Endogenous kink threshold regression

Jianhan Zhang* Chaoyi Chen†‡ Yiguo Sun§ Thanasis Stengos ¶

October 16, 2023

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 nonparametric 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 impact of COVID-19 cases on labor markets in the US and Canada.

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

*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

The threshold regression (TR) model is extensively used to capture potential shifts in economic relationships; e.g., Tong (1990) and Hansen (2000). However, the conventional TR model requires a discontinuous regression function at the true threshold level, yet in many empirical applications, this discontinuity may not be warranted. As an alternative, Chan and Tsay (1998) introduces a continuous threshold autoregressive model, allowing for a piece-wise linear function of the threshold variable. This model permits a continuous threshold regression but retains a slope discontinuity at the true threshold level, making it a specific case within the broader class of threshold autoregressive models. Building on Chan and Tsay (1998), Hansen (2017) extends the concept by introducing testing for a threshold effect and inference on the regression parameters for a continuous threshold model with an unknown threshold parameter value, referred to as the kink threshold regression (KTR) model. It is well established that the least-squares estimator for the TR model has a nonstandard limiting distribution and is super consistent. For instance, Chan (1993) establishes, under a "fixed threshold effect" assumption, that the threshold parameter estimator converges to a function of a compound Poisson process. In contrast, under a "diminishing threshold effect" assumption, Hansen (2000) shows that the limiting distribution involves two independent Brownian motions. However, 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 continuity.

The above mentioned studies assume strict exogeneity for both slope regressors and the threshold variable. As real-world nonlinear asymmetric mechanisms often involve endogeneity, the literature on the TR model has evolved to account for this. Under Hansen (2000)'s diminishing threshold effect framework, Caner and Hansen (2004) permit endogenous slope regressors by employing generalized method of moments (GMM) and two-stage least squares (2SLS) to estimate the slope parameters. 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, introducing an inverse Mills ratio as a bias correction term. Following Kourtellos et al. (2016), Christopoulos et al. (2021) use a copula method to handle the endogenous threshold variable. Yu et al. (2023) 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. One group extends the 2SLS method of Caner and Hansen (2004), while the other is a natural extension of the conventional CF approach of Newey et al. (1999). It is important to note that both CF methods cannot be directly used to estimate the KTR model with endogeneity. In fact, continuity makes the inference of the least squares estimator for the KTR model quite different from the conventional TR model even without endogeneity. Hidalgo et al. (2019) emphasize that attempting to estimate a KTR model under the TR framework of Hansen (2000), ignoring continuity of the true model, results in an irregular Hessian matrix. [1] This makes the least squares estimator of the threshold parameter to converge at a slower cube root-n rate, in contrast to the root-n rate for the KTR model (Hansen (2017)). Consequently, both CF methods proposed by Yu et al. (2023), designed for the TR framework, cannot apply to the KTR model without modification.[2] More recently, Kourtellos et al. (2022) extend Yu et al. (2023) to allow for an unknown endogenous form, introducing a nonparametric bias correction term into the TR model. The proposed semiparametric model avoids misspecification issues but remains within the TR model framework. Seo and Shin (2016) consider a dynamic panel TR model with endogeneity and develop a first-differenced GMM estimator that accommodates both endogenous threshold variables and

[1] Note 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., Assumption 1.7 in Hansen (2000).

[2] The KTR model violates Assumption I.9 for CF-I and II.9 for CF-II in the Yu et al. (2023).

regressors. Under a fixed threshold effect framework of Chan (1993) and assuming i.i.d. samples, Yu and Phillips (2018) construct an integrated difference kernel estimator(IDKE) for the threshold parameter. The IDKE offers consistency without requiring instrumental variables and is super-consistent for the TR model with an endogenous threshold variable and exogenous slope regressors. However, the i.i.d. assumption limits its applicability.

In contrast to many TR model studies, surprisingly, to our knowledge, no estimation and asymptotic results for the least squares estimator of the KTR model with endogeneity have been developed.[3] Therefore, this paper aims to fill this gap in the literature. Building upon the work of Yu et al. (2023) and Kourtellos et al. (2022), our study introduces a two-step semiparametric CF approach to handle endogeneity in a KTR model, allowing for both slope regressors and the threshold variable to be endogenous. Our main theoretical contributions can be summarized in three aspects. Firstly, in the spirit of Kourtellos et al. (2022), we employ a nonparametric error correction term to control for potential endogeneity, extending from a TR model to a KTR model. Our proposed estimator exhibits a joint normal limiting distribution with a standard root-n convergence rate. Secondly, in the first step, we adopt a nonparametric IV regression approach, differing from Kourtellos et al. (2022) who employ a linear IV regression method. Consequently, our proposed two-step nonparametric CF approach is more general, fitting within the framework of non/semi-parametric estimation of the structural kink regression model. Thirdly, our method shares similarities with Ozabaci et al. (2014), which address endogeneity in a nonparametric additive regression model and can be expanded to a partially linear model, both through the sieve approximation method.[4] Yet, their results are not directly applicable to our situation

[3] We notice that the first-differenced GMM estimator proposed by Seo and Shin (2016) is applicable to the KTR panel data model with endogeneity. In their work, they also introduce a two-stage least squares (2SLS) estimator in the appendix; however, it is distinct from our approach. Their 2SLS estimator specifically accommodates an endogenous regressor while maintaining the threshold variable as exogenous.

[4] Note that, to mitigate the curse of dimensionality, we adopt their approach by imposing additive

since our proposed semiparametric KTR model intentionally lacks smoothness at the kink point. Technically, this necessitates the proof to verify a stochasticity equicontinuity condition, extending Hansen (2017) from the finite dimension case to the infinite dimension case. We establish this condition by following Chen (2007).

We develop the KTR model for both time series and panel data settings. In the time series model, we focus on estimating and studying the asymptotic properties of the least squares estimator with weakly dependent data. In the panel data context, we address time-invariant fixed effects using the first-differencing (FD) method. We derive the asymptotic results for our proposed estimator with a large number of cross-sections (N) and fixed time periods (T) observations. We then apply our model to assess the threshold effect of COVID-19 cases on the labor markets of the United States and Canada. Since the beginning of 2020, the global economy has been significantly affected by the COVID-19 pandemic. A multitude of studies have investigated its consequences and spread, analyzing both linear and nonlinear aspects. For instance, Karavias et al. (2022) examines the structural effect of COVID-19 on stock returns using a linear panel model with an unknown structural break time. Considering the possible kink relationship between contact rate and the outbreak of COVID-19, Lee et al. (2021) construct a Susceptible–Infected–Recovered (SIR) model to monitor the outbreak via using reported cases of COVID-19. Regarding the labor market, a body of literature explores the indirect effects of COVID-19, such as the impact of government Stay-at-Home/Lockdown policies on the labor market (e.g., Baek et al. (2021), Kong and Prinz (2020)). Other studies focus on investigating the effect of COVID-19 on the labor market of some particular groups (e.g., Lee et al. (2021)). Surprisingly, few studies examine the overall impact of COVID-19 on the entire labor market. Given the prolonged duration and multiple waves of COVID-19 cases, we hypothesize the presence of a threshold effect or structural break in the relationship between COVID-19 cases and the

constraints in both two regression stages.

unemployment rate. Thus, we apply our proposed KTR model with endogeneity to explore this potential nonlinearity. Our findings indicate that while the impact of COVID-19 on unemployment is consistently positive in both regimes, it becomes more pronounced when the number of cases exceeds a certain threshold.[5]

The rest of the paper is organized as follows. Section 2 introduces the times-series KTR model with endogeneity, presenting the estimation method and asymptotic properties of our proposed estimators. Section 3 extends the model to the panel data 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. Additional materail and the mathematical proofs are provided in the appendix/supplementary material.

To proceed, we adopt the following notation throughout the paper. We use subscript 0 to denote the true parameters and the accent $\hat{}$ to denote the estimators. We define $\|\cdot\|$ as the Euclidean norm and $\|\cdot\|_\infty$ as the sup-norm. The operators $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and distribution, respectively. $\mathbf{0}_{A\times B}$ denotes a $A \times B$ matrix of zeros, while I_m denotes an identity matrix of size m. $\nabla M(a)$ denotes the first order partial derivative of function $M(\cdot)$ with respect to a.

2 Time series model

2.1 Model and estimation

Following Hansen (2017), we consider a KTR model

$$y_t = \beta_{10}(x_t - \gamma_0)I(x_t < \gamma_0) + \beta_{20}(x_t - \gamma_0)I(x_t \geq \gamma_0) + z_t'\beta_{30} + u_t, \tag{1}$$

[5]We use the number of COVID-19 tests performed as the instrumental variable since it is strongly correlated with the number of cases and has no relevance to the unemployment rate.

where x_t is a scalar and plays the role of the threshold variable, z_t is an $\ell \times 1$ vector of covariates and includes an intercept term, $I(\cdot)$ is the indicator function and u_t is the error term with zero mean and finite variance. In order to capture a potential dynamic feature in the dependent variable, we allow to include the lagged dependent variable y_{t-1} in either x_t or z_t. When $x_t = y_{t-1}$, eq.(1) becomes a self-exciting continuous threshold autoregressive model, as in Chan and Tsay (1998).[6] In model (1), we have $d = 3 + \ell$ parameters to be estimated, including an unknown threshold value $\gamma_0 \in \Gamma$, where Γ is a compact set.

In eq.(1), we allow either an endogenous threshold variable x_t or endogenous regressors $z_{1,t}$, or both of them. [7] Note $z_{1,t} = [z_{1,t}^1, \cdots, z_{1,t}^{d_1}]'$ is a subset of $z_t = [z_{1,t}', z_{2,t}']'$. In order not to lose generality, our theory is derived for a general case that both x_t and z_t are endogenous. For $k_1 = 1, \cdots, d_1$, denote p_{xt} and $p_{zt}^{k_1}$, as the vector of instrument variables for x_t and $z_{1t}^{k_1}$, respectively, where p_{xt} and $p_{zt}^{k_1}$, may include the lagged terms of (x_t, z_t), and are allowed to have duplicate variables. To have enough instrument variables, we require the dimension of p_{xt}, $d_{px} \geq 1$, and the dimension of $p_{zt}^{k_1}$, $d_{pz}^{k_1} \geq 1$ for each k_1. To simplify notation, we collapse all instrumental variables as a vector p_t, with elements including all non-overlapping terms in p_{xt} and $p_{zt}^{k_1}$, for $k_1 = 1, \cdots, d_1$.

To avoid the possibility of model misspecification, we assume a nonparametric structure of the reduced-form equations.[8] Specifically, the reduced-form equations of x_t and $z_{1,t}$ are

$$x_t = g_{x0}(p_{xt}) + v_{xt}, \tag{2}$$

$$z_{1,t}^{k_1} = g_{z0}^{k_1}(p_{zt}^{k_1}) + v_{zt}^{k_1}, \quad \text{for} \quad k_1 = 1, \cdots, d_1, \tag{3}$$

where $g_{x0}(\cdot)$ and $g_{z0}^{k_1}(\cdot)$ are an unknown function of p_{xt} and $p_{zt}^{k_1}$, respectively.

The endogeneity of the threshold variable x_t and regressors $z_{1,t}$ comes from the contem-

[6]Note that Chan and Tsay (1998) consider a more general setup by considering y_{t-d} as the threshold variable for some positive integer d. Here we keep $d = 1$.

[7]Note when $x_t = y_{t-1}$, y_{t-1} is sequentially exogenous under Assumption T1.2.

[8]Note that both Kourtellos et al. (2022) and Yu et al. (2023) assume linear reduced forms.

poraneous correlation between u_t and v_t, where $v_t = [v_{xt}, v_{zt}^1, \cdots, v_{zt}^{d_1}]'$. Here we assume each element of v_t is independent of each other to simplify our analysis. Using the control function approach, we assume $E(u_t | \mathcal{F}_{t-1}, x_t, z_{1,t}) = E(u_t|v_t) = h_0(v_t)$, almost surely, where $\mathcal{F}_t$ is the smallest sigma-field generated from $\{x_s, z_{1,s}, z_{2,s+1}, u_s, p_{s+1}) : 1 \leq s \leq t \leq n\}$, and $h_0(\cdot)$ is an unknown function of v_t. Therefore, we have

$$E(y_t|\mathcal{F}_{t-1}, x_t, z_{1,t}) = \beta_{10}(x_t - \gamma_0)I(x_t < \gamma_0) + \beta_{20}(x_t - \gamma_0)I(x_t \geq \gamma_0) + z_t'\beta_{30} + h_0(v_t). \tag{4}$$

Denoting $\delta_0 = \beta_{20} - \beta_{10}$, we rewrite model (1) as

$$y_t = \beta_{10}(x_t - \gamma_0) + \delta_0(x_t - \gamma_0)I(x_t \geq \gamma_0) + z_t'\beta_{30} + h_0(v_t) + \varepsilon_t, \tag{5}$$

where $\varepsilon_t = u_t - h_0(v_t)$. Note that, since $E(\varepsilon_t|\mathcal{F}_{t-1}, x_t, z_{1,t}) = 0$ almost surely, model (5) is free of the endogeneity problem.

Below, we outline the steps taken to estimate model (5). Before doing so, it is important to note that in our reduced-form equations, $g_{x0}(p_{xt})$, $g_{z0}^{k_1}(p_{zt}^{k_1})$, for $k_1 = 1, \cdots, d_1$, and $h_0(v_t)$ are all multi-factor nonparametric functions. To address the curse of dimensionality problem inherent in nonparametric estimation, we assume a nonparametric additive structure among different factors in our reduced-form equations, as demonstrated in Ozabaci et al. (2014).

To do that, we first define $\Psi(v) = \{\psi_1(v), \psi_2(v), \cdots\}$ as a typical sequence of orthonormal basis functions in L_2 space.[9] For a vector of variables $A_t = [A_{1,t}, \cdots, A_{d_2,t}]'$, let

$$\Psi_{\vartheta_n}(A_t) = [\psi_1(A_{1,t}), \cdots, \psi_{\vartheta_n}(A_{1,t}), \cdots, \psi_1(A_{d_2,t}), \cdots, \psi_{\vartheta_n}(A_{d_2,t})]', \tag{6}$$

where $\Psi_{\vartheta_n}(A_t)$ is a $(\vartheta_n d_2) \times 1$ vector. Then, we can approximate $g_{x0}(p_{xt})$, $g_{z0}(p_{zt})$, and

[9] In this paper, we use the normalized Hermite orthonormal basis functions to approximate the nonparametric functions, which theoretically allows unbounded support for p_t and v_t.

$h_0(v_t)$ by

$$g_{x0}^*(p_{xt}) = \Psi_{\vartheta_{1n}}(p_{xt})'\beta_{x0}, \quad (7)$$

$$g_{z0}^{k_1*}(p_{zt}) = \Psi_{\vartheta_{1n}}(p_{zt}^{k_1})'\beta_{z0}^{k_1}, \quad \text{for} \quad k_1 = 1, \cdots, d_1, \quad (8)$$

$$h_0^*(v_t) = \Psi_{\vartheta_{2n}}(v_t)'\beta_{h0}, \quad (9)$$

where β_{x0}, $\beta_{z0}^{k_1}, k_1 = 1, \cdots, d_1$, and β_{h0} are vectors of coefficients, with dimension $(\vartheta_{1n}d_{px}) \times 1$, $(\vartheta_{1n}d_{pz}^{k_1}) \times 1$, for $k_1 = 1, \cdots, d_1$, and $[\vartheta_{2n}(d_1+1)] \times 1$, respectively. Note that ϑ_{1n} and ϑ_{2n} control the complexity of sieve space to approximate the unknown functions $g_{x0}(p_{xt})$ and $g_{z0}^{k_2}(p_{zt})$ in our first-step regressions (i.e. eqs.(2)-(3)), and $h_0(v_t)$ in our augmented regression (eq.(5)). In sieve estimation, both ϑ_{1n} and ϑ_{2n} increase slowly with n. We assume that ϑ_{1n} and ϑ_{2n} grow at different rates to better observe the effect of our first-step estimates on our second-step estimates.

Now, we start to estimate the model. Among the commonly used semi-/nonparametric kernel and sieve estimation methods, we specifically focus on the method of sieves for both steps, as this method is particularly convenient for estimating the additive structure.

First step: By applying the OLS estimation to models (2) and (3), with more specific expressions provided in eq.(7) and eq.(8), we obtain the linear sieve least squares estimator

$$\hat{g}_x(p_{xt}) = \Psi_{\vartheta_{1n}}(p_{xt})'[\sum_{s=1}^{n} \Psi_{\vartheta_{1n}}(p_{xs})\Psi_{\vartheta_{1n}}(p_{xs})']^{-1} \sum_{s=1}^{n} \Psi_{\vartheta_{1n}}(p_{xs})x_s,$$

$$\hat{g}_z^{k_1}(p_{zt}^{k_1}) = \Psi_{\vartheta_{1n}}(p_{zt}^{k_1})'[\sum_{s=1}^{n} \Psi_{\vartheta_{1n}}(p_{zs}^{k_1})\Psi_{\vartheta_{1n}}(p_{zs}^{k_1})']^{-1} \sum_{s=1}^{n} \Psi_{\vartheta_{1n}}(p_{zs}^{k_1})z_{1s}^{k_1}, \quad \text{for} \quad k_1 = 1, \cdots, d_1.$$

Then, we collect the residuals $\hat{v}_{xt} = x_t - \hat{g}_x(p_{xt})$ and $\hat{v}_{zt}^{k_1} = z_{1,t}^{k_1} - \hat{g}_z^{k_1}(p_{zt}^{k_1})$, for each $k_1 = 1, \cdots, d_1$. We denote $\hat{v}_t = [\hat{v}_{xt}, \hat{v}_{zt}^1, \cdots, \hat{v}_{zt}^{d_1}]'$.

Second step: Let $\beta = [\beta_1, \delta, \beta_3']'$, which is a $(d-1) \times 1$ vector. Then, by replacing $h_0(\cdot)$ with $h^*(\cdot)$, where $h^*(\cdot) = \Psi_{\vartheta_{2n}}(\cdot)'\beta_h$, and v_t with $\hat{v}_t$, we construct the least squares objective function for model (5) as follows:

$$\hat{S}_n(\beta, \gamma, \beta_h) = \frac{1}{n} \sum_{t=1}^{n} [y_t - \beta_1(x_t - \gamma) - \delta(x_t - \gamma)I(x_t \geq \gamma) - z_t'\beta_3 - \Psi_{\vartheta_{2n}}(\hat{v}_t)'\beta_h]^2, \quad (10)$$

and the least squares estimator of model (5) solves the following optimization problem:

$$(\hat{\beta}, \hat{\gamma}, \hat{\beta}_h) = \underset{(\beta,\gamma,\beta_h)\in B\times\Gamma\times B_h}{\operatorname{argmin}} \hat{S}_n(\beta, \gamma, \beta_h). \tag{11}$$

Note that $\hat{S}_n(\beta, \gamma, \beta_h)$ is non-smooth in γ. Therefore, we use a grid search method in practice. For a given $\gamma \in \Gamma$, we obtain the least squares estimator of (β_0, β_{h0}) as follows:

$$\left[\hat{\beta}(\gamma)', \hat{\beta}_h(\gamma)'\right]' = \left[\hat{X}(\gamma)'\hat{X}(\gamma)\right]^{-1} \hat{X}(\gamma)'y, \tag{12}$$

where $y = [y_1, y_2, \ldots, y_n]'$, $\hat{X}(\gamma) = [\hat{x}_1(\gamma), \hat{x}_2(\gamma), \ldots, \hat{x}_n(\gamma)]'$, and $\hat{x}_t(\gamma) = [x_t - \gamma, (x_t - \gamma)I(x_t \geq \gamma), z_t', \Psi_{\vartheta_{2n}}(\hat{v}_t)']'$ for $t = 1, \cdots, n$.

Next, we substitute (β, β_h) by $\left(\hat{\beta}(\gamma), \hat{\beta}_h(\gamma)\right)$ into $\hat{S}_n(\beta, \gamma, \beta_h)$ and obtain the estimator of γ_0 as follows:

$$\hat{\gamma} = \underset{\gamma\in\Gamma}{\operatorname{argmin}} \hat{S}_n\left(\hat{\beta}(\gamma), \gamma, \hat{\beta}_h(\gamma)\right). \tag{13}$$

Then, the profiled estimator for (β_0, β_{h0}) is given by $(\hat{\beta}, \hat{\beta}_h) = \left(\hat{\beta}(\hat{\gamma}), \hat{\beta}_h(\hat{\gamma})\right)$.

2.2 Assumptions and limiting results

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

Assumptions-time series.

Assumption T1:

T1.1. For some $r > 1$, (a) $\{(y_t, x_t, z_t, p_t)\}$ is a strictly stationary, β-mixing sequence with mixing coefficients $\alpha(m) = O(m^{-A})$ for some $A > r/(r-1)$; (b) $E|y_t|^{4r} < \infty$, $E|x_t|^{4r} < \infty$, $E\left\|z_t\right\|^{4r} < \infty$.

T1.2. (a) $E(u_t|\mathcal{F}_{t-1}, x_t, z_{1,t}) = E(u_t|v_t) = h_0(v_t)$ almost surely for all t, where $\mathcal{F}_t$ is the smallest sigma-field generated from $\{(x_s, z_{1,s}, z_{2,s+1}, u_s, p_{s+1}) : 1 \leq s \leq t \leq n\}$; (b) $\{(v_t, \mathcal{F}_{t-1})\}$ is a martingale difference sequence with $E(v_t|\mathcal{F}_{t-1}) = 0$ almost surely; (c) $E[u_t^2|\mathcal{F}_{t-1}, x_t, z_{1,t}] < \infty$.

Remark: Assumption T1.1(a) assumes a β-mixing sequence with a sufficiently fast decaying dependence over t, where r involves a trade-off between series correlation and the number of finite order of moments, see e.g., Hansen (2017). Assumption T1.1(b) provides regular moment conditions. Assumptions T1.2(a)(b) define the endogeneity and ensure the specification of eq.(5). Assumption T1.2(c) is a bounded conditional variance assumption used to derive the convergence rate, see Newey (1997).

Assumption T2:

T2.1. $g_{x0}(\cdot)$, $g_{z0}(\cdot)$, and $h_0(\cdot)$ all belong to $\mathcal{H}$, a subset of Hölder functional space, $\Lambda^{\eta}(\cdot)$, with $\eta > \max\{(1+d_1)/2, 2\}$. [10] All these unknown functions and their first-order derivatives are uniformly bounded over $\mathcal{R}$.

T2.2. $\Psi(\cdot) = \{\psi_1(\cdot), \psi_2(\cdot), \cdots\}$ are uniformly bounded sequences of orthonormal basis functions in $\mathcal{H}_n$, a subset of $\Lambda^{\eta}(\cdot)$.

T2.3. (a) $g_{x0}(\cdot)$ and $g_{z0}(\cdot)$ are squared integrable, and there exist β_{x0}, $\beta_{z0}^{k_2}$, and a finite constant C_g satisfying

$\sup_{p_x \in \mathcal{R}^{d_{px}}} |g_{x0}(p_x) - \Psi_{\vartheta_{1n}}(p_x)'\beta_{x0}| \leq C_g \vartheta_{1n}^{-\eta}$,

$\sup_{p_z^{k_1} \in \mathcal{R}^{d_{pz}^{k_1}}} |g_{z0}^{k_1}(p_z^{k_1}) - \Psi_{\vartheta_{1n}}(p_z^{k_1})'\beta_{z0}^{k_1}| \leq C_g \vartheta_{1n}^{-\eta}$, for $k_1 = 1, .., d_1$;

(b) $h_0(\cdot)$ is squared integrable, and there exist β_{h0} and a finite constant C_h satisfying $\sup_{v \in \mathcal{R}^{1+d_1}} |h_0(v) - \Psi_{\vartheta_{2n}}(v)'\beta_{h0}| \leq C_h \vartheta_{2n}^{-\eta}$.

T2.4. For a sufficiently large ϑ_{1n}, there exist a set of constants $(\underline{c}, \overline{c})$, such that:

(a) $-\infty < \underline{c} \leq \lambda_{min}\{E[\Psi_{\vartheta_{1n}}(p_{xt})\Psi_{\vartheta_{1n}}(p_{xt})']\} \leq \lambda_{max}\{E[\Psi_{\vartheta_{1n}}(p_{xt})\Psi_{\vartheta_{1n}}(p_{xt})']\} \leq \overline{c} < \infty$,

$-\infty < \underline{c} \leq \lambda_{min}\{E[v_{xt}^2\Psi_{\vartheta_{1n}}(p_{xt})\Psi_{\vartheta_{1n}}(p_{xt})']\} \leq \lambda_{max}\{E[v_{xt}^2\Psi_{\vartheta_{1n}}(p_{xt})\Psi_{\vartheta_{1n}}(p_{xt})']\} \leq \overline{c} < \infty$;

(b) $-\infty < \underline{c} \leq \lambda_{min}\left\{E[\Psi_{\vartheta_{1n}}(p_{zt}^{k_1})\Psi_{\vartheta_{1n}}(p_{zt}^{k_1})']\right\} \leq \lambda_{max}\left\{E[\Psi_{\vartheta_{1n}}(p_{zt}^{k_1})\Psi_{\vartheta_{1n}}(p_{zt}^{k_1})']\right\} \leq \overline{c} < \infty$,

$-\infty < \underline{c} \leq \lambda_{min}\left\{E[(v_{zt}^{k_1})^2\Psi_{\vartheta_{1n}}(p_{zt}^{k_1})\Psi_{\vartheta_{1n}}(p_{zt}^{k_1})']\right\} \leq \lambda_{max}\left\{E[(v_{zt}^{k_1})^2\Psi_{\vartheta_{1n}}(p_{zt}^{k_1})\Psi_{\vartheta_{1n}}(p_{zt}^{k_1})']\right\} \leq \overline{c} < \infty$, for $k_1 = 1, \cdots, d_1$;

(c) L is full rank in column, where L is defined beneath eq.(15).

[10] The Hölder functional space is widely used in semiparametric estimation. Any function belonging to this space can be well approximated by the sieve method. For details, see, e.g., Section 2.3.1 in Chen (2007).

Remark: Assumption T2 provides the necessary conditions for our nonparametric functions, $g_{x0}(\cdot)$, $g_{z0}(\cdot)$, $h_0(\cdot)$ and the set of basis functions, $\Psi(\cdot)$. Specifically, Assumption T2.1 requires that all our nonparametric functions belong to the Hölder functional space $\Lambda^{\eta}(\cdot)$, a standard requirement in sieve estimation. Assumption T2.2 imposes properties of the basis functions. Assumptions T2.3 (a)-(b) control the sieve approximation bias. To allow the infinite support of variables in our unknown functions, we restrict the space of unknown functions for our analysis.[11] For a nonparametric function with bounded support, we can directly apply Theorem 1.1 of Dzyadyk and Shevchuk (2008), which indicates Assumptions T2.3 (a)-(b) hold given $g_{x0}(\cdot)$, $g_{z0}(\cdot)$ and $h_0(\cdot)$ all η-smooth. In cases where variables have unbounded support, as in our use of the normalized Hermite orthonormal basis functions, we can apply the result of Xiang (2012). In that, the hold of Assumptions T2.3(a)-(b) requires $g_{x0}(\cdot)$, $g_{z0}(\cdot)$ and $h_0(\cdot)$ have the lowest level of smoothness ϱ, with $\varrho > 2(\eta + 1)$. Assumption T2.4 is a full rank condition.

Assumption T3:

T3.1. $\delta_0 \neq 0$ and $h_0(\cdot) \neq 0$ holds over at least one non-empty interval of its domain.

T3.2. (a) $\phi = (\beta, \gamma, h) \in (B, \Gamma, \mathcal{H}) = \Phi$, $\phi_0 = (\beta_0, \gamma_0, h_0) \in (B, \Gamma, \mathcal{H}) = \Phi$,

$\phi^* = (\beta, \gamma, h^*) \in (B, \Gamma, \mathcal{H}_n) = \Phi_n$, $\phi_0^* = (\beta_0, \gamma_0, h_0^*) \in (B, \Gamma, \mathcal{H}_n) = \Phi_n$,

and $\beta_{h0} \in B_h$, where B, Γ and B_h are all compact sets;

(b) ϕ_0 is the unique minimizer of $E[\varepsilon_t(\phi)^2]$ over the space Φ, where

$\varepsilon_t(\phi) = y_t - \beta_1(x_t - \gamma) - \delta(x_t - \gamma)I(x_t \geq \gamma) - z_t'\beta_3 - h(v_t)$.

T3.3. For any ϑ_{2n}, there exist constants $\underline{c_2}$ and $\overline{c_2}$ such that $-\infty < \underline{c_2} \leq \lambda_{min}\{E[x_t(\gamma)x_t'(\gamma)]\} \leq \lambda_{max}\{E[x_t(\gamma)x_t'(\gamma)]\} \leq \bar{c}_2 < \infty$, and $-\infty < \underline{c}_2 \leq \lambda_{min}\{E[\varepsilon_t^2 x_t(\gamma)x_t'(\gamma)]\} \leq \lambda_{max}\{E[\varepsilon_t^2 x_t(\gamma)x_t'(\gamma)]\} \leq$

[11] For sieve approximation of a nonparametric function with unbounded support variables, another possible solution is to apply the results of Chen et al. (2005). They introduce a weighted sup-norm metric distance between the nonparametric function and its sieve approximation, similar to the trim method used in the kernel approximation.

$\bar{c}_2 < \infty$ hold uniformly over $\gamma \in \Gamma$, where $x_t(\gamma)$ equals $\hat{x}_t(\gamma)$ with $\hat{v}_t$ being replaced with v_t.

T3.4. x_t has a density function $f(x)$ and $f(x) \le \bar{f} < \infty$ over its domain for some finite constant $\bar{f}$.

Remark: Assumption T3 is a prerequisite for establishing the asymptotic properties of the estimator for the parameters (β, γ). Assumption T3.1 ensures the existence of the threshold effect. Assumption T3.2 (a) assumes the compactness of the parameter space, while Assumption T3.2 (b) provides an identification assumption similar to Assumption 2.1 in Hansen (2017). Assumption T3.3 ensures the existence of $(\hat{\beta}(\gamma), \hat{\beta}_h(\gamma))$ for any $\gamma \in \Gamma$.

Denote $\|\Psi_{\vartheta_n}\|_{\mathcal{P}}^2 = \max_{s \le \mathcal{P}} \sup_{v \in \mathcal{R}} \|\nabla^s \Psi_{\vartheta_n}(v)\|^2$, where $\nabla^s \Psi_{\vartheta_n}(\cdot)$ is the sth derivative of $\Psi_{\vartheta_n}(\cdot)$. We then have $\|\nabla \Psi_{\vartheta_n}\|_{\mathcal{P}} = O(\vartheta_n^{\mathcal{P}+1/2})$ (see, e.g., the normalized Hermite functions and wavelet functions defined in Blundell et al. (2007)).

Assumption T4: As $n \to \infty$, $\vartheta_{1n} \to \infty$, $\vartheta_{2n} \to \infty$, and $\|\Psi_{\vartheta_{2n}}\|_1 \left(\vartheta_{1n}^{-\eta} + \sqrt{\vartheta_{1n}/n}\right) \sqrt{\vartheta_{2n}} = o(1)$.

Remark: Assumption T4 imposes restrictions on the smoothing parameters ϑ_{1n} and ϑ_{2n} that are needed to derive the consistency in Theorem 1-time series. Note that the convergence rate of our first-step sieve estimator is $O_p(\vartheta_{1n}^{-\eta} + \sqrt{\vartheta_{1n}/n})$, which is standard in the literature (see, e.g., Newey (1997)). Thus, $\|\Psi_{\vartheta_{2n}}\|_1 \left(\vartheta_{1n}^{-\eta} + \sqrt{\vartheta_{1n}/n}\right) \sqrt{\vartheta_{2n}} = o(1)$ is needed to derive the consistency results of $\hat{\theta}$. Here, we assume $\vartheta_{1n} \ne \vartheta_{2n}$ to demonstrate the effect of the first-step estimator. However, in practice, it is convenient to set $\vartheta_{1n} = \vartheta_{2n}$, simplifying Assumption T4 to $\vartheta_{2n}^4/n = o(1)$, as seen in, for example, Assumption $A.5^*$ in Ozabaci et al. (2014).

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

Theorem 1-time series. Denote $\theta_0 = (\beta_0', \gamma_0)'$, $\hat{\theta} = \left(\hat{\beta}', \hat{\gamma}\right)'$, $\hat{h}(\cdot) = \Psi_{\vartheta_{2n}}(\cdot)'\hat{\beta}_h$, and $\hat{\phi}_n = (\hat{\theta}, \hat{h})$. Under Assumptions T1-T4, as $n \to \infty$, we have

$$d\left(\hat{\phi}_n, \phi_0\right) = O_p\left(\vartheta_{2n}^{-\eta} + \sqrt{\frac{\vartheta_{2n}}{n}}\right), \tag{14}$$

where $d\left(\hat{\phi}_n, \phi_0\right) = \left\|\hat{\theta} - \theta_0\right\| + \left\|\hat{h} - h_0\right\|_\infty$.

Theorem 1-time series establishes the consistency and the convergence rate of our proposed estimator, where the proof follows Theorems 3.1 and 3.2 of Chen (2007). For any $\phi \in \Phi$, denoting

$$H_t(\theta) = -\frac{\partial}{\partial \phi}\varepsilon_t(\phi) = \begin{pmatrix} (x_t - \gamma) \\ (x_t - \gamma)I(x_t \geq \gamma) \\ z_t \\ \beta_1 + \delta I(x_t \geq \gamma) \\ 1 \end{pmatrix}, \tag{15}$$

$H_t = H_t(\theta_0)$, and $m_t = H_t\varepsilon_t$, we obtain the limiting distribution of our proposed estimator blow.

Theorem 2-time Series. Under Assumptions T1-T4, as $n \to \infty$, we have

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left[0, (L'L)^{-1}L'VL(L'L)^{-1}\right], \tag{16}$$

where $V = \lim_{n\to\infty} Var(m_t)$ and $L = E[\partial H_t\varepsilon_t/\partial\theta']$.

Remark: The proof of Theorem 2-time series is given in the supplementary appendix, in which the proof follows Theorem 2 of Chen et al. (2003). The slope and threshold value estimators converge at the root-n rate and are jointly normally distributed with a non-zero asymptotic covariance. To make an inference, given the sieve estimates $\hat{\phi}_n$, the asymptotic variance-covariance matrix can be consistently estimated by using $\hat{V}_n = n^{-1}\sum_{t=1}^{n} m_t(\hat{\phi}_n)m_t(\hat{\phi}_n)'$ and $\hat{L}_n = n^{-1}\sum_{t=1}^{n}\partial[H_t(\hat{\theta})\hat{\varepsilon}_t(\hat{\phi}_n)]/\partial\theta'$, where $m_t(\hat{\phi}) = H_t(\hat{\theta})\hat{\varepsilon}_t(\hat{\phi}_n)$ and $\hat{\varepsilon}_t(\hat{\phi}_n) = y_t - \hat{\beta}_1(x_t - \hat{\gamma}) - \hat{\delta}(x_t - \hat{\gamma})I(x_t \geq \hat{\gamma}) - z_t'\hat{\beta}_3 - \Psi_{\vartheta_{2n}}(\hat{v}_t)'\hat{\beta}_h$. The full expression of $\hat{L}_n$ is presented in the supplementary appendix.

3 Panel data 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 which we will discuss more in section 5. Therefore, we extend our baseline time series model to a panel data endogenous kink threshold panel model with unknown fixed effects. Below, we present our model, the estimation strategy, and the asymptotic results.

3.1 Model and estimation

Following the general setup in the panel data literature, we consider the panel data with sufficiently large numbers of cross-sectional units N and a small time period T. Our panel kink threshold regression model is defined as follows:

$$y_{i,t} = \beta_{10}(x_{i,t} - \gamma_0)I(x_{i,t} < \gamma_0) + \beta_{20}(x_{i,t} - \gamma_0)I(x_{i,t} \geq \gamma_0) + z'_{i,t}\beta_{30} + b_i + u_{i,t}, \quad (17)$$

for $i = 1, \cdots, n,$ and $t = t_0, \cdots, T$,[12] where $y_{i,t}$ is the dependent variable, $x_{i,t}$ is a scalar threshold variable, $z_{i,t}$ is an $\ell \times 1$ vector of time-varying regressors,[13] b_i is the i^{th} unobserved individual fixed effect, and $u_{i,t}$ is the error term.

Denote the vector of coefficients $\beta_0 = (\beta_{10}, \beta_{20}, \beta'_{30})' \in R^{k-1}$, with $k = 3 + \ell$. The unknown

[12] Here we set t starts at t_0 to avoid the missing value problem caused by taking the first differences and possible lagged variables in our regression.

[13] Here, we focus on the static panel data KTR model. For a dynamic panel data KTR model, one could apply the GMM method introduced by Seo and Shin (2016), which is something we did not consider in the present paper. The reason for not including the dynamic version of the panel model is that, within a control function approach to solve endogeneity, we would need to derive an expression for the first-step regression. In a dynamic panel data KTR model, the lagged dependent variables cannot be used as instrumental variables, as we would encounter a recursive endogeneity problem. Our method can be extended to accommodate the dynamic model by constructing the first-step regression using exogenous variables other than its own lagged term. It's important to note that these exogenous variables must have a strong explanatory or predictive power for the dependent variable. However, in practice, finding such exogenous variables can be a challenging task, and we do not see any advantages compared to the existing GMM method introduced by Seo and Shin (2016).

threshold value γ_0 is an interior point of a compact set, Γ. Again, we allow the endogenous threshold variable, $x_{i,t}$ and endogenous regressors $z_{1,i,t}$, where $z_{1,i,t}$ is a $d_1 \times 1$ vector, which is a subset of $z_{i,t} = [z'_{1,i,t}, z'_{2,i,t}]'$. Naturally, we also allow endogeneity arising from $\text{Cov}(x_{i,t}, b_i) \neq 0$ and $\text{Cov}(z_{i,t}, b_i) \neq 0$. To remove the time-invariant fixed effects, we apply the first-differencing method to model (17), and by denoting $\delta_0 = \beta_{20} - \beta_{10}$, we obtain

$$\Delta y_{i,t} = \beta_{10}\Delta x_{i,t} + \delta_0(X_{i,t} - \gamma_0\tau_2)'\mathbf{I}_{it}(\gamma_0) + \Delta z'_{i,t}\beta_{30} + \Delta u_{i,t}, \tag{18}$$

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

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

Endogeneity in model (18) arises from the contemporaneous correlation between $(x_{i,t}, z_{1,i,t})$ and $u_{i,t}$. The reduced form equations for $x_{i,t}$ and $z_{1,i,t}$ are given by

$$x_{i,t} = g_{x0}(p_{x,i,t}) + v_{x,i,t}, \tag{19}$$

$$z_{1,i,t} = g_{z0}^{k_1}(p_{z,i,t}^{k_1}) + v_{z,i,t}^{k_1}, \quad \text{for} \quad k_1 = 1, \cdots, d_1, \tag{20}$$

where $p_{x,i,t}$ and $p_{z,i,t}^{k_1}$, for $k_1 = 1, \cdots, d_1$ are allowed to share common variables. To simplify notation, we denote all instrumental variables, including $p_{x,i,t}$ and $p_{z,i,t}^{k_1}$, for $k_1 = 1, .., d_1$, as $p_{i,t}$. Additionally, we define $v_{i,t} = \left[v_{x,i,t}, v_{z,i,t}^1, ..., v_{z,i,t}^{d_1}\right]'$ as a $(1 + d_1) \times 1$ vector. Note that, since we assume endogeneity arises from the correlation between $v_{x,i,t}$ and $v_{z,i,t}$ with the error term $u_{i,t}$, this implies $\text{Cov}(x_{i,t}, u_{i,t}) \neq 0$ and $\text{Cov}(z_{1,i,t}, u_{i,t}) \neq 0$. To simplify our analysis, we assume $\text{Cov}(v_{x,i,t}, v_{z,i,t}) = 0$.

Using the control function approach and denoting $g_{v0}(\cdot)$ as an unknown function of $v_{i,t}$, for each i, we assume $E\left(u_{i,t}|\mathcal{F}_{i,t-1}, x_{i,t}, z_{1,i,t}\right) = E\left(u_{i,t}|v_{i,t}\right) = g_{v0}(v_{i,t})$ almost surely, where $\mathcal{F}_{i,t}$ is the smallest sigma-field generated from $\{(x_{i,s}, z_{1,i,s}, z_{2,i,s+1}, u_{i,s}, p_{i,s+1}) : 1 \leq s \leq t \leq T\}$.

Under the law of iterative expectation, we have

$$
\begin{aligned}
& E(u_{i,t}|\mathcal{F}_{i,t-2}, x_{i,t}, x_{i,t-1}, z_{i,t}, z_{1,i,t-1}, p_{i,t}) \\
& = E[E(u_{i,t}|\mathcal{F}_{i,t-1}, x_{i,t}, z_{1,i,t})|\mathcal{F}_{i,t-2}, x_{i,t}, x_{i,t-1}, z_{i,t}, z_{1,i,t-1}, p_{i,t}] \\
& = E(g_{v0}(v_{i,t})|\mathcal{F}_{i,t-2}, x_{i,t}, x_{i,t-1}, z_{i,t}, z_{1,i,t-1}, p_{i,t}) = g_{v0}(v_{i,t})
\end{aligned}
$$

and $E(u_{i,t-1}|\mathcal{F}_{i,t-2}, x_{i,t}, x_{i,t-1}, z_{i,t}, z_{1,i,t-1}, p_{i,t}) = E(u_{i,t-1}|\mathcal{F}_{i,t-2}, x_{i,t-1}, z_{1,i,t-1}) = g_{v0}(v_{i,t-1})$ since future information does not affect past information. Therefore, $E(\Delta u_{i,t}|\mathcal{F}_{i,t-2}, x_{i,t}, x_{i,t-1}, z_{i,t}, z_{1,i,t-1}, p_{i,t}) = h_0(v_{i,t}, v_{i,t-1})$, where $h_0(v_{i,t}, v_{i,t-1}) = g_{v0}(v_{i,t}) - g_{v0}(v_{i,t-1})$. It then follows that

$$
\begin{aligned}
& E(\Delta y_{i,t}|\mathcal{F}_{i,t-2}, x_{i,t}, x_{i,t-1}, z_{i,t}, z_{1,i,t-1}, p_{i,t}) \\
& = \beta_{10}\Delta x_{i,t} + \delta_0(X_{i,t} - \gamma_0\tau_2)'\mathbf{I}_{it}(\gamma_0) + \Delta z_{i,t}'\beta_{30} + h_0(v_{i,t}, v_{i,t-1}).
\end{aligned} \tag{21}
$$

Given above equation, we can rewrite model (18) as

$$
\Delta y_{i,t} = \beta_{10}\Delta x_{i,t} + \delta_0(X_{i,t} - \gamma_0\tau_2)'\mathbf{I}_{it}(\gamma_0) + \Delta z_{i,t}'\beta_{30} + h_0(v_{i,t}, v_{i,t-1}) + \Delta\varepsilon_{i,t}, \tag{22}
$$

where $\Delta\varepsilon_{i,t} = \Delta u_{i,t} - h_0(v_{i,t}, v_{i,t-1})$. Note that, since $E(\Delta\varepsilon_{i,t}|\mathcal{F}_{i,t-2}, x_{i,t}, x_{i,t-1}, z_{i,t}, z_{1,i,t-1}, p_{i,t}) = 0$ almost surely, model (22) is free of the endogeneity problem.

Similar to the time-series model, for a vector of variables $A_{i,t} = [A_{1,i,t}, \cdots, A_{d_2,i,t}]'$, let

$$
\Psi_{\vartheta_N}(p_{i,t}) = [\psi_1(A_{1,i,t}), \cdots, \psi_{\vartheta_n}(A_{1,i,t}), \cdots, \psi_1(A_{d_2,i,t}), \cdots, \psi_{\vartheta_N}(A_{d_2,i,t})]',
$$

which is a $(\vartheta_N d_p) \times 1$ vector of orthonormal basis functions. Then the series approximations of $g_{x0}(p_{x,i,t})$, $g_{z0}^{k_1}(p_{z,i,t}^{k_1})$, and $g_{v0}(v_{i,t})$ are given as folows:

$$
\begin{aligned}
g_{x0}^*(p_{x,i,t}) &= \Psi_{\vartheta_{1N}}(p_{x,i,t})'\beta_{x0}, && (23) \\
g_{z0}^{k_1 *}(p_{z,i,t}) &= \Psi_{\vartheta_{1N}}(p_{z,i,t})'\beta_{z0}^{k_1}, \quad \text{for} \quad k_1 = 1, \cdots, d_1, && (24) \\
g_{v0}^*(v_{i,t}) &= \Psi_{\vartheta_{2N}}(v_{i,t})'\beta_{h0},
\end{aligned}
$$

where β_{x0}, $\beta_{z0}^{k_1}, k_1 = 1, \cdots, d_1$, and β_{h0} are vectors of coefficients, with dimension $(\vartheta_{1N} d_{px}) \times 1$, $(\vartheta_{1N} d_{pz}^{k_1}) \times 1$, for $k_1 = 1, \cdots, d_1$, and $[\vartheta_{2N}(d_1 + 1)] \times 1$, respectively. Thus, we can express the series approximation for $h_0(v_{i,t}, v_{i,t-1})$ as

$$h_0^*(v_{i,t}, v_{i,t-1}) = g_{v0}^*(v_{i,t}) - g_{v0}^*(v_{i,t-1}) = \Delta\Psi_{\vartheta_{2N}}(v_{i,t})'\beta_{h0}, \tag{25}$$

where $\Delta\Psi_{\vartheta_{2N}}(v_{i,t}) = \Psi_{\vartheta_{2N}}(v_{i,t}) - \Psi_{\vartheta_{2N}}(v_{i,t-1})$.

Next, we proceed to show the estimation strategy for our LS sieve estimator.

First step: By applying the OLS estimation to models (19) and (20) with the more specific expressions in eq.(23) and eq.(24), we obtain

$$\begin{aligned}
\hat{g}_x^*(p_{x,i,t}) &= \Psi_{\vartheta_{1N}}(p_{x,i,t})'[\sum_{i=1}^{N}\sum_{s=t_0}^{T}\Psi_{\vartheta_{1N}}(p_{x,i,s})\Psi_{\vartheta_{1N}}(p_{x,i,s})']^{-1}\sum_{i=1}^{N}\sum_{s=t_0}^{T}\Psi_{\vartheta_{1N}}(p_{x,i,s})x_{i,s}, \\
\hat{g}_z^{k_1*}(p_{z,i,t}^{k_1}) &= \Psi_{\vartheta_{1N}}(p_{z,i,t}^{k_1})'[\sum_{i=1}^{N}\sum_{s=t_0}^{T}\Psi_{\vartheta_{1N}}(p_{z,i,s}^{k_1})\Psi_{\vartheta_{1N}}(p_{z,i,s}^{k_1})']^{-1}\sum_{i=1}^{N}\sum_{s=t_0}^{T}\Psi_{\vartheta_{1N}}(p_{z,i,s}^{k_1})z_{1,i,s}^{k_1},
\end{aligned}$$

for $k_1 = 1, \cdots, d_1$. Similarly, we get $\hat{g}_x^*(p_{x,i,t-1})$ and $\hat{g}_z^{k_1*}(p_{z,i,t-1}^{k_1})$, for $k_1 = 1, \cdots, k_1$. Then, we collect the residuals $\hat{v}_{x,i,t} = x_{i,t} - \hat{g}_x^*(p_{x,i,t})$, $\hat{v}_{x,i,t-1} = x_{i,t-1} - \hat{g}_x^*(p_{x,i,t-1})$ and $\hat{v}_{z,i,t}^{k_1} = z_{1,i,t}^{k_1} - \hat{g}_z^{k_1*}(p_{z,i,t}^{k_1})$, $\hat{v}_{z,i,t-1}^{k_1} = z_{1,i,t-1}^{k_1} - \hat{g}_z^{k_1*}(p_{z,i,t-1}^{k_1})$ for each $k_1 = 1, \cdots, d_1$.

Second step: By replacing $(v_{i,t}, v_{i,t-1})$ with $(\hat{v}_{i,t}, \hat{v}_{i,t-1})$ in eq.(22) and eq.(25), we obtain the following least squares criterion function

$$S_N(\beta, \gamma, \beta_h) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=t_0}^{T}\left\{\Delta y_{i,t} - \beta_1\Delta x_{i,t} - \delta(X_{i,t} - \gamma\tau_2)\mathbf{I}_{i,t}(\gamma) - \Delta z_{i,t}'\beta_3 - \Delta\Psi_{\vartheta_{2N}}(\hat{v}_{i,t})'\beta_h\right\}^2, \tag{26}$$

and our least squares estimator is the minimizer of $S_N(\beta, \gamma, \beta_h)$; i.e.,

$$(\hat{\beta}, \hat{\gamma}, \hat{\beta}_h) = \underset{(\beta,\gamma,\beta_h)\in B\times\Gamma\times B_h}{\operatorname{argmin}} S_N(\beta, \gamma, \beta_h). \tag{27}$$

For a given $\gamma \in \Gamma$, we obtain the conditional least squares estimator of (β_0, β_{h0}),

$$[\hat{\beta}(\gamma)', \hat{\beta}_h(\gamma)']' = [\sum_{i=1}^{N}\sum_{t=t_0}^{T}\widehat{\Delta x_{i,t}}(\gamma)\widehat{\Delta x_{i,t}}(\gamma)']^{-1}\sum_{i=1}^{N}\sum_{t=t_0}^{T}\widehat{\Delta x_{i,t}}(\gamma)\Delta y_{i,t}, \tag{28}$$

where $\widehat{\Delta x_{i,t}}(\gamma) = [\Delta x_{i,t}, (X_{i,t} - \gamma\tau_2)'\mathbf{I}_{i,t}(\gamma), \Delta z_{i,t}', \Delta\Psi_{\vartheta_{2N}}(\hat{v}_{i,t})']'$.

Next, by substituting (β, β_h) with $\left(\hat{\beta}(\gamma), \hat{\beta}_h(\gamma)\right)$ into $\hat{S}_N(\beta, \gamma, \beta_h)$, we obtain the estimator of γ_0,

$$\hat{\gamma} = \underset{\gamma\in\Gamma}{\operatorname{argmin}}\, \hat{S}_N\left(\hat{\beta}(\gamma), \gamma, \hat{\beta}_h(\gamma)\right). \tag{29}$$

Then, the least squares estimator for $(\beta_0, \gamma_0, \beta_{h0})$ is given by $(\hat{\beta}, \hat{\gamma}, \hat{\beta}_h) = \left(\hat{\beta}(\hat{\gamma}), \hat{\gamma}, \hat{\beta}_h(\hat{\gamma})\right)$.

3.2 Assumptions and limiting results

In this subsection, we will derive the limiting result of our proposed estimator for the panel data model. Below, we outline the necessary regularity conditions.

Assumptions -panel.

Assumption P1: For some $\xi > 1$,

P1.1. (a) $\{(y_{it}, x_{it}, z_{it}, p_{it}) : t = 1, 2, \cdots\}$ are independently identically distributed (i.i.d.) across index i; (b) $E|\sum_{t=t_0}^{T} \Delta y_{it}|^{4\xi} < \infty$, $E|\sum_{t=t_0}^{T} \Delta x_{it}|^{4\xi} < \infty$, $E\left\|\sum_{t=t_0}^{T} \Delta z_{it}\right\|^{4\xi} < \infty$.

P1.2. For all $1 \leq i \leq N$, (a) $E\left(u_{i,t}|\mathcal{F}_{i,t-1}, x_{i,t}, z_{1,i,t}\right) = E\left(u_{i,t}|v_{i,t}\right) = g_{v0}(v_{i,t})$ almost surely for all $t_0 \leq t \leq T$, where $\mathcal{F}_{i,t}$ is the smallest sigma-field generated from $\{(x_{i,s}, z_{1,i,s}, z_{2,i,s+1}, u_{i,s}, p_{i,s+1}) : 1 \leq s \leq t \leq N\}$; (b) $\{(v_{i,t}, \mathcal{F}_{i,t-1})\}$ is a martingale difference sequence with $E(v_{i,t}|\mathcal{F}_{i,t-1}) = 0$ almost surely; (c) $E[\Delta u_{i,t}^2|\mathcal{F}_{i,t-1}, x_{i,t}, z_{1,i,t}] < \infty$.

Assumption P2:

P2.1. $g_{x0}(\cdot)$, $g_{z0}(\cdot)$, and $h_0(\cdot)$ belong to $\mathcal{H}$, a subset of Hölder functional space, $\Lambda^{\eta}(\cdot)$, with $\eta > \max\{(1 + d_1)/2, 2\}$, all unknown functions and their first-order derivatives are uniformly bounded over $\mathcal{R}$.

P2.2. $\Psi = \{\psi_1, \psi_2, \cdots\}$ are uniformly bounded, sequences of orthonormal basis functions in $\mathcal{H}_N$, a subset of $\Lambda^{\eta}(\cdot)$.

P2.3. $g_{x0}(\cdot)$, $g_{z0}(\cdot)$ and $g_{v0}(\cdot)$ are squared integrable, and there exist β_{x0}, β_{z0}, β_{h0} and finite constant C, that:

$\sup_{p_x \in \mathcal{R}^{d_{px}}} |g_{x0}(p_x) - \Psi_{\vartheta_{1N}}(p_x)'\beta_{x0}| \leq C\vartheta_{1N}^{-\eta}$,

$\sup_{p_z^{k_1} \in \mathcal{R}^{d_{pz}^{k_1}}} |g_{z0}^{k_1}(p_z^{k_1}) - \Psi_{\vartheta_{1N}}(p_z^{k_1})'\beta_{z0}^{k_1}| \leq C\vartheta_{1N}^{-\eta}, \quad \text{for} \quad k_1 = 1, \cdots, d_1$,

$\sup_{v \in \mathcal{R}^{1+d_1}} |g_{v0}(v) - \Psi_{\vartheta_{2N}}(v)'\beta_{h0}| \leq C\vartheta_{2N}^{-\eta}$.

P2.4. for a sufficiently large ϑ_{1n}, there exist a set of constant $(\underline{c}, \overline{c})$,

(a) $-\infty < \underline{c} \leq \lambda_{min}\left\{E[\sum_{t=t_0}^{T} \Psi_{\vartheta_{1N}}(p_{x,i,t})\Psi_{\vartheta_{1N}}(p_{x,i,t})']\right\} \leq \lambda_{max}\left\{E[\sum_{t=t_0}^{T} \Psi_{\vartheta_{1N}}(p_{x,i,t})\Psi_{\vartheta_{1N}}(p_{x,i,t})']\right\} \leq \overline{c} < \infty$,

$-\infty < \underline{c} \le \lambda_{min}\left\{E[\sum_{t=t_0}^{T} v_{x,i,t}^2 \Psi_{\vartheta_{1N}}(p_{x,i,t})\Psi_{\vartheta_{1n}}(p_{x,i,t})']\right\} \le \lambda_{max}\left\{E[\sum_{t=t_0}^{T} v_{x,i,t}^2 \Psi_{\vartheta_{1N}}(p_{x,i,t})\Psi_{\vartheta_{1N}}(p_{x,i,t})']\right\} \le \overline{c} < \infty$;

(b) $-\infty < \underline{c} \le \lambda_{min}\left\{E[\sum_{t=t_0}^{T} \Psi_{\vartheta_{1N}}(p_{z,i,t}^{k_1})\Psi_{\vartheta_{1N}}(p_{z,i,t}^{k_1})']\right\} \le \lambda_{max}\left\{E[\sum_{t=t_0}^{T} \Psi_{\vartheta_{1N}}(p_{z,i,t}^{k_1})\Psi_{\vartheta_{1N}}(p_{z,i,t}^{k_1})']\right\} \le \overline{c} < \infty$,

$-\infty < \underline{c} \le \lambda_{min}\left\{E[\sum_{t=t_0}^{T} (v_{z,i,t}^{k_1})^2 \Psi_{\vartheta_{1N}}(p_{z,i,t}^{k_1})\Psi_{\vartheta_{1N}}(p_{z,i,t}^{k_1})']\right\} \le \lambda_{max}\left\{E[\sum_{t=t_0}^{T} (v_{z,i,t}^{k_1})^2 \Psi_{\vartheta_{1N}}(p_{z,i,t}^{k_1})\Psi_{\vartheta_{1N}}(p_{z,i,t}^{k_1})']\right\} \le \overline{c} < \infty$, for $k_1 = 1, \cdots, d_1$.

(c) $\mathcal{L}$ is full rank in column, where $\mathcal{L}$ is defined under eq.(31).

Assumption P3:

P3.1. $\delta_0 \ne 0$ and $h_0(\cdot) \ne 0$ holds over at least one non-empty interval of its domain.

T3.2 (a) $\phi_0 = (\beta_0, \gamma_0, h_0) \in (B, \Gamma, \mathcal{H}) = \Phi$, $\beta_{h0} \in B_h \subset \mathcal{R}^{1+d_1}$, $\phi_N = (\beta_0, \gamma_0, h^*) \in (B, \Gamma, \mathcal{H}_N) = \Phi_N$, both B, Γ and B_h are compact set;

(b) ϕ_0 is the unique minimizer of $E[\sum_{t=t_0}^{T} \Delta\varepsilon_{i,t}(\phi)^2]$ over the space Φ, where $\Delta\varepsilon_{i,t}(\phi) = \Delta y_{i,t} - \beta_1 \Delta x_{i,t} - \delta(X_{i,t} - \gamma)\mathbf{I}_{i,t}(\gamma) - \Delta z_{i,t}'\beta_3 - h(v_{i,t}, v_{i,t-1})$ with $\phi = (\beta, \gamma, h) \in \Phi$.

P3.3. for any ϑ_{2N}, there exist constants $\underline{c_2}$ and $\overline{c_2}$ such that $-\infty < \underline{c_2} \le \lambda_{min}\left\{E[\sum_{t=t_0}^{T} \Delta x_{i,t}(\gamma)\Delta x_{i,t}'(\gamma)]\right\} \le \lambda_{max}\left\{E[\sum_{t=t_0}^{T} \Delta x_{i,t}(\gamma)\Delta x_{i,t}'(\gamma)]\right\} \le \overline{c}_2 < \infty$, and $-\infty < \underline{c}_2 \le \lambda_{min}\left\{E[\sum_{t=t_0}^{T} \Delta\varepsilon_{i,t}^2 \Delta x_{i,t}^*(\gamma)\Delta x_{i,t}^{*\prime}(\gamma)]\right\} \le \lambda_{max}\left\{E[\sum_{t=t_0}^{T} \Delta\varepsilon_{i,t}^2 \Delta x_{i,t}^*(\gamma)\Delta x_{i,t}^{*\prime}(\gamma)]\right\} \le \overline{c}_2 < \infty$ hold uniformly over $\gamma \in \Gamma$, where $\Delta x_{i,t}(\gamma)$ equals $\widehat{\Delta x_{i,t}}(\gamma)$ with $(\hat{v}_{i,t}, \hat{v}_{i,t-1})$ being replaced with $(v_{i,t}, v_{i,t-1})$.

P3.4. $x_{i,t}$ has a density function $f(x)$ and $f(x) \le \bar{f} < \infty$ over its domain for some finite constant $\bar{f}$.

For a scalar v, let $\|\Psi_{\vartheta_N}\|_{\mathcal{P}}^2 = \max_{s \le \mathcal{P}} \sup_{v \in \mathcal{R}} \|\nabla^s \Psi_{\vartheta_N}(v)\|^2$, where $\nabla^s \Psi_{\vartheta_N}(\cdot)$ is the sth derivative of $\Psi_{\vartheta_N}(\cdot)$. We then have $\|\nabla \Psi_{\vartheta_N}\|_{\mathcal{P}} = O(\vartheta_N^{\mathcal{P}+1/2})$ (see, e.g., the normalized Hermite functions and wavelet functions defined in Blundell et al. (2007)).

Assumption P4: $\vartheta_{1N} \to \infty$, $\vartheta_{2N} \to \infty$; $\|\Psi_{\vartheta_{2N}}\|_1 (\vartheta_{1N}^{-\eta} + \sqrt{\vartheta_{1N}/N})\sqrt{\vartheta_{2N}} = o(1)$.

Remark: Note that the assumption of i.i.d. across i can be relaxed to allow for independent but not identically distributed (inid) observations. Discussions on other assumptions of the panel KTR model mirrors that of the time series KTR model; thus, we will not repeat it here.

Theorem 1- panel. Denote $\theta_0 = (\beta_0', \gamma_0)'$, $\hat{\theta} = (\hat{\beta}', \hat{\gamma})'$, $\hat{h}(\cdot) = \Delta\Psi_{\vartheta_{2N}}(\cdot)'\hat{\beta}_h$, and $\hat{\phi}_N = (\hat{\theta}, \hat{h})$. Then, under Assumptions P1-P4, as $N \to \infty$, with a fixed T, we have

$$d(\hat{\phi}_N, \phi_0) = O_p\left(\vartheta_{2N}^{-\eta} + \sqrt{\frac{\vartheta_{2N}}{N}}\right), \tag{30}$$

where $d(\hat{\phi}_N, \phi_0) = \|\hat{\theta} - \theta_0\| + \|\hat{h} - h_0\|_\infty$.

Let

$$H_{i,t}(\beta, \gamma) = -\frac{\partial}{\partial \phi}\Delta\varepsilon_{i,t}(\phi) = \begin{pmatrix} \Delta x_{it} \\ (X_{i,t} - \tau_2\gamma)\mathbf{I}_{i,t}(\gamma) \\ \Delta z_{i,t} \\ -\delta\tau_2'\mathbf{I}_{i,t}(\gamma) \\ 1 \end{pmatrix}, \tag{31}$$

$H_{i,t} = H_{i,t}(\theta_0)$, $m_{i,t} = H_{i,t}\Delta\varepsilon_{i,t}$, $\mathcal{V} = \lim_{N\to\infty}\sum_{t=t_0}^{T} Var\left[m_{i,t}\right]$, and $\mathcal{L} = \sum_{t=t_0}^{T} E\left[\partial H_{i,t}\varepsilon_{i,t}/\partial\theta'\right]$.

Theorem 2-panel. Under Assumptions P1-P4, as $N \to \infty$, we have

$$\sqrt{N}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left[0, (\mathcal{L}'\mathcal{L})^{-1}\mathcal{L}'\mathcal{V}\mathcal{L}(\mathcal{L}'\mathcal{L})^{-1}\right]. \tag{32}$$

Remark: The proof is provided in the appendix. Similar to the time series model, our slope and threshold estimators are jointly normally distributed with root-N convergence rate and they have a non-zero asymptotic covariance matrix. To make an inference, given the sieve estimate $\hat{\phi}_N$, the asymptotic variance-covariance matrix can be consistently estimated by using $\hat{\mathcal{V}}_N = N^{-1}\sum_{i=1}^{N}\sum_{t=t_0}^{T}\left[m_{i,t}(\hat{\phi}_N)m_{i,t}(\hat{\phi}_N)'\right]$, $\hat{\mathcal{L}}_N = N^{-1}\sum_{i=1}^{N}\sum_{t=t_0}^{T}\partial[H_{i,t}(\hat{\theta})\varepsilon_{i,t}(\hat{\phi}_N)]/\partial\theta'$, and $m_{i,t}(\hat{\phi}_N) = H_{i,t}(\hat{\theta})\Delta\hat{\varepsilon}_{i,t}(\hat{\phi}_N)$, with $\Delta\hat{\varepsilon}_{i,t}(\hat{\phi}_N) = \Delta y_{i,t} - \hat{\beta}_1\Delta x_{i,t} - \hat{\delta}(X_{i,t} - \tau_2\hat{\gamma})\mathbf{I}_{i,t}(\hat{\gamma}) - \Delta z_{i,t}'\hat{\beta}_3 - \Delta\Psi_{\vartheta_{2N}}(\hat{v}_{i,t})'\hat{\beta}_h$. The full expression of $\hat{\mathcal{L}}_N$ is presented in the supplementary appendix.

4 Monte Carlo simulations

This section contains Monte Carlo simulations to evaluate the finite sample performance of our proposed estimator. Below, we list the following data-generating processes (DGPs).

$$\textbf{DGP1:} \quad y_t = c_0 + \beta_{10} x_t + \delta_0 (x_t - \gamma_0) I(x_t \geq \gamma_0) + \beta_{30} y_{t-1} + u_t, \quad u_t = 0.1\varepsilon_t + \kappa \sin(v_t),$$

$$x_t = 0.7 + 0.5 \sin(x_{t-1}) + v_t, \quad t = 1, \cdots, n. \tag{33}$$

In the time series setup, DGP1 considers the endogeneity of x_t, which comes from the common factor v_t between x_t and u_t. We set $(\varepsilon_t, v_t) \sim i.i.d.\mathcal{N}(0, I_2)$, $x_1 = 0$ and remove the first two observations to avoid the effect of starting value. The unknown true parameter values are $c_0 = \beta_{10} = \delta_0 = 1$, $\beta_{30} = 0.5$, and $\gamma_0 = 1$. We use κ to control the severity of endogeneity. The MC results for DGP1 are presented in Table 1.

[Table 1]

In Table(1), we let κ equal $2, 1, 0.05$, and set $n = [100, 200, 400]$. We compare the results of our proposed estimator with those of the least squares estimator, without considering the endogeneity issue, under different sample sizes. In this context, we deliberately keep the polynomial order at 6 to place our emphasis on tracking the convergence of our proposed estimator as the sample size (n) increases.[14]. First, we find that under different levels of endogeneity (i.e., $\kappa = 2, 1$), the estimator of $(\beta_1, \delta, \gamma)$ using the control function approach provides a consistent estimate. In contrast, the least squares estimator ignoring the endogeneity shows inconsistency. For the coefficient for the exogenous variables z_t, β_3, while the least squares estimator without CF appears to be barely affected by the endogeneity of x_t, we still observe that our proposed estimator employing CF outperforms the one without

[14] We also investigate the impact of varying the order of polynomials. We adjust the order of the basis functions, ϑ_{1n} and ϑ_{2n}, across $[3, 4, 5, 6]$. Under our DGP1, we observe that the RMSE of the estimators for all different orders of polynomials show consistency. We display that in the appendix.

CF. Turning to the weak endogeneity case (i.e., $\kappa = 0.05$), we note that both estimators - those employing the control function approach and those not utilizing it - perform well. Interestingly, the estimator without the control function approach exhibits a smaller RMSE. This is likely because, in cases with a relatively small sample size and weak endogeneity, the sieve estimator produces a larger variance.[15]

$$\begin{aligned} \textbf{DGP2:} \quad y_{i,t} &= c_0 + \beta_{10} x_{i,t} + \delta(x_{i,t} - \gamma_0) I(x_{i,t} \geq \gamma_0) + \beta_{30} z_{i,t} + u_{i,t}, \\ u_{i,t} &= 0.1\varepsilon_{i,t} + \kappa[\sin(v_{1,i,t}) + \sin(v_{2,i,t})], \\ x_{i,t} &= 0.7 + 0.5\sin(x_{i,t-1}) + v_{1,i,t}, \quad z_{i,t} = 0.7 + 0.5\sin(Z_{i,t-1}) + v_{2,i,t}, \\ & \quad i = 1, \cdots, N, t = 1, \cdots, T. \end{aligned} \tag{34}$$

For the panel data model, we consider DGP2, which involves an endogenous threshold variable, $x_{i,t}$, and an endogenous regressor, $z_{i,t}$. The endogeneity of $x_{i,t}$ comes from the common factor $v_{1,i,t}$, between $x_{i,t}$ and the error term $u_{i,t}$, and the endogeneity of $z_{i,t}$ comes form the common factor $v_{2,i,t}$ sharing with $u_{i,t}$. We set $(\varepsilon_{1,i,t}, v_{1,i,t}, v_{2,i,t}) \sim i.i.d.\mathcal{N}(0, I_3)$. The unknown parameters are $c_0 = 0, \beta_{10} = -0.5, \delta_0 = 1.2, \beta_{30} = 0.4$, and $\gamma_0 = 1$. With a fixed $T = 10$, we let $N = [20, 40, 80]$.

The MC results for DGP2 are reported in Table 2.

[Table 2]

In Table 2, to modulate the level of endogeneity, we choose κ from 2, 1, and 0.05. Notably, the findings from our panel Monte Carlo simulations align with those observed in the time series analysis [16].

[15]With a weak endogeneity($\kappa = 0.05$), as we expand the dataset to $n = 800$, the estimator employing the control function approach continues to outperform the one without using it.

[16]Similar as in the time series model, under our DGP2, we test the effect of the order of polynomials by choosing $\vartheta_{1N} = \vartheta_{2N}$ among $[3, 4, 5, 6]$. The MC results of the estimators with all different orders of polynomials show consistency and convergence. We present the results in the supplementary appendix.

5 Empirical application: The effect of COVID-19 on the US and Canadian labour markets

Since the worldwide outbreak in early 2020, all countries have suffered tremendously from the COVID-19 pandemic. For the labor market, there is a strand 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)). 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 was still ongoing. This fact motivated us to investigate the potential nonlinear relationship between COVID-19 cases and labor market performance, whereas the potential nonlinearity is tied to the occasional lockdown policies introduced by governments to ease up pandemic pressure on hospitals as cases surge. One difficulty in estimating the nonlinear effect of the case numbers on unemployment is endogeneity since there is strong evidence that COVID-19 case numbers are endogenous. Extending the canonical epidemiology model, Eichenbaum et al. (2021) find that during COVID-19, people cut back on working to avoid being affected. On the other hand, the increase in unemployment reduced the possible increases in contamination at the workplace and as such may have helped reduce the spread of the disease. Thus, we can expect the relationship to have a two-way causality endogeneity. 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 July 2020 to September 2021, while the US data spans from July 2020 to Dec 2021. As we use a two-period lagged variable in our regression, we drop the data for the first few months to avoid zero values. The covered periods are long enough to capture multiple waves of COVID-19 outbreaks, which provide an overall picture of this relationship[17]. We propose to use the following KTR model to examine our hypothesis

$$\begin{cases} Une_{it} = & \beta_0 + \beta_{low}(Case_{it} - \gamma_0)I(Case_{it} < \gamma_0) + \beta_{high}(Case_{it} - \gamma_0)I(Case_{it} \geq \gamma_0) \\ & +b_i + u_{it}, \\ Case_{it} = & \beta_{10} + g_1(Test_{i,t}) + g_2(Case_{i,t-1}) + v_{it}, \end{cases} \tag{35}$$

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 natural logarithm of the number of cases confirmed for COVID-19 in the t^{th} month.[18] Also, $Test_{it}$ equals the natural logarithm of the number of tests conducted in the t^{th} month. And, b_i is the individual fixed effect, which captures 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 the model (35). In particular, we use the lagged term of the endogenous variable, $Case_{i,t-1}$ and $Test_{i,t}$ as the instrument variable[19]. The functions $g_1(\cdot)$ and $g_2(\cdot)$ are unknown. In our estimation, we approximate them using the $6th$-order Hermite basis functions.

As a comparison, we also estimate and report the linear panel regression model, which

[17] To save space, we merge the data description and the table containing information about provinces/states in our dataset in the appendix.

[18] Note that we remove all log(0) by 0 to avoid the calculation problem. The same procedure is applied to the $Test$ variable.

[19] It is intuitive to assume $Test_{i,t}$ has no direct effect on the unemployment rate and can be viewed as an exogenous variable. At the same time, it is highly associated with the endogenous variable $Case_{i,t}$. Therefore, $Test_{i,t}$ is a valid IV.

is in the following form

$$\begin{cases} Une_{it} & = \beta_0 + \beta_{linear} Case_{it} + b_i + u_{it}, \\ Case_{it} & = \beta_{10} + g_1(Test_{i,t}) + g_2(Case_{i,t-1}) + v_{it}. \end{cases} \tag{36}$$

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 (36) is similar to the estimation strategy introduced in Section 3.1, except it does not require a grid search over γ.

Table 3 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 positive and statistically significant. For the KTR model, although the coefficient estimates for both the low and high regimes are positive - with the impact on the unemployment rate being more pronounced in the higher regime when the number of COVID-19 cases exceeds 32604 - we do not find any significance in the model for either regime. The results from using a control function approach to address the endogeneity are presented in the last two columns of Table 3. We observe that the coefficient estimate of COVID-19 for the linear model with endogeneity correction is similar to that without controlling for endogeneity, yet it remains insignificant. In the KTR model, when comparing to results without accounting for endogeneity, we observe a more pronounced threshold effect: The coefficient estimate in the low regime diminishes but remains insignificant, while the coefficient in the high regime increases and becomes significant. We also apply the linearity and endogeneity test for Regression (4).

To test for nonlinearity, we perform a bootstrapping test for the existence of a threshold effect following Hansen (1996, 2017). Our null hypothesis of interest is $\beta_{low,0} = \beta_{high,0}$. We repeat 10,000 simulations for the bootstrapping and obtain a p-value equal to 0.2447. The test fails to reject the null of linearity. For the endogeneity test, we apply the Wald test[20] and the test statistic equals 15.3589, which is greater than the critical value of 12.592 at the 0.05 significance level. This implies the existence of endogeneity.

Table 4 summarizes the estimation results for the US data. Similar to Table 3, regression (1) provides the results for the linear model, while regression (2) offers those for the KTR model, with neither controlling for endogeneity. Unlike the results from the Canadian data, the coefficient estimate from the linear model is negative, albeit still insignificant. In the KTR model, we split into two regimes based on a threshold level of 39,344 cases for that month. The coefficient for the low regime (β_{low}) is negative but not significant. Similar to our findings in the Canadian dataset, we observe a non-significant positive effect in the high regime. Regressions (3) and (4) present the estimation results using the CF approach. After correcting for endogeneity, we observe that the magnitude of the coefficient estimate for both the linear and the KTR model becomes more pronounced. Additionally, the level of the threshold estimate rises dramatically from 39,344 to 116,891. Interestingly, while the impact in the low regime remains negative, it now becomes significant. A potential reason for this might be the inherent stickiness of the labor market. If employers believe that the impact of the pandemic will be short-lived and not overly severe, the demand for labor remains steady. However, as some employees fall ill and others may play a wait-and-see strategy, there emerges a shortfall in the labor supply, indicating a tightening labor market. As a result, the influence of COVID-19 on the unemployment rate remains negative until the number of cases surpasses a certain threshold. In the high regime, we observe a positive

[20] To save space, we relegate the endogenous test to our online appendix. For further details, refer to Section G in the Supplementary Material.

and significant effect of COVID-19 cases on the unemployment rate, thus our conclusion mirrors what we found with the Canada dataset. The adverse impact of COVID-19 on the labor market becomes evident only when the number of confirmed cases reaches a significant magnitude. Focusing on Regression (4), we also implement the threshold effect test and obtain the bootstrap p-value= 0.0288. We reject the null hypothesis of linearity at 5% significant level, favoring the KTR model. The Wald test for endogeneity yields a statistic equal to 66.9861, which is greater than the critical value at the 5% significant level, 15.3589. This suggests the existence of endogeneity in $Case_{i,t}$. Both test statistics support our hypotheses.

6 Conclusion

Extending Hansen (2017), we consider a kink threshold model with endogeneity. Following Kourtellos et al. (2016) and Yu et al. (2023), we employ the nonparametric control function approach to tackle the endogeneity and propose a two-step semiparametric estimator. Compared to other methods that address endogeneity in the context of a threshold regression model, our method is both easier to apply and more reliable, especially with a small sample size and our Monte Carlo simulations support that. We apply our model to the potential nonlinear effect of COVID-19 cases on the unemployment rate in Canada and the US and we find that COVID-19 cases have a significant negative effect on labor market activity, but only when the number of confirmed cases surpasses a certain threshold. Below that level, the impact is positive, possibly due to the stickiness of labor demand.

Table 1: DGP1-Main

		β_1		δ		β_3		γ	
		bias	rmse	bias	rmse	bias	rmse	bias	rmse
NO CF/κ = 2	n=100	0.0227	0.6005	0.8707	1.171	-0.0963	0.1017	-0.9138	1.0315
	n=200	-0.1153	0.4408	1.0269	1.1208	-0.0924	0.0951	-1.0199	1.0494
	n=400	-0.1516	0.3436	1.0585	1.0964	-0.0899	0.0913	-1.0485	1.0576
CF/κ = 2	n=100	0.2527	0.5264	0.2304	0.5929	-0.0557	0.0737	-0.4081	0.6009
	n=200	0.1688	0.319	0.0561	0.2674	-0.0295	0.0498	-0.2706	0.4353
	n=400	0.0977	0.1974	-0.0199	0.1443	-0.0148	0.0341	-0.1278	0.2332
NO CF/κ = 1	n=100	0.1976	0.2875	0.209	0.2768	-0.0532	0.0609	-0.538	0.5705
	n=200	0.1643	0.2299	0.2286	0.2672	-0.05	0.0541	-0.5645	0.5832
	n=400	0.1611	0.2004	0.2253	0.2469	-0.048	0.0502	-0.5701	0.5806
CF/κ = 1	n=100	0.1858	0.2429	0.0344	0.1785	-0.0301	0.051	-0.1553	0.2411
	n=200	0.1114	0.1579	-0.0156	0.1176	-0.0149	0.0375	-0.0862	0.1389
	n=400	0.0525	0.1016	-0.0227	0.0792	-0.0077	0.0273	-0.0441	0.0804
NO CF/κ = 0.05	n=100	0.0229	0.035	-0.0001	0.0349	0.0049	0.0226	-0.0143	0.0403
	n=200	0.0255	0.0314	-0.0003	0.024	0.0066	0.0172	-0.0074	0.0275
	n=400	0.028	0.0303	-0.0004	0.0171	0.0068	0.0129	-0.002	0.0143
CF/κ = 0.05	n=100	0.1858	0.2429	0.0344	0.1785	-0.0301	0.051	-0.1553	0.2411
	n=200	0.1114	0.1579	-0.0156	0.1176	-0.0149	0.0375	-0.0862	0.1389
	n=400	0.0525	0.1016	-0.0227	0.0792	-0.0077	0.0273	-0.0441	0.0804

Note: This table presents bias and root mean squared error(rmse) of our proposed estimator. We use $6th$-order Hermite basis functions for both first-step and second-step estimation(i.e.,$\vartheta_{1n} = \vartheta_{2n} = 6$). κ controls the level of endogeneity and CF denotes the use of the control function approach, see eq.(33) for a detailed description.

Table 2: DGP2-Main

		β_1		δ		β_3		γ	
	T=10	bias	rmse	bias	rmse	bias	rmse	bias	rmse
NO CF/κ = 2	N=20	-0.1443	0.9956	0.876	1.7408	1.237	1.2422	-1.0162	1.3142
	N=40	-0.2739	0.7644	1.1063	1.5121	1.2343	1.2369	-1.1885	1.3187
	N=80	-0.342	0.588	1.2281	1.3548	1.2312	1.2325	-1.2884	1.3213
CF/κ = 2	N=20	0.0984	0.4585	0.0696	0.5318	0.3123	0.3809	-0.405	0.6754
	N=40	0.0705	0.2432	-0.0223	0.217	0.1457	0.2088	-0.1863	0.3786
	N=80	0.0126	0.1365	-0.0305	0.1219	0.0551	0.1139	-0.0849	0.1821
NO CF/κ = 1	N=20	0.1654	0.3478	0.1937	0.3296	0.6184	0.6211	-0.7334	0.779
	N=40	0.1666	0.2764	0.1842	0.2661	0.617	0.6183	-0.7491	0.7734
	N=80	0.1713	0.2319	0.1748	0.2206	0.6156	0.6163	-0.753	0.7662
CF/κ = 1	N=20	0.1447	0.2157	-0.0554	0.1902	0.1567	0.1928	-0.1198	0.2275
	N=40	0.0664	0.1178	-0.0407	0.1072	0.0728	0.1062	-0.0527	0.1028
	N=80	0.0144	0.0691	-0.0228	0.0692	0.0276	0.0584	-0.0295	0.061
NO CF/κ = 0.05	N=20	0.0303	0.0387	-0.0105	0.0321	0.0308	0.0323	-0.0139	0.038
	N=40	0.0329	0.0368	-0.0099	0.0232	0.0309	0.0316	-0.0071	0.0267
	N=80	0.035	0.0366	-0.0098	0.0179	0.0308	0.0311	-0.0018	0.0135
CF/κ = 0.05	N=20	0.0092	0.0475	-0.0029	0.072	0.0085	0.0333	-0.0028	0.0397
	N=40	0.0036	0.033	-0.0024	0.0514	0.0036	0.0224	-0.0007	0.0214
	N=80	0.0009	0.0225	-0.0009	0.0346	0.0016	0.015	-0.0001	0.0079

Note: This table presents bias and root mean squared error(rmse) of our proposed estimator. We use $6th$-order Hermite basis functions for both first-step and second-step estimation(i.e.,$\vartheta_{1n} = \vartheta_{2n} = 6$). κ controls the level of endogeneity and CF denotes the use of the control function approach, see eq.(34) for a detailed description.

Table 3: Correlation between the unemployment rate and COVID-19 cases(Canada dataset)

	(1)	(2)	(3)	(4)
Model	Linear	Threshold	Linear	Threshold
γ(Case)		10.3922***		10.3922***
		(0.1317)		(0.7257)
β_{linear}	0.1442**		0.1401	
	(0.0707)		(0.0871)	
β_{low}		0.1209		0.0323
		(0.1015)		(0.2379)
β_{high}		0.8269		1.6587***
		(0.5815)		(0.6199)
Control function			✓	✓
N_{low}		122		122
N_{high}		14		14
N_{total}	136	136	136	136

NOTE: ***, **, * indicate significance at 1% level, 5% level, 10% level, respectively;

Wald endogeneity test: $H_0 : \beta_{h0} = \mathbf{0}_{\vartheta_{\mathbf{2N}}}$, $W_n = 15.3589 > 12.592 (\alpha = 0.05)$;

Linearity test: $\beta_{low} = \beta_{high}$, $P_n = 0.2447$.

Table 4: Correlation between unemployment rate and COVID-19 cases(US dataset)

	(1)	(2)	(3)	(4)
Model	Linear	Threshold	Linear	Threshold
γ(Case)		10.580****		11.669***
		(0.663)		(0.374)
β_{linear}	−0.012		−0.157***	
	(0.045)		(0.035)	
β_{low}		−0.054		−0.245***
		(0.119)		(0.075)
β_{high}		0.056		0.249***
		(0.292)		(0.084)
Control function			✓	✓
N_{low}		551		783
N_{high}		333		101
N_{total}	884	884	884	884

NOTE: ***, **, * indicate significance at 1% level, 5% level, 10% level, respectively;

Wald endogeneity test: $H_0 : \beta_{h0} = \mathbf{0}_{\vartheta_{2N}}$, $W_n = 66.9861 > 12.592 (\alpha = 0.05)$;

Linearity test: $\beta_{low} = \beta_{high}$, $P_n = 0.029$.

References

Baek, C., P. B. McCrory, T. Messer, and P. Mui (2021). Unemployment effects of stay-at-home orders: Evidence from high-frequency claims data. *The Review of Economics and Statistics 103*(5), 979–993.

Blundell, R., X. Chen, and D. Kristensen (2007). Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica 75*(6), 1613–1669.

Caner, M. and B. E. Hansen (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 R. S. Tsay (1998, 06). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika 85*(2), 413–426.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Chapter 76 in Handbook of Econometrics*, 5549–5632.

Chen, X., H. Hong, and E. Tamer (2005). Measurement error models with auxiliary data. *The Review of Economic Studies 72*(2), 343–366.

Chen, X., O. Linton, and I. Van Keilegom (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica 71*(5), 1591–1608.

Christopoulos, D., P. McAdam, and E. Tzavalis (2021). Dealing with endogeneity in threshold models using copulas. *Journal of Business & Economic Statistics 39*(1), 166–178.

Dzyadyk, V. K. and I. A. Shevchuk (2008). Theory of uniform approximation of functions by polynomials. In *Theory of Uniform Approximation of Functions by Polynomials*. de Gruyter.

Eichenbaum, M. S., S. Rebelo, and M. Trabandt (2021). The macroeconomics of epidemics. *The Review of Financial Studies 34*(11), 5149–5187.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica 64*(2), 413–430.

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.

Hidalgo, J., J. Lee, and M. H. Seo (2019). Robust inference for threshold regression models. *Journal of Econometrics 210*(2), 291–309.

Karavias, Y., P. K. Narayan, and J. Westerlund (2022). Structural breaks in interactive effects panels and the stock market reaction to covid-19. *Journal of Business & Economic Statistics*, 1–14.

Kong, E. and D. Prinz (2020). Disentangling policy effects using proxy data: Which shutdown policies affected unemployment during the covid-19 pandemic? *Journal of Public Economics 189*, 104257.

Kourtellos, A., T. Stengos, and Y. Sun (2022). Endogeneity in semiparametric threshold regression. *Econometric Theory 38*(3), 562–595.

Kourtellos, A., T. Stengos, and C. M. Tan (2016). Structural threshold regression. *Econometric Theory 32*(4), 827–860.

Lee, S., Y. Liao, M. H. Seo, and Y. Shin (2021). Sparse hp filter: Finding kinks in the covid-19 contact rate. *Journal of econometrics 220*(1), 158–180.

Lee, S. Y. T., M. Park, and Y. Shin (2021, January). Hit harder, recover slower? unequal employment effects of the covid-19 shock. Working Paper 28354, National Bureau of Economic Research.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of econometrics 79*(1), 147–168.

Newey, W. K., J. L. Powell, and F. Vella (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica 67*(3), 565–603.

Ozabaci, D., D. J. Henderson, and L. Su (2014). Additive nonparametric regression in the presence of endogenous regressors. *Journal of Business & Economic Statistics 32*(4), 555–575.

Seo, M. H. and Y. Shin (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.

Xiang, S. (2012). Asymptotics on laguerre or hermite polynomial expansions and their applications in gauss quadrature. *Journal of Mathematical Analysis and Applications 393*(2), 434–444.

Yu, P., Q. Liao, and P. C. B. Phillips (2023). New control function approaches in threshold regression with endogeneity. *Econometric Theory*, 1–55.

Yu, P. and P. C. Phillips (2018). Threshold regression with endogeneity. *Journal of Econometrics 203*(1), 50–68.

SUPPLEMENTARY MATERIAL for "Endogenous kink threshold regression"

Jianhan Zhang* Chaoyi Chen†‡ Yiguo Sun§ Thanasis Stengos ¶

October 16, 2023

This online supplement is composed of eight parts. Section A contains the proof of Theorem 1-time series, including the consistency and the convergence rate. Section B gives the proof of Theorem 2-time series. Section C contains the proof of Theorem 1-panel. Section D shows the proof of the Theorem 2-panel. Section E shows the proof of Lemmas. Section F gives the omitted expression of the variance-covariance matrix in the main text. Section G proposes a Wald test to test the endogeneity. Section H reports the MC results of changing the order of polynomials (for time-series, $\vartheta_{2n}, \vartheta_{2n}$, for the panel data model, $\vartheta_{1N}, \vartheta_{2N}$) in our basis functions; Section I contains the summary statistics of our dataset.

A Proof of Theorem 1-time series

A.1 Consistency

To establish the consistency of our proposed estimator, we apply the results from Theorem 3.1 in Chen (2007), which provides a general consistency result for sieve extreme estimators.

*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.

In doing so, we must verify Conditions 3.1-3.5 as outlined in Theorem 3.1 of Chen (2007). Notably, Condition 3.1 aligns with our Assumption T3.2(b), which presupposes ϕ_0 as the unique minimizer of our objective function. Condition 3.2 corresponds to our Assumption T2.3, which posits the existence of an appropriate sieve approximation for our unknown functions, denoted as $h_0(\cdot)$. Condition 3.3 is satisfied due to the continuity property of the KTR model. Condition 3.4 is met through the assumption of the compactness of the sieve space, in accordance with our Assumption T3.2(a). In summary, to apply Theorem 3.1 from Chen (2007), our task is to demonstrate Condition 3.5, namely, the uniform convergence of the objective function over the sieve space. We will elucidate this in the following steps.

Denote $\hat{S}_n(\phi^*) = 1/n \sum_{t=1}^{n} \hat{\varepsilon}_t(\phi^*)^2$, where $\hat{\varepsilon}_t(\phi^*)$ equals $\hat{\varepsilon}_t(\hat{\phi}_n)$ defined in Remark under Theorem 2-time series by replacing $\hat{\phi}_n$ with ϕ^*; $S_n(\phi^*) = 1/n \sum_{t=1}^{n} \varepsilon_t(\phi^*)^2$, where $\varepsilon_t(\phi^*)$ equals $\hat{\varepsilon}_t(\phi^*)$ by replacing $\hat{v}_t$ with v_t.

For a kink model with a continuous objective function, we only need to show the uniform convergence of $\hat{S}_n(\phi^*)$ to $E[S_n(\phi^*)]$ for $\phi^* \in \Phi_n$ as $n \to \infty$, which equivalent to Condition 3.5 of Chen (2007). In other words, we need to prove:

$$plim_{n\to\infty} \sup_{\phi^* \in \Phi_n} |\hat{S}_n(\phi^*) - E[S_n(\phi^*)]| = 0. \tag{A.1}$$

To show that, by the Triangular inequality, we have

$$\begin{aligned}\sup_{\phi^* \in \Phi_n} |\hat{S}_n(\phi^*) - E[S_n(\phi^*)]| &\leq \sup_{\phi^* \in \Phi_n} |\hat{S}_n(\phi^*) - S_n(\phi^*)| + \sup_{\phi^* \in \Phi_n} |S_n(\phi^*) - E[S_n(\phi^*)]| \\ &= S_1 + S_2. \end{aligned} \tag{A.2}$$

(S_1:) We first show S_1 is $o_p(1)$. We have the expression

$$\begin{aligned}
\sup_{\phi^*\in\Phi_n}\left[\hat{S}_n(\phi^*)-S_n(\phi^*)\right] &= \sup_{\phi^*\in\Phi_n}\frac{1}{n}\sum_{t=1}^{n}[\hat{\varepsilon}_t(\phi^*)^2-\varepsilon_t(\phi^*)^2] \\
&= \sup_{\phi^*\in\Phi_n}\frac{1}{n}\sum_{t=1}^{n}[\hat{\varepsilon}_t(\phi^*)-\varepsilon_t(\phi^*)]^2+\sup_{\phi^*\in\Phi_n}\frac{2}{n}\sum_{t=1}^{n}\varepsilon_t(\phi^*)[\hat{\varepsilon}_t(\phi^*)-\varepsilon_t(\phi^*)] \\
&= \sup_{\beta_h\in B_h}\frac{1}{n}\sum_{t=1}^{n}\left\{[\Psi_{\vartheta_{2n}}(v_t)-\Psi_{\vartheta_{2n}}(\hat{v}_t)]'\beta_h\right\}^2 \\
&\quad+\sup_{\phi^*\in\Phi_n}\frac{2}{n}\sum_{t=1}^{n}\varepsilon_t(\phi^*)[\Psi_{\vartheta_{2n}}(v_t)-\Psi_{\vartheta_{2n}}(\hat{v}_t)]'\beta_h \\
&= A_1+2A_2. \qquad \text{(A.3)}
\end{aligned}$$

Next, we prove the convergence of A_1 and A_2. By simple calculation, under Lemma 3, T2.2, and T3.2(a), applying the Cauchy-Schwarz inequality and Taylor expansion, we have

$$\begin{aligned}
|A_1| &= \sup_{\beta_h\in B_h}\frac{1}{n}\sum_{t=1}^{n}\left\{[\Psi_{\vartheta_{2n}}(v_t)-\Psi_{\vartheta_{2n}}(\hat{v}_t)]'\beta_h\right\}^2 \\
&\le \sup_{\beta_h\in B_h}\frac{1}{n}\sum_{t=1}^{n}\|[\Psi_{\vartheta_{2n}}(v_t)-\Psi_{\vartheta_{2n}}(\hat{v}_t)]\|^2\|\beta_h\|^2 \\
&= \sup_{\beta_h\in B_h}\frac{1}{n}\sum_{t=1}^{n}\|\nabla\Psi_{\vartheta_{2n}}(\bar{v}_t)(v_t-\hat{v}_t)\|^2\|\beta_h\|^2 \\
&\le \sup_{\beta_h\in B_h}\frac{1}{n}\sum_{t=1}^{n}\|\nabla\Psi_{\vartheta_{2n}}(\bar{v}_t)\|^2\|v_t-\hat{v}_t\|^2\|\beta_h\|^2 \\
&= O_p\left[\|\Psi_{\vartheta_{2n}}\|_1^2(\vartheta_{1n}^{-2\eta}+\vartheta_{1n}/n)\vartheta_{2n}\right], \qquad \text{(A.4)}
\end{aligned}$$

where denote $\bar{v}_t$ as a specific vector which lies in between v_t and $\hat{v}_t$, $\nabla\Psi_{\vartheta_{2n}}(\bar{v}_t)$ is the partial derivative with respect to $\bar{v}_t$, which is a $[\vartheta_{2n}(1+d_1)]\times(1+d_1)$ matrix. Under Assumption T4, $|A_1|=o_p(1)$.

Next, we show A_2. To show A_2 is bounded, first we show $\sup_{\phi^*\in\Phi_n}1/n\sum_{t=1}^{n}\varepsilon_t(\phi^*)^2$ is bounded. Note that by simple calculation, we have

$$\sup_{\phi^*\in\Phi_n}\frac{1}{n}\sum_{t=1}^n \varepsilon_t(\phi^*)^2 = \sup_{\phi^*\in\Phi_n}\frac{1}{n}\sum_{t=1}^n[\varepsilon_t(\phi^*)-\varepsilon_t(\phi_0^*)+\varepsilon_t(\phi_0^*)-\varepsilon_t+\varepsilon_t]^2$$
$$\leq \sup_{\phi^*\in\Phi_n}\frac{2}{n}\sum_{t=1}^n[\varepsilon_t(\phi^*)-\varepsilon_t(\phi_0^*)]^2+\frac{2}{n}\sum_{t=1}^n[\varepsilon_t(\phi_0^*)-\varepsilon_t(\phi_0)]^2+\frac{2}{n}\sum_{t=1}^n\varepsilon_t^2$$
$$= O_p(1)+O_p(\vartheta_{2n}^{-2\eta})+O_p(1)=O_p(1), \tag{A.5}$$

where the boundness of the first term is given by

$$\sup_{\phi^*\in\Phi_n}\frac{1}{n}\sum_{t=1}^n\left[\varepsilon_t(\phi^*)-\varepsilon_t(\phi_0^*)\right]^2$$
$$\leq \sup_{\phi^*\in\Phi_n}\frac{2}{n}\sum_{t=1}^n x_t^2[(\beta_{10}-\beta_1)^2+(\delta_0-\delta)^2]+2\sup_{\phi^*\in\Phi_n}[\beta_{10}\gamma_0-\beta_1\gamma]^2+2\sup_{\phi^*\in\Phi_n}[\delta_0\gamma_0-\delta\gamma]^2$$
$$+\sup_{\phi^*\in\Phi_n}\frac{2}{n}\sum_{t=1}^n(\beta_{30}-\beta_3)'z_tz_t'(\beta_{30}-\beta_3)+\sup_{\phi^*\in\Phi_n}\frac{2}{n}\sum_{t=1}^n[h_0^*(v_t)-\Psi_{\vartheta_{2n}}(v_t)'\beta_h]^2$$
$$=O_p(1), \tag{A.6}$$

which holds under Assumption T1.1(b), T2.3(b) and T3.2(a). The second term is given by $|\varepsilon_t(\phi_0^*)-\varepsilon_t(\phi_0)|^2\leq \sup_{v\in\mathcal{R}^{1+d_1}}|h_0(v_t)-\Psi_{\vartheta_{2n}}(v_t)'\beta_{h0}|^2=O_p(\vartheta_{2n}^{-2\eta})$, under Assumption T2.3(b). The last term uses the fact that our true error term is bounded under Assumptions T1.1(b) and T2.3(b).

Following results in (A.4) and (A.5), and applying the Hölder inequality gives

$$|A_2| \leq \left(\sup_{\phi^*\in\Phi_n}\frac{1}{n}\sum_{t=1}^n|\varepsilon_t(\phi^*)|^2\right)^{1/2}\left(\sup_{\phi^*\in\Phi_n}\frac{1}{n}\sum_{t=1}^n|[\Psi_{\vartheta_{2n}}(v_t)-\Psi_{\vartheta_{2n}}(\hat{v}_t)]'\beta_h|^2\right)^{1/2}$$
$$=O_p(1)\left\{O_p\left[\|\nabla\Psi_{\vartheta_{2n}}\|_1(\vartheta_{1n}^{-\eta}+\sqrt{\vartheta_{1n}/n})\sqrt{\vartheta_{2n}}\right]\right\}$$
$$=o_p(1). \tag{A.7}$$

To sum up, we have $S_1=\sup_{\phi^*\in\Phi_n}|\hat{S}_n(\phi^*)-S_n(\phi^*)|=o_p(1)$, as $n\to\infty$.

(S_2:) To show the uniform convergence of S_2, we check the conditions of Corollary 2.2 in Newey (1991). The Corollary needs three conditions: (1) the compactness of the parameter space; (2) the point-wise convergence of the objective function $S_n(\phi^*)$(i.e., for any $\phi^*\in\Phi_n$, $\lim_{n\to\infty}|S_n(\phi^*)-E[S_n(\phi^*)]|=0$); (3) the stochastic equicontinuity of $S_n(\phi^*)$. Note the compactness of the parameter space is given by our Assumption T3.2. Next, we show the latter two conditions hold in our case.

First, we demonstrate the point-wise convergence for $S_n(\phi^*)$, which we establish using the Weak Law of Large Numbers (WLLN) for a β-mixing sequence as in Hansen (2019). It's worth noting that Hansen's WLLN only necessitates that our $E[\varepsilon_t(\phi^*)^2]$ is uniformly bounded over $\phi^* \in \Phi_n$, a condition satisfied under Assumptions T1.1 and T2.3(b). Having established the point-wise convergence result, our next objective is to demonstrate stochastic equicontinuity. To achieve this, we apply Condition 3A from Newey (1991), a Lipschitz continuity condition that provides a sufficient criterion for stochastic equicontinuity. [1]. Next, we show Condition 3A holds in our case. We will now proceed to show that Condition 3A holds in our specific case.

Without loss of generality, we denote $\dot{\phi}^* = (\dot{\beta}, \dot{\gamma}, \dot{h}^*) \in \Phi_n$ and $\ddot{\phi}^* = (\ddot{\beta}, \ddot{\gamma}, \ddot{h}^*) \in \Phi_n$, with $\dot{\gamma} \leq \ddot{\gamma}$ and $d(\dot{\phi}^*, \ddot{\phi}^*) \leq 1$, where $d(\cdot, \cdot)$ is a measure of distance, defined in Theorem 1-time series. By simple calculation, under the assumption of a compact parameter space (Assumption T3.2(a)), we need to show

$$|S_n(\dot{\phi}^*) - S_n(\ddot{\phi}^*)]| \leq B_n K\left(d(\dot{\phi}^*, \ddot{\phi}^*)\right), \tag{A.8}$$

where B_n needs to be bounded, $K(\cdot)$ denotes a function with $K : [0, \infty) \to [0, \infty)$, $K(0) = 0$ and it is continuous at 0. Without loss of generality, we assume $S_n(\phi^{**}) \leq S_n(\phi^*)$. To show (A.8), we need to rely on the absolute value inequality. The inequality indicates for any real number a, b, c, d, we have $|a - b| - |c - d| \leq |a - b - c + d| \leq |a - c| + |b - d|$. Thus,

[1] For similar assumption, see A4 of Andrews (1994).

applying that inequality and using the point-wise convergence result of $S_n(\phi^*)$, we have

$$
\begin{aligned}
& |S_n(\dot{\phi}^*) - S_n(\ddot{\phi}^*)]| \\
= \quad & \left\{ |S_n(\dot{\phi}^*) - S_n(\ddot{\phi}^*)]| - E|S_n(\dot{\phi}^*) - S_n(\ddot{\phi}^*)]| \right\} + E|S_n(\dot{\phi}^*) - S_n(\ddot{\phi}^*)]| \\
\leq \quad & |S_n(\dot{\phi}^*) - E[S_n(\dot{\phi}^*)]| + |S_n(\ddot{\phi}^*) - E[S_n(\ddot{\phi}^*)]| + E|S_n(\dot{\phi}^*) - S_n(\ddot{\phi}^*)]| \\
= \quad & o_p(1) + E|S_n(\dot{\phi}^*) - S_n(\ddot{\phi}^*)]|. \qquad \text{(A.9)}
\end{aligned}
$$

Now, combining (A.8) with (A.9), we only need to show $E|S_n(\dot{\phi}^*) - S_n(\ddot{\phi}^*)|$ is bounded, by simple calculation, we have

$$
\begin{aligned}
E|S_n(\dot{\phi}^*) - S_n(\ddot{\phi}^*)]| \quad & \leq \underbrace{E\{[\varepsilon_t(\dot{\phi}^*) - \varepsilon_t(\ddot{\phi}^*)]^2\}}_{D_{v1}} + 2\underbrace{E|\varepsilon_t(\dot{\phi}^*)[\varepsilon_t(\dot{\phi}^*) - \varepsilon_t(\ddot{\phi}^*)]|}_{D_{v2}} \qquad \text{(A.10)} \\
& \leq constant \times E\,\|w_t\| \times d(\dot{\phi}^*, \ddot{\phi}^*), \qquad \text{(A.11)}
\end{aligned}
$$

where w_t is a vector whose elements contain any pairwise inner products of $(y_t, 1, x_t, z_t, h_0(v_t))$. To be more precisely, for D_{v1}, we have

$$
\begin{aligned}
D_{v1} \quad & = E\left\{[\varepsilon_t(\dot{\phi}^*) - \varepsilon_t(\ddot{\phi}^*)]^2\right\} \\
& \leq 2\underbrace{E[(x_t - \ddot{\gamma})\ddot{\beta} - (x_t - \dot{\gamma})\dot{\beta}]^2}_{D_x} + 2\underbrace{E[\ddot{\delta}(x_t - \ddot{\gamma})I(x_t \geq \ddot{\gamma}) - \dot{\delta}(x_t - \dot{\gamma})I(x_t \geq \dot{\gamma})]^2}_{D_\gamma} \\
& + 2E\,\|z_t\|^2 \left\|\ddot{\beta}_3 - \dot{\beta}_3\right\|^2 + 2E\left\{[\ddot{h}^*(v_t) - \dot{h}^*(v_t)]^2\right\} \\
& \leq constant \times E\,\|w_t\| \times d(\dot{\phi}^*, \ddot{\phi}^*)^2, \qquad \text{(A.12)}
\end{aligned}
$$

where we only need to show D_x and D_γ. For D_x, with a compact parameter space(Assumption T3.2(a)), we have

$$
\begin{aligned}
D_x \quad & \leq 2E|x_t|^2(\ddot{\beta} - \dot{\beta})^2 + 2(\dot{\beta}(\dot{\gamma} - \ddot{\gamma}))^2 + 2(\dot{\beta} - \ddot{\beta})^2\ddot{\gamma}^2 \\
& \leq constant \times E\,\|w_t\| \times d(\dot{\phi}^*, \ddot{\phi}^*)^2. \qquad \text{(A.13)}
\end{aligned}
$$

For D_γ, under Assumptions T3.2(a) and T3.4, we have

$$\begin{aligned} D_\gamma &\leq 2\ddot{\delta}^2 E\left\{[(x_t-\ddot{\gamma})I(x_t\geq\ddot{\gamma})-(x_t-\dot{\gamma})I(x_t\geq\ddot{\gamma})]^2\right\}+2(\ddot{\delta}-\dot{\delta})^2E\left[(x_t-\ddot{\gamma})^2I(x_t\geq\ddot{\gamma})\right] \\ &\leq 4\dot{\delta}^2E[(x_t-\dot{\gamma})^2I(\dot{\gamma}\leq x_t\leq\ddot{\gamma})]+4\dot{\delta}^2(\ddot{\gamma}-\dot{\gamma})^2E[I(x_t\geq\ddot{\gamma})]+constant_1\times d(\dot{\phi}^*,\ddot{\phi}^*)^2 \\ &\leq constant_2\times\max\{\bar{f},E\left\|w_t\right\|\}\times d(\dot{\phi}^*,\ddot{\phi}^*)^2\leq constant_3\times E\left\|w_t\right\|\times d(\dot{\phi}^*,\ddot{\phi}^*)^2. \quad (A.14)\end{aligned}$$

For D_{v2}, we use a similar method as used to prove D_{v1}. Then, we have

$$\begin{aligned} D_{v2} &= E|\varepsilon_t(\dot{\phi}^*)[\varepsilon_t(\dot{\phi}^*)-\varepsilon_t(\ddot{\phi}^*)]| \\ &\leq constant\times E\left\|w_t\right\|\times d(\dot{\phi}^*,\ddot{\phi}^*). \end{aligned} \quad (A.15)$$

Combining (A.12) and (A.15) gives (A.11). Under Assumptions T1.1 and T2.2, we have $E\left\|w_t\right\|=O(1)$ by applying the Hölder inequality [2] [3]. We then combine these findings with (A.9) and (A.11) to conclude our proof for (A.8), with $K\left(d(\dot{\phi}^*,\ddot{\phi}^*)\right)=d(\dot{\phi}^*,\ddot{\phi}^*)$. In summary, we demonstrate that Condition 3A, as presented in Newey (1991), holds within our model. This establishes all necessary conditions for Corollary 2.2 from the same source, ultimately leading to the conclusion that S_2 converges to zero in probability, denoted as $S_2=o_p(1)$.

Given that both S_1 and S_2 are of order $o_p(1)$, we establish the uniform convergence results presented in equation (A.1). This uniform convergence implies that as the sample size n approaches infinity, the distance $d(\hat{\phi}_n,\phi_0)$ converges to zero in probability, which can be inferred from the results outlined in Theorem 3.1 of Chen (2007).

A.2 Convergence rate

To establish the convergence rate of our proposed estimator, we apply Theorem 3.2 in Chen (2007), which permits the time series data to be β-mixing. In doing so, under our

[2] note here we only require the moment condition up to order 2.

[3] Note the soundness of $\|w_t\|$ and $h_0(v_t)$ also imply the requirement of $E[\varepsilon_t^2]$ to be equicontinuous in Corollary 2.2 Newey (1991) is automatically satisfied.

β-mixing sequence assumption (Assumption T1.1(a)), it suffices to verify Condition 3.7 and Condition 3.8 in Theorem 3.2 in her work.

Firstly, we check Condition 3.7 of Chen (2007). In our case, it is equivalent to verify

$$\sup_{\phi^{**}\in\Phi_n, d(\phi^{**},\phi_0)\leq\omega} Var\left(\varepsilon_t^2(\phi^{**})-\varepsilon_t^2\right) \leq constant \times \omega^2 \tag{A.16}$$

for any small $0 < \omega < 1$. By definition and $\varepsilon_t^2(\phi^{**}) - \varepsilon_t^2 = 2\varepsilon_t[\varepsilon_t(\phi^{**}) - \varepsilon_t] + [\varepsilon_t(\phi^{**}) - \varepsilon_t]^2$, we have

$$\begin{aligned} Var(\varepsilon_t^2(\phi^{**}) - \varepsilon_t^2) &\leq E\left\{[\varepsilon_t^2(\phi^{**}) - \varepsilon_t^2]^2\right\} \\ &\leq 8E\left\{\varepsilon_t^2[\varepsilon_t(\phi^{**}) - \varepsilon_t]^2\right\} + 2E\left\{[\varepsilon_t(\phi^{**}) - \varepsilon_t]^4\right\} \\ &\leq constant \times E\left\{[\varepsilon_t(\phi^{**}) - \varepsilon_t]^2\right\} + 2E\left\{[\varepsilon_t(\phi^{**}) - \varepsilon_t]^4\right\}, \\ &= constant_1 \times \omega^2 + constant_2 \times \omega^4, \end{aligned} \tag{A.17}$$

since $E[\varepsilon_t^2(\varepsilon_t(\phi^{**}) - \varepsilon_t)^2] = E\left[E\left[\varepsilon_t^2|\mathcal{F}_{t-1}, x_t, z_{1t}\right](\varepsilon_t(\phi^{**}) - \varepsilon_t)^2\right] \leq constant \times E\left[(\varepsilon_t(\phi^{**}) - \varepsilon_t)^2\right]$ under Assumptions T1.1.(b)(c) and T2.3 (b), and

$$\begin{aligned} |\varepsilon_t(\phi^{**}) - \varepsilon_t]| &\leq |x_t(\beta_0 - \beta^{**}) + (x_t - \gamma_0)\delta_0 - (x_t - \gamma^{**})\delta^{**} + z_t'(\beta_{30} - \beta_3^{**})| \\ &+ |h_0(v_t) - h^{**}(v_t)| \end{aligned}$$

depends on the distance between ϕ^{**} and ϕ_0.

Secondly, we check Condition 3.8 of Chen (2007), which is equivalently to show

$$\sup_{\phi^{**}\in\Phi_n, d(\phi^{**},\phi_0)\leq\delta} |\varepsilon_t^2(\phi^{**}) - \varepsilon_t^2| \leq constant \times \delta^s \|w_t\|, \tag{A.18}$$

for any small $0 < \delta < 1$ and some $s \in (0, 2)$. Using the results above we have

$$\begin{aligned} |\varepsilon_t(\phi^{**})^2 - \varepsilon_t^2| &= [\varepsilon_t(\phi^{**}) - \varepsilon_t]^2 + 2|\varepsilon_t[\varepsilon_t(\phi^{**}) - \varepsilon_t]| \\ &\leq constant_1 \times \delta^2 \times \|w_t\| + constant_2 \times \delta \times \|w_t\| \\ &\leq constant \times \delta \times \|w_t\|, \end{aligned}$$

which verifies (A.18) with $s = 1$.

Now, we are in a position to obtain the convergence rate, which equals $O_p(\max\left\{\delta_n, \vartheta_{2n}^{-\eta}\right\})$ in our case following Theorem 3.2 of Chen (2007). Next, we solve δ_n. Specifically, by the definition of Condition A.3 of Chen and Shen (1998), δ_n solves the optimizing inequality of the metric entropy with bracket, where

$$\delta_n = \sup\left\{\delta > 0 : \delta^{-2}\int_{b\delta^2}^{\delta}\sqrt{H_{[\ ]}(\omega_1, G_n, \|\cdot\|)}d\omega_1\right\},$$

where $H_{[\ ]}(\omega_1, G_n, \|\cdot\|)$ denotes the metric entropy with bracketing and $G_n = \{\varepsilon_t(\phi^{***})^2 - \varepsilon_t^2 : d(\phi^{***}, \phi_0) \leq \omega_1, \phi^{***} \in \Phi_n\}$, for any given number $w_1 > 0$.

Let $C_u = \sqrt{E\|w_t\|}$, for all $0 < \omega_1/C_u \leq \delta < 1$, we have $H_{[\ ]}(\omega_1, G_n, \|\cdot\|) \leq \log N(\omega_1/C_u, B\times\Gamma, \|\cdot\|_2) + \log N(\omega_1/C_u, \mathcal{H}_n, \|\cdot\|_\infty)$. Note the first part is the L_2 metric entropy of the parametric part which equals $|\log(\omega_1/C_u)|$ following Hansen (2017). For the second part, by Lemma 2.5 in Geer (2000), we have the inequality, $\log N(\omega_1/C_u, \mathcal{H}_n, \|\cdot\|_\infty) \leq constant \times (1 + d_1)\vartheta_{2n} \times log(1 + 4/\omega_1)$. Note the first part is bounded by the second part for any $\omega_1/C_u > 0$ and some sufficiently large ϑ_{2n}.

Next, following the proof of Proposition 3.3 in Chen (2007) and solving

$$\begin{aligned}
&\frac{1}{\sqrt{n}\delta_n^2}\int_{b\delta_n^2}^{\delta_n}\sqrt{H_{[\ ]}(\omega_1, G_n, \|\cdot\|)}d\omega_1 \\
&\leq \frac{1}{\sqrt{n}\delta_n^2}\int_{b\delta_n^2}^{\delta_n}\sqrt{logN(\omega_1/C_u, B\times\Gamma, \|\cdot\|_2) + logN(\omega_1/C_u, \mathcal{H}_n, \|\cdot\|_\infty)}d\omega_1 \\
&\leq \frac{1}{\sqrt{n}\delta_n^2}\int_{b\delta_n^2}^{\delta_n}\sqrt{|log(\omega_1/C_u)| + constant \times (1 + d_1)\vartheta_{2n} \times log(1 + 4/\omega_1)}d\omega_1 \\
&\leq constant_1 + \frac{1}{\sqrt{n}\delta_n^2}\int_{b\delta_n^2}^{\delta_n}\sqrt{constant \times (1 + d_1)\vartheta_{2n} \times log(1 + 4/\omega_1)}d\omega_1 \\
&\leq constant_2 + constant_3 \times \frac{1}{\sqrt{n}\delta_n^2}\sqrt{\vartheta_{2n}} \times \delta_n \leq constant_4, \qquad \text{(A.19)}
\end{aligned}$$

gives $\delta_n \asymp \sqrt{\vartheta_{2n}/n}$, where $\asymp$ indicates "asymptotic equivalent". Following Theorem 3.2 of Chen (2007), we obtain the convergence rate of our proposed estimator, which equals $\max\left\{\sqrt{\vartheta_{2n}/n}, \vartheta_{2n}^{-\eta}\right\}$.

B Proof of Theorem 2-time series

In this section, we establish the asymptotic normality of our sieve estimator by verifying the conditions (2.1)-(2.6) as outlined in Theorem 2 of Chen et al. (2003), building upon the consistency results of our proposed estimator[4]. Given the proofs above and under Assumptions T1.1(b) and T2.4(c), it is readily seen that Conditions (2.1)-(2.3) hold. Condition (2.4) is not explicitly required here as in our case h enters m_t linearly[5]. Below, we only need to check Conditions (2.5) and (2.6) in details.

First, we check Condition (2.5)′ in Chen et al. (2003) since Chen et al. (2003) in Remark (ii) states that Condition (2.5)′ is a sufficient condition for Condition (2.5). That is, we need to check the following stochastic equicontinuity condition:

$$\sup_{\phi,\tilde{\phi}\in\Phi,d(\phi,\tilde{\phi})\leq\tilde{\delta}} \left\| 1/n\sum_{t=1}^{n} m_t(\theta,h) - E[m_t(\theta,h)] - 1/n\sum_{t=1}^{n} m_t(\theta_0,h_0) \right\| = o_p(n^{-1/2}), \quad \text{(B.1)}$$

with $0 < \tilde{\delta} < 1$. To show that, we apply the results of Lemma 4.2 in Chen (2007), which gives a sufficient condition of establishing Condition (2.5)′ of Chen et al. (2003) for a β-mixing sequence[6]. Next, we apply the results of Lemma 4.2 of Chen (2007) by establishing Conditions (4.2.1)-(4.2.3) in that paper.

Specifically, we show Condition (4.2.1) of Theorem 4.2 in Chen (2007), which in our case requires $1/n\sum_{t=1}^{n} m_t(\theta,h)$ to be locally uniformly L_2 continuous with respect to (θ,h). To prove Condition (4.2.1) of Chen (2007), it is suffice to show $E\left\|m_t(\theta,h) - m_t(\tilde{\theta},\tilde{h})\right\|^2 \leq constant\times\tilde{\delta}^2$. In our case, under Assumptions T1.1(b)(finite moment conditions), T2.3(b)(unknown functions are squared integrable) and T3.2(a)(compact parameter space), we use C-R in-

[4]Note although the paper Chen et al. (2003) focus on an i.i.d. sequence, her Theorem 2 also works with β-mixing data

[5]Chen et al. (2003) Remark 2(iii)

[6]also see Theorem 3 of Chen et al. (2003) for a similar analysis for an $i.i.d$ sequence.

equality and obtain

$$
\begin{aligned}
& E\left\|m_t(\theta,h)-m_t(\tilde{\theta},\tilde{h})\right\|^2 \\
&= E\left\|[m_t(\theta,h)-m_t(\theta,\tilde{h})]+[m_t(\theta,\tilde{h})-m_t(\tilde{\theta},\tilde{h})]\right\|^2 \\
&\leq 2E\left\|H_t(\theta)[\varepsilon_t(\theta,h)-\varepsilon_t(\tilde{\theta},\tilde{h})]\right\|^2+2E\left\|[H_t(\theta)-H_t(\tilde{\theta})]\varepsilon_t(\theta,\tilde{h})+H_t(\tilde{\theta})[\varepsilon_t(\theta,\tilde{h})-\varepsilon_t(\tilde{\theta},\tilde{h})]\right\|^2 \\
&\leq constant_1\times E\left\{\left\|H_t(\theta)(\tilde{h}-h)\right\|^2+E\left\|w_{1,t}(\theta-\tilde{\theta})\right\|^2\right\} \\
&+constant_2\times E\left\{\left\|w_{1,t}\times\tilde{h}\times(\theta-\tilde{\theta})\right\|^2+E\left\|w_{1,t}(\theta-\tilde{\theta})\right\|^2\right\} \\
&\leq constant_3\times E\left\|w_{1,t}\right\|^2\times\tilde{\delta}^2, \qquad \text{(B.2)}
\end{aligned}
$$

where $w_{1,t}$ equals w_t by removing $h_0(v_t)$ in w_t with w_t being defined in (A.11), and the last step holds by simply applying the Hölder inequality. Note $E\left\|w_{1,t}\right\|^2$ is bounded under Assumption T1.1(b), as it requires a moment condition up to order 4. Thus, (B.2) can be written as $constant\times\tilde{\delta}^2$, and (B.1) holds which implies Condition (4.2.1) of Lemma 4.2 Chen (2007). Note their Condition (4.2.2) holds as h belongs to a subset of Hölder functional space with $\eta>(1+d_1)/2$, and Condition (4.2.3) is our Assumption T1.1. Thus, we verify Conditions (4.2.1)-(4.2.3) of Lemma 4.2 in Chen (2007).

Second, we prove Condition(2.6) of Chen et al. (2003) by applying the CLT for a β-mixing sequence. To show that, it is sufficient to apply the result of Lemma 5.1 in Newey (1994), which gives the asymptotic normality of $1/\sqrt{n}\sum_{t=1}^n m_t(\theta_0,\hat{h})$. As in our case $h(\cdot)$ enters $Em_t(\theta,h)$ linearly, the proof can be greatly simplified. Under the linearization property, his Conditions (5.1) and (5.3) are satisfied in our case, and we only need to establish his Condition (5.2), which in our case only requires $H_t(\theta_0)[\hat{h}(v_t)-h_0(v_t)]$ to be stochastic equicontinuity, we need to show

$$
\frac{1}{\sqrt{n}}\sum_{t=1}^n\left\{H_t(\theta_0)[\hat{h}(v_t)-h_0(v_t)]-E\left\{H_t(\theta_0)[\hat{h}(v_t)-h_0(v_t)]\right\}\right\}\xrightarrow{p}0. \qquad \text{(B.3)}
$$

Given $H_t(\theta_0)$ to be bounded under Assumption T1.1(b), (B.3) is satisfied by applying the stochastic equicontinuity result in (B.1). Now, we show that all the conditions of Lemma

5.1 in Newey and Mcfadden (1994) hold in our case.

C Proof of Theorem 1-panel

C.1 Consistency

Similar to the proof of Theorem 1-time Series, to establish the consistency of our proposed estimate, we apply the results from Theorem 3.1 in Chen (2007).

To apply Chen's results, we check the Conditions 3.1- 3.5 listed in Theorem 3.1 of her paper. Notably, Condition 3.1 aligns with our Assumption P3.2(b), which presupposes ϕ_0 as the unique minimizer of our objective function. Condition 3.2 corresponds to our Assumption P2.3, which posits the existence of an appropriate sieve approximation for our unknown functions, denoted as $h_0(\cdot)$. Condition 3.3 is satisfied due to the continuity property of the KTR model. Condition 3.4 is met through the assumption of the compactness of the sieve space, in accordance with our Assumption P3.2(a). In summary, to apply Theorem 3.1 from Chen (2007), our task is to demonstrate Condition 3.5, namely, the uniform convergence of the objective function over the sieve space, which is the same as in a time series KTR model. We will elucidate this in the following steps.

Denote $\phi^* = (\theta, h^*) = (\beta, \gamma, h^*)$, where $\phi^* \in \Phi_N$, $\hat{S}_N(\phi^*) = 1/N \sum_{i=1}^N \sum_{t=t_0}^T \Delta\hat{\varepsilon}_{i,t}(\phi^*)^2$, where $\Delta\hat{\varepsilon}_{i,t}(\phi^*)$ equals $\Delta\hat{\varepsilon}_{i,t}(\hat{\phi}^*_N)$ defined in Remark under Theorem 2-panel by replacing $\hat{\phi}^*_N$ with ϕ^*; $S_N(\phi^*) = 1/N \sum_{i=1}^N \sum_{t=t_0}^T \Delta\varepsilon_{i,t}(\phi^*)^2$, where $\Delta\varepsilon_{i,t}(\phi^*)$ equals $\Delta\hat{\varepsilon}_{i,t}(\phi^*)$ by replacing $\hat{v}_t$ by v_t. To establish Condition 3.5 of Chen (2007), we need to show

$$plim_{N\to\infty} \sup_{\phi^*\in\Phi_n} |\hat{S}_N(\phi^*) - E[S_N(\phi^*)]| = 0, . \tag{C.1}$$

Firstly, by simple calculation, we decompose (C.1) and get

$$\begin{aligned} \sup_{\phi^*\in\Phi_n} |\hat{S}_N(\phi^*) - E[S^*_N(\phi^*)]| &\leq \sup_{\phi^*\in\Phi_n} |\hat{S}_N(\phi^*) - S^*_N(\phi^*)| + \sup_{\phi^*\in\Phi_n} |S^*_N(\phi^*) - E[S^*_N(\phi^*)]| \\ &= P_1 + P_2. \end{aligned} \tag{C.2}$$

Next, we prove P_1 and P_2 to be $o_p(1)$, respectively.

(P_1:) for all $\phi^* \in \Phi_n$, we have

$$
\begin{aligned}
& \sup_{\phi^* \in \Phi_n} \left| \hat{S}_N(\phi^*) - S_N(\phi^*) \right| \\
&= \sup_{\phi^* \in \Phi_n} \left| \frac{1}{N} \sum_{i=1}^{N} \sum_{t=t_0}^{T} \left[\Delta\hat{\varepsilon}_{i,t}(\phi^*)^2 - \Delta\varepsilon_{i,t}(\phi^*)^2 \right] \right| \\
&= \sup_{\phi^* \in \Phi_n} \left| \frac{1}{N} \sum_{i=1}^{N} \sum_{t=t_0}^{T} [\Delta\hat{\varepsilon}_{i,t}(\phi^*) - \Delta\varepsilon_{i,t}(\phi^*)]^2 \right| + \sup_{\phi^* \in \Phi_n} \left| \frac{2}{N} \sum_{i=1}^{N} \sum_{t=t_0}^{T} \Delta\varepsilon_{i,t}(\phi^*) \left[\Delta\hat{\varepsilon}_{i,t}(\phi^*) - \Delta\varepsilon_{i,t}(\phi^*) \right] \right| \\
&\leq \sup_{\beta_h^* \in B_h} \left| \frac{1}{N} \sum_{i=1}^{N} \sum_{t=t_0}^{T} \{ [\Psi_{\vartheta_{2N}}(v_{i,t}) - \Psi_{\vartheta_{2N}}(\hat{v}_{i,t})]'\beta_h - [\Psi_{\vartheta_{2N}}(v_{i,t-1}) - \Psi_{\vartheta_{2N}}(\hat{v}_{i,t-1})]'\beta_h \}^2 \right| \\
&+ \sup_{\phi^* \in \Phi_n} \left| \frac{2}{N} \sum_{i=1}^{N} \sum_{t=t_0}^{T} \Delta\varepsilon_{i,t}(\phi^*) \left\{ [\Psi_{\vartheta_{2N}}(v_{i,t}) - \Psi_{\vartheta_{2N}}(\hat{v}_{i,t})]'\beta_h - [\Psi_{\vartheta_{2N}}(v_{i,t-1}) - \Psi_{\vartheta_{2N}}(\hat{v}_{i,t-1})]'\beta_h \right\} \right| \\
&= D_1 + 2D_2. \qquad \text{(C.3)}
\end{aligned}
$$

Given

$$
\begin{aligned}
D_1 \quad \leq \quad & \sup_{\beta_h \in B_h} \left| \frac{1}{N} \sum_{i=1}^{N} \sum_{t=t_0}^{T} \{ [\Psi_{\vartheta_{2N}}(v_{i,t}) - \Psi_{\vartheta_{2N}}(\hat{v}_{i,t})]'\beta_h \}^2 \right| \\
& + \sup_{\beta_h^* \in B_h} \left| \frac{1}{N} \sum_{i=1}^{N} \sum_{t=t_0}^{T} \{ [\Psi_{\vartheta_{2N}}(v_{i,t-1}) - \Psi_{\vartheta_{2N}}(\hat{v}_{i,t-1})]'\beta_h \}^2 \right|,
\end{aligned}
$$

which each part yields a similar structure as A_1 in Theorem 1-time series, thus we can directly apply the results of (A.4) and obtain $D_1 = O_p \left[\|\Psi_{\vartheta_{2N}}\|_1^2 \, (\vartheta_{1N}^{-2\eta} + \vartheta_{1N}/N)\vartheta_{2N} \right]$.

Similarly, for D_2, we apply the results of A_2 (eq. A.7)) in the times series part, we have

$$
\begin{aligned}
D_2 \quad \leq \quad & \sup_{\phi^* \in \Phi_n} \left| \frac{2}{N} \sum_{i=1}^{N} \sum_{t=t_0}^{T} \Delta\varepsilon_{i,t}(\phi^*)[\Psi_{\vartheta_{2N}}(v_{i,t}) - \Psi_{\vartheta_{2N}}(\hat{v}_{i,t})]'\beta_h \right| \\
& + \sup_{\phi^* \in \Phi_n} \left| \frac{2}{N} \sum_{i=1}^{N} \sum_{t=t_0}^{T} \Delta\varepsilon_{i,t}(\phi^*)[\Psi_{\vartheta_{2N}}(v_{i,t-1}) - \Psi_{\vartheta_{2N}}(\hat{v}_{i,t-1})]'\beta_h \right| \\
= \quad & O_p \left[\|\nabla\Psi_{\vartheta_{2N}}\|_1 \, (\vartheta_{1N}^{-\eta} + \sqrt{\vartheta_{1N}/N})\sqrt{\vartheta_{2N}} \right], \qquad \text{(C.4)}
\end{aligned}
$$

given $1/N \sum_{i=1}^{N} \sum_{t=t_0}^{T} \Delta\varepsilon_{i,t}^2 \leq \infty$, under Assumptions P1.1 and P2.3(implies $E[\Delta\varepsilon_{i,t}(\phi^*)]$ is bounded over all $\phi^* \in \Phi_n$) and the Uniform Law of Large Numbers(ULLN).

(P_2 :) Note that P_2 yields a uniform convergence condition of our objective function. To establish P_2, similar to the time series model, we rely on the conclusion of Corollary 2.2 in Newey (1991). Again, we check three conditions required by the Corollary, (1) the compactness of the parameter space; (2) the point-wise convergence of the objective function $S_N(\phi^*)$, and (3) the stochastic equicontinuity of $S_N(\phi^*)$. Note the compactness of Φ_n is given by our Assumption P3.2(a). We only need to show Conditions (2) and (3).

First, we prove the point-wise convergence(Condition (2)). With a fixed time period and an i.i.d. assumption over i, the point-wise convergence holds by directly applying the WLLN for the i.i.d. sequence.

Then, based on the point-wise convergence, in our case, we show the uniform convergence (Condition (3)). To do that, it is sufficient to apply the results of Condition 3A of Newey (1991). Next, we show that Condition 3A holds in our case. Viewing $S_N(\phi)$ as the average of the sum squared differences between error term in time t and $t-1$, each part has a similar structure as in the time series model, thus the Condition 3A of Newey (1991) also holds here and we obtain $P_2 = o_p(1)$.

Given P_1 and P_2 both $o_p(1)$, we establish the prove the uniform convergence results (C.2), which implies (C.1). Then, we finish proving all conditions required by Theorem 3.1 of Chen (2007) and obtain the consistency of our proposed estimator, as $N \to \infty$, $d(\hat{\phi}^*_N, \phi_0) = o_p(1)$.

C.2 Convergence rate

In this section, we establish the convergence rate of our proposed sieve estimate in the panel KTR model. Similar to the time-series KTR model, we establish the convergence by applying the result of Theorem 3.2 in Chen (2007). To do that, we need to check Conditions (3.7) and (3.8) in that theorem.

First, we check Condition 3.7 of theorem 3.2 in Chen (2007). In our case, it is equivalent to verify

$$\sup_{\phi^{**}\in\Phi_n, d(\phi^{**},\phi)\leq\omega} Var\left(\Delta\varepsilon_{i,t}(\phi^{**})^2 - \Delta\varepsilon_{i,t}^2\right), \tag{C.5}$$

for any small $0 < \omega < 1$. By definition and $\Delta\varepsilon_{i,t}(\phi^{**})^2 - \Delta\varepsilon_{i,t}^2 = 2\Delta\varepsilon_{i,t}[\Delta\varepsilon_{i,t}(\phi^{**}) - \Delta\varepsilon_{i,t}(\phi)] + [\Delta\varepsilon_{i,t}(\phi^{**}) - \varepsilon_t]^2$, we have

$$\begin{aligned}
Var\left(\Delta\varepsilon_{i,t}(\phi^{**})^2 - \Delta\varepsilon_{i,t}^2\right) &\leq E\left\{[\Delta\varepsilon_{i,t}(\phi^{**})^2 - \Delta\varepsilon_{i,t}^2]^2\right\} \\
&\leq 8E\left\{\Delta\varepsilon_{i,t}^2[\Delta\varepsilon_{i,t}(\phi^{**}) - \Delta\varepsilon_t]^2\right\} + E\left\{[\Delta\varepsilon_{i,t}(\phi^{**}) - \Delta\varepsilon_{i,t}]^4\right\} \\
&\leq constant \times E\left\{[\Delta\varepsilon_{i,t}(\phi^{**}) - \Delta\varepsilon_{i,t}]^2\right\} + 2\left\{[\Delta\varepsilon_{i,t}(\phi^{**}) - \Delta\varepsilon_{i,t}]^4\right\} \\
&constant_1 \times \omega^2 + constant_2 \times \omega^4, \qquad \text{(C.6)}
\end{aligned}$$

since $E[\Delta\varepsilon_{i,t}(\Delta\varepsilon_{i,t}(\phi^{**}) - \Delta\varepsilon_{i,t})] = E[E[\Delta\varepsilon_{i,t}^2|\mathcal{F}_{i,t-2}, x_{i,t}, x_{i,t-1}, z_{1,i,t}, z_{1,i,t-1}, p_{i,t}, p_{i,t-1}](\Delta\varepsilon_{i,t}^2(\phi^{**}) - \Delta\varepsilon_{i,t})^2] \leq constant \times E[(\Delta\varepsilon_{i,t}(\phi^{**}) - \Delta\varepsilon_{i,t})^2]$ under Assumptions P1.1(b)(c) and P2.3(b), and

$$\begin{aligned}
E\left|\varepsilon_{i,t}(\phi^*) - \Delta\varepsilon_{i,t}\right| \;\leq\;& E\left|\Delta x_{i,t}(\beta_0 - \beta)\right| \\
&+E\left|\delta_0(X_{i,t} - \tau_2\gamma_0)\mathbf{I}_{i,t}(\gamma_0) - \delta(X_{i,t} - \tau_2\gamma)\mathbf{I}_{i,t}(\gamma)\right| \\
&+E\left\|\Delta z_{i,t}\right\|\left\|\beta_{30} - \beta_3\right\| + E\left|g_{v0}(v_{i,t}) - g_v^*(v_{i,t})\right| \\
&+2E\left|g_v^*(v_{i,t-1}) - g_{v0}(v_{i,t-1})\right|,
\end{aligned}$$

depends on the distance between ϕ^{**} and ϕ_0.

Next, we show Condition 3.8 of Chen (2007), which in our case is equivalent to show

$$\sup_{\phi^{**}\in\Phi_n, d(\phi^{**},\phi_0)\leq\delta} |\Delta\varepsilon_{i,t}(\phi^{**})^2 - \Delta\varepsilon_t^2| \leq constant \times \delta^s \left\|w_{i,t}\right\|, \tag{C.7}$$

for any small $0 < \delta < 1$ and some $s \in (0, 2)$. Using the results above we obtain

$$\begin{aligned}
|\Delta\varepsilon_{i,t}(\phi^{**})^2 - \Delta\varepsilon_{i,t}^2| &= [\Delta\varepsilon_{i,t}(\phi^{**}) - \Delta\varepsilon_{i,t}]^2 + 2|\Delta\varepsilon_{i,t}[\Delta\varepsilon_{i,t}(\phi^{**}) - \Delta\varepsilon_{i,t}]| \\
&\leq constant_1 \times \delta^2 \times \left\|w_{i,t}\right\| + constant_2 \times \delta \times \left\|w_{i,t}\right\| \\
&\leq constant \times \delta \times \left\|w_{i,t}\right\| \qquad \text{(C.8)}
\end{aligned}$$

which verifies (C.7) with $s = 1$.

Last, we calculate the convergence rate. Note that $\Delta\varepsilon_{i,t}(\phi) = \Delta u_{i,t}(\theta) - h_0(v_{i,t}, v_{i,t-1}) = [u_{i,t} - g_{v0}(v_{i,t})] - [u_{i,t-1} - g_{v0}(v_{i,t-1})]$. Similar as in the time-series model, denote $\mathcal{G}_N = \left\{\Delta\varepsilon_{i,t}(\phi^{***})^2 - \Delta\varepsilon_{i,t}^2 : d(\phi^{***}, \phi_0) \le \omega_1, \phi^{***} \in \Phi_N\right\}$, and $\mathcal{G}_N$ has a metric entropy with bracketing $H_{[\,]}(\omega_1, \mathcal{G}_N, \|\cdot\|)$. It is clear that $H_{[\,]}(\omega_1, \mathcal{G}_N, \|\cdot\|)$ can be decomposed as a summation which each part has the same structure as in the time series model (A.2), thus we can directly apply that result and obtain the convergence rate of our proposed estimator, which equals $\max\left\{\sqrt{\vartheta_{2N}/N}, \vartheta_{2N}^{-\eta}\right\}$.

D Proof of Theorem 2-panel

In this section, we establish the asymptotic normality of our sieve estimator by verifying the conditions (2.1)-(2.6) as outlined in Theorem 2 of Chen et al. (2003), building upon the consistency results of our proposed estimator[7]. Given the proofs above and under Assumptions P1.1(b) and P2.4(c), it is readily seen that Conditions (2.1)-(2.3) hold. Condition (2.4) is not explicitly required here as in our case h enters $m_{i,t}$ linearly[8]. Similar as in the time series model, below, we only need to check Conditions (2.5) and (2.6) in details.

To show Condition 2.5(2.5′) of Chen et al. (2003) holds in our case with a β-mixing sequence, we need to check following stochastic equicontinuity condition:

$$\sup_{\phi,\tilde{\phi}\in\Phi, d(\phi,\tilde{\phi})\le\tilde{\delta}} \left\| 1/N \sum_{i=1}^{N}\sum_{t=t_0}^{T} m_{i,t}(\theta, h) - E[\sum_{t=t_0}^{T} m_{i,t}(\theta,h)] - 1/N\sum_{i=1}^{N}\sum_{t=t_0}^{T} m_{i,t}(\theta_0, h_0)\right\| = o_p(N^{-1/2}), \tag{D.1}$$

note (D.1) has a similar structure ‘as (B.1) in the time series part. Again, we apply the

[7]Note although the paper Chen et al. (2003) focus on an *i.i.d.* sequence, her Theorem 2 also works with β-mixing data

[8]Chen et al. (2003) Remark 2(iii)

result of Lemma 4.2 in Chen (2007), which holds under the Conditions (4.2.1)-(4.2.3).

In our case, Condition (4.2.1) is satisfied given $1/N\sum_{i=1}^{N}\sum_{t=t_0}^{T}m_{i,t}(\phi)$ to be locally uniformly L_2 continuous with respect to (θ,h), which it is suffice to show $E\left\|m_{i,t}(\theta,h)-m_{i,t}(\tilde{\theta},\tilde{h})\right\|^2 \le constant \times \tilde{\delta}^2$. Recall the expression of $H_{i,t}(\theta)$ and $\Delta\varepsilon_{i,t}(\phi)$, we have

$$\begin{aligned}
&E\left\|m_{i,t}(\theta,h)-m_{i,t}(\tilde{\theta},\tilde{h})\right\|^2\\
&=E\left\|[m_{i,t}(\theta,h)-m_{i,t}(\theta,\tilde{h})]+[m_{i,t}(\theta,\tilde{h})-m_{i,t}(\tilde{\theta},\tilde{h})]\right\|^2\\
&\le 2E\left\|H_{i,t}(\theta)[\Delta\varepsilon_{i,t}(\theta,h)-\Delta\varepsilon_{i,t}(\tilde{\theta},\tilde{h})]\right\|^2\\
&+2E\left\|[H_{i,t}(\theta)-H_{i,t}(\tilde{\theta})]\Delta\varepsilon_{i,t}(\theta,\tilde{h})+H_{i,t}(\tilde{\theta})[\Delta\varepsilon_{i,t}(\theta,\tilde{h})-\Delta\varepsilon_{i,t}(\tilde{\theta},\tilde{h})]\right\|^2\\
&\le constant_1\times\left\{E\left\|H_{i,t}(\theta)(\tilde{h}-h)\right\|^2+E\left\|w_{i,t}(\theta-\tilde{\theta})\right\|^2\right\}\\
&+constant_2\times\left\{E\left\|w_{i,t}\times\tilde{h}(\theta-\tilde{\theta})\right\|^2+E\left\|w_{i,t}(\theta-\tilde{\theta})\right\|^2\right\}\\
&\le constant\times E\left\|w_{i,t}\right\|^2\times\tilde{\delta}^2, \qquad\qquad (D.2)
\end{aligned}$$

the results is similar as (B.2), where we use Assumption P1.1(b)(finite moment conditions), P2.3(b)(unknown functions are squared integrable) and P3.2(compact parameter space). Note their Condition (4.2.2) holds as h belongs to a subset of Hölder functional space with $\eta > (1+d_1)/2$, and Condition (4.2.3) is our Assumption P1.1. Thus, we verify Conditions (4.2.1)-(4.2.3) of Lemma 4.2 in Chen (2007).

Next, we establish Condition (2.6) of Chen et al. (2003), which we need to prove the asymptotic normality of $1/\sqrt{N}\sum_{i=1}^{N}\sum_{t=t_0}^{T}m_{i,t}(\theta_0,\hat{h})$. Here we apply the CLT following the results of Lemma 5.1 in Newey and Mcfadden (1994) by checking his Conditions (5.1)-(5.3). Again, as $h(v_{i,t},v_{i,t-1})$ enters $E[m_{i,t}(\theta,h)]$ linearly, his Conditions (5.1)(5.3) is automatically satisfied. Thus, we only need to show his Condition (5.2). In our case, that requires $H_{i,t}(\theta_0)[\hat{h}(v_{i,t},v_{i,t-1})]$ to be stochasitic equicontinuity, and it directly follows the result of (D.1).

E Lemma

Lemma 1. *Denote* $\Omega_{n,xx} = \frac{1}{n}\sum_{t=1}^{n} \Psi_{\vartheta_{1n}}(p_{xt})\Psi_{\vartheta_{1n}}(p_{xt})'$, $\Omega_{n,zz} = \frac{1}{n}\sum_{t=1}^{n} \Psi_{\vartheta_{1n}}(p_{zt})\Psi_{\vartheta_{1n}}(p_{zt})'$ *and* $\Omega_{n,vv} = \frac{1}{n}\sum_{t=1}^{n} \Psi_{\vartheta_{2n}}(v_t)\Psi_{\vartheta_{2n}}(v_t)'$. *Under Assumptions T1.1, T2.2,T2.4 and T4 we have*

(a)

$$E\left\|\Omega_{n,xx} - E(\Omega_{n,xx})\right\| = O(\vartheta_{1n}/\sqrt{n}), \tag{E.1}$$

$$E\left\|\Omega_{n,zz} - E(\Omega_{n,zz})\right\| = O(\vartheta_{1n}/\sqrt{n}). \tag{E.2}$$

(b)

$$\lambda_{\min}(\Omega_{n,xx}) = \lambda_{\min}E(\Omega_{n,xx}) + o_p(1), \quad \lambda_{\max}(\Omega_{n,xx}) = \lambda_{\max}E(\Omega_{n,xx}) + o_p(1), \tag{E.3}$$

$$\lambda_{\min}(\Omega_{n,zz}) = \lambda_{\min}E(\Omega_{n,zz}) + o_p(1), \quad \lambda_{\max}(\Omega_{n,zz}) = \lambda_{\max}E(\Omega_{n,zz}) + o_p(1), \tag{E.4}$$

(c)

$$\left\|\Omega_{n,xx}^{-1} - E(\Omega_{n,xx})^{-1}\right\|_{sp} = O_p(\vartheta_{1n}/\sqrt{n}), \tag{E.5}$$

$$\left\|\Omega_{n,zz}^{-1} - E(\Omega_{n,zz})^{-1}\right\|_{sp} = O_p(\vartheta_{1n}/\sqrt{n}).. \tag{E.6}$$

Proof. To save space, here we only prove (E.1), (E.3) and (E.5), other parts in Lemma 1 hold following similar process.

(a) First we prove (E.1), we need to show that

$$E\left\|\Omega_{n,xx} - E(\Omega_{n,xx})\right\|^2 = O(\vartheta_{1n}^2/n). \tag{E.7}$$

Note that the i-th row, j-th block element of matrix $\Omega_{n,xx}$ equals $\frac{1}{n}\sum_{t=1}^{n} \Psi_i(p_{xt})\Psi_j(p_{xt})'$, we denote $\Psi_i(p_{xt})\Psi_j(p_{xt})'$ as $\Omega_{xx}^{ij,t}$. Similarly, denote $E(\Omega_{xx}^{ij,t})$ as the i-th row, j-th element of matrix $E(\Omega_{n,xx})$. Following the definition of matrix norm, we can decompose (E.7) as

$$
\begin{aligned}
& E\left\|\Omega_{n,xx}-E(\Omega_{n,xx})\right\|^2 \\
& =\sum_{i=1}^{\vartheta_{1n}}\sum_{j=1}^{\vartheta_{1n}} E\Big(\frac{1}{n}\sum_{t=1}^{n}\Omega_{xx}^{ij,t}-E(\Omega_{xx}^{ij,t})\Big)^2 \\
& +\frac{2}{n}\sum_{i=1}^{\vartheta_{1n}}\sum_{j=1}^{\vartheta_{1n}}\sum_{\tau=1}^{n-1}(1-\frac{\tau}{n})Cov(\Omega_{xx}^{ij,1},\Omega_{xx}^{ij,1+\tau}) \\
& = L_1+2L_2, \qquad \text{(E.8)}
\end{aligned}
$$

where we get L_2 by applying a stationary covariance results implied by the β-mixing. Note L_1 captures the correlation within every certain t and L_2 gives the time series dependence. Under Assumption T2.2 and by the Triangular inequality and Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
L_1 \quad & \leq \frac{2}{n}\sum_{t=1}^{n}\sum_{i=1}^{\vartheta_{1n}}\sum_{j=1}^{\vartheta_{1n}} E[\Omega_{xx}^{ij,t}]^2+\frac{2}{n}\sum_{i=1}^{\vartheta_{1n}}\sum_{j=1}^{\vartheta_{1n}}[E(\Omega_{xx}^{ij,t})]^2 \\
& \leq \frac{2}{n}\sum_{t=1}^{n}\sum_{i=1}^{\vartheta_{1n}}\sum_{j=1}^{\vartheta_{1n}}\left\{E[\Psi_i(p_{xt})]^4\right\}^{1/2}\left\{E[\Psi_j(p_{xt})]^4\right\}^{1/2}+\frac{2}{n}\sum_{i=1}^{\vartheta_{1n}}\sum_{j=1}^{\vartheta_{1n}} E[\Psi_i(p_{xt})]^2E[\Psi_j(p_{xt})]^2 \\
& = O(\vartheta_{1n}^2/n). \qquad \text{(E.9)}
\end{aligned}
$$

For L_2, we can apply the Davydov inequality for a β-mixing process with mixing coefficients $\alpha(m)$, under Assumption T1.1, we have

$$
\begin{aligned}
L_2 \quad & =\frac{1}{n}\sum_{i=1}^{\vartheta_{1n}}\sum_{j=1}^{\vartheta_{1n}}\sum_{\tau=1}^{n-1}(1-\frac{\tau}{n})Cov(\Omega_{xx}^{ij,1},\Omega_{xx}^{ij,1+\tau}) \\
& \leq \frac{C}{n}\sum_{i=1}^{\vartheta_{1n}}\sum_{j=1}^{\vartheta_{1n}}\sum_{\tau=1}^{n-1}(1-\frac{\tau}{n})a(\tau)^{1-1/2r} \\
& \leq \frac{C\vartheta_{1n}^2}{n}\sum_{\tau=1}^{n-1}a(\tau)^{1-1/2r} \\
& = O(\vartheta_{1n}^2/n), \qquad \text{(E.10)}
\end{aligned}
$$

as $a(\tau)^{1-1/2r}$ is a stationary process under Assumption T1.1. Combing L_1 and L_2, we prove eq.(E.7) thus E.1.

(b) Then, (E.3) holds by applying the Weyl's inequality in Seber (2008), we have

$$\lambda_{\min}(E(\Omega_{n,xx}))+\lambda_{\min}(\Omega_{n,xx}-E(\Omega_{n,xx}))\le\lambda_{\min}(\Omega_{n,xx})\le\lambda_{\min}(E(\Omega_{n,xx}))+\lambda_{\max}(\Omega_{n,xx}-E(\Omega_{n,xx}))\tag{E.11}$$

Note that for a symmetric matrix $(\Omega_{n,xx}-E(\Omega_{n,xx}))$, we have

$\lambda_{\max}(\Omega_{n,xx}-E(\Omega_{n,xx}))=\|\Omega_{n,xx}-E(\Omega_{n,xx})\|_{sp}\le\|\Omega_{n,xx}-E(\Omega_{n,xx})\|$, and

$-\|\Omega_{n,xx}-E(\Omega_{n,xx})\|\le\lambda_{\min}[\Omega_{n,xx}-E(\Omega_{n,xx})]$.

Given that, and combining with the results we get in Lemma 1(a), we can rewrite (E.11) as

$$\begin{aligned}\lambda_{\min}(E(\Omega_{n,xx}))-\|\Omega_{n,xx}-E(\Omega_{n,xx})\| &\le\lambda_{\min}(\Omega_{n,xx})\\ &\le\lambda_{\min}(E(\Omega_{n,xx}))+\|\Omega_{n,xx}-E(\Omega_{n,xx})\| &\text{(E.12)}\\ \lambda_{\min}(E(\Omega_{n,xx}))-O_p(\vartheta_{1n}/\sqrt{n}) &\le\lambda_{\min}(\Omega_{n,xx})\\ &\le\lambda_{\min}(E(\Omega_{n,xx}))+O_p(\vartheta_{1n}/\sqrt{n}) &\text{(E.13)}\end{aligned}$$

Thus, under Assumption T4, we can prove that $\lambda_{\min}(\Omega_{n,xx})=\lambda_{\min}(E(\Omega_{n,xx}))+o_p(1)$. The second part of (E.3) and other parts in Lemma 1(b) can be proved following a similar process, we do not repeat it here.

(c) Applying the sub-multiplicative property of the spectral norm, we apply the result $\|\Omega_{n,xx}-E(\Omega_{n,xx})\|_{sp}\le\|\Omega_{n,xx}-E(\Omega_{n,xx})\|$ shown in Lemma1(b) and obtain

$$\begin{aligned}\left\|\Omega_{n,xx}^{-1}-E(\Omega_{n,xx})^{-1}\right\|_{sp} &=\left\|\Omega_{n,xx}^{-1}(\Omega_{n,xx}-E(\Omega_{n,xx}))E(\Omega_{n,xx})^{-1}\right\|_{sp}\\ &\le\left\|\Omega_{n,xx}^{-1}\right\|_{sp}\|\Omega_{n,xx}-E(\Omega_{n,xx})\|_{sp}\left\|E(\Omega_{n,xx})^{-1}\right\|_{sp}\\ &\le\left\|\Omega_{n,xx}^{-1}\right\|_{sp}\|\Omega_{n,xx}-E(\Omega_{n,xx})\|\left\|E(\Omega_{n,xx})^{-1}\right\|_{sp}\\ &=O_p(1)O_p(\vartheta_{1n}/\sqrt{n})O_p(1), &\text{(E.14)}\end{aligned}$$

where $\left\|\Omega_{n,xx}^{-1}\right\|_{sp}=[\lambda_{\min}(\Omega_{n,xx})]^{-1}=[\lambda_{\min}(E(\Omega_{n,xx}))+o_p(1)]^{-1}=O_p(1)$ under Assumption T2.4(a).

Lemma 2. *Under Assumption T2.4 and Lemma 1, we have*

$$\left\| \frac{1}{n} \sum_{t=1}^{n} \Psi_{\vartheta_{1n}}(p_{xt}) v_{xt} \right\|^2 = O_p(\vartheta_{1n}/n), \tag{E.15}$$

$$\left\| \frac{1}{n} \sum_{t=1}^{n} \Psi_{\vartheta_{1n}}(p_x) v_z^{k_2} \right\|^2 = O_p(\vartheta_{1n}/n), \quad \text{for} \quad k_2 = 1, ..., d_1. \tag{E.16}$$

Proof. Following the definition, under Assumption T2.4(b), we have

$$\begin{aligned}
\left\| \frac{1}{n} \sum_{t=1}^{n} \Psi_{\vartheta}(p_{xt}) v_{xt} \right\|^2 &= \frac{1}{n} tr \left\{ \frac{1}{n} \sum_{t=1}^{n} [\Psi_{\vartheta_{1n}}(p_{xt}) \Psi_{\vartheta_{1n}}(p_{xt})' v_{xt}^2] \right\} \\
&= \frac{\vartheta_{1n}}{n} \lambda_{max} \left\{ \frac{1}{n} \sum_{t=1}^{n} [\Psi_{\vartheta_{1n}}(p_{xt}) \Psi_{\vartheta_{1n}}(p_{xt})' v_{xt}^2] \right\} = O_p(\vartheta_{1n}/n)
\end{aligned} \tag{E.17}$$

where (E.17) holds if $\lambda_{max} \left\{ \frac{1}{n} \sum_{t=1}^{n} [\Psi_{\vartheta_{1n}}(p_{xt}) \Psi_{\vartheta_{1n}}(p_{xt})' v_{xt}^2] \right\} = \lambda_{max} \left\{ E[\Psi_{\vartheta_{1n}}(p_{xt}) \Psi_{\vartheta_{1n}}(p_{xt})' v_{xt}^2] \right\} + o_p(1)$.

The proof is similar as Lemma 1(b), it is suffice to show

$\left\| \left\{ \frac{1}{n} \sum_{t=1}^{n} [\Psi_{\vartheta_{1n}}(p_{xt}) \Psi_{\vartheta_{1n}}(p_{xt})' v_{xt}^2] \right\} - E[\Psi_{\vartheta_{1n}}(p_{xt}) \Psi_{\vartheta_{1n}}(p_{xt})' v_{xt}^2] \right\|^2 = o_p(1)$.

Following (E.8), (E.9) and (E.10) by replacing $\Omega_{n,xx}$ and $E(\Omega_{n,xx})$ by $\frac{1}{n} \sum_{t=1}^{n} [\Psi_{\vartheta_{1n}}(p_{xt}) \Psi_{\vartheta_{1n}}(p_{xt})' v_{xt}^2]$ and $E[\Psi_{\vartheta_{1n}}(p_{xt}) \Psi_{\vartheta_{1n}}(p_{xt})' v_{xt}^2]$, respectively. (E.16) can be proved following the same process.

Lemma 3. *Under Assumptions T2.2, T2.3, T2.4 and Lemma 1, 2, we have*

$$\frac{1}{n} \sum_{t=1}^{n} \|\hat{v}_t - v_t\|^2 = O_p(\vartheta_{1n}^{-2\eta} + \vartheta_{1n}/n). \tag{E.18}$$

Proof. To simplify our analysis, we establish the convergence of v_{xt}, and then, the convergence of other parts of $\hat{v}_t$ hold following a similar routine. Recall the definition of v_{xt} and

$\hat{v}_{xt}$ in our main text, applying the Triangular inequality, we have

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}(\hat{v}_{xt}-v_{xt})^2 &= \frac{1}{n}\sum_{t=1}^{n}[g_{x0}(p_{xt})-\hat{g}_x^*(p_{xt})]^2 \\ &\leq \frac{2}{n}\sum_{t=1}^{n}[g_{x0}(p_{xt})-\Psi_{\vartheta_{1n}}(p_{xt})'\beta_{x0}]^2+\frac{2}{n}\sum_{t=1}^{n}[\Psi'_{\vartheta_{1n}}(\beta_{x0}-\hat{\beta}_x)]^2 \\ &= O_p(\vartheta_{1n}^{-2\eta})+O_p\left\|\beta_{x0}-\hat{\beta}_x\right\|^2, \qquad \text{(E.19)}\end{aligned}$$

under Assumption T2.3, and $\frac{1}{n}\sum_{t=1}^{n}\|\Psi_{\vartheta_{1n}}(p_{xt})\|^2 = tr\left\{\frac{1}{n}\sum_{t=1}^{n}[\Psi_{\vartheta_{1n}}(p_{xt})\Psi_{\vartheta_{1n}}(p_{xt})']\right\} = O_p(1)$(Assumption T4.2(a)). The last part is $\left\|\hat{\beta}_x-\hat{\beta}_{x0}\right\|$. Following the expression of the control function, we have

$$\begin{aligned}&\left\|\hat{\beta}_x-\beta_{x0}\right\| \leq \left\|\left[\frac{1}{n}\sum_{t=1}^{n}\Psi_{\vartheta_{1n}}(p_{xt})\Psi_{\vartheta_{1n}}(p_{xt})\right]^{-1}\frac{1}{n}\sum_{t=1}^{n}\Psi_{\vartheta_{1n}}(p_{xt})[x_t-\Psi_{\vartheta_{1n}}(p_{xt})'\beta_{x0}]\right\| \\ &\leq \frac{1}{\lambda_{min}(\Omega_{n,xx})}\left\|\frac{1}{n}\sum_{t=1}^{n}\Psi_{\vartheta_{1n}}(p_{xt})\left[v_{xt}+g_{x0}(p_{xt})-\Psi_{\vartheta_{1n}}(p_{xt})'\beta_{x0}\right]\right\| \\ &\leq \frac{1}{\lambda_{min}(\Omega_{n,xx})}\left\{\left\|\frac{1}{n}\sum_{t=1}^{n}\Psi_{\vartheta_{1n}}(p_{xt})v_{xt}\right\|+\left\|\frac{1}{n}\sum_{t=1}^{n}\Psi_{\vartheta_{1n}}[g_{x0}(p_{xt})-\Psi_{\vartheta_{1n}}(p_{xt})'\beta_{x0}]\right\|\right\} \\ &= O_p(1)O_p(\sqrt{\vartheta_{1n}/n}+\vartheta_{1n}^{-\eta})=O_p(\sqrt{\vartheta_{1n}/n}+\vartheta_{1n}^{-\eta}), \qquad \text{(E.20)}\end{aligned}$$

which holds under Lemma 1, 2 and Assumption T2.4(a).

Combining (E.19) and (E.20), we conclude that $\frac{1}{n}\sum_{t=1}^{n}(\hat{v}_{xt}-v_{xt})^2 = O_p((\vartheta_{1n}^{-2\eta}+\vartheta_{1n}/n)$. Similarly, we can prove that $\frac{1}{n}\sum_{t=1}^{n}(\hat{v}_{zt}^{k_2}-v_{zt}^{k_2})^2 = O_p((\vartheta_{1n}^{-2\eta}+\vartheta_{1n}/n)$, for $k_2=1,..,d_1$. In summary, we have $\frac{1}{n}\sum_{t=1}^{n}\|\hat{v}_t-v_t\|^2 = (1+d_1)O_p((\vartheta_{1n}^{-2\eta}+\vartheta_{1n}/n) = O_p((\vartheta_{1n}^{-2\eta}+\vartheta_{1n}/n)$[9].

F Variance-covariance matrix

In this section, we give the expression of the sample variance-covariance matrix in practice.

(1) In Theorem 2-Time series, the estimator of the variance-covariance matrix takes the form, $(\hat{L}_n'\hat{L}_n)^{-1}\hat{L}_n'\hat{V}_n\hat{L}_n(\hat{L}_n'\hat{L}_n)^{-1}$, where $\hat{L}_n = \frac{1}{n}\sum_{t=1}^{n}\hat{L}_t$ with

[9] As we assume elements of v_t are pairwise independent.

$\hat{L}_t =$

$$\left[\begin{array}{cccc} -(x_t-\hat{\gamma})^2 & -(x_t-\hat{\gamma})^2 I(x_t \geq \hat{\gamma}) & -(x_t-\hat{\gamma})z'_{i,t} & -[\hat{\beta}+\hat{\delta}I(x_t\geq\hat{\gamma})](x_t-\hat{\gamma}) \\ -(x_t-\hat{\gamma})^2 I(x_t \geq \hat{\gamma}) & (x_t-\hat{\gamma})^2 I(x_t \geq \hat{\gamma}) & -(x_t-\hat{\gamma})I(x_t\geq\hat{\gamma})z'_t & -(\hat{\beta}+\hat{\delta})(x_t-\hat{\gamma})I(x_t\geq\hat{\gamma}) \\ -z_t(x_t-\hat{\gamma}) & z_t(x_t-\hat{\gamma})I(x_t\geq\hat{\gamma}) & z_t z'_t & z_t[\hat{\beta}+\hat{\delta}I(x_t\geq\hat{\gamma})] \\ -[\hat{\beta}+\hat{\delta}I(x_t\geq\hat{\gamma})](x_t-\hat{\gamma}) & -[\varepsilon_t(\hat{\phi}^*)-(\hat{\beta}+\hat{\delta})(x_t-\hat{\gamma})]I(x_t\geq\hat{\gamma}) & -[\hat{\beta}+\hat{\delta}I(x_t\geq\hat{\gamma})]z'_t & -[\hat{\beta}+\hat{\delta}I(x_t\geq\hat{\gamma})]^2 \\ -(x_t-\hat{\gamma}) & -(x_t-\hat{\gamma})I(x_t\geq\hat{\gamma}) & z'_t & -[\hat{\beta}+\hat{\delta}I(x_t\geq\hat{\gamma})]x_t \end{array}\right],$$

note $\hat{L}_t$ has a similar structure to $\hat{Q}$ in Hansen (2017), except we extended the last row with the partial derivatives of the unknown function $h(\cdot)$.

(2) Following Theorem 2-panel, with parametric estimator $\hat{\theta} = (\hat{\beta}', \hat{\gamma})'$, the variance-covariance matrix has the expression $(\hat{\mathcal{L}}'_N \hat{\mathcal{L}}_N)^{-1} \hat{\mathcal{L}}'_N \mathcal{V}_N \hat{\mathcal{L}}_N (\hat{\mathcal{L}}'_N \hat{\mathcal{L}}_N)^{-1}$, where $\hat{\mathcal{L}}_N = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=t_0}^{T} \hat{\mathcal{L}}_{i.t}$,

with

$$\hat{\mathcal{L}}_{i,t} =$$

$$\left[\begin{array}{cccc} -\Delta x_{i,t}^2 & -\Delta x_{i,t}(X_{i,t}-\tau_2\hat{\gamma})\mathbf{I}_{i,t}(\hat{\gamma}) & -\Delta x_{i,t}\Delta z_{i,t}' & \hat{\delta}\Delta x_{i,t}\tau_2'\mathbf{I}_{i,t}(\hat{\gamma}) \\ -(X_{i,t}-\tau_2\hat{\gamma})\mathbf{I}_{i,t}(\hat{\gamma})\Delta x_{i,t} & [(X_{i,t}-\tau_2\hat{\gamma})\mathbf{I}_{i,t}(\hat{\gamma})]^2 & -(X_{i,t}-\tau_2\hat{\gamma})\mathbf{I}_{i,t}(\hat{\gamma})\Delta z_{i,t}' & (X_{i,t}-\tau_2\hat{\gamma})\mathbf{I}_{i,t}(\hat{\gamma})\hat{\delta}\tau_2'\mathbf{I}_{i,t}(\hat{\gamma}) \\ -\Delta z_{i,t}\Delta x_{i,t} & -\Delta z_{i,t}(X_{i,t}-\tau_2\hat{\gamma})\mathbf{I}_{i,t}(\hat{\gamma}) & \Delta z_{i,t}\Delta z_{i,t}' & \hat{\delta}\Delta z_{i,t}\tau_2'\mathbf{I}_{i,t}(\hat{\gamma}) \\ -\hat{\delta}\tau_2'\mathbf{I}_{i,t}(\hat{\gamma})\Delta x_{i,t} & -\tau_2'\mathbf{I}_{i,t}(\hat{\gamma})\varepsilon_{i,t}(\hat{\phi}^*)+\hat{\delta}\tau_2'\mathbf{I}_{i,t}(\hat{\gamma})(X_{i,t}-\tau_2\hat{\gamma})\mathbf{I}_{i,t}(\hat{\gamma}) & -\hat{\delta}\tau_2'\mathbf{I}_{i,t}(\hat{\gamma})\Delta z_{i,t}' & \hat{\delta}^2\mathbf{I}_{i,t}(\hat{\gamma})'\mathbf{I}_{i,t}(\hat{\gamma}) \\ -\Delta x_{i,t} & -(X_{i,t}-\tau_2\hat{\gamma})\mathbf{I}_{i,t}(\hat{\gamma}) & \Delta z_{i,t}' & \hat{\delta}\tau_2'\mathbf{I}_{i,t}(\hat{\gamma}) \end{array}\right].$$

G Endogeneity test

In this section, we construct a Wald-type test to test the potential endogeneity of our threshold variable, x, and regressors, z. For a time-series KTR model, under our setup, we can express the null hypothesis H_0 of no endogeneity and the alternative hypothesis H_1 as

$$H_0 : h(\cdot) = 0; \qquad H_1 : h(\cdot) \neq 0. \tag{G.1}$$

For sieve approximation, we can equivalently rewrite the above hypothesis testing as

$$H_0 : \beta_{h0} = \mathbf{0}_{\vartheta_{2n}}; \qquad H_1 : \beta_{h0} \neq \mathbf{0}_{\vartheta_{2n}}.$$

Recall that in the estimation of the time-series model, we can express the estimator $\hat{\beta}_h$ by a partial linear regression. Denote $\tilde{X}(\hat{\gamma}) = [x_1(\hat{\gamma}), ..., x_n(\hat{\gamma})]'$, $\tilde{x}_t(\hat{\gamma}) = [x_t - \hat{\gamma}, (x_t - \hat{\gamma})I(x_t \geq \hat{\gamma}), z_t']$, and $\tilde{M}(\hat{\gamma}) = I_n - \tilde{X}(\hat{\gamma})[\tilde{X}(\hat{\gamma})'\tilde{X}(\hat{\gamma})]^{-1}\tilde{X}(\hat{\gamma})$, we have

$$\hat{\beta}_h = [\Psi_{\vartheta_{2n}}(\hat{v})\tilde{M}(\hat{\gamma})\Psi_{\vartheta_{2n}}(\hat{v})]^{-1}\Psi_{\vartheta_{2n}}(\hat{v})'\tilde{M}(\hat{\gamma})y \tag{G.2}$$

Following the covariance estimator introduced by Andrews (1991), we construct a Wald statistic:

$$
\begin{aligned}
W_n = \quad & \hat{\beta}_h{}' \Psi_{\vartheta_{2n}}(\hat{v}) \tilde{M}(\hat{\gamma}) \Psi_{\vartheta_{2n}}(\hat{v})' [\Psi_{\vartheta_{2n}}(\hat{v})' \tilde{M}(\hat{\gamma}) J_n(\hat{\gamma}) \tilde{M}(\hat{\gamma}) \Psi_{\vartheta_{2n}}(\hat{v})]^{-1} \\
& \times \Psi_{\vartheta_{2n}}(\hat{v}) \tilde{M}(\hat{\gamma}) \Psi_{\vartheta_{2n}}(\hat{v})' \hat{\beta}_h, \qquad \text{(G.3)}
\end{aligned}
$$

where $J_n(\hat{\gamma})$ is an n by n diagonal matrix with typical elements equal to $\hat{\varepsilon}_t^2/(1-\hat{Q}_{tt})$, $\hat{Q}_{tt}$ is the $(t,t)th$ element of $\tilde{M}(\hat{\gamma})\Psi_{\vartheta_{2n}}(\hat{v})'[\Psi_{\vartheta_{2n}}(\hat{v})'\tilde{M}(\hat{\gamma})J_n(\hat{\gamma})\tilde{M}(\hat{\gamma})\Psi_{\vartheta_{2n}}(\hat{v})]^{-1}\Psi_{\vartheta_{2n}}(\hat{v})\tilde{M}(\hat{\gamma})$, and $\hat{\varepsilon}_t = y_t - \hat{\beta}_1(x_t - \hat{\gamma}) - \hat{\delta}(x_t - \hat{\gamma})I(x_t \geq \hat{\gamma}) - z_t'\hat{\beta}_3 - \Psi'_{\vartheta_{2n}}\hat{\beta}_h$. Then for n large enough, W_n convergence to a Chi-squared distribution with ϑ_{2n} degree of freedom under the null hypothesis. For the panel data model, the setup is similar, so we do not repeat it here.

H Monte Carlo results

Table 1: DGP1- Polynomials order changes

		β_1		δ		β_3		γ	
		bias	rmse	bias	rmse	bias	rmse	bias	rmse
$\vartheta_{1n} = \vartheta_{2n} = 6$	n=100	0.2527	0.5264	0.2304	0.5929	-0.0557	0.0737	-0.4081	0.6009
	n=200	0.1688	0.319	0.0561	0.2674	-0.0295	0.0498	-0.2706	0.4353
	n=400	0.0977	0.1974	-0.0199	0.1443	-0.0148	0.0341	-0.1278	0.2332
$\vartheta_{1n} = \vartheta_{2n} = 5$	n=100	0.1873	0.5022	0.1992	0.6053	-0.0492	0.0711	-0.3908	0.6024
	n=200	0.1066	0.2928	0.049	0.2878	-0.0241	0.0485	-0.2504	0.429
	n=400	0.0539	0.1824	-0.0173	0.1382	-0.0112	0.0336	-0.1153	0.2191
$\vartheta_{1n} = \vartheta_{2n} = 4$	n=100	0.0709	0.5289	0.2756	0.8135	-0.0271	0.0379	-0.4686	0.7476
	n=200	0.1068	0.2576	0.1169	0.3543	-0.0149	0.0247	-0.2595	0.5235
	n=400	0.104	0.1647	0.0511	0.2223	-0.0095	0.0171	-0.1166	0.3124
$\vartheta_{1n} = \vartheta_{2n} = 3$	n=100	0.0319	0.4545	0.264	0.7322	-0.0221	0.0353	-0.3921	0.7265
	n=200	0.0687	0.2327	0.1447	0.4117	-0.0114	0.0232	-0.1899	0.5224
	n=400	0.078	0.1453	0.0762	0.2751	-0.0072	0.0161	-0.0689	0.3233

Note: This table presents the effect of the order of polynomials using DGP1, we change ϑ_{1n} and ϑ_{2n} among $3, 4, 5, 6$, where ϑ_{1n} and ϑ_{2n} are the order of Hermite basis functions for our first step and second step estimation, respectively;

Table 2: DGP2-Polynomials order changes

		β_1		δ		β_3		γ	
	T=10	bias	rmse	bias	rmse	bias	rmse	bias	rmse
$\vartheta_{2N} = \vartheta_{1N} = 6$	N=20	0.0984	0.4585	0.0696	0.5318	0.3123	0.3809	-0.405	0.6754
	N=40	0.0705	0.2432	-0.0223	0.217	0.1457	0.2088	-0.1863	0.3786
	N=80	0.0126	0.1365	-0.0305	0.1219	0.0551	0.1139	-0.0849	0.1821
$\vartheta_{2N} = \vartheta_{1N} = 5$	N=20	0.0218	0.4483	0.0799	0.5241	0.267	0.3377	-0.385	0.6571
	N=40	0.0118	0.2283	-0.0099	0.2156	0.136	0.1967	-0.1765	0.3703
	N=80	-0.0284	0.1376	-0.0216	0.1211	0.0673	0.1172	-0.0827	0.1821
$\vartheta_{2N} = \vartheta_{1N} = 4$	N=20	0.0936	0.3553	0.0795	0.481	0.3225	0.3483	-0.2762	0.5837
	N=40	0.1096	0.1946	0.0082	0.2356	0.2587	0.2722	-0.1074	0.3273
	N=80	0.0936	0.1296	-0.0132	0.1251	0.2251	0.2324	-0.0583	0.149
$\vartheta_{2N} = \vartheta_{1N} = 3$	N=20	0.0319	0.3287	0.1272	0.5009	0.3541	0.3799	-0.2172	0.5867
	N=40	0.0658	0.1689	0.0354	0.2581	0.31	0.3243	-0.0801	0.3203
	N=80	0.0613	0.1085	0.0014	0.1492	0.2866	0.2941	-0.047	0.1469

Note: This table presents the effect of the order of polynomials using DGP2, we change ϑ_{1N} and ϑ_{2N} among $3, 4, 5, 6$, where ϑ_{1N} and ϑ_{2N} are the order of Hermite basis functions for our first step and second step estimation, respectively;

I Dataset description

Table 3: Summary Statistics

Canada data (July 2020-Sep 2021)

Variable	Obs	Mean	Std. Dev.	Min	Max
Case	126	7.2135	2.7987	0	11.6392
Test	126	11.4824	1.4961	8.5067	14.3842
Une	126	8.6537	1.9709	5.6	15.3

US data (July 2020-Dec 2021)

Variable	Obs	Mean	Std. Dev.	Min	Max
Case	884	10.0783	1.4022	4.8283	13.9584
Test	884	12.4150	1.4469	7.7998	16.0631
Une	884	5.5887	1.9557	1.9	14.8

NOTE: *Case* = natural logarithm of the number of COVID-19 cases confirmed; *Test* = natural logarithm of the number of COVID-19 test performed; *Une* =Unemployment rate (Seasonal adjusted).

Table 4: 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
	New York	North Dakota	Virginia
	Hawaii	Nebraska	Vermont
	Iowa	New Hampshire	Washington
	Idaho	New Jersey	Wisconsin
	Illinois	New Mexico	West Virginia
	Indiana	Nevada	Wyoming
	Kansas		
	52		

References

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, 817–858.

Andrews, D. W. (1994). Chapter 37 empirical process methods in econometrics. *Handbook of Econometrics*, 2247–2294.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Chapter 76 in Handbook of Econometrics*, 5549–5632.

Chen, X., O. Linton, and I. Van Keilegom (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica 71*(5), 1591–1608.

Chen, X. and X. Shen (1998). Sieve extremum estimates for weakly dependent data. *Econometrica*, 289–314.

Geer, S. A. (2000). *Empirical Processes in M-estimation*, Volume 6. Cambridge university press.

Hansen, B. (2019). A weak law of large numbers under weak mixing.

Hansen, B. E. (2017). Regression kink with an unknown threshold. *Journal of Business & Economic Statistics 35*(2), 228–240.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, 1161–1167.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica: Journal of the Econometric Society*, 1349–1382.

Newey, W. K. and D. Mcfadden (1994). Chapter 36 large sample estimation and hypothesis testing. *Handbook of Econometrics*, 2111–2245.

Seber, G. A. F. (2008). *A matrix handbook for statisticians.* J. Wiley.