SUPPLEMENTARY MATERIAL for "Endogenous kink threshold regression"

October 14, 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. 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

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

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

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

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

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

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

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

where the boundness of the first term is given by

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[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.

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 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

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

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

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

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

[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)

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 inequality and obtain

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

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

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

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

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

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

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

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

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

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

Given

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

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

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

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

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

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

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

since $E[\Delta\varepsilon_{i,t}(\Delta\varepsilon_{i,t}(\phi^{**})-\Delta\varepsilon_{i,t})] = E[E[\Delta\varepsilon_{i,t}^2|\mathcal{F}_{i,t-2}, x_{i,t}, x_{i,t-1}, z_{1,i,t}, z_{1,i,t-1}, p_{i,t}, p_{i,t-1}](\Delta\varepsilon_{i,t}^2(\phi^{**})-\Delta\varepsilon_{i,t})^2] \le constant \times E[(\Delta\varepsilon_{i,t}(\phi^{**}) - \Delta\varepsilon_{i,t})^2]$ 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

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

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 $\mathcal{G}_N=\left\{\Delta\varepsilon_{i,t}(\phi^{***})^2-\Delta\varepsilon_{i,t}^2 : d(\phi^{***},\phi_0)\leq\omega_1, \phi^{***}\in\Phi_N\right\}$, 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:

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

note (D.1) has a similar structure ‘as (B.1) in the time series part. Again, we apply the 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 $E\left\|m_{i,t}(\theta,h) - m_{i,t}(\tilde{\theta},\tilde{h})\right\|^2 \leq$

[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)

constant $\times \tilde{\delta}^2$. Recall the expression of $H_{i,t}(\theta)$ and $\Delta\varepsilon_{i,t}(\phi)$, we have

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\|\Omega_{n,xx} - E(\Omega_{n,xx})\|_{sp} \leq \|\Omega_{n,xx} - E(\Omega_{n,xx})\|$ shown in Lemma1(b) and obtain

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

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 Lemma1, we have*

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

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

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

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

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 Lemma1(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

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

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 $\frac{1}{n}\sum_{t=1}^{n}\left\|\hat{v}_t-v_t\right\|^2=(1+d_1)O_p((\vartheta_{1n}^{-2\eta}+\vartheta_{1n}/n)=O_p((\vartheta_{1n}^{-2\eta}+\vartheta_{1n}/n)$[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.

$\hat{L}_t =$

$$
\left[\begin{matrix}
-(x_t-\hat{\gamma})^2 & -(x_t-\hat{\gamma})^2 I(x_t\geq\hat{\gamma}) \\
-(x_t-\hat{\gamma})^2 I(x_t\geq\hat{\gamma}) & (x_t-\hat{\gamma})^2 I(x_t\geq\hat{\gamma}) \\
-z_t(x_t-\hat{\gamma}) & z_t(x_t-\hat{\gamma}) I(x_t\geq\hat{\gamma}) \\
-[\hat{\beta}+\hat{\delta}I(x_t\geq\hat{\gamma})](x_t-\hat{\gamma}) & -[\varepsilon_t(\hat{\phi}^*)-(\hat{\beta}+\hat{\delta})(x_t-\hat{\gamma})]I(x_t\geq\hat{\gamma}) \\
-(x_t-\hat{\gamma}) & -(x_t-\hat{\gamma}) I(x_t\geq\hat{\gamma})
\end{matrix}\right.
$$

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

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:

$$W_n = \hat{\beta_h}' \Psi_{\vartheta_{2n}}(\hat{v}) \tilde{M}(\hat{\gamma}) \Psi_{\vartheta_{2n}}(\hat{v})' [\Psi_{\vartheta_{2n}}(\hat{v})' \tilde{M}(\hat{\gamma}) J_n(\hat{\gamma}) \tilde{M}(\hat{\gamma}) \Psi_{\vartheta_{2n}}(\hat{v})]^{-1} \times \Psi_{\vartheta_{2n}}(\hat{v}) \tilde{M}(\hat{\gamma}) \Psi_{\vartheta_{2n}}(\hat{v})' \hat{\beta}_h, \tag{G.3}$$

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.