Chan and Tsay \(1998\)](#), [Hansen \(2017\)](#) provides testing for a threshold effect and inference on the regression parameters for a continuous threshold model with an unknown threshold parameter value (hereinafter, kink threshold regression (KTR) model). It is well established the limiting distribution of the least-squares estimator for the TR model is nonstandard and the estimator is super consistent. For example, [Chan \(1993\)](#) establishes that the threshold parameter estimator converges to a functional of a compound Poisson process. Adopting a “diminishing threshold effect” assumption, [Hansen \(2000\)](#) shows the limiting distribution involves two independent Brownian motions. By contrast, as shown in [Hansen \(2017\)](#), the limiting distribution of the least-squares estimator for the KTR model is normal, and the convergence rate is standard root-n due to the nature of continuity.

All the above studies assume strict exogeneity in both slope regressors and the threshold variable. As many practical issues of nonlinear asymmetric mechanisms are endogenously determined, a growing body of literature has developed for the TR model to allow for endogeneity. Under [Hansen \(2000\)](#)’s diminishing threshold effect framework, [Caner and Hansen \(2004\)](#) allow the slope regressors to be endogenous by using the generalized method of moments (GMM) and two-stage least squares (2SLS) method to estimate the slope parameters and the threshold parameter, respectively. Inspired by the sample selection method of [Heckman \(1979\)](#), [Kourtellos et al. \(2016\)](#) employ a control function (CF) approach to estimate the TR model with endogeneity, where they introduce an inverse Mills ratio as a bias correction term into the regression. Following [Kourtellos et al. \(2016\)](#), [Christopoulos et al. \(2021\)](#) use a copula method to deal with the endogenous threshold variable. [Yu et al. \(2020\)](#) generalize the CF approach of [Kourtellos et al. \(2016\)](#) and classify

two groups of CF methods for the TR model with endogeneity based on the choice of variables in the conditional set. Specifically, the first group is an extension of the 2SLS method proposed by [Caner and Hansen \(2004\)](#), while another is a natural extension to the conventional CF approach documented in [Newey et al. \(1999\)](#). It is worth noticing that both CF methods cannot be directly used to estimate the KTR model with endogeneity. Essentially, even without the endogeneity, the continuity makes the inference of the least squares estimator for the KTR model quite different from the conventional TR model. [Hidalgo et al. \(2019\)](#) underscore that if we wrongly estimate a KTR model with a TR framework of [Hansen \(2000\)](#), ignoring the continuity of the true model, the Hessian matrix becomes irregular¹. This causes the least squares estimator of the threshold parameter converges at a cube root-n convergence rate, slower than the root-n convergence rate for KTR model as shown by [Hansen \(2017\)](#). As a result, both CF methods proposed by [Yu et al. \(2020\)](#), designing for the TR framework, cannot apply to the KTR model without deviation.² More recently, [Kourtellis et al. \(2021\)](#) extend [Yu et al. \(2020\)](#) to allow for the unknown endogenous form by introducing a nonparametric bias correction term into the model. The proposed semiparametric model bypasses any misspecification problem, but still in the framework of the TR model. [Seo and Shin \(2016\)](#) consider a dynamic panel TR model with endogeneity and develop a first-differenced GMM estimator, which allows both threshold variable and regressors to be endogenous. Yet, the GMM method is notorious for its poor small sample performance. Under a fixed threshold effect framework of [Chan \(1993\)](#) and assuming i.i.d. sample, [Yu and Phillips \(2018\)](#) construct an integrated difference kernel estimator(IDKE) for the threshold parameter. The most appealing feature of the IDKE is the consistency of the estimator holds without requiring the instrumental variables. Also, the IDKE is super-consistent for the TR model with endogenous threshold variable and exogenous slope regressors. Nevertheless, the i.i.d. assumption broadly limits the scope of applications for this method.

In contrast to the proliferate studies on the TR model, surprisingly, to our knowledge, no estimation and asymptotic result has been developed for the least squares estimator of the KTR model with endogeneity.³ Thus, this paper aims to fill this gap in the literature. Following [Yu](#)

¹Note that estimating the KTR model under the TR model framework violates the full rank condition that is required for a non-degenerated asymptotic distribution of threshold estimator, see, e.g., [Hansen \(2000\)](#) Assumption 1.7.

²The KTR model violates the Assumption I.9 for CF-I and II.9 for CF-II in the [Yu et al. \(2020\)](#).

³We notice the first-differenced GMM estimator proposed by [Seo and Shin \(2016\)](#) works for the KTR model with endogeneity. However, it is widely regarded that the GMM method has a poorer finite sample performance than

et al. (2020) and Kourtellos et al. (2021), our paper employs the CF approach to correct the endogeneity in a KTR model. Our proposed method allows both slope regressors and the threshold variable to be endogenous. We develop the model both in a time-series and a panel data context. Specifically, we explore the estimation and study the asymptotic properties of the least squares estimator for the time-series model with the weakly dependent data. For the panel, we eliminate the time-invariant fixed effects by using the first-differencing (FD) method and derive the asymptotic results of our proposed estimator with both large numbers of cross-section (N) and time-series (T) observations. Similar to Hansen (2017), our proposed estimator exhibits a joint normal distribution with a standard root- n convergence rate.

We then apply our model to test the threshold effect of COVID-19 cases on the US and Canadian labor markets. The COVID-19 pandemic has been plaguing most economies since early 2020. Since then, much literature has worked on measuring its (linear/nonlinear) effect on the economy. Among others, for example, Karavias et al. (2022) consider a linear panel model with an unknown structural break time to examine the structural effect of COVID-19 on stock returns. They find the COVID-19 pandemic is detrimental to the stock market before the break, and, interestingly, the negative impact vanishes afterward. For the labor market, there is a stream of literature that examines the indirect effect of COVID-19 on the labor market, for example, measuring the impact of the government Stay at Home/Lockdown policy on the labor market (e.g., Baek et al. (2021), Kong and Prinz (2020)). On the other hand, another stream of literature focuses on investigating the effect of COVID-19 on the labor market of some particular groups (e.g., Lee et al. (2021)). Yet, surprisingly, few studies address exploring the effects of COVID-19 on the labor market integration. The fact that the unemployment rates for most advanced economies have recovered to the pre-COVID level while the pandemic is still ongoing indicates the potential nonlinear relationship between the COVID-19 cases and the labor market performance. Considering the nature of long-lasting and multiple waves of COVID-19 cases, we conjecture there is a threshold effect (or a structural break) of COVID-19 cases on the unemployment rate. As such, we apply our proposed KTR model with endogeneity to explore this potential nonlinearity. We demonstrate that, although the impact of COVID-19 on unemployment is significantly positive in both regimes, the magnitude is more prominent if the

the least squares estimator.

case number exceeds a certain level.⁴

The rest of the paper is organized as follows. Section 2 introduces the times-series KTR model with endogeneity, presenting our proposed estimators' estimation method and asymptotic properties. Section 3 extends the model to the panel context. Section 4 reports Monte Carlo simulation results, suggesting our proposed estimator has a good small sample performance. Section 5 provides our empirical application results, while section 6 concludes the paper. We relegate all the mathematical proofs to the Appendix.

To proceed, we adopt the following notation throughout the paper. We use subscript 0 to denote the true parameters and the accent $\hat{\cdot}$ to denote the estimators. We define $\|\cdot\|$ as the Euclidean norm. The operators \xrightarrow{p} and \xrightarrow{d} denote convergence in probability and distribution, respectively. We denote $(N, T) \rightarrow \infty$ as the joint convergence of N and T , when N and T pass to infinity simultaneously. $\mathbf{0}_{A \times B}$ denotes a $A \times B$ matrix of ones and I_m denotes identical matrix of size m .

2 Time series model

2.1 Model and estimation

Following Hansen (2017), we consider a KTR model

$$y_t = \beta_{10}(x_t - \gamma_0)I(x_t < \gamma_0) + \beta_{20}(x_t - \gamma_0)I(x_t \geq \gamma_0) + \beta'_{30}z_t + u_t, t = 1, \dots, n, \quad (2.1)$$

where x_t is the threshold variable, a scalar. $I(\cdot)$ is the indicator function and z_t is an $\ell \times 1$ vector of regressors, including an intercept term. Model (2.1) has $k = 3 + \ell$ parameters to be estimated, including an unknown threshold value γ_0 , which is an interior point of a compact set, Γ . Denote the true value $\beta_0 = (\beta_{10}, \beta_{20}, \beta'_{30})'$ and we have $\beta \in \mathcal{B} \subset \mathcal{R}^{k-1}$, which are both $(k-1) \times 1$ vectors.

In the kink threshold regression framework, we allow both an endogenous threshold variable x_t and endogenous regressors z_{1t} , where z_{1t} is a $d_{z1} \times 1$ vector and it is a subset of $z_t = [z'_{1t}, z'_{2t}]'$. The

⁴Considering the number of COVID-19 tests performed is highly associated with the cases and is irrelevant to the unemployment rate, we use the number of COVID-19 tests performed as the instrumental variable.

reduced form equations of x_t and z_{1t} are

$$x_t = \pi'_{x0} p_{xt} + v_{xt} \quad (2.2)$$

$$z_{1t} = \pi'_{z0} p_{zt} + v_{zt} \quad (2.3)$$

where p_{xt} and p_{zt} allow to have duplicate variables, p_{xt} is a $d_{px} \times 1$ vector with $d_{px} \geq 1$ and p_{zt} is a $d_{pz} \times 1$ vector with $d_{pz} \geq d_{z1}$. To simplify notation, we denote all instrumental variables as p_t , which includes the non-overlapping terms in p_{xt} and p_{zt} and p_{xt} and p_{zt} are allowed to share common variables. The endogeneity of the threshold variable x_t and regressors z_{1t} come from the contemporaneous correlation between u_t and v_t , where $v_t = [v_{xt}, v'_{zt}]'$ is a $(1 + d_{z1}) \times 1$ vector and $\text{Cov}(v_{xt}, v_{zt}) \neq 0$. Using the control function approach, we assume $E(u_t | \mathcal{F}_{t-1}, x_t, z_{1t}) = E(u_t | v_t) = \beta'_{40} v_t$ almost surely, where \mathcal{F}_t is the smallest sigma-field generated from $\{(x_s, z_{1s}, z_{2,s+1}, u_s, p_{s+1}) : 1 \leq s \leq t \leq n\}$ and β_{40} is a $(1 + d_{z1}) \times 1$ vector. Therefore, we have

$$E(y_t | \mathcal{F}_{t-1}, x_t, z_{1t}) = \beta_{10}(x_t - \gamma_0)I(x_t < \gamma_0) + \beta_{20}(x_t - \gamma_0)I(x_t \geq \gamma_0) + \beta'_{30} z_t + \beta'_{40} v_t. \quad (2.4)$$

Let $\delta_0 = \beta_{20} - \beta_{10}$. We can rewrite model (2.1) as

$$y_t = \beta_{10}(x_t - \gamma_0) + \delta_0(x_t - \gamma_0)I(x_t \geq \gamma_0) + \beta'_{30} z_t + \beta'_{40} v_t + \varepsilon_t, \quad (2.5)$$

where $\varepsilon_t = u_t - \beta'_{40} v_t$. Note that, since $E(\varepsilon_t | x_t, z_{1t}, \mathcal{F}_{t-1}) = 0$ almost surely, the integrated model (2.5) is free of the endogenous problem. Thus, we can estimate model (2.5) by the least squares method.

Below, we outline the steps that are taken in the estimation procedure for model (2.5).

First step: Applying the OLS estimation to model (2.2) and (2.3), we obtain the least squares estimator $\hat{\pi}_x = (\sum_{t=1}^n p_{xt} p'_{xt})^{-1} \sum_{t=1}^n p_{xt} x_t$, $\hat{\pi}_z = (\sum_{t=1}^n p_{zt} p'_{zt})^{-1} \sum_{t=1}^n p_{zt} z_t$ and collect the residuals $\hat{v}_{xt} = x_t - \hat{\pi}'_{xt} p_{xt}$, $\hat{v}_{zt} = z_t - \hat{\pi}'_{zt} p_{zt}$. Then we have $\hat{v}_t = [\hat{v}_{xt}, \hat{v}'_{zt}]'$.

Second step: Let $\theta = (\beta_1, \delta, \beta'_3, \beta'_4)'$, which is a $(k + d_{z1}) \times 1$ vector. Then, by replacing v_t

with \hat{v}_t in (2.5), the least squares objective function of model (2.5) becomes

$$S_n(\theta, \gamma) = \frac{1}{n} \sum_{t=1}^n [y_t - \beta_1(x_t - \gamma) - \delta(x_t - \gamma)I(x_t \geq \gamma) - \beta_3'z_t - \beta_4'\hat{v}_t]^2, \quad (2.6)$$

and the least squares estimator of model (2.5) solves the following optimization problem:

$$(\hat{\theta}, \hat{\gamma}) = \underset{(\theta, \gamma) \in B \times \Gamma}{\operatorname{argmin}} S_n(\theta, \gamma). \quad (2.7)$$

Note that $S_n(\theta, \gamma)$ is non-smooth in γ . Therefore, we use a grid search method empirically. For a given $\gamma \in \Gamma$, we obtain the conditional least squares estimator of θ

$$\hat{\theta}(\gamma) = [X(\gamma)'X(\gamma)]^{-1} X(\gamma)'y, \quad (2.8)$$

where $y = [y_1, y_2, \dots, y_n]'$, $X(\gamma) = [x_1(\gamma), x_2(\gamma), \dots, x_n(\gamma)]'$, and $x_t(\gamma) = [x_t - \gamma, (x_t - \gamma)I(x_t \geq \gamma), z_t', \hat{v}_t']'$ for $t = 1, \dots, n$.

Next, we substitute θ by $\hat{\theta}(\gamma)$ into $S_n(\theta, \gamma)$ and obtain the least squares estimator of γ_0 as

$$\hat{\gamma} = \underset{\gamma \in \Gamma}{\operatorname{argmin}} S_n(\hat{\theta}(\gamma), \gamma) = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \frac{1}{n} [y - X(\gamma)\hat{\theta}(\gamma)]'[y - X(\gamma)\hat{\theta}(\gamma)]. \quad (2.9)$$

Then, the least squares estimator for θ_0 is given by $\hat{\theta} = \hat{\theta}(\hat{\gamma})$.

2.2 Assumptions and limiting results

Below, we list regularity assumptions used to derive the consistency and asymptotic distribution of our proposed estimators.

Assumptions-time series. For some $r > 1$,

T1. (y_t, x_t, z_t, p_t) is a strictly stationary, ergodic, and absolutely regular sequence with mixing coefficients $\alpha(m) = O(m^{-\xi})$ for some $\xi > r/(r-1)$;

T2. (a) $E|y_t|^{4r} < \infty$, $E|x_t|^{4r} < \infty$, $E\|z_t\|^{4r} < \infty$; (b) $E\|v_t\|^{4r} < \infty$, and $E\|p_t\|^{4r} < \infty$, $E(p_t p_t')$ is nonsingular;

T3. $\inf_{r \in \Gamma} \det Q(\gamma) > 0$, where $Q(\gamma) = E[x_t^*(\gamma)x_t^{*'}(\gamma)]$, and $x_t^*(\gamma)$ equals $x_t(\gamma)$ with \hat{v}_t being replaced with v_t ;

- T4. x_t has a density function $f(x)$ and $f(x) \leq \bar{f} < \infty$ over its domain for some finite constant \bar{f} ;
- T5. (a) $E(u_t|\mathcal{F}_{t-1}, x_t, z_t) = E(u_t|v_t) = \beta'_{40}v_t$ almost surely for all t , where \mathcal{F}_t is the smallest sigma-field generated from $\{(x_s, z_s, u_s, p_{s+1}) : 1 \leq s \leq t \leq n\}$; (b) $\{(v_t, \mathcal{F}_{t-1})\}$ is a martingale difference sequence with $E(v_t|\mathcal{F}_{t-1}) = 0$ almost surely;
- T6. $\delta_0 \neq 0$ and $\theta \in B \subset \mathcal{R}^{k+d_{z1}}$, where B is compact;
- T7. $\gamma_0 = \underset{\gamma \in \Gamma}{\operatorname{argmin}} L^*(\theta^*(\gamma), \gamma)$ is unique, where $\theta^*(\gamma) = E[x_t^*(\gamma)x_t^{*'}(\gamma)]^{-1}E[x_t^*(\gamma)y_t]$, $L^*(\theta, \gamma) = E[S_n^*(\theta, \gamma)]$, $S_n^*(\theta, \gamma)$ equals $S_n(\theta, \gamma)$ with \hat{v}_t being replaced with v_t , and Γ is compact.

In Assumptions T1, we assume a β -mixing sequence, where the choice of r involves a trade-off between the allowable degree of serial dependence and the number of finite moments; see discussions given in Remark 2.3 of [Chan and Tsay \(1998\)](#) and Assumption 1.1 of [Hansen \(2017\)](#). Assumption T2 contains unconditional moment conditions. Assumption T2(a) is the regular moment conditions required and T2(b) and T5(b) ensure that the OLS estimators of the reduced form models (2.2)-(2.3) exist and converge to the true parameter vector at the root-n rate. Assumption T3 ensures that the parameter estimation is well defined for all $\gamma \in \Gamma$. Assumption T4 makes sure our x_t has a bounded density function. Assumption T5(a) is the assumption for a linear endogenous structure, which can be easily extended to a non-linear endogenous structure. By Assumption T6, we consider a kink regression model. Assumption T7 is an identification assumption, similar to Assumption 2.1 of [Hansen \(2017\)](#).

Next, denote $\phi = (\theta', \gamma)'$, a $(k + 1 + d_{z1}) \times 1$ vector, and Let $H_t = H_t(\phi_0)$ with

$$H_t(\phi) = -\frac{\partial}{\partial \phi} [y_t - x_t^*(\gamma)\theta] = \begin{pmatrix} x_t^*(\gamma) \\ -\beta_1 - \delta I(x_t \geq \gamma) \end{pmatrix}. \quad (2.10)$$

Below, we present the limiting results of our proposed estimator.

Theorem 1-Time Series. Under Assumptions T1-T7, as $n \rightarrow \infty$, we have

$$\sqrt{n}(\hat{\phi} - \phi_0) \xrightarrow{d} N(0, V), \quad (2.11)$$

where $V = Q^{-1}SQ^{-1}$, $S = E(H_t H_t' \varepsilon_t^2)$, and $Q = E(H_t H_t')$, here $E(H_t H_t')$ is a $(k + 1 + d_{z1}) \times (k + 1 + d_{z1})$ matrix.

Remark 1: The proof of Theorem 1-Time series is given in the appendix. The slope and threshold estimators converge at the square-root- n rate and are jointly normally distributed with a non-zero asymptotic covariance matrix. By contrast, for the discontinuous TR model, the threshold estimator converges faster than square-root- n , and the distribution is non-standard distributed. Thus, the TR model's threshold estimator is asymptotically independent of the slope estimators. These stark differences originate from the continuous nature of the KTR function. To make inference, we suggest to use the following as the estimator for the asymptotic variance-covariance matrix

$$\hat{V} = \hat{Q}^{-1} \hat{S} \hat{Q}^{-1},$$

where $\hat{Q} = n^{-1} \sum_{t=1}^n \hat{H}_t(\hat{\phi}) \hat{H}_t'(\hat{\phi})$, $\hat{S} = n^{-1} \sum_{t=1}^n \hat{H}_t(\hat{\phi}) \hat{H}_t'(\hat{\phi}) \hat{\varepsilon}_t^2(\hat{\phi})$ with $\hat{\varepsilon}_t(\hat{\phi}) = y_t - \hat{\beta}_1(x_t - \hat{\gamma}) - \hat{\delta}(x_t - \hat{\gamma})I(x_t \geq \hat{\gamma}) - \hat{\beta}_3'z_t - \hat{\beta}_4'\hat{v}_t$, and

$$\hat{H}_t(\phi) = -\frac{\partial}{\partial \phi} [y_t - x_t'(\gamma)\theta] = \begin{pmatrix} x_t(\gamma) \\ -\beta_1 - \delta I(x_t \geq \gamma) \end{pmatrix}. \quad (2.12)$$

3 Panel model extension

Many empirical problems of nonlinear asymmetric mechanisms are explicitly within a panel data context, including but not limited to the potential threshold effect of COVID-19 on the unemployment rate that we will discuss more in section 5. Therefore, we extend our baseline time-series model to an endogenous kink threshold panel model with unknown fixed-effects and cross-sectional independence. Below, we present our model, the estimation strategy, and the asymptotic results.

3.1 Model and estimation

We consider the panel data with both sufficiently large number of cross sectional units N and number of time periods T . To remove the time-invariant fixed effects, we apply the first-differencing method, twisting from the within-transformation that used in [Zhang et al. \(2017\)](#). Our panel kink

threshold regression model is as follows

$$y_{it} = \beta_{10}(x_{it} - \gamma_0)I(x_{it} < \gamma_0) + \beta_{20}(x_{it} - \gamma_0)I(x_{it} \geq \gamma_0) + \beta'_{30}z_{it} + b_i + u_{it}, \quad (3.1)$$

for $i = 1, \dots, N, t = 1, \dots, T$, where y_{it} is the dependent variable, x_{it} is a scalar threshold variable, z_{it} is an $\ell \times 1$ vector of time varying regressors, which may include the time effect. b_i is the i^{th} unobserved individual effect, which is independent of the errors u_{it} for all t . We denote $\beta_0 = (\beta_{10}, \beta_{20}, \beta'_{30})' \in R^{k-1}$, where $k = 3 + \ell$. The unknown threshold value γ_0 is an interior point of a compact set, Γ . Again, we have the endogenous threshold variable and endogenous regressors $z_{1,it}$, where $z_{1,it}$ is a $d_{z1} \times 1$ vector and it is a subset of $z_{it} = [z'_{1,it}, z'_{2,it}]'$. The reduced form equations of x_{it} and $z_{1,it}$ are given by

$$x_{it} = \Pi'_{x0}p_{x,it} + v_{x,it}, \quad (3.2)$$

$$z_{1,it} = \Pi'_{z0}p_{z,it} + v_{z,it}, \quad (3.3)$$

where $p_{x,it}$ and $p_{z,it}$ allow to have common variables, $p_{x,it}$ is a $d_{px} \times 1$ vector with $d_{px} \geq 1$ and $p_{z,it}$ is a $d_{pz} \times 1$ vector with $d_{pz} \geq d_{z1}$. To simplify notation, we denote all instrumental variables by p_{it} , including $p_{x,it}$, $p_{z,it}$ and $v_{it} = [v_{x,it}, v'_{z,it}]'$, a $(1 + d_{z1}) \times 1$ vector. In addition, we allow $\text{Cov}(v_{x,it}, v_{z,it}) \neq 0$. Using the control function approach, for each i , we assume $E(u_{it}|\mathcal{F}_{i,t-1}, x_{it}, z_{1,it}) = E(u_{it}|v_{it}) = \beta'_{40}v_{it}$ almost surely, where $\mathcal{F}_{i,t}$ is the smallest sigma-field generated from $\{(x_{is}, z_{1,is}, z_{2,i,s+1}, u_{is}, p_{i,s+1}) : 1 \leq s \leq t \leq T\}$ and β_{40} is a $(1 + d_{z1}) \times 1$ vector. The endogeneity of the threshold variable x_{it} and regressors $z_{1,it}$ come from the contemporaneous correlation between u_{it} and v_{it} .

Applying the first-differencing to model (3.1) and denoting $\delta_0 = \beta_{20} - \beta_{10}$ yields

$$\Delta y_{it} = \beta_{10}\Delta x_{it} + \delta_0(X_{it} - \gamma_0\tau_2)'I_{it}(\gamma_0) + \beta'_{30}\Delta z_{it} + \Delta u_{it}, \quad (3.4)$$

where $\Delta a_{it} = a_{it} - a_{i,t-1}$ denotes the first difference of variable a , τ_m is an $m \times 1$ vector of ones, and

$$X_{it} - \gamma_0\tau_2 = \begin{pmatrix} x_{it} - \gamma_0 \\ x_{i,t-1} - \gamma_0 \end{pmatrix} \quad \text{and} \quad I_{it}(\gamma_0) = \begin{pmatrix} I(x_{it} \geq \gamma_0) \\ -I(x_{i,t-1} \geq \gamma_0) \end{pmatrix}.$$

Next, we have

$$E(\Delta y_{it} | \mathcal{F}_{i,t-2}, x_{it}, x_{i,t-1}, z_{it}, z_{1,i,t-1}, p_{it}) = \beta_{10} \Delta x_{it} + \delta_0 (X_{it} - \gamma_0 \tau_2)' \mathbf{I}_{it}(\gamma_0) + \beta'_{30} \Delta z_{it} + \beta'_{40} \Delta v_{it}, \quad (3.5)$$

where applying the law of iterative expectation and using the reduced form equations (3.2) and (3.3) gives

$$\begin{aligned} & E(u_{it} | \mathcal{F}_{i,t-2}, x_{it}, x_{i,t-1}, z_{it}, z_{1,i,t-1}, p_{it}) \\ &= E[E(u_{it} | \mathcal{F}_{i,t-1}, x_{it}, z_{1,it}) | \mathcal{F}_{i,t-2}, x_{it}, x_{i,t-1}, z_{it}, z_{1,i,t-1}, p_{it}] \\ &= \beta'_{40} E(v_{it} | \mathcal{F}_{i,t-2}, x_{it}, x_{i,t-1}, z_{it}, z_{1,i,t-1}, p_{it}) = \beta'_{40} v_{it} \end{aligned} \quad (3.6)$$

and $E(u_{i,t-1} | \mathcal{F}_{i,t-2}, x_{it}, x_{i,t-1}, z_{it}, z_{1,i,t-1}, p_{it}) = E(u_{i,t-1} | \mathcal{F}_{i,t-2}, x_{i,t-1}, z_{1,i,t-1}) = \beta'_{40} v_{i,t-1}$, since future information does not affect past information.

Thus, combining (3.4) with (3.5) gives

$$\Delta y_{it} = \beta_{10} \Delta x_{it} + \delta_0 (X_{it} - \gamma_0 \tau_2)' \mathbf{I}_{it}(\gamma_0) + \beta'_{30} \Delta z_{it} + \beta'_{40} \Delta v_{it} + \Delta \varepsilon_{it}, \quad (3.7)$$

where $\Delta \varepsilon_{it} = \Delta u_{it} - \beta'_{40} \Delta v_{it}$ and $E(\Delta \varepsilon_{it} | \mathcal{F}_{i,t-2}, x_{it}, x_{i,t-1}, z_{it}, z_{1,i,t-1}, p_{it}) = 0$. Hence, by including the auxiliary regressor Δv_{it} into the regression, our model (3.7) has no endogenous issue.

Next, we proceed to show the estimation strategy

First step: Applying the OLS estimation to model (3.2) and (3.3), we obtain the least squares estimators:

$$\hat{\Pi}_x = \left(\sum_{i=1}^N \sum_{t=1}^T p_{x,it} p'_{x,it} \right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T p_{x,it} x_{it} \right), \quad \hat{\Pi}_z = \left(\sum_{i=1}^N \sum_{t=1}^T p_{z,it} p'_{z,it} \right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T p_{z,it} z_{1,it} \right).$$

Then, we collect the residuals $\Delta \hat{v}_{x,it} = \Delta x_{it} - \Delta p'_{x,it} \hat{\Pi}_x$ and $\Delta \hat{v}_{z,it} = \Delta z_{1,it} - \Delta p'_{z,it} \hat{\Pi}_z$. Let $\hat{v}_{it} = [\hat{v}_{x,it}, \hat{v}_{z,it}]'$.

Second step: Let $\theta = (\beta_1, \delta, \beta'_3, \beta'_4)' \in R^{k+d_{x1}}$, which is a $(k + d_{z1}) \times 1$ vector. Replacing Δv_{it}

by $\Delta\hat{v}_{it}$ in (3.5), we obtain the least squares criterion function

$$S_{NT}(\theta, \gamma) = \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T [\Delta y_{it} - \beta_1 \Delta x_{it} - \delta(X_{it} - \gamma\tau_2) \mathbf{I}_{it}(\gamma) - \beta_3' \Delta z_{it} - \beta_4' \Delta\hat{v}_{it}]^2. \quad (3.8)$$

Our least square estimator is the joint minimizer of $S_{NT}(\theta, \gamma)$,

$$(\hat{\theta}, \hat{\gamma}) = \underset{(\theta, \gamma) \in B \times \Gamma}{\operatorname{argmin}} S_{NT}(\theta, \gamma). \quad (3.9)$$

For a given $\gamma \in \Gamma$, we get the conditional least squares estimator of θ ,

$$\hat{\theta}(\gamma) = \underset{\theta \in B}{\operatorname{argmin}} \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T (\Delta y_{it} - \Delta x'_{it}(\gamma) \theta)^2, \quad (3.10)$$

where $\Delta x_{it}(\gamma) = [\Delta x_{it}, (X_{it} - \gamma\tau_2)' \mathbf{I}_{it}(\gamma), \Delta z'_{it}, \Delta\hat{v}'_{it}]'$.

By solving (3.10), we have

$$\hat{\theta}(\gamma) = \left[\sum_{i=1}^N \sum_{t=2}^T \Delta x_{it}(\gamma) \Delta x'_{it}(\gamma) \right]^{-1} \left[\sum_{i=1}^N \sum_{t=2}^T \Delta x_{it}(\gamma) \Delta y_{it} \right]. \quad (3.11)$$

Empirically, we can use a grid search method to obtain $\hat{\gamma}$ by minimizing the sum squared error criterion function

$$\hat{\gamma} = \underset{\gamma \in \Gamma}{\operatorname{argmin}} S_{NT}(\hat{\theta}(\gamma), \gamma) = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T [\Delta y_{it} - \Delta x'_{it}(\gamma) \hat{\theta}(\gamma)]^2. \quad (3.12)$$

Then, we obtain the estimator of θ_0 with $\hat{\theta} = \hat{\theta}(\hat{\gamma})$.

3.2 Assumptions and limiting results

The assumptions needed for the panel model and its asymptotic theory are collected below.

Assumptions-panel. For some $r > 1$,

P1. (a) $\{(y_{it}, x_{it}, z_{it}, p_{it}) : t = 1, 2, \dots\}$ are independently identically distributed (i.i.d.) across index i ; (b) for each i , $\{(y_{it}, x_{it}, z_{it}, p_{it}) : t = 1, 2, \dots\}$ is a strictly stationary, ergodic, and absolutely regular sequence with mixing coefficients $\alpha(m) = O(m^{-\xi})$ for some $\xi > r/(r-1)$;

P2. (a) $E|y_{it}|^{4r} < \infty$, $E|x_{it}|^{4r} < \infty$, $E\|z_{it}\|^{4r} < \infty$; (b) $E\|v_{it}\|^{4r} < \infty$, and $E\|p_{it}\|^{4r} < \infty$, $E(p_{it}p'_{it})$

is non-singular;

P3. $\inf_{\gamma \in \Gamma} \det Q(\gamma) > 0$, where $Q(\gamma) = E[\Delta x_{it}^*(\gamma) \Delta x_{it}^{*'}(\gamma)]$ and Δx_{it}^* equals $\Delta x_{it}(\gamma)$ with $\Delta \hat{v}_{it}$ being replaced with Δv_{it} ;

P4. x_{it} has a density function $f(x)$ and $f(x) \leq \bar{f} < \infty$ over its domain for a finite real number \bar{f} ;

P5. For each i , (a) $\{v_{it}, \mathcal{F}_{i,t-1}\}$ is a martingale difference sequence with $E(v_{it} | \mathcal{F}_{i,t-1}) = 0$, where \mathcal{F}_{it} is the smallest sigma-field generated from $\{(x_{is}, z_{1,is}, z_{2,i,s+1}, u_{is}, p_{i,s+1}) : 1 \leq s \leq t \leq T\}$; (b) $E(u_{it} | \mathcal{F}_{i,t-1}, x_{it}, z_{1,it}) = E(u_{it} | v_{it}) = \beta'_{40} v_{it}$ almost surely; (c) a_t and b_i are independent of the error term u_{it} for all t ;

P6. $\delta_0 \neq 0$ and $\theta \in B \subset \mathcal{R}^{k+d_z1}$, where B is compact;

P7. $\gamma_0 = \underset{\gamma \in \Gamma}{\operatorname{argmin}} L^*(\theta^*(\gamma), \gamma)$ is unique, where $\theta^*(\gamma) = E[\Delta x_{it}^*(\gamma) \Delta x_{it}^{*'}(\gamma)]^{-1} E[\Delta x_{it}^*(\gamma) \Delta y_{it}]$, $L^*(\theta^*(\gamma), \gamma) = E[S_{NT}^*(\theta^*(\gamma), \gamma)]$, $S_{NT}^*(\theta, \gamma)$ equals $S_{NT}(\theta, \gamma)$ with $\Delta \hat{v}_{it}$ being replaced with Δv_{it} , and Γ is compact.

In Assumption P1(a), we assume independency across index i . Assumption P1(b) assumes a β -mixing sequence across index t , where the choice of r involves a trade-off between the allowable degree of serial dependence and the number of finite moments. And the asymptotics are taken in large N and large T . Note [Zhang et al. \(2017\)](#) only allow N goes to infinity and treat T as fixed. Assumption P2 contains unconditional moment conditions. Assumption P2(a) is the regularity moment conditions required and P2(b) and P5(b) ensure that the OLS estimators of the reduced form models (3.2)-(3.3) exist and converge to the true parameter vector at the root- NT rate. Assumption P3 ensures that the parameter estimation is well defined for all $\gamma \in \Gamma$. Assumption P4 makes sure our x_{it} has a bounded density function. Assumption P5(a) is the assumption for a linear endogenous structure, which can be easily extended to a non-linear endogenous structure. Assumption P5(c) assumes the unobserved individual effect b_i and unobserved time fixed effect a_t are independent of the errors u_{it} for all t , which is a standard assumption in panel data model. By Assumption P6, we consider a kink regression model. Assumption P7 is an identification assumption, similar to Assumption 2.1 of [Hansen \(2017\)](#).

Denote $\phi = (\theta', \gamma)'$ and let

$$\Delta H_{it}(\phi) = -\frac{\partial}{\partial \phi} [\Delta y_{it} - \Delta x_{it}^{*'}(\gamma) \theta] = \begin{pmatrix} \Delta x_{it}^*(\gamma) \\ -\delta [I(x_{it} \geq \gamma) - I(x_{it-1} \geq \gamma)] \end{pmatrix}, \quad (3.13)$$

and $\Delta H_{it} = \Delta H_{it}(\phi_0)$.

Theorem 1-panel. Under Assumptions P1-P7, as $(N, T) \rightarrow \infty$, we have

$$\sqrt{NT} \left(\hat{\phi} - \phi_0 \right) \xrightarrow{d} N(0, \mathcal{V}), \quad (3.14)$$

where $\mathcal{V} = \mathcal{Q}^{-1} \mathcal{S} \mathcal{Q}^{-1}$, $\mathcal{S} = E(\Delta H_{it} \Delta H'_{it} \Delta \varepsilon_{it}^2)$, $\mathcal{Q} = E(\Delta H_{it} \Delta H'_{it})$, and $E(\Delta H_{it} \Delta H'_{it})$ is a $(k+1+d_{z1}) \times (k+1+d_{z1})$ matrix.

Remark 2: The proof is provided in the appendix. Similar to the time-series model, our slope and threshold estimators are jointly normally distributed with root- NT convergence rate and they have a non-zero asymptotic covariance matrix. To make inference, we estimate the asymptotic variance covariance matrix by

$$\hat{\mathcal{V}} = \hat{\mathcal{Q}}^{-1} \hat{\mathcal{S}} \hat{\mathcal{Q}}^{-1}$$

where $\hat{\mathcal{Q}} = \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T \Delta \hat{H}_{it}(\hat{\phi}) \Delta \hat{H}'_{it}(\hat{\phi})$ and $\hat{\mathcal{S}} = \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T \Delta \hat{H}_{it}(\hat{\phi}) \Delta \hat{H}'_{it}(\hat{\phi}) \Delta \hat{\varepsilon}_{it}^2(\hat{\phi})$. Here $\Delta \hat{\varepsilon}_{it}(\hat{\phi}) = \Delta y_{it} - \hat{\beta}_1 \Delta x_{it} - \hat{\delta}(X_{it} - \hat{\gamma} \tau_2) \mathbf{I}_{it}(\hat{\gamma}) - \hat{\beta}'_3 \Delta z_{it} - \hat{\beta}'_4 \Delta \hat{v}_{it}$ and

$$\Delta \hat{H}_{it}(\phi) = -\frac{\partial}{\partial \phi} [\Delta y_{it} - \Delta x'_{it}(\gamma) \theta] = \begin{pmatrix} \Delta x_{it}(\gamma) \\ -\delta[I(x_{it} \geq \gamma) - I(x_{it-1} \geq \gamma)] \end{pmatrix}. \quad (3.15)$$

4 Monte Carlo simulations

This section contains Monte Carlo simulations to evaluate the finite sample performance of our proposed estimator. Below, we list the four data generating processes (DGPs)—two give time series data and two yield panel data.

DGP1:

$$y_t = c_0 + \beta_{10} x_t + \delta_0(x_t - \gamma_0) I(x_t \geq \gamma_0) + u_t, \quad u_t = 0.1 \varepsilon_t + \kappa v_t, \quad (4.1)$$

$$x_t = 2 + v_t + p_t, \quad t = 1, \dots, n. \quad (4.2)$$

DGP2:

$$y_t = c_0 + \beta_{10}x_t + \delta_0(x_t - \gamma_0)I(x_t \geq \gamma_0) + \beta_{30}z_t + u_t, \quad u_t = 0.1\varepsilon_t + \kappa(v_{1t} + v_{2t}), \quad (4.3)$$

$$x_t = 2 + (0.9v_{1t} + 0.1v_{2t}) + p_{1t}, \quad z_t = 2 + (0.1v_{1t} + 0.9v_{2t}) + p_{2t}, \quad (4.4)$$

$$t = 1, \dots, n.$$

DGP3:

$$y_{it} = c_0 + \beta_{10}x_{it} + \delta_0(x_{it} - \gamma_0)I(x_{it} \geq \gamma_0) + b_i + u_{it}, \quad u_{it} = 0.1v_{it} + \kappa v_{it}, \quad (4.5)$$

$$x_{it} = 2 + v_{it} + p_{it}, \quad u_{it} = 0.1v_{it} + \kappa v_{it}, \quad i = 1, \dots, N, t = 1, \dots, T. \quad (4.6)$$

DGP4:

$$y_{it} = c_0 + \beta_{10}x_{it} + \delta_0(x_{it} - \gamma_0)I(x_{it} \geq \gamma_0) + \beta_{30}z_{it} + b_i + u_{it}, \quad u_{it} = 0.1v_{it} + \kappa(v_{1,it} + v_{2,it}), \quad (4.7)$$

$$x_{it} = 2 + (0.9v_{1,it} + 0.1v_{2,it}) + p_{1t} \quad z_{it} = 2 + (0.1v_{1,it} + 0.9v_{2,it}) + p_{2t} \quad (4.8)$$

$$i = 1, \dots, N, t = 1, \dots, T.$$

In the time-series setup, we consider two different data generating processes, DGP1, and DGP2. In DGP1, we only allow the threshold variable to be endogenous, while in DGP2, we allow both the threshold variable x_t and slope regressor z_t to be endogenous. The endogeneity of x_t in DGP1 comes from the common factor v_t between x_t and u_t . In DGP2, the endogeneity of (x_t, z_t) comes from the common factors v_{1t} and v_{2t} shared with u_t . DGP3 and DGP4 are designed for the panel KTR context. Specifically, DGP3 allows the threshold variable x_t to be endogenous, and DGP4 allows both the threshold variable and regressors to be endogenous. In DGP3, the endogeneity of x_{it} comes from the common factor v_{it} between x_{it} and u_{it} . In DGP4, the endogeneity of (x_{it}, z_{it}) comes from the common factors $v_{1,it}$ and $v_{2,it}$ shared with u_{it} . For all data generating processes, we use κ to control the severity of endogeneity and we set $c_0 = \beta_{10} = \delta_0 = \beta_{30} = 1$, and $\gamma_0 = 2$.

In DGP1, $(v_t, p_t, \varepsilon_t) \sim i.i.d.N(0, I_3)$, where p_t is our instrumental variable. In DGP2,

$(v_{1t}, v_{2t}, p_{1t}, p_{2t}, \varepsilon_t) \sim i.i.d.N(0, I_4)$ and p_{1t} and p_{2t} are the instrumental variable for x_t and z_t , respectively. In DGP3, we generate $(v_{it}, p_{it}, \varepsilon_{it}) \sim i.i.d.N(0, I_3)$, and the unknown fixed effects, $b_i \sim i.i.d.N(0, 1)$. In DGP4, we have $(v_{1,it}, v_{2,it}, p_{1,it}, p_{2,it}, \varepsilon_{it}) \sim i.i.d.N(0, I_5)$ and b_i represents individual fixed effects with distribution $N(0, 1)$ across i . With $\kappa \in \{0.05, 0.5, 0.95, 2\}$, we check the performance of our estimator under low, moderate, and high endogenous severity. We set the sample size $n = 100, 200, 300$, and 400 for GDP1 and DGP2, and $N = 10, 20, 30$, and 40 and $T = 10, 20, 30$, and 40 for DGP3 and DGP4. The number of Monte Carlo replications is 5,000. Tables 2, 3, 4 and 5 report the root mean squared errors (RMSEs) for our proposed estimator for DGP1, DGP2, DGP3, and DGP4, respectively. To save space, we only report the panel model results with severe endogeneity (i.e. $\kappa = 2$).⁵

[Table 2; Table 3]

Table 2 and Table 3 display the Monte Carlo simulation results for our DGP1 and DGP2. We compare the results of our proposed estimator and the least squares estimator ignoring endogeneity issue under different sample sizes. We have the following observations. First, we find that, as the number of observations increases, the RMSE without control functions remains large as the endogenous severity rises (κ increases). For example, without using the control function correction approach, the RMSE for β_1 barely decreases, even with a mild degree of endogeneity. By contrast, with the control function correction, the RMSEs for all parameters decrease rapidly as the sample size increases, confirming the validity of our CF approach to tackling endogeneity.

[Table 4; Table 5]

Tables 4 and 5 give the Monte Carlo simulation results for DGP3 and DGP4, respectively. With severe endogeneity, the findings are similar to those in the time-series model.

5 Empirical study

Since the worldwide outbreak in early 2020, global countries have suffered tremendously from the COVID-19 pandemic. Recently, there has been a growing interest in the literature examining the

⁵The results for other cases are available from the authors upon request.

COVID-19 impact on the labor market. For example, [Baek et al. \(2021\)](#) study the impact of the Stay-at-Home order on the US labor market. They consider the first wave of COVID-19 and measure the fluctuation of the labor market by State Initial Claims for Unemployment Insurance, where they find that the Stay-at-Home policy only accounts for a small fraction of the total negative effect of COVID-19 on the labor market. Using individual-level data, [Lee et al. \(2021\)](#) find that the negative impact of COVID-19 on the labor market spread unequally across the population. Among other interesting findings, we observe that the unemployment rates for most advanced economies have recovered to the pre-COVID level while the pandemic is still ongoing. This unconventional fact motivates us to investigate the potential nonlinear relationship between the COVID-19 cases and the labor market performance. We extrapolate the potential nonlinearity ties-up with the occasional lockdown policy that the government imposed aimed at easing pandemic pressure for hospitals as cases surge. For that reason, in this section, we study the effect of COVID-19 on the Canadian and US labor markets by using our proposed endogenous kink threshold panel model. We collect monthly data for each province/state. Canadian data spans from January 2020 to September 2021, while the US data spans from March 2020 to September 2021. The covered periods are long enough to capture multiple waves of COVID-19 outbreaks, which provide an overall picture of this relationship. We propose to use the following KTR model to examine our hypothesis

$$\begin{cases} une_{it} &= \beta_0 + \beta_{low}(case_{it} - \gamma_0)I(case_{it} < \gamma_0) + \beta_{high}(case_{it} - \gamma_0)1(case_{it} \geq \gamma_0) + \lambda_t + b_i + u_{it}, \\ case_{it} &= \beta_{20} + \beta_{30}test_{it} + v_{it}, \end{cases} \quad (5.1)$$

where i represents a province for Canadian data and a state for US data, and t refers to the time. The dependent variable of interest, une_{it} , is the monthly seasonally adjusted unemployment rate, and $case_{it}$ is the log of the number of cases confirmed for COVID-19 in the t^{th} month.⁶ Also, $test_{it}$ equals the log of the number of tests conducted in the t^{th} month. And, λ_t denotes the time effect, and b_i is the individual fixed effect, which capture the idiosyncratic characteristics of provinces/states. Considering the potential bidirectional causality between une_{it} and $case_{it}$, we thereby apply the CF approach, given in Section 3.1, to estimate model (5.1). In particular, we use $test_{it}$ as the instrument variable since this variable is highly associated with the number of cases

⁶Note that we replace $\log(0)$ by 0 to avoid missing data. The same procedure is applied to the tests variable.

but has no direct effect on the unemployment rate.

For comparison purposes, we also estimate and report the linear panel regression model, which is in the following form

$$\begin{cases} une_{it} &= \beta_0 + \beta_{linear} case_{it} + \lambda_t + b_i + u_{it}, \\ case_{it} &= \beta_{20} + \beta_{30} test_{it} + v_{it}. \end{cases} \quad (5.2)$$

Similar to the KTR model, we also employ a CF method to deal with the endogeneity in the linear panel model by taking the following steps. We first take the first differencing to remove the individual fixed effects to estimate the model. Then, we obtain the OLS residuals from the reduced form equation of $case_{it}$ and include it as an additional regressor in the first-differenced model to correct for endogeneity. Last, we apply the OLS method to estimate the augmented first-differenced unemployment rate model. In short, the estimation procedure for model (5.2) is similar to the estimation strategy introduced in Section 3.1, except it does not require a grid search over γ .

As mentioned above, this study uses two data sets: one for Canada and one for the US. The descriptive statistics for all variables are shown in Table 1.⁷ Note that we take log bases on 10. The regions covered in our data set are shown in Table 6.

Table 1: Summary Statistics

Canada data(Jan 2020-Sep 2021)						
	Variable	Obs	Mean	Std. Dev.	Min	Max
Log cases confirmed	<i>case</i>	210	2.5953	1.4732	0	5.0548
Log test performed	<i>test</i>	210	4.3490	1.5687	0	6.2470
Unemployment rate (Seasonal adjusted)	<i>une</i>	210	9.2681	2.6692	4.5	17
US data(Mar 2020-Sep 2021)						
	Variable	Obs	Mean	Std. Dev.	Min	Max
Log cases confirmed	<i>case</i>	988	4.1622	0.7253	0	6.0620
Log test performed	<i>test</i>	988	5.3963	0.6660	0	6.9761
Unemployment rate (Seasonal adjusted)	<i>une</i>	988	6.8249	3.2931	2	29.5

⁷Canada Data source: the number of COVID-19 cases and tests performed from Government of Canada; unemployment rate is taken from Statistics Canada. The US Data source: the number of COVID-19 cases and tests performed are taken from the Centers for Disease and Control and Prevention America; unemployment rate is taken from the U.S. Bureau of Labor Statistics.

Table 7 reports the estimation results for Canadian data. Regressions (1) and (2) report the results from the linear and KTR models without controlling for endogeneity, respectively. Specifically, the estimate for β_{linear} of the linear model is significantly positive, confirming that the surge in the (log) number of confirmed COVID-19 cases leads to a higher unemployment rate. However, the findings of the KTR model suggest that, although the impact of COVID-19 on the unemployment rate is positive for both regimes, the size is more substantial if the (log) number of confirmed COVID-19 cases surpasses a certain threshold level (for Canada, this number is 12,882). The last two columns of Table 7 report the results when we use a control function approach to correct the endogeneity, assuming a linear endogenous function. We observe that the coefficient estimate of the COVID-19 for the linear model with endogenous correction is relatively more prominent than the ones we obtain without controlling for the endogeneity. Next, for the KTR model with endogeneity, our estimation results are consistent with the findings in the KTR model without controlling for the endogeneity, underpinning the positive effect is more significant if the number of confirmed COVID-19 cases exceeds a certain level. Besides, we observe that the impact magnitude of the COVID-19 cases on the unemployment rate is also more substantial compared to the ones without the endogenous correction in both regimes. Interestingly, the size of the coefficient estimate for the linear regression model is always in-between those of the low and high regime for the KTR model, regardless of controlling for the endogeneity. To test the nonlinearity, we perform a test for the existence of a threshold effect. Our null hypothesis of interest is $\beta_{low,0} = \beta_{high,0}$. Employing the Wald test introduced by Hansen (2017), we conduct the threshold effect test and compute the asymptotically valid p -value via a multiplier bootstrap. We repeat 10,000 simulations for the bootstrapping and obtain a p -value equal to 0.0001. Thus, the test strongly rejects the null hypothesis at 1% significance level, suggesting that the linear model fails to capture the nonlinearity.

Table 8 summarizes the estimation results for the US data. Again, regressions (1) and (2) report the results for the linear model and the KTR model without controlling for endogeneity, respectively. In general, we draw similar conclusions as we obtained from examining Canadian data. The estimated coefficient of the (log) number of confirmed COVID-19 cases on the unemployment rate for the linear model, β_{linear} , is positive and significant at 5% level. Turning to the KTR model, the estimation results show that the impact of COVID-19 on the unemployment rate is

significantly positive if and only if the number of confirmed COVID-19 cases is above the threshold level, 16,596. If the number is below this level, the effect turns out to be negative. Regressions (3) and (4) present the estimation results using the CF approach and assuming a linear form of endogeneity. We observe that the magnitude of the coefficient estimate for both the linear and the KTR model is more extensive, and the level of the threshold estimate is lower compared to ones before correcting the endogeneity. Interestingly, the impact of the low regime switches the sign from negative to positive after controlling for the endogeneity, despite both impacts are inconsiderable and close to zero. Our results offer a word of caution to the notion that increasing the number of confirmed COVID-19 cases unorthodoxly facilitates to clear the labor market if the number of confirmed COVID-19 cases is not large enough. The main reason for the negative sign in the low regime of the regression (2) is the result of endogenous distortion. We also implement the threshold effect test and obtain the bootstrap p -value = 0.0003. Similar to the result with Canadian data, we strongly reject the null hypothesis of linearity at 1% significant level, favoring the KTR model.

We also apply a t -test to test for the endogeneity of $case_{it}$ for the KTR model. Compared with the complexity of threshold variable endogeneity test in TR model, endogeneity test of threshold variable in the KTR model is quite standard. Testing the endogeneity of threshold variable in the TR model need to consider the existence of a threshold effect; see, e.g., [Kourtellis et al. \(2021\)](#). While, in the KTR model, testing the endogeneity of threshold variable is free of the existence of the threshold effect, since the threshold variable is also part of regressors which stay in the KTR model. Via using control function approach, testing for the endogeneity of $case$ is equivalent to testing $\beta_{40} = 0$, where β_{40} is the coefficient of endogeneity bias correction term (Δv_{it}) and Δv_{it} is the first difference of v_{it} in model (5.1); see, e.g., equation (3.7) for detailed definition. In our case, the t statistic in the absolute value for Canada is 19.76, for the US is 13.27. As both test statistics are higher than the 1% critical value, $t_{0.01} = 2.617$, we strongly reject the null-hypothesis of non-endogeneity even at 1% level, which supports the existence of endogeneity.

Finally, as was mentioned earlier, when we estimate the KTR model, following [Hansen \(2017\)](#), we use a grid search method to go through the parameter space of the threshold value and obtain the global minimum least square estimator, using an interval of length 0.1. Figure 1 plots the least-square criterion $S_n(\gamma)$ as a function of γ . Both for the Canadian and the US plots we find our

criteria functions are reasonably smooth and have a well-defined global minimum, which suggests the interval we choose for grid search is sufficiently enough.

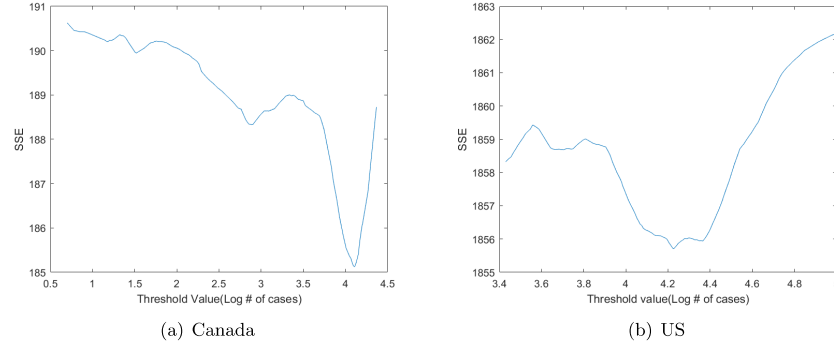


Figure 1: Least Square Criteria

6 Conclusion

As in Hansen (2017), we consider a kink threshold model, extending the model to allow for endogeneity, following Kourtellis et al. (2016) and Yu et al. (2020) and applying a control function approach to tackle the problem. Monte Carlo simulations show that the small sample performance of our proposed estimator is quite satisfactory both for time series and panel cases. Last, we apply our model to examine the effect of COVID-19 cases on unemployment rate in Canada and the US. We find COVID-19 cases above certain thresholds both for Canada and the US have had significant negative effects on labor market activity, while below that threshold the effect overall was found to be moderate.

Our method has several possible extensions. Instead of introducing bias correction terms linearly ($\beta'_{40}v_t$ in (2.5)), one can introduce a nonparametric endogeneity correction term($g(v_t)$, where $g(\cdot)$ is an unknown function), like Kourtellis et al. (2021) in threshold regression model. Also, one may want to relax the linear specification in reduced-form functions (2.2)-(2.3) to a more flexible semi-/nonparametric specification.

References

- Andrews, D. W. (1994). Chapter 37 empirical process methods in econometrics. *Handbook of Econometrics*, pages 2247–2294.
- Baek, C., McCrory, P. B., Messer, T., and Mui, P. (2021). Unemployment effects of stay-at-home orders: Evidence from high-frequency claims data. *The Review of Economics and Statistics*, 103(5):979–993.
- Caner, M. and Hansen, B. E. (2004). Instrumental variable estimation of a threshold model. *Econometric Theory*, 20(05):813–843.
- Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics*, 21(1):520–533.
- Chan, K. S. and Tsay, R. S. (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika*, 85(2):413–426.
- Christopoulos, D., McAdam, P., and Tzavalis, E. (2021). Dealing with endogeneity in threshold models using copulas. *Journal of Business & Economic Statistics*, 39(1):166–178.
- Doukhan, P., Massart, P., and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Annales de l’I.H.P. Probabilités et statistiques*, 31(2):393–427.
- Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68:575–603.
- Hansen, B. E. (2017). Regression kink with an unknown threshold. *Journal of Business & Economic Statistics*, 35(2):228–240.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.
- Herrndorf, N. (1984). A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables. *The Annals of Probability*, 12(1):141 – 153.
- Hidalgo, J., Lee, J., and Seo, M. H. (2019). Robust inference for threshold regression models. *Journal of Econometrics*, 210(2):291–309.

- Karavias, Y., Narayan, P. K., and Westerlund, J. (2022). Structural breaks in interactive effects panels and the stock market reaction to covid-19. *Journal of Business & Economic Statistics*, pages 1–14.
- Kong, E. and Prinz, D. (2020). Disentangling policy effects using proxy data: Which shutdown policies affected unemployment during the covid-19 pandemic? *Journal of Public Economics*, 189:104257.
- Kourtellis, A., Stengos, T., and Sun, Y. (2021). Endogeneity in semiparametric threshold regression. *Econometric Theory*.
- Kourtellis, A., Stengos, T., and Tan, C. M. (2016). Structural threshold regression. *Econometric Theory*, 32(4):827–860.
- Lee, S. Y. T., Park, M., and Shin, Y. (2021). Hit harder, recover slower? unequal employment effects of the covid-19 shock. Working Paper 28354, National Bureau of Economic Research.
- Newey, W. K. and Mcfadden, D. (1994). Chapter 36 large sample estimation and hypothesis testing. *Handbook of Econometrics*, pages 2111–2245.
- Newey, W. K., Powell, J. L., and Vella, F. (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica*, 67(3):565–603.
- Seber, G. A. F. (2008). *A matrix handbook for statisticians*. J. Wiley.
- Seo, M. H. and Shin, Y. (2016). Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics*, 195(2):169–186.
- Tong, H. (1990). *Non-linear time series: a dynamical system approach*. Oxford university press.
- Yu, P., Liao, Q., and Phillips, P. C. B. (2020). New control function approaches in threshold regression with endogeneity. *mimeo*.
- Yu, P. and Phillips, P. C. (2018). Threshold regression with endogeneity. *Journal of Econometrics*, 203(1):50–68.
- Zhang, Y., Zhou, Q., and Jiang, L. (2017). Panel kink regression with an unknown threshold. *Economics Letters*, 157:116–121.

7 Appendix

7.1 Proof of Theorem 1-time series

We first show that $\sup_{\phi \in B \times \Gamma} |S_n(\phi) - S_n^*(\phi)| \xrightarrow{P} 0$, which implies that minimizing $S_n(\phi)$ with respect to ϕ is equivalent to minimizing $S_n^*(\phi)$, where $S_n(\phi)$ and $S_n^*(\phi)$ are defined in (2.6) and Assumption T7, respectively. Then, closely following the mathematical proof of Hansen (2017), we can show that the optimization problem of $S_n^*(\phi)$ with respect to ϕ satisfies all the required conditions in Section 3.2 of Andrews (1994), which completes the proof of this theorem. Hence, below, we only need to prove $\sup_{\phi \in B \times \Gamma} |S_n(\phi) - S_n^*(\phi)| \xrightarrow{P} 0$.

Now, we denote $P_\gamma = X(\gamma)[X'(\gamma)X(\gamma)]^{-1}X'(\gamma)$ and $P_\gamma^* = X^*(\gamma)[X^{*'}(\gamma)X^*(\gamma)]^{-1}X^{*'}(\gamma)$, where $X^*(\gamma)$ is defined in the same form as $X(\gamma)$ by replacing $x_t(\gamma)$ with $x_t^*(\gamma)$. For any $\phi \in B \times \Gamma$, we have

$$S_n(\phi) = \frac{1}{n} \sum_{t=1}^n [y_t - x_t'(\gamma)\theta]^2 = \frac{1}{n} y' (I_n - P_\gamma) y. \quad (7.1)$$

Let $X - \gamma_0 \tau_n$ and $I_n(\gamma_0)(X - \gamma_0 \tau_n)$ be the matrix form of $x_t - \gamma_0$ and $(x_t - \gamma_0)I(x_t \geq \gamma_0)$ respectively, where τ_n is an $n \times 1$ vector of ones, $I_n(\gamma_0) = \text{diag}\{I(x_1 \geq \gamma_0), \dots, I(x_n \geq \gamma_0)\}$. Also, let ε stack up ε_i , v stack up v_i , and \hat{v} stack up \hat{v}_i for $i = 1, \dots, n$. Below, we decompose $S_n(\phi)$. By simple calculation, we have

$$y = X^*(\gamma_0)\theta_0 + \varepsilon = X(\gamma)\theta_0 + [X^*(\gamma) - X(\gamma)]\theta_0 + [X^*(\gamma_0) - X^*(\gamma)]\theta_0 + \varepsilon,$$

where we have

$$\begin{aligned} X^*(\gamma) - X(\gamma) &= [\mathbf{0}_{n \times 1}, \mathbf{0}_{n \times 1}, \mathbf{0}_{n \times \ell}, v - \hat{v}], \\ X^*(\gamma_0) - X^*(\gamma) &= [(\gamma - \gamma_0)\tau_n, I_n(\gamma_0)(X - \gamma_0 \tau_n) - I_n(\gamma)(X - \gamma \tau_n), \mathbf{0}_{n \times \ell}, \mathbf{0}_{n \times (1+d_{z1})}]. \end{aligned}$$

Note that $X'(\gamma)(I_n - P_\gamma) = \mathbf{0}_{(k+d_{z1}) \times n}$ and $\hat{\varepsilon} = (v - \hat{v})\beta_{40} + \varepsilon$, where $\hat{\varepsilon}$ stacks up $\hat{\varepsilon}_i$ for $i = 1, \dots, n$.

Thus, we can show

$$\begin{aligned}
S_n(\phi) &= \frac{1}{n} \{ \tau_n(\gamma - \gamma_0)\beta_{10} + [I_n(\gamma_0)(X - \gamma_0\tau_n) - I_n(\gamma)(X - \gamma\tau_n)]\delta_0 + \hat{\varepsilon} \}' (I_n - P_\gamma) \times \\
&\quad \{ \tau_n(\gamma - \gamma_0)\beta_{10} + [I_n(\gamma_0)(X - \gamma_0\tau_n) - I_n(\gamma)(X - \gamma\tau_n)]\delta_0 + \hat{\varepsilon} \} \\
&= \frac{1}{n} \hat{\varepsilon}'(I_n - P_\gamma)\hat{\varepsilon} + \frac{2}{n} \delta_0 [I_n(\gamma)\tau_n(\gamma - \gamma_0) + d(\gamma_0, \gamma)(X - \gamma_0\tau_n)]' (I_n - P_\gamma)\hat{\varepsilon} \\
&\quad + \frac{1}{n} \delta_0^2 [I_n(\gamma)\tau_n(\gamma - \gamma_0) + d(\gamma_0, \gamma)(X - \gamma_0\tau_n)]' \\
&\quad \times (I_n - P_\gamma) [I_n(\gamma)\tau_n(\gamma - \gamma_0) + d(\gamma_0, \gamma)(X - \gamma_0\tau_n)],
\end{aligned} \tag{7.2}$$

where $d(\gamma_0, \gamma) = I_n(\gamma_0) - I_n(\gamma)$.

Next, we decompose $S_n^*(\phi)$. By simple calculation, we have

$$y = X^*(\gamma_0)\theta_0 + \varepsilon = X^*(\gamma)\theta_0 + [X^*(\gamma_0) - X^*(\gamma)]\theta_0 + \varepsilon.$$

As $X^{*'}(\gamma)(I_n - P_\gamma^*) = \mathbf{0}_{(k+d_{z1}) \times n}$, we obtain

$$\begin{aligned}
S_n^*(\phi) &= \frac{1}{n} \{ \tau_n(\gamma - \gamma_0)\beta_{10} + [I_n(\gamma_0)(X - \gamma_0\tau_n) - I_n(\gamma)(X - \gamma\tau_n)]\delta_0 + \varepsilon \}' (I_n - P_\gamma^*) \times \\
&\quad \{ \tau_n(\gamma - \gamma_0)\beta_{10} + [I_n(\gamma_0)(X - \gamma_0\tau_n) - I_n(\gamma)(X - \gamma\tau_n)]\delta_0 + \varepsilon \} \\
&= \frac{1}{n} \varepsilon'(I_n - P_\gamma^*)\varepsilon + \frac{2}{n} \delta_0 [I_n(\gamma)\tau_n(\gamma - \gamma_0) + d(\gamma_0, \gamma)(X - \gamma_0\tau_n)]' (I_n - P_\gamma^*)\varepsilon \\
&\quad + \frac{1}{n} \delta_0^2 [I_n(\gamma)\tau_n(\gamma - \gamma_0) + d(\gamma_0, \gamma)(X - \gamma_0\tau_n)]' (I_n - P_\gamma^*) \\
&\quad \times [I_n(\gamma)\tau_n(\gamma - \gamma_0) + d(\gamma_0, \gamma)(X - \gamma_0\tau_n)].
\end{aligned} \tag{7.3}$$

Subtracting (7.3) from (7.2), we have

$$\begin{aligned}
& S_n(\phi) - S_n^*(\phi) \\
= & \frac{1}{n}(\hat{\varepsilon}'\hat{\varepsilon} - \varepsilon'\varepsilon) + \frac{2}{n}\delta_0[I_n(\gamma)\tau_n(\gamma - \gamma_0) + d(\gamma_0, \gamma)(X - \gamma_0\tau_n)]'(\hat{\varepsilon} - \varepsilon) \\
& - \frac{1}{n}\delta_0^2[I_n(\gamma)\tau_n(\gamma - \gamma_0) + d(\gamma_0, \gamma)(X - \gamma_0\tau_n)]'(P_\gamma - P_\gamma^*)[I_n(\gamma)\tau_n(\gamma - \gamma_0) + d(\gamma_0, \gamma)(X - \gamma_0\tau_n)] \\
& - \frac{1}{n}(\hat{\varepsilon}'P_\gamma\hat{\varepsilon} - \varepsilon'P_\gamma^*\varepsilon) + \frac{2}{n}\delta_0[I_n(\gamma)\tau_n(\gamma - \gamma_0) + d(\gamma_0, \gamma)(X - \gamma_0\tau_n)]'(P_\gamma^*\varepsilon - P_\gamma\hat{\varepsilon}) \\
= & S_1 + 2S_2 - S_{31} - S_{32} + 2S_{33}.
\end{aligned}$$

Below, we show that S_1, S_2, S_{31}, S_{32} , and S_{33} are all $o_p(1)$ uniformly over the domain of ϕ .

S₁: Denoting $\hat{\pi} = [\hat{\pi}'_x, \hat{\pi}'_z]'$, $\pi_0 = [\pi'_{x0}, \pi'_{z0}]'$ and $\rho_t = [p'_{xt}, p'_{zt}]'$, we have $\hat{\pi} - \pi_0 = O_p(n^{-1/2})$ under Assumptions T2(b) and T5(b) and

$$\begin{aligned}
S_1 &= \frac{1}{n}(\hat{\varepsilon}'\hat{\varepsilon} - \varepsilon'\varepsilon) \\
&= \frac{2}{n}\beta'_{40}(v - \hat{v})'\varepsilon + \frac{1}{n}\beta'_{40}(v - \hat{v})'(v - \hat{v})\beta_{40} \\
&= O_p(n^{-1/2}) + O_p(n^{-1}) = o_p(1),
\end{aligned} \tag{7.4}$$

since $n^{-1}\varepsilon'\varepsilon = \sigma_\varepsilon^2 + o_p(1)$ under Assumption T5 and

$$\|v - \hat{v}\|^2 = (\hat{\pi} - \pi_0)' \sum_{t=1}^n \rho_t \rho'_t (\hat{\pi} - \pi_0) \leq n \|\hat{\pi} - \pi_0\|^2 \lambda_{\max}(n^{-1} \sum_{t=1}^n \rho_t \rho'_t) = O_p(1) \tag{7.5}$$

under Assumption T2(b), where $\lambda_{\max}(A)$ denotes the largest eigenvalue of a symmetric matrix A .

S₂: By Assumptions T2 (a), T7 and applying $\frac{1}{n} \|\hat{\varepsilon} - \varepsilon\| = \frac{1}{n} \|(v - \hat{v})\beta_{40}\| \leq \frac{1}{n} \|\beta_{40}\| \|v - \hat{v}\| = O_p(n^{-1})$, for all $\gamma \in \Gamma$, we can show

$$S_2 = \frac{1}{n}\delta_0(\gamma - \gamma_0)\tau'_n I_n(\gamma)(\hat{\varepsilon} - \varepsilon) + \frac{1}{n}\delta_0(X - \gamma_0\tau_n)'d(\gamma_0, \gamma)(\hat{\varepsilon} - \varepsilon) = O_p(n^{-1/2}) + O_p(n^{-1/2}) = o_p(1).$$

S₃₁, S₃₂, S₃₃: By showing S_{31}, S_{32}, S_{33} are all $o_p(1)$ for any $\phi \in B \times \Gamma$, we only need to prove:

$$\max_{\gamma \in \Gamma} \frac{1}{n} \left\| [I_n(\gamma)\tau_n(\gamma - \gamma_0) + d(\gamma_0, \gamma)(X - \gamma_0\tau_n)]' [X(\gamma) - X^*(\gamma)] \right\| = o_p(1), \quad (7.6)$$

$$\max_{\gamma \in \Gamma} \frac{1}{n} \left\| \hat{\varepsilon}' X(\gamma) - \varepsilon' X^*(\gamma) \right\| = o_p(1), \quad (7.7)$$

$$\max_{\gamma \in \Gamma} \frac{1}{n} \left\| X'(\gamma)X(\gamma) - X^{*'}(\gamma)X^*(\gamma) \right\| = o_p(1). \quad (7.8)$$

First, applying (7.5), we have

$$\max_{\gamma \in \Gamma} \frac{1}{n} \|X(\gamma) - X^*(\gamma)\| = \frac{1}{n} \|\hat{v} - v\| = O_p(n^{-1}) = o_p(1). \quad (7.9)$$

Next, by Assumptions T2(a) and T7, we have

$$\begin{aligned} & \max_{\gamma \in \Gamma} \frac{1}{n} \left\| [I_n(\gamma)\tau_n(\gamma - \gamma_0) + d(\gamma_0, \gamma)(X - \gamma_0\tau_n)]' [X(\gamma) - X^*(\gamma)] \right\| \\ & \leq \max_{\gamma \in \Gamma} \frac{1}{n} \|I_n(\gamma)\tau_n(\gamma - \gamma_0) + d(\gamma_0, \gamma)(X - \gamma_0\tau_n)\| \|X(\gamma) - X^*(\gamma)\| = O_p(n^{-1/2}) = o_p(1), \end{aligned}$$

which verifies (7.6).

Next, we show (7.7). Applying triangle inequality gives

$$\begin{aligned} & \max_{\gamma \in \Gamma} \frac{1}{n} \left\| \hat{\varepsilon}' X(\gamma) - \varepsilon' X^*(\gamma) \right\| \\ & \leq \max_{\gamma \in \Gamma} \frac{1}{n} \left\| \varepsilon' [X(\gamma) - X^*(\gamma)] \right\| + \max_{\gamma \in \Gamma} \frac{1}{n} \left\| \beta'_{40}(v - \hat{v})' X(\gamma) \right\| \end{aligned} \quad (7.10)$$

$$\begin{aligned} & = \frac{1}{n} \left\| \varepsilon'(\hat{v} - v) \right\| + \max_{\gamma \in \Gamma} \frac{1}{n} \left\| \beta'_{40}(v - \hat{v})' X(\gamma) \right\| \\ & = O_p(n^{-1/2}) + O_p(n^{-1/2}) = o_p(1), \end{aligned} \quad (7.11)$$

where $\frac{1}{n} \|\varepsilon'(\hat{v} - v)\| = O_p(n^{-1/2})$ by (7.4), and $\max_{\gamma \in \Gamma} \frac{1}{n} \|(v - \hat{v})' X(\gamma)\| \leq \frac{1}{n} \|v - \hat{v}\| \|X(\gamma)\| = O_p(n^{-1/2})$ by (7.5) and under Assumption T2.

Finally, by Assumption P2 and (7.9), we can show

$$\begin{aligned}
& \max_{\gamma \in \Gamma} \frac{1}{n} \|X'(\gamma)X(\gamma) - X^{*'}(\gamma)X^*(\gamma)\| \\
&= \max_{\gamma \in \Gamma} \frac{2}{n} \|(X - \gamma\tau_n)'(\hat{v} - v)\| + \max_{\gamma \in \Gamma} \frac{2}{n} \|[I_n(\gamma)(X - \gamma\tau_n)]'(\hat{v} - v)\| + \frac{2}{n} \|z'(\hat{v} - v)\| + \frac{1}{n} \|\hat{v}'\hat{v} - v'v\| \\
&= O_p(n^{-1/2}) + O_p(n^{-1/2}) + O_p(n^{-1/2}) + \frac{1}{n} \|\hat{v}\hat{v}' - vv'\|, \tag{7.12}
\end{aligned}$$

where z is an $n \times \ell$ matrix, and $z = [z_1, z_2, \dots, z_n]'$. Then, we only left to show $\frac{1}{n} \|\hat{v}'\hat{v} - v'v\| = o_p(1)$.

Note that,

$$\begin{aligned}
\frac{1}{n} \|\hat{v}'\hat{v} - v'v\| &= \frac{1}{n} \|(\hat{v} - v)'(\hat{v} - v)\| + \frac{2}{n} \|(\hat{v} - v)'v\| \\
&= O_p(n^{-1}) + A, \tag{7.13}
\end{aligned}$$

where, under Assumption T2, for any bounded v_t , we have

$$A = \frac{1}{n} \|(\hat{v} - v)'v\| \leq \frac{1}{n} \|(\hat{v} - v)\| \|v\| = O_p(n^{-1/2}). \tag{7.14}$$

Thus, combining (7.12), (7.13) with (7.14), we obtain

$$\frac{1}{n} \max_{\gamma \in \Gamma} \|X'(\gamma)X(\gamma) - X^{*'}(\gamma)X^*(\gamma)\| = o_p(1).$$

To sum up, we have $\sup_{\phi \in B \times \Gamma} |S_n(\phi) - S_n^*(\phi)| \xrightarrow{P} 0$, which completes the proof of Theorem 1-time series. ■

7.2 Proof of Theorem 1-panel

Similar to the proof of Theorem 1-Time Series, we first prove $\sup_{\phi \in B \times \Gamma} |S_{NT}(\phi) - S_{NT}^*(\phi)| \xrightarrow{P} 0$, which implies that the minimizer of $S_{NT}(\phi)$ is also the minimizer of $S_{NT}^*(\phi)$, where $\phi = (\theta', \gamma)'$, and the definition of $S_{NT}(\phi)$ and $S_{NT}^*(\phi)$ are given by (3.8) and Assumption P7, respectively. Then, by showing that, for $\phi \in B \times \Gamma$, the optimization problem of $S_{NT}^*(\phi)$ satisfies all the four required conditions in Section 3.2 of Andrews (1994), we verify Theorem 1-panel.

We first show that $\sup_{\phi \in B \times \Gamma} |S_{NT}(\phi) - S_{NT}^*(\phi)| \xrightarrow{P} 0$.

Define $P(\gamma) = \Delta x(\gamma)[\Delta x(\gamma)' \Delta x(\gamma)]^{-1} \Delta x(\gamma)'$ and $P^*(\gamma)$ is in the same form of $P(\gamma)$ with $\Delta x(\gamma)$ being replaced by $\Delta x^*(\gamma)$, where $\Delta x(\gamma) = [\Delta x_{12}(\gamma), \dots, \Delta x_{1T}(\gamma), \dots, \Delta x_{N2}, \dots, \Delta x_{NT}(\gamma)]'$ is an $[N(T-1)] \times (k + d_{z1})$ matrix, and $\Delta x^*(\gamma)$ equals $\Delta x(\gamma)$ with $\Delta x_{it}(\gamma)$ being replaced by $\Delta x_{it}^*(\gamma)$. Then, for any $\phi \in B \times \Gamma$, $S_{NT}(\phi)$ can be expressed as

$$S_{NT}(\phi) = \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T (\Delta y_{it} - \Delta x_{it}(\gamma)' \theta)^2 = \frac{1}{N(T-1)} \Delta y' P(\gamma) \Delta y,$$

where $\Delta y = [\Delta y_{12}, \dots, \Delta y_{1T}, \dots, \Delta y_{N2}, \dots, \Delta y_{NT}]'$ is an $[N(T-1)] \times 1$ vector. Note that, we can rewrite Δy as

$$\Delta y = \Delta x^*(\gamma_0) \theta_0 + \Delta \varepsilon = [\Delta x^*(\gamma_0) - \Delta x(\gamma)] \theta_0 + \Delta x(\gamma) \theta_0 + \Delta \varepsilon,$$

where $\Delta \varepsilon = [\Delta \varepsilon_{12}, \dots, \Delta \varepsilon_{1T}, \dots, \Delta \varepsilon_{N2}, \dots, \Delta \varepsilon_{NT}]'$ an $[N(T-1)] \times 1$ vector.

As $\Delta x(\gamma)' [I_{NT} - P(\gamma)] = \mathbf{0}_{(1+d_{z1}) \times N(T-1)}$, we have

$$\begin{aligned} S_{NT}(\phi) &= \frac{1}{N(T-1)} \{ [\Delta x^*(\gamma_0) - \Delta x(\gamma)] \theta_0 + \Delta \varepsilon \}' [I_N - P(\gamma)] \\ &\quad \times \{ [\Delta x^*(\gamma_0) - \Delta x(\gamma)] \theta_0 + \Delta \varepsilon \}. \end{aligned}$$

where I_{NT} is an identity matrix with dimension $N(T-1)$.

Similarly, we can also decompose Δy as

$$\Delta y = \Delta x^*(\gamma_0) \theta_0 + \Delta \varepsilon = [\Delta x^*(\gamma_0) - \Delta x^*(\gamma)] \theta_0 + \Delta x^*(\gamma) \theta_0 + \Delta \varepsilon.$$

Since $\Delta x^{*'}(\gamma) [I_{NT} - P^*(\gamma)] = \mathbf{0}_{(k+d_{z1}) \times N(T-1)}$, we have

$$\begin{aligned} S_{NT}^*(\phi) &= \frac{1}{N(T-1)} \{ [\Delta x^*(\gamma_0) - \Delta x^*(\gamma)] \theta_0 + \Delta \varepsilon \}' [I_{NT} - P^*(\gamma)] \\ &\quad \times \{ [\Delta x^*(\gamma_0) - \Delta x^*(\gamma)] \theta_0 + \Delta \varepsilon \}. \end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
& S_{NT}(\phi) - S_{NT}^*(\phi) \\
&= \frac{1}{N(T-1)} \{[\Delta x^*(\gamma) - \Delta x(\gamma)]\theta_0\}'[I_{NT} - P(\gamma)]\{[\Delta x^*(\gamma) - \Delta x(\gamma)]\theta_0\} \\
&\quad + \frac{1}{N(T-1)} \{[\Delta x^*(\gamma_0) - \Delta x^*(\gamma)]\theta_0 + \Delta\varepsilon\}'[P^*(\gamma) - P(\gamma)]\{[\Delta x^*(\gamma_0) - \Delta x^*(\gamma)]\theta_0 + \Delta\varepsilon\} \\
&\quad + \frac{2}{N(T-1)} \{[\Delta x^*(\gamma) - \Delta x(\gamma)]\theta_0\}'[I_{NT} - P(\gamma)]\{[\Delta x^*(\gamma_0) - \Delta x^*(\gamma)]\theta_0 + \Delta\varepsilon\} \\
&= S_1 + S_2 + 2S_3,
\end{aligned}$$

where denoting $\Delta w = [\Delta w'_{12}, \dots, \Delta w'_{1T}, \dots, \Delta w'_{N2}, \dots, \Delta w'_{NT}]'$ for $w = v$ and \hat{v} and $\chi(\gamma) = [(X_{12} - \gamma\tau_2)'\mathbf{I}_{12}(\gamma), \dots, (X_{1T} - \gamma\tau_2)'\mathbf{I}_{1T}(\gamma), \dots, (X_{N2} - \gamma\tau_2)'\mathbf{I}_{N2}(\gamma), \dots, (X_{NT} - \gamma\tau_2)'\mathbf{I}_{NT}(\gamma)]'$, we have

$$\begin{aligned}
\Delta x^*(\gamma) - \Delta x(\gamma) &= [\mathbf{0}_{N(T-1) \times 1}, \mathbf{0}_{N(T-1) \times 1}, \mathbf{0}_{N(T-1) \times 1}, \Delta v - \Delta \hat{v}], \\
\Delta x^*(\gamma_0) - \Delta x^*(\gamma) &= [\mathbf{0}_{N(T-1) \times 1}, \chi(\gamma_0) - \chi(\gamma), \mathbf{0}_{N(T-1) \times 1}, \mathbf{0}_{N(T-1) \times (1+d_{z1})}].
\end{aligned}$$

Hence, we obtain

$$\begin{aligned}
S_1 &= \frac{1}{N(T-1)} \beta'_{40} (\Delta v - \Delta \hat{v})' (I_{NT} - P(\gamma)) (\Delta v - \Delta \hat{v}) \beta_{40} \\
&\leq \frac{1}{N(T-1)} \lambda_{\max}(I_{NT} - P(\gamma)) \beta'_{40} (\Delta v - \Delta \hat{v})' (\Delta v - \Delta \hat{v}) \beta_{40} \\
&= O_p([N(T-1)]^{-1})
\end{aligned} \tag{7.15}$$

since $I_{NT} - P(\gamma)$ is an idempotent matrix with $\lambda_{\max}(I_{NT} - P(\gamma)) = 1$ and denoting $\hat{\Pi} = [\hat{\Pi}'_x, \hat{\Pi}'_z]'$, $\Pi_0 = [\Pi'_{x0}, \Pi'_{z0}]'$, and $\mathcal{P}_{it} = [p'_{x,it}, p'_{z,it}]'$, we have $\hat{\Pi} - \Pi_0 = O_p([N(T-1)]^{-1/2})$ under Assumptions P2(b) and P5(b) and

$$\begin{aligned}
\|\Delta v - \Delta \hat{v}\|^2 &= (\hat{\Pi} - \Pi_0)' \sum_{i=1}^N \sum_{t=2}^T \Delta \mathcal{P}_{it} \Delta \mathcal{P}'_{it} (\hat{\Pi} - \Pi_0) \\
&\leq N(T-1) \|\hat{\Pi} - \Pi_0\|^2 \lambda_{\max} \left(\frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T \Delta \mathcal{P}_{it} \Delta \mathcal{P}'_{it} \right) \\
&= O_p(1)
\end{aligned} \tag{7.16}$$

under Assumption P2(b).

Next, we consider

$$\begin{aligned}
\mathcal{S}_2 &= \frac{\delta_0^2}{N(T-1)} [\chi(\gamma_0) - \chi(\gamma)]' [P^*(\gamma) - P(\gamma)] [\chi(\gamma_0) - \chi(\gamma)] + \frac{1}{N(T-1)} \Delta \varepsilon' [P^*(\gamma) - P(\gamma)] \Delta \varepsilon \\
&\quad + \frac{2\delta_0}{N(T-1)} [\chi(\gamma_0) - \chi(\gamma)]' [P^*(\gamma) - P(\gamma)] \Delta \varepsilon, \\
&= \mathcal{S}_{21} + \mathcal{S}_{22} + 2\mathcal{S}_{23}
\end{aligned} \tag{7.17}$$

where $\Delta \varepsilon = [\Delta \varepsilon_{12}, \dots, \Delta \varepsilon_{1T}, \dots, \Delta \varepsilon_{N2}, \dots, \Delta \varepsilon_{NT}]'$. First, we establish \mathcal{S}_{21} . Under Assumption P2(a), we have $\frac{1}{N(T-1)} \|\chi(\gamma_0) - \chi(\gamma)\|^2 = O_p(1)$. Then, by Lemma-1 and Assumption P2, we have

$$\begin{aligned}
\mathcal{S}_{21} &= \frac{\delta_0^2}{N(T-1)} \left\| [\chi(\gamma_0) - \chi(\gamma)]' [P^*(\gamma) - P(\gamma)] [\chi(\gamma_0) - \chi(\gamma)] \right\|_{sp} \\
&\leq \frac{\delta_0^2}{N(T-1)} \|\chi(\gamma_0) - \chi(\gamma)\|_{sp}^2 \|P^*(\gamma) - P(\gamma)\|_{sp} \\
&= o_p(1)
\end{aligned} \tag{7.18}$$

where $\|A\|_{sp}$ is the spectral norm of a square matrix A and $\|A\|_{sp} = \lambda_{max}^{1/2}(A'A)$. Then by Lemma-1 and closely following the proof of (7.18), we can show that \mathcal{S}_{22} and \mathcal{S}_{23} are also $o_p(1)$.

Last, by simple calculation, we can express \mathcal{S}_3 as

$$\begin{aligned}
\mathcal{S}_3 &= \frac{2\beta_{10}\delta_0}{N(T-1)} (\Delta v - \Delta \hat{v})' [I_{NT} - P(\gamma)] [\chi(\gamma_0) - \chi(\gamma) + \Delta \varepsilon] \\
&= \frac{2\beta_{10}\delta_0}{N(T-1)} (\Delta v - \Delta \hat{v})' [\chi(\gamma_0) - \chi(\gamma) + \Delta \varepsilon] - \frac{2\beta_{10}\delta_0}{N(T-1)} (\Delta v - \Delta \hat{v})' P(\gamma) [\chi(\gamma_0) - \chi(\gamma) + \Delta \varepsilon] \\
&= \mathcal{S}_{31} - 2\mathcal{S}_{32}
\end{aligned} \tag{7.19}$$

where $\mathcal{S}_{31} = o_p(1)$ under Assumption P2 and

$$\begin{aligned}
\mathcal{S}_{32} &= \frac{2\beta_{10}\delta_0}{N(T-1)} (\Delta v - \Delta \hat{v})' P(\gamma) [\chi(\gamma_0) - \chi(\gamma) + \Delta \varepsilon] \\
&\leq 2\beta_{10}\delta_0 \max_{\gamma \in \Gamma} \left\| \frac{1}{N(T-1)} (\Delta v - \Delta \hat{v})' \Delta x(\gamma) \right\| \max_{\gamma \in \Gamma} \left\| \left[\frac{1}{N(T-1)} \Delta x(\gamma)' \Delta x(\gamma) \right]^{-1} \right\| \\
&\quad \times \max_{\gamma \in \Gamma} \left\| \frac{1}{N(T-1)} \Delta x(\gamma)' [\chi(\gamma_0) - \chi(\gamma) + \Delta \varepsilon] \right\| \\
&= o_p(1) O_p(1) o_p(1)
\end{aligned} \tag{7.20}$$

by (7.27) in Lemma-1 and Assumption P2.

To sum up, we have $\sup_{\phi \in B \times \Gamma} |S_{NT}(\phi) - S_{NT}^*(\phi)| \xrightarrow{P} 0$.

Next, we show that the four conditions required by Section 4.3 [Andrews \(1994\)](#) also hold in our case. Denote $\frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T M_{it}(\phi)$ as the first order partial derivative of $S_{NT}^*(\phi)$ and the minimizer of $S_{NT}^*(\phi)$ is given by solving $\frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T M_{it}(\phi) = 0$. By definition, $M_{it}(\phi) = \Delta H_{it}(\phi) \Delta \varepsilon_{it}(\phi)$, and $\Delta \varepsilon_{it}(\phi) = \Delta y_{it} - \Delta x_{it}^*(\gamma) \theta$. Under Assumption P7, the true threshold value γ_0 is the unique minimizer of $L^*(\theta^*(\gamma), \gamma)$. And, by Assumption P3, $\theta^*(\gamma)$ is uniquely defined for all $\gamma \in \Gamma$, where Γ is a compact set. Combining Assumptions T3 and T7, we have the true parameter ϕ_0 minimizes $L^*(\phi) = L^*(\theta, \gamma)$ and is the unique solution of $M(\phi_0) = E[M_{it}(\phi_0)] = 0$, where $L^*(\phi) = E[S_{NT}^*(\phi)]$. Following we establish the four conditions one by one.

Condition 1: $\hat{\phi} \xrightarrow{P} \phi_0$.

For the kink regression model, $\Delta x_{it}^*(\gamma)$ is continuous in γ and we have $\Delta \varepsilon_{it}(\phi) = \Delta y_{it} - \Delta x_{it}^*(\gamma) \theta$. Therefore, $\Delta \varepsilon_{it}(\phi)$ and $\Delta \varepsilon_{it}(\phi)^2$ are continuous in ϕ . Using the Cauchy-Schwarz inequality, we have,

$$\Delta \varepsilon_{it}^2(\phi) \leq 2\Delta y_{it}^2 + 2|\Delta x_{it}^*(\gamma) \theta|^2 \leq 2\Delta y_{it}^2 + 2\bar{\theta}^2 \|\Delta x_{it}^*(\gamma)\|^2, \quad (7.21)$$

where $\bar{\theta} = \sup\{\|\theta\| : \theta \in B\}$ and is bounded under Assumption P6. Recall the definition of $\Delta x_{it}^*(\gamma)$, under Assumptions P2 and P7, we have the finite bound $\|\Delta x_{it}^*(\gamma)\|^2 \leq \Delta x_{it}^2 + 2(x_{it} - \gamma)^2 + \Delta z_{it}' \Delta z_{it} + \Delta v_{it}' \Delta v_{it}$. Thus, for $\phi \in B \times \Gamma$, $E[\Delta \varepsilon_{it}^2(\phi)] = O(1)$. Applying Lemma 2.4 of [Newey and Mcfadden \(1994\)](#), we can show $\sup_{\phi \in B \times \Gamma} |S_n^*(\phi) - L^*(\phi)| \xrightarrow{P} 0$ as $NT \rightarrow \infty$, where $L^*(\phi) = E[S_n^*(\phi)]$.

Finally, by Assumptions P3, P6, and P7, $B \times \Gamma$ is compact and ϕ_0 is the unique minimizer of $L^*(\phi)$. Thus, we conclude the proof of Condition 1 by applying Theorem 2.1 of [Newey and Mcfadden \(1994\)](#).

Condition 2: $\frac{1}{\sqrt{N(T-1)}} \sum_{i=1}^N \sum_{t=2}^T \Delta H_{it} \Delta \varepsilon_{it} \xrightarrow{d} N(0, S)$.

Following [Herrndorf \(1984\)](#), we complete the proof for Condition 2 by applying the CLT for the strong mixing process under Assumptions P1 and P2.

Condition 3: $\mathcal{Q}(\phi)$ is continuous in ϕ for $\phi \in B \times \Gamma$ and $\mathcal{Q}(\phi_0) = \mathcal{Q}$, where

$$\begin{aligned} \mathcal{Q}(\phi) &= -\frac{\partial}{\partial \phi} E[\Delta H_{it}(\phi) \Delta \varepsilon_{it}] = E[\Delta H_{it}(\phi) \Delta H'_{it}(\phi)] \\ &+ E \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \Delta \varepsilon_{it}(\phi)[I(x_t \geq \gamma) - I(x_{it-1} \geq \gamma)] \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \Delta \varepsilon_{it}(\phi)[I(x_t \geq \gamma) - I(x_{it-1} \geq \gamma)] & 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

Obviously, $\mathcal{Q}(\phi)$ is continuous w.r.t. θ . For γ , note that the parameter γ enters $\mathcal{Q}(\phi)$ through one of the following forms: $W_{i,t-D_1}I(x_{i,t-D_2} \geq \gamma)$, $x_{i,t-D_1}I(x_{i,t} \geq \gamma)I(x_{i,t-1} \geq \gamma)$, or $I(x_{i,t} \geq \gamma)I(x_{i,t-1} \geq \gamma)$, where $D_j \in \{0, 1\}$, $j = 1, 2$ and $W_{i,t-D_j}$ can be any vector whose elements are pairwise products of $(1, y_{it}, y_{i,t-1}, x_{it}, x_{i,t-1}, z_{it}, z_{i,t-1}, v_{it}, v_{i,t-1})$. By Assumptions P2, P4 and applying the law of iterated expectation, we can show that $\mathcal{Q}(\phi)$ is also continuous w.r.t γ . By definition, the second term of $\mathcal{Q}(\phi)$ equals zero when we evaluate it at $\phi = \phi_0$, which implies $\mathcal{Q}(\phi_0) = \mathcal{Q}$. Hence, we conclude our proof for Condition 3.

Condition 4: $g_{NT}(\phi) = \frac{1}{\sqrt{N(T-1)}} \sum_{i=1}^N \sum_{t=2}^T (M_{it}(\phi) - M(\phi))$ is stochastic equicontinuous, where $M_{it}(\phi) = \Delta H_{it}(\phi) \Delta \varepsilon_{it}(\phi)$ and $M(\phi) = E[M_{it}(\phi)]$.

Note that θ enters $M_{it}(\phi)$ with linear or quadratic forms. Therefore, we only need to establish the stochastic equicontinuity w.r.t. γ . Hence, without loss of generality, we temporarily ignore θ and focus on γ in $M_{it}(\phi)$ and write $M_{it}(\phi)$ as $M_{it}(\gamma)$. Then, we have

$$\begin{aligned} M_{it}(\gamma) &= \Delta H_{it}(\gamma) \Delta \varepsilon_{it}(\gamma) \\ &= [\Delta x_{it}^*(\gamma)(\Delta y_{it} - \Delta x_{it}^{*'}(\gamma)\theta), \delta[I(x_{it} \geq \gamma) - I(x_{it-1} \geq \gamma)(\Delta y_{it} - \Delta x_{it}^{*'}(\gamma)\theta)]. \end{aligned} \quad (7.22)$$

Note that the parameter γ enters $M_{it}(\gamma)$ through one of the following forms: $W_{i,t-D_1}I(x_{i,t-D_2} \geq \gamma)$, $x_{i,t-D_1}I(x_{i,t} \geq \gamma)I(x_{i,t-1} \geq \gamma)$, or $I(x_{i,t} \geq \gamma)I(x_{i,t-1} \geq \gamma)$, where $D_j \in \{0, 1\}$, $j = 1, 2$, which defined in Condition 3. Then we construct a bound which helps establishing Condition 4.

We denote $\mathbf{F}(\cdot)$ as the cumulative distribution of x_{it} . Then, for any $\gamma_2 \geq \gamma_1$ and $\gamma_1, \gamma_2 \in \Gamma$, by Assumption P4 and employing Taylor expansion, we have $\mathbf{F}(\gamma_2) - \mathbf{F}(\gamma_1) \leq \bar{f}|\gamma_2 - \gamma_1|$. Thus, the

bound equation for $W_{i,t-D_1}I(x_{i,t-D_2} \geq \gamma)$, $x_{i,t-D_1}I(x_{i,t} \geq \gamma)I(x_{i,t-1} \geq \gamma)$ and $I(x_{i,t} \geq \gamma)I(x_{i,t-1} \geq \gamma)$ are given as follows by applying the Hölder's inequality:

$$\begin{aligned}
& E|W_{i,t-D_1}I(\gamma_1 \leq x_{i,t} \leq \gamma_2)I(\gamma_1 \leq x_{i,t-1} \leq \gamma_2)|^2 \\
& \leq E(|W_{i,t-D_1}|^{2r})^{1/r} (E|I(\gamma_1 \leq x_{i,t} \leq \gamma_2)|)^{1/q_1} (E|I(\gamma_1 \leq x_{i,t-1} \leq \gamma_2)|)^{1/q_2} \\
& \leq \mathcal{C}(\mathbf{F}(\gamma_2) - \mathbf{F}(\gamma_1))^{(1/q_1+1/q_2)} \\
& \leq \mathcal{C}\bar{f}^{(1/q_1+1/q_2)}|\gamma_2 - \gamma_1|^{(1/q_1+1/q_2)}, \tag{7.23}
\end{aligned}$$

where $\frac{1}{q_1} + \frac{1}{q_2} + \frac{1}{r} = 1$ and $E(\|W_{i,t-D_1}\|^{2r})^{1/r} \leq \mathcal{C}$.

Following we establishing Condition 4 by applying [Doukhan et al. \(1995\)](#) Theorem 1. In our case, with martingale differences sequences, their condition (2.15) holds. Let γ_k be an equally spaced grid search point on Γ , where $k = 1, \dots, \mathcal{N}$. Then, for any $\gamma \in \Gamma$, there exist a bracket such that $\min[M_{it}(\gamma_{k-1}), M_{it}(\gamma_k)] \leq M_{it}(\gamma) \leq \max[M_{it}(\gamma_{k-1}), M_{it}(\gamma_k)]$. Denote μ as any positive number and $\mathcal{N}(\mu) = \mu^{-2/q}$. By the bound equation (7.23), we have $E\|M_{it}(\gamma_k) - M_{it}(\gamma_{k-1})\|^2 \leq \mathcal{C}\bar{f}^{1/q}|\gamma_k - \gamma_{k-1}|^{1/q} \leq O(\mathcal{N}(\mu)^{-q}) = O(\mu^2)$, where $O(\mathcal{N}(\mu)^{-1})$ is the distance between grid points and $\frac{1}{q} = \frac{1}{q_1} + \frac{1}{q_2}$. Thus, we have $\mathcal{N}(\mu)$ are the L^2 bracketing numbers and $H_2(\mu) = \ln \mathcal{N}(\mu) = |\log \mu|$ is the metric entropy for the class $\{m_{it}(\gamma)|\gamma \in \Gamma\}$. Combining this with Assumption P1, we can apply Theorem 1 of [Doukhan et al. \(1995\)](#)⁸ to establish the stochastic equicontinuity of g_{NT} w.r.t γ , which finish the proof of condition 4.

As Conditions 1-4 are proved above, it is sufficient for us to establish Theorem 1-panel.■.

Lemma 1 *Under Assumptions P2, P3, P5(b) and P7, we have $\max_{\gamma \in \Gamma} \|P^*(\gamma) - P(\gamma)\|_{sp} = o_p(1)$.*

⁸Note that, by collecting the bounded condition we show above and the Assumption P1, condition (2.15) in [Doukhan et al. \(1995\)](#) is satisfied. Thus, Theorem 1 of [Doukhan et al. \(1995\)](#) holds here.

Proof. By triangular inequality and simple calculations, we have

$$\begin{aligned}
& \max_{\gamma \in \Gamma} \|P(\gamma) - P^*(\gamma)\|_{sp} \\
&= \max_{\gamma \in \Gamma} \|\Delta x(\gamma) [\Delta x(\gamma)' \Delta x(\gamma)]^{-1} \Delta x(\gamma)' - \Delta x^*(\gamma) [\Delta x^*(\gamma)' \Delta x^*(\gamma)]^{-1} \Delta x^*(\gamma)'\|_{sp} \\
&\leq \max_{\gamma \in \Gamma} \|\Delta x(\gamma) \{ [\Delta x(\gamma)' \Delta x(\gamma)]^{-1} - [\Delta x^*(\gamma)' \Delta x^*(\gamma)]^{-1} \} \Delta x(\gamma)'\|_{sp} \\
&\quad + \max_{\gamma \in \Gamma} \|[\Delta x(\gamma) - \Delta x^*(\gamma)] [\Delta x^*(\gamma)' \Delta x^*(\gamma)]^{-1} [\Delta x(\gamma) - \Delta x^*(\gamma)]'\|_{sp} \\
&\quad + 2 \max_{\gamma \in \Gamma} \|[\Delta x(\gamma) - \Delta x^*(\gamma)] [\Delta x^*(\gamma)' \Delta x^*(\gamma)]^{-1} \Delta x^*(\gamma)'\|_{sp} \\
&= \mathcal{S}_{p1} + \mathcal{S}_{p2} + 2\mathcal{S}_{p3} = o_p(1). \tag{7.24}
\end{aligned}$$

First, we verify that \mathcal{S}_{p2} and \mathcal{S}_{p3} are $o_p(1)$. Under Assumptions P2, P3 and P7, by equation (7.16), we have

$$\begin{aligned}
\mathcal{S}_{p2} &= \max_{\gamma \in \Gamma} \|[\Delta x(\gamma) - \Delta x^*(\gamma)] [\Delta x^*(\gamma)' \Delta x^*(\gamma)]^{-1} [\Delta x(\gamma) - \Delta x^*(\gamma)]'\|_{sp} \\
&\leq \frac{1}{N(T-1)} \max_{\gamma \in \Gamma} \|\Delta x(\gamma) - \Delta x^*(\gamma)\| \max_{\gamma \in \Gamma} \left\| \left[\frac{1}{N(T-1)} \Delta x^*(\gamma)' \Delta x^*(\gamma) \right]^{-1} \right\|_{sp} \\
&\quad \times \max_{\gamma \in \Gamma} \|\Delta x(\gamma) - \Delta x^*(\gamma)\| \\
&= \frac{1}{N(T-1)} \|\Delta v - \Delta \hat{v}\|^2 \max_{\gamma \in \Gamma} \left\| \left[\frac{1}{N(T-1)} \Delta x^*(\gamma)' \Delta x^*(\gamma) \right]^{-1} \right\|_{sp} \\
&= O_p([N(T-1)]^{-1}) O_p(1) = o_p(1), \tag{7.25}
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{S}_{p3} &= \max_{\gamma \in \Gamma} \|[\Delta x(\gamma) - \Delta x^*(\gamma)] [\Delta x^*(\gamma)' \Delta x^*(\gamma)]^{-1} \Delta x^*(\gamma)'\|_{sp} \\
&\leq \|\Delta \hat{v} - \Delta v\| \max_{\gamma \in \Gamma} \left\| \left[\frac{1}{N(T-1)} \Delta x^*(\gamma)' \Delta x^*(\gamma) \right]^{-1} \right\|_{sp} \max_{\gamma \in \Gamma} \frac{1}{N(T-1)} \|\Delta x^*(\gamma)\| \\
&= O_p(1) O_p(1) o_p(1) = o_p(1), \tag{7.26}
\end{aligned}$$

where we have

$$\begin{aligned}
& \left\| \left[\frac{1}{N(T-1)} \Delta x^*(\gamma)' \Delta x^*(\gamma) \right]^{-1} \right\|_{sp} \\
= & \left\| \left[\frac{1}{N(T-1)} \Delta x^*(\gamma)' \Delta x^*(\gamma) \right]^{-1} \right\|_{sp} = \lambda_{min}^{-1} \left[\frac{1}{N(T-1)} \Delta x^*(\gamma)' \Delta x^*(\gamma) \right] \\
= & \lambda_{min}^{-1} (E[\Delta x_{it}^*(\gamma) \Delta x_{it}^{*'}(\gamma)]) + O_p \left(\left\| \frac{1}{N(T-1)} \Delta x^*(\gamma)' \Delta x^*(\gamma) - E[\Delta x_{it}^*(\gamma) \Delta x_{it}^{*'}(\gamma)] \right\| \right) \\
= & O_p(1). \tag{7.27}
\end{aligned}$$

by Assumption P3 and applying Weyl's theorem in [Seber \(2008\)](#). $\lambda_{min}(A)$ denotes the smallest eigenvalue of a symmetric matrix A . Next, we show \mathcal{S}_{p1} is $o_p(1)$. Under Assumption P2, we have

$$\begin{aligned}
\mathcal{S}_{p1} &= \max_{\gamma \in \Gamma} \left\| \Delta x(\gamma) \{ [\Delta x(\gamma)' \Delta x(\gamma)]^{-1} - [\Delta x^*(\gamma)' \Delta x^*(\gamma)]^{-1} \} \Delta x(\gamma)' \right\|_{sp} \\
&\leq \max_{\gamma \in \Gamma} \frac{1}{N(T-1)} \left\| \Delta x(\gamma) \right\|^2 N(T-1) \max_{\gamma \in \Gamma} \left\| [\Delta x(\gamma)' \Delta x(\gamma)]^{-1} - [\Delta x^*(\gamma)' \Delta x^*(\gamma)]^{-1} \right\|_{sp} \\
&= O_p(1) o_p(1) \tag{7.28}
\end{aligned}$$

where by (7.27) and closely following the proof of Theorem 1-time series (7.12), (7.13) and (7.14), we obtain

$$\begin{aligned}
& N(T-1) \max_{\gamma \in \Gamma} \left\| [\Delta x(\gamma)' \Delta x(\gamma)]^{-1} - [\Delta x^*(\gamma)' \Delta x^*(\gamma)]^{-1} \right\|_{sp} \\
= & N(T-1) \max_{\gamma \in \Gamma} \left\| [\Delta x(\gamma)' \Delta x(\gamma)]^{-1} \{ [\Delta x(\gamma)' \Delta x(\gamma)] - [\Delta x^*(\gamma)' \Delta x^*(\gamma)] \} [\Delta x^*(\gamma)' \Delta x^*(\gamma)]^{-1} \right\|_{sp} \\
\leq & \max_{\gamma \in \Gamma} \left\| \left[\frac{1}{N(T-1)} \Delta x(\gamma)' \Delta x(\gamma) \right]^{-1} \right\|_{sp} \max_{\gamma \in \Gamma} \frac{1}{N(T-1)} \left\| [\Delta x(\gamma)' \Delta x(\gamma)] - [\Delta x^*(\gamma)' \Delta x^*(\gamma)] \right\| \\
& \times \max_{\gamma \in \Gamma} \left\| \left[\frac{1}{N(T-1)} \Delta x^*(\gamma)' \Delta x^*(\gamma) \right]^{-1} \right\|_{sp} \\
= & O_p(1) o_p(1) O_p(1) = o_p(1) \tag{7.29}
\end{aligned}$$

Given \mathcal{S}_{p1} , \mathcal{S}_{p2} and \mathcal{S}_{p3} are both $o_p(1)$, we have $\max_{\gamma \in \Gamma} \|P^*(\gamma) - P(\gamma)\|_{sp} = o_p(1)$ ■

Table 2: Estimation Results for DGP1

	MEAN- β ($\beta_{10} = 1$)	RMSE- β	MEAN- δ ($\delta_0 = 1$)	RMSE- δ	MEAN- γ ($\gamma_0 = 2$)	RMSE- γ
$\kappa = 0.05$	No CF					
T=100	1.0248	0.0313	1.0006	0.0261	2.0005	0.0405
T=200	1.025	0.0279	1	0.018	2.0001	0.0232
T=300	1.025	0.0267	0.9999	0.0145	1.9998	0.0137
T=400	1.0252	0.0263	0.9997	0.0123	2.0001	0.0071
$\kappa = 0.05$	CF					
T=100	0.9965	0.0252	1.0004	0.0248	1.9999	0.0382
T=200	0.9984	0.0158	1	0.017	1.9998	0.0202
T=300	0.9989	0.0121	0.9999	0.0137	1.9998	0.0115
T=400	0.9994	0.0101	0.9997	0.0116	2.0001	0.0057
$\kappa = 0.5$	No CF					
T=100	1.2449	0.2541	1.0103	0.0907	1.9999	0.1477
T=200	1.2488	0.2531	1.0036	0.0631	2.002	0.0993
T=300	1.2482	0.2511	1.0022	0.051	1.997	0.0818
T=400	1.2494	0.2515	1.0008	0.0432	1.9998	0.0717
$\kappa = 0.5$	CF					
T=100	0.9685	0.1564	1.0004	0.0248	1.9999	0.0382
T=200	0.9851	0.0923	1	0.017	1.9998	0.0202
T=300	0.9891	0.0715	0.9999	0.0137	1.9998	0.0115
T=400	0.993	0.0618	0.9997	0.0116	2.0001	0.0057
$\kappa = 0.95$	No CF					
T=100	1.4534	0.4738	1.0404	0.175	1.9938	0.3185
T=200	1.4686	0.4771	1.0149	0.1176	2.0038	0.2023
T=300	1.469	0.4746	1.0092	0.0947	1.9943	0.1599
T=400	1.4722	0.4759	1.0052	0.08	1.9998	0.1342
$\kappa = 0.95$	CF					
T=100	0.9406	0.2959	1.0004	0.0248	1.9999	0.0382
T=200	0.9717	0.174	1	0.017	1.9998	0.0202
T=300	0.9793	0.1349	0.9999	0.0137	1.9998	0.0115
T=400	0.9866	0.1166	0.9997	0.0116	2.0001	0.0057
$\kappa = 2$	No CF					
T=100	1.9103	0.9654	1.1788	0.4275	1.9975	0.6629
T=200	1.9581	0.9822	1.0872	0.2698	2.0054	0.4974
T=300	1.9689	0.9845	1.0567	0.2098	1.9906	0.4067
T=400	1.9808	0.9907	1.0362	0.1732	1.9972	0.3354
$\kappa = 2$	CF					
T=100	0.8754	0.6224	1.0004	0.0248	1.9999	0.0382
T=200	0.9406	0.3655	1	0.017	1.9998	0.0202
T=300	0.9565	0.2835	0.9999	0.0137	1.9998	0.0115
T=400	0.9716	0.2449	0.9997	0.0116	2.0001	0.0057

Note: CF=Control function approach

Table 3: Estimation Results for DGP2

	MEAN- β_1 ($\beta_{10} = 1$)	RMSE- β_1	MEAN- δ ($\delta_0 = 1$)	RMSE- δ	MEAN- β_3 ($\beta_{30} = 1$)	RMSE- β_3	MEAN- γ ($\gamma_0 = 2$)	RMSE- γ
$\kappa = 0.05$	NO CF							
n=100	1.0249	0.0330	1.0002	0.0288	1.0249	0.0263	2	0.0448
n=200	1.0249	0.0287	0.9994	0.0202	1.0250	0.0257	1.9994	0.0266
n=300	1.0247	0.0269	1.0004	0.0160	1.0250	0.0255	1.9999	0.0163
n=400	1.0248	0.0263	1.0001	0.0138	1.0250	0.0253	1.9998	0.0101
$\kappa = 0.05$	CF							
n=100	0.9970	0.0268	1.0004	0.0261	0.9973	0.0240	2.0003	0.0402
n=200	0.9986	0.0159	0.9995	0.0181	0.9989	0.0113	1.9998	0.0200
n=300	0.9989	0.0120	1.0002	0.0144	0.9992	0.0091	1.9998	0.0107
n=400	0.9990	0.0102	1.0001	0.01239	0.9991	0.0078	1.9999	0.0054
$\kappa = 0.5$	NO CF							
n=100	1.2416	0.2639	1.0196	0.1362	1.2495	0.2526	2.0015	0.2336
n=200	1.2459	0.2556	1.0058	0.0920	1.2498	0.2513	1.9984	0.1487
n=300	1.2471	0.2531	1.0081	0.0745	1.2501	0.2511	2.0019	0.1169
n=400	1.2483	0.2527	1.0034	0.0634	1.2496	0.2503	2.0001	0.0995
$\kappa = 0.5$	CF							
n=100	0.9707	0.1554	1.0004	0.0261	0.9729	0.1832	2.0003	0.0402
n=200	0.9857	0.0918	0.9995	0.0181	0.9879	0.0859	1.9998	0.0200
n=300	0.9908	0.0697	1.0002	0.0144	0.9911	0.0705	1.9998	0.0107
n=400	0.9923	0.0591	1.0001	0.0123	0.9906	0.06	1.9999	0.0054
$\kappa = 0.95$	NO CF							
n=100	1.4327	0.4870	1.0910	0.2836	1.4741	0.4798	2.0030	0.4857
n=200	1.4554	0.4778	1.0353	0.1789	1.4747	0.4774	1.9984	0.3251
n=300	1.4639	0.4766	1.0286	0.1427	1.4752	0.4770	2.0059	0.2472
n=400	1.4667	0.4758	1.0154	0.1201	1.4741	0.4755	1.9974	0.2036
$\kappa = 0.95$	CF							
n=100	0.9445	0.2926	1.0004	0.0261	0.9484	0.3451	2.0003	0.0402
n=200	0.9727	0.1732	0.9995	0.0181	0.9769	0.1630	1.9998	0.0200
n=300	0.9827	0.1315	1.0002	0.0144	0.9831	0.1338	1.9998	0.0107
n=400	0.9855	0.1116	1.0001	0.0123	0.9822	0.1137	1.9999	0.0054
$\kappa = 2$	NO CF							
n=100	1.8730	1.0006	1.2777	0.7323	1.9984	1.0101	2.0203	0.7903
n=200	1.9085	0.9673	1.1748	0.4336	1.9996	1.0051	1.9936	0.6681
n=300	1.9393	0.9748	1.1374	0.3429	2.0004	1.0041	2.0124	0.5622
n=400	1.9538	0.9791	1.0908	0.2753	1.9982	1.0011	1.9952	0.4914
$\kappa = 2$	CF							
n=100	0.8834	0.6139	1.0004	0.0261	0.8914	0.7232	2.0003	0.0402
n=200	0.9425	0.3640	0.9995	0.0181	0.9513	0.3430	1.9998	0.0200
n=300	0.9638	0.2762	1.0002	0.0144	0.9644	0.2816	1.9998	0.0107
n=400	0.9698	0.2344	1.0001	0.0123	0.9626	0.2393	1.9999	0.0054
Note: CF=Control function approach								

Table 4: Estimation Results for DGP3

		MEAN- β_1 ($\beta_{10} = 1$)	RMSE- β_1	MEAN- δ ($\delta_0 = 1$)	RMSE- δ	MEAN- γ ($\gamma_0 = 2$)	RMSE- γ
No	CF FD						
T=10	N=10	1.8529	0.9965	1.2912	0.745	1.9985	0.8329
	N=20	1.9137	0.9692	1.1745	0.4255	2.0045	0.6624
	N=30	1.9394	0.9739	1.1244	0.3292	2.0066	0.581
	N=40	1.9539	0.9782	1.0926	0.2705	1.9993	0.5053
T=20	N=10	1.9239	0.9703	1.1484	0.3751	1.9947	0.6415
	N=20	1.9605	0.9805	1.0748	0.2448	1.9887	0.4519
	N=30	1.9755	0.9877	1.0454	0.1905	1.9955	0.3698
	N=40	1.9823	0.9905	1.0309	0.1603	1.9964	0.2961
T=30	N=10	1.9444	0.9726	1.106	0.2879	1.9888	0.5154
	N=20	1.9749	0.9864	1.0465	0.1847	1.9979	0.3527
	N=30	1.9886	0.9955	1.0223	0.1474	2.0013	0.2737
	N=40	1.9908	0.9955	1.0171	0.1247	1.9979	0.2278
T=40	N=10	1.9653	0.9839	1.071	0.2336	2	0.4497
	N=20	1.9838	0.9912	1.0302	0.1547	1.9947	0.2851
	N=30	1.9907	0.9954	1.0202	0.1246	2.0033	0.2235
	N=40	1.9947	0.998	1.0125	0.1069	2.0031	0.1856
CF	FD						
T=10	N=10	0.8509	0.7733	1.01	0.063	2.0062	0.0974
	N=20	0.9404	0.4249	1.0049	0.0312	2.0023	0.0505
	N=30	0.9582	0.334	1.0025	0.0234	2.0009	0.0348
	N=40	0.9687	0.2777	1.0014	0.0191	2.0003	0.0256
T=20	N=10	0.9444	0.3906	1.0015	0.0249	2	0.0385
	N=20	0.9725	0.2574	1.0008	0.0163	2.0005	0.0182
	N=30	0.9795	0.2022	1.0005	0.0131	1.9999	0.008
	N=40	0.9863	0.1739	1.0001	0.011	2	0.004
T=30	N=10	0.9643	0.2902	1.0006	0.0182	2	0.0237
	N=20	0.9804	0.2009	1.0002	0.0125	2	0.0075
	N=30	0.9893	0.16	1.0002	0.0102	2	0.0014
	N=40	0.9917	0.1375	1.0001	0.0089	2	0.0014
T=40	N=10	0.9728	0.2459	1.0005	0.015	2	0.0148
	N=20	0.9853	0.1679	1.0001	0.0107	2	0.0035
	N=30	0.9908	0.1372	1.0001	0.0086	2	0
	N=40	0.9948	0.1178	1.0001	0.0075	2	0
FD=first difference; CF=control function approach;							

Table 5: Estimation Results for DGP4

		MEAN- β_1 ($\beta_{10} = 1$)	RMSE- β_1	MEAN- δ ($\delta_0 = 1$)	RMSE- δ	MEAN- β_3 ($\beta_{30} = 1$)	RMSE- β_3	MEAN- γ ($\gamma_0 = 2$)	RMSE- γ
NO	CF FD								
T=10	N=10	1.8287	1.0494	1.3452	0.9967	1.9996	1.0183	1.9938	0.8768
	N=20	1.8725	0.9687	1.2520	0.5777	2.0000	1.0087	1.9983	0.7566
	N=30	1.9114	0.9685	1.1816	0.4439	1.9990	1.0046	2.0093	0.6678
	N=40	1.9318	0.9730	1.1447	0.3591	1.9981	1.0024	2.0105	0.5908
T=20	N=10	1.8755	0.9707	1.2434	0.5786	2.0034	1.0120	2.0005	0.7403
	N=20	1.9306	0.9725	1.1493	0.3599	2.0015	1.0059	2.0092	0.5913
	N=30	1.9548	0.9816	1.0875	0.2802	1.9993	1.0021	1.9951	0.4993
	N=40	1.9669	0.9849	1.0660	0.2283	1.9995	1.0016	2.0002	0.4272
T=30	N=10	1.9063	0.9632	1.1925	0.4410	2.0009	1.0067	2.0047	0.6625
	N=20	1.9523	0.9783	1.0962	0.2774	2.0004	1.0031	2.0019	0.4969
	N=30	1.9689	0.9853	1.0617	0.2162	2.0005	1.0024	2.0001	0.4056
	N=40	1.9807	0.9920	1.0433	0.1855	1.9989	1.0003	2.0035	0.3388
T=40	N=10	1.9331	0.9728	1.1368	0.3575	2.0001	1.0043	2.0001	0.5849
	N=20	1.9576	0.9769	1.0763	0.2355	1.9988	1.0009	1.9884	0.4281
	N=30	1.9785	0.9897	1.0421	0.1823	2.0000	1.0014	2.0000	0.3414
	N=40	1.9860	0.9937	1.0282	0.1561	2.0001	1.0011	1.9976	0.2817
CF	FD								
T=10	N=10	0.9686	0.2487	1.0088	0.0889	0.9715	0.2412	2.0014	0.1411
	N=20	0.9851	0.1701	1.0027	0.0587	0.9908	0.1613	1.9978	0.0935
	N=30	0.9869	0.1375	1.0016	0.0498	0.9902	0.1316	1.9995	0.0759
	N=40	0.9948	0.1168	1.0022	0.0423	0.9930	0.1113	1.9998	0.0654
T=20	N=10	0.9870	0.1659	1.0017	0.0449	0.9875	0.1640	1.9997	0.0711
	N=20	0.9939	0.1161	1.0019	0.0320	0.9950	0.1139	2.0012	0.0501
	N=30	0.9974	0.0922	1.0001	0.0258	0.9976	0.0919	2.0004	0.0401
	N=40	0.9992	0.0803	1.0004	0.0219	0.9982	0.0786	2.0007	0.0317
T=30	N=10	0.9926	0.1347	1.0008	0.0316	0.9915	0.1330	1.9996	0.0492
	N=20	0.9968	0.0923	1.0004	0.0216	0.9969	0.0934	2.0003	0.0313
	N=30	0.9974	0.0748	1.0000	0.0180	0.9980	0.0748	2.0000	0.0203
	N=40	0.9984	0.0648	1.0002	0.0157	0.9962	0.0657	2.0002	0.0150
T=40	N=10	0.9932	0.1147	1.0005	0.0247	0.9951	0.1145	1.9997	0.0359
	N=20	0.9975	0.0796	1.0001	0.0175	0.9981	0.0806	2.0003	0.0190
	N=30	0.9996	0.0649	1.0000	0.0141	0.9976	0.0643	2.0002	0.0098
	N=40	0.9980	0.0558	0.9999	0.0122	0.9993	0.0556	2.0001	0.0071
FD=first difference; CF=control function approach									

Table 6: Province(Canada) or State(US) in our data set

Canada		US	
Alberta	Alaska	Kentucky	Ohio
British Columbia	Alabama	Louisiana	Oklahoma
Manitoba	Arkansas	Massachusetts	Oregon
New Brunswick	Arizona	Maryland	Pennsylvania
Newfoundland and Labrador	California	Maine	Puerto Rico
Nova Scotia	Colorado	Michigan	Rhode Island
Ontario	Connecticut	Minnesota	South Carolina
Prince Edward Island	District of Columbia	Missouri	South Dakota
Quebec	Delaware	Mississippi	Tennessee
Saskatchewan	Florida	Montana	Texas
10	Georgia	North Carolina	Utah
	Guam	North Dakota	Virginia
	Hawaii	Nebraska	Vermont
	Iowa	New Hampshire	Washington
	Idaho	New Jersey	Wisconsin
	Illinois	New Mexico	West Virginia
	Indiana	Nevada	Wyoming
	Kansas	New York	
		53	

Table 7: Correlation between unemployment rate and COVID-19 cases
Canada dataset

	(1)	(2)	(3)	(4)
Model	OLS	Threshold	OLS	Threshold
$\gamma(\text{cases})$		4.11		4.11
β_{linear}	0.5523*** (0.1657)		0.6728*** (0.3360)	
β_{low}		0.4918*** (0.1678)		0.5955*** (0.3375)
β_{high}		2.0915*** (0.6906)		2.2618*** (0.7461)
Control function			Linear	Linear
Fixed effect	✓	✓	✓	✓
N_{low}		170		170
N_{high}		40		40
N_{total}	210	210	210	210

***, **, * indicate significant at 1% level, 5% level, 10% level, respectively.

Linearity test: $H_0 : \beta_{low,0} = \beta_{high,0}$, bootstrap p - value = 0.0001.

Endogeneity test: $H_0 : \beta_{40} = 0$, $t = 19.76 > t_{0.01} = 2.617$.

(Note β_{40} is the coefficient of Δv_{it} . See Section 5 for detailed discussion.)

Table 8: Correlation between unemployment rate and COVID-19 cases
US dataset

	(1)	(2)	(3)	(4)
Model	OLS	Threshold	OLS	Threshold
$\gamma(\text{cases})$		4.23		4.22
β_{linear}	0.0136*** (0.0789)		0.1013*** (0.093)	
β_{low}		-0.0726*** (0.0848)		0.0178*** (0.0978)
β_{high}		0.5354*** (0.2037)		0.6207*** (0.2113)
Control function			Linear	Linear
Fixed effect	✓	✓	✓	✓
N_{low}		488		482
N_{high}		500		506
N_{total}	988	988	988	988

***, **, * indicate significant at 1% level, 5% level, 10% level, respectively.

Linearity test: $H_0 : \beta_{low,0} = \beta_{high,0}$, bootstrap p -value = 0.0003.

Endogeneity test: $H_0 : \beta_{40} = 0$, $t = 13.27 > t_{0.01} = 2.617$.

(Note β_{40} is the coefficient of Δv_{it} . See Section 5 for detailed discussion.)