Manuscript submitted to *The Econometrics Journal*, pp.

Online Supplementary: Threshold Nonlinearities and the Democracy-Growth Nexus

Chaoyi Chen[†‡] and Thanasis Stengos[‡]

[†] *Magyar Nemzeti Bank (Central Bank of Hungary), Budapest, 0165, Hungary.*
E-mail: `chenc@mnb.hu`
[‡] *MNB Institute, John von Neumann University, Budapest, 1117, Hungary*
[‡] *Department of Economics and Finance, University of Guelph, Guelph, Ontario N1G 2W1, Canada.*
E-mail: `tstengos@uoguelph.ca`

Summary This paper investigates the relationship between democracy and economic growth in the context of a linear index threshold regression model. We first introduce the baseline model with endogeneity and propose a two-step smoothed GMM estimation method. We establish the consistency and derive the asymptotic distributions of the proposed estimators. We then extend the proposed approach to encompass the dynamic panel context and employ the model to delve into the impact of democratization on economic growth. Our findings reveal that democratization's impact on growth is nonlinear and depends on the country's current institutional quality level. We find a significantly positive effect in both regimes, but countries with higher institutional quality benefit more from democratization. Furthermore, democracy's impact on economic growth is more pronounced in countries with higher education levels than others, suggesting that education also plays a crucial role in enhancing the positive effects of democracy on economic growth. Our proposed estimator can be used in other situations that require more than one threshold variable. In these cases, our hybrid estimator has less stringent data requirements than an alternative scenario where the thresholds would enter separately, especially when the threshold variables are correlated.

Keywords: *Endogenous threshold effects and regressors, GMM, Index model, Democratization, Threshold Regression.*

This supplementary material consists of seven sections. In Section S1, a comprehensive analysis of the non-smoothed GMM estimator for the linear index threshold model is presented. The proofs of Lemmas 2.1 to 2.4 can be found in Section S2. Section S3 delves into the Monte Carlo simulation study of the proposed estimators' finite sample performance. This includes a comparison between the smoothed GMM estimator and the smoothed least squares estimator by Seo and Shin (2016), considering both exogenous and endogenous threshold variable designs. The simulation results align with theoretical expectations. Section S4 showcases the Monte Carlo simulation results for the linearity test. The issue of weak instruments due to a highly persistent regressor in a first-differenced estimator is addressed in Section S5. Here, we propose a system GMM estimator to mitigate the weak instrument challenge. The simulation outcomes suggest that the proposed smoothed system-GMM estimator outperforms the smoothed FD-GMM estimator in handling the persistent regressor within the dynamic panel context. Additional findings related to the democracy-growth application are presented in Section S6. A heuristic example demonstrating the smoothness of the GMM estimator for the threshold model within a fixed threshold effect framework is illustrated in Section S7.

S1. THE GMM ESTIMATOR FOR THE LINEAR INDEX THRESHOLD MODEL

This section considers the standard (non-smoothed) GMM estimator for model (2.1) in the paper. We provide the estimation strategy and discuss the limiting results. Similar to the smoothed GMM estimator, we also propose the a statistic for the test of linearity.

S1.1. Estimation

For the moment condition (2.2), the natural sample analogue to $E(g_t(\theta))$ is,

$$g_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} g_t(\theta). \tag{1.0.1}$$

The GMM estimators can be obtained as

$$\hat{\theta}^{GMM} = \underset{\theta\in\Theta}{argmin}\ Q_n(\theta), \tag{1.0.2}$$

where

$$Q_n(\theta) = g_n(\theta)^T W_n g_n(\theta) = \left[\frac{1}{n}\sum_{t=1}^{n} g_t(\theta)\right]^T W_n \left[\frac{1}{n}\sum_{t=1}^{n} g_t(\theta)\right], \tag{1.0.3}$$

and W_n is a positive definite matrix with $W_n \xrightarrow{p} \Omega^{-1}$, where $\Omega = E\left(g_t(\theta_n) g_t(\theta_n)^T\right)$.

As $Q_n(\theta)$ is not continuous in ψ, it is more practical to use a grid search empirically for low-dimensional q_{2t} [1]. Note that the model is linear in β and δ for a given ψ. Thus, for a given ψ and a weight matrix W_n, we have

$$\left(\hat{\beta}_{(\psi)}^T, \hat{\delta}_{(\psi)}^T\right)^T = \left[\hat{G}(\psi)^T W_n \hat{G}(\psi)\right]^{-1} \hat{G}(\psi)^T W_n \left[-\frac{1}{n}\sum z_t y_t\right], \tag{1.0.4}$$

where $\underset{m\times 2k}{\hat{G}(\psi)} = \left[\hat{G}_\beta^T, \hat{G}_\delta^T(\psi)\right]^T$, $\underset{m\times k}{\hat{G}_\beta} = -\frac{1}{n}\sum z_t x_t^T$ and $\underset{m\times k}{\hat{G}_\delta(\psi)} = -\frac{1}{n}\sum z_t \tilde{x}_t^T I\left(q_{1t} + q_{2t}^T\psi > 0\right)$.

Then, the threshold index estimators can be obtained by

$$\hat{\psi}^{GMM} = \underset{\psi\in\Theta_\psi}{\text{argmin}}\ Q_n(\psi) = \left[\frac{1}{n}\sum_{t=1}^{n} g_t(\hat{\beta}_{(\psi)}, \hat{\delta}_{(\psi)}, \psi)\right]^T W_n \left[\frac{1}{n}\sum_{t=1}^{n} g_t(\hat{\beta}_{(\psi)}, \hat{\delta}_{(\psi)}, \psi)\right], \tag{1.0.5}$$

and

$$\left(\hat{\beta}^{GMM^T}, \hat{\delta}^{GMM^T}\right)^T = \left(\hat{\beta}^T(\hat{\psi}), \hat{\delta}^T(\hat{\psi})\right)^T. \tag{1.0.6}$$

Therefore, the 2-step method can be obtained as:

Step 1: Estimate the model with $W_n = I_m$, where I_m is an $m\times m$ identity matrix, and get residual $\hat{e}$.

Step 2: Estimate the model with $W_n = \left[\frac{1}{n}\sum_{t=1}^{n}(z_t z_t^T \hat{e}_t^2)\right]^{-1}$.

S1.2. Asymptotic Results

THEOREM 1.1. *(a) Under Assumption 4.1 in the paper, as $n\to\infty$, we have*

[1] For the high-dimensional q_{2t}, we may consider to use the MIQ algorithm (e.g. Lee et al. (2021)).

$$\hat{\theta}^{GMM} \xrightarrow{p} \theta_n. \tag{1.1.1}$$

(b) Under Assumption 4.1 in the paper, as n→ ∞, choosing a weighting matrix such that $W_n \xrightarrow{p} W = \Omega^{-1}$*, we have*

$$\begin{bmatrix} \sqrt{n} & 0 & 0 \\ 0 & \sqrt{n} & 0 \\ 0 & 0 & n^{\frac{1}{2}-\alpha} \end{bmatrix} \begin{bmatrix} \hat{\beta} - \beta_0 \\ \hat{\delta} - \delta_n \\ \hat{\psi} - \psi_0 \end{bmatrix} \xrightarrow{d} N\left(0, \left(G^T \Omega^{-1} G\right)^{-1}\right), \tag{1.1.2}$$

where Ω *and* G *are defined in Assumption 4.1 (d) and Assumption 4.1 (g).*

Proof. Closely following the proof of Theorem 4.1, we conclude the proof of part (a). For part (b), we follow Theorem 7.1 of Newey and Mcfadden (1994).

First, by the Central Limit Theorem (CLT) for strong mixing data, we have $\sqrt{n} g_n(\theta_n) \xrightarrow{d} N(0, \Omega)$, where $\Omega = E(g_t(\theta_n) g_t(\theta_n)^T)$.

Next, let $W_n \xrightarrow{p} W = \Omega^{-1}$, and

$$D_n = k_n^{-1} G^T W_n g_n(\theta_n),$$
$$H = k_n^{-1} G^T W G k_n^{-1},$$
$$R(\theta) = \frac{Q_n(\theta) - Q_n(\theta_n) - Q(\theta) - D_n^T(\theta - \theta_n)}{\|\theta - \theta_n\|}.$$

we next show the stochastic differentiability condition hold. That is, for any $\gamma_n \to 0$, we have

$$\underset{\|\theta-\theta_n\| \le \gamma_n}{Sup} \left| \frac{\sqrt{n} R(\theta)}{1 + \sqrt{n}\, \|\theta - \theta_n\|} \right| = o_p(1).$$

Define $\varepsilon_n(\theta) = \frac{g_n(\theta) - g_n(\theta_n) - g(\theta)}{1+\sqrt{n}\|\theta-\theta_n\|}$. For $\gamma_n \to 0$ and $U = \{\|\theta - \theta_n\| \le \gamma_n\}$, $\underset{\theta \in U}{Sup}\{\sqrt{n}\, \|\varepsilon_n(\theta)\|\} \xrightarrow{p} o_p(1)$ if empirical process $\sqrt{n}\,(g_n(\theta) - g(\theta))$ is stochastically equicontinuous. Note that $g_t(\theta)$ is linear in β and δ, which are bounded by Assumption 4.1 (d). Therefore, we only need to check the stochastic equicontinuity of the empirical process $\frac{1}{\sqrt{n}} \sum_{t=1}^n \Big[z_t \tilde{x}_t I(q_{1t} + q_{2t}^T \psi > 0) - E\left\{ z_t \tilde{x}_t I(q_{1t} + q_{2t}^T \psi > 0) \right\} \Big]$. Let $F = (\|z_t \tilde{x}_t\|) \underset{||\psi-\psi_0|| \le \gamma_n}{sup} I\left(q_{1t} > -q_{2t}^T \psi \bigwedge -q_{2t}^T \psi_0\right)$ be the envelope function, where $\bigwedge$ denotes the minimum operator. Since the indicator functions of half intervals constitute a *type I* class or a Vapnik– Chervonenkis (VC) class, by Assumptions 4.1 (a)-(b) and (d), the stochastic equicontinuity follows the Theorem 1 of Andrews (1994) and the Theorem 2.14.1 of Vaart and Wellner (2000). Evidently, $\varepsilon_n(\theta_n) = 0$.

Following proof of Theorem 7.2 of Newey and Mcfadden (1994), we decompose $|\frac{\sqrt{n}R(\theta)}{1+\sqrt{n}\|\theta-\theta_n\|}|$ into five terms,

$$\left| \frac{\sqrt{n} R(\theta)}{1 + \sqrt{n}\, \|\theta - \theta_n\|} \right| \le \sum_{j=1}^{5} r_{nj}(\theta),$$

where

$$
\begin{aligned}
r_{n1}(\theta) &= \frac{\sqrt{n}\left(2\sqrt{n}\,\|\theta-\theta_n\| + \|\theta-\theta_n\|^2\right)\left|\varepsilon_n(\theta)^T W_n \varepsilon_n(\theta)\right|}{\|\theta-\theta_n\|\,(1+\sqrt{n}\,\|\theta-\theta_n\|)} \\
&= \frac{(2n+\sqrt{n}\,\|\theta-\theta_n\|)\left|\varepsilon_n(\theta)^T W_n \varepsilon_n(\theta)\right|}{(1+\sqrt{n}\,\|\theta-\theta_n\|)}, \\
r_{n2}(\theta) &= \frac{\sqrt{n}\left|\left[g(\theta)-Gk_n^{-1}(\theta-\theta_n)\right]^T W_n g_n(\theta_n)\right|}{\|\theta-\theta_n\|\,(1+\sqrt{n}\,\|\theta-\theta_n\|)}, \\
r_{n3}(\theta) &= \frac{n\left|\left[g(\theta)+g_n(\theta_n)\right]^T W_n \varepsilon_n(\theta)\right|}{(1+\sqrt{n}\,\|\theta-\theta_n\|)}, \\
r_{n4}(\theta) &= \frac{\sqrt{n}\left|g(\theta)^T W_n \varepsilon_n(\theta)\right|}{\|\theta-\theta_n\|}, \\
r_{n5}(\theta) &= \frac{\sqrt{n}\left|g(\theta)^T\left[W_n - W\right]g(\theta)\right|}{\|\theta-\theta_n\|\,(1+\sqrt{n}\,\|\theta-\theta_n\|)}.
\end{aligned}
$$

By the consistency of θ and $\underset{\theta\in U}{Sup}\{\sqrt{n}||\varepsilon_n,(\theta)||\} \xrightarrow{p} o_p(1)$, we have

$$
\underset{\theta\in U}{Sup}\{r_{n1}(\theta)\} = \underset{\theta\in U}{Sup}\left\{\frac{(2+\frac{\|\theta-\theta_n\|}{\sqrt{n}})\left|\left(\sqrt{n}\varepsilon_n(\theta)\right)^T W_n \sqrt{n}\varepsilon_n(\theta)\right|}{(1+\sqrt{n}\,\|\theta-\theta_n\|)}\right\} = o_p(1).
$$

Note that, by the differentiability of $g(\theta)$, we can show

$$
\underset{\theta\in U}{Sup}\left\{\frac{\|\sqrt{n}g(\theta)\|}{(1+\sqrt{n}\,\|\theta-\theta_n\|)}\right\} \le \underset{\theta\in U}{Sup}\left\{\frac{\|g(\theta)\|}{\|\theta-\theta_n\|}\right\} \le \underset{\theta\in U}{Sup}\left\{\frac{\left\|g(\theta_n)+Gk_n^{-1}(\theta-\theta_n)+o(\|\theta-\theta_n\|)\right\|}{\|\theta-\theta_n\|}\right\} = O(1),
$$

and

$$
\begin{aligned}
&\underset{\theta\in U}{Sup}\left\{\frac{\left\|g(\theta)-Gk_n^{-1}(\theta-\theta_n)\right\|}{\|\theta-\theta_n\|\,(1+\sqrt{n}\,\|\theta-\theta_n\|)}\right\} \le \underset{\theta\in U}{Sup}\left\{\frac{\left\|g(\theta)-Gk_n^{-1}(\theta-\theta_n)\right\|}{\|\theta-\theta_n\|}\right\} \\
&= \underset{\theta\in U}{Sup}\left\{\frac{\left\|g(\theta)-g(\theta_n)-Gk_n^{-1}(\theta-\theta_n)\right\|}{\|\theta-\theta_n\|}\right\} = o(1).
\end{aligned}
$$

Applying Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
\underset{\theta\in U}{Sup}\{r_{n2}(\theta)\} &= \underset{\theta\in U}{Sup}\left\{\frac{\left|\left[g(\theta)-Gk_n^{-1}(\theta-\theta_n)\right]^T W_n \sqrt{n}g_n(\theta_n)\right|}{\|\theta-\theta_n\|\,(1+\sqrt{n}\,\|\theta-\theta_n\|)}\right\} = o_p(1), \\
\underset{\theta\in U}{Sup}\{r_{n3}(\theta)\} &\le \underset{\theta\in U}{Sup}\left\{\frac{\sqrt{n}\,\|g(\theta)+g_n(\theta_n)\|\ \ \|W_n\|\ \ \|\sqrt{n}\varepsilon_n(\theta)\|}{(1+\sqrt{n}\,\|\theta-\theta_n\|)}\right\} = o_p(1),
\end{aligned}
$$

and

$$\underset{\theta\in U}{Sup}\{r_{n4}(\theta)\} = \underset{\theta\in U}{Sup}\left\{\frac{\sqrt{n}\left\|g(\theta)^T W_n \varepsilon_n(\theta)\right\|}{\|\theta-\theta_n\|}\right\} \le \underset{\theta\in U}{Sup}\left\{\frac{\|g(\theta)\|}{\|\theta-\theta_n\|}\|W_n\|\sqrt{n}\,\|\varepsilon_n(\theta)\|\right\} = o_p(1).$$

$$\underset{\theta\in U}{Sup}\{r_{n5}(\theta)\} = \underset{\theta\in U}{Sup}\left\{\frac{\sqrt{n}\left|g(\theta)^T\left[W_n-W\right]g(\theta)\right|}{\|\theta-\theta_n\|\left(1+\sqrt{n}\,\|\theta-\theta_n\|\right.}\right\} \le \underset{\theta\in U}{Sup}\left\{\frac{\|g(\theta)\|}{\|\theta-\theta_n\|}\|W_n-W\|\frac{\|g(\theta)\|}{\|\theta-\theta_n\|}\right\} = o_p(1).$$

To sum up, we obtain

$$\underset{\theta\in U}{Sup}\left\{\left|\frac{\sqrt{n}R(\theta)}{1+\sqrt{n}\,\|\theta-\theta_n\|}\right|\right\} \le \sum_{j=1}^{5}\underset{\theta\in U}{Sup}\{r_{nj}(\theta)\} \le \sum_{j=1}^{5} o_p(1) = o_p(1).$$

Next, by replacing $(\widehat{\theta}-\theta_n)$ with $k_n^{-1}(\hat{\theta}-\theta_n)$ and applying the same arguments as in the proof of Theorem 7.1 of Newey and Mcfadden (1994), we have $k_n^{-1}(\hat{\theta}-\theta_n) = O_p(n^{-1/2})$. Similarly, let $\tilde{\theta} = \theta_n - \left[k_n^{-1}G^T W G k_n^{-1}\right]^{-1}\left(k_n^{-1}G^T W_n\right)g_n(\theta_n)$. Closely following the same lines proof of Theorem 7.1 of Newey and Mcfadden (1994), we can show $\left\|k_n^{-1}(\hat{\theta}-\theta_n) - k_n^{-1}(\tilde{\theta}-\theta_n)\right\| = o_p(n^{-1/2})$. Hence, $\sqrt{n}k_n^{-1}\left\|\hat{\theta}-\tilde{\theta}\right\| \xrightarrow{p} 0$. By following $\sqrt{n}k_n^{-1}\left(\tilde{\theta}-\theta_n\right) \xrightarrow{d} N\left(0,\left(G^T\Omega^{-1}\right)^{-1}\right)$, we have $\sqrt{n}k_n^{-1}\left(\hat{\theta}-\theta_n\right) \xrightarrow{d} N\left(0,\left(G^T\Omega^{-1}G\right)^{-1}\right)$.

Q.E.D.

The convergence rate for the estimator of the slope parameter is standard root-n. The convergence rate for the thresholds depends on the unknown α, which determines the decaying rate of the threshold effect. Intuitively, unlike the smoothed least square of Seo and Linton (2007), where the smoothness results from the objective funtion, the smoothness of the GMM estimator relies on the nature of the sample averaging [2].

G_β and G_δ can be estimated as

$$\hat{G}_\beta = -\frac{1}{n}\sum_{t=1}^{n} z_t x_t^T,$$

$$\hat{G}_\delta = -\frac{1}{n}\sum_{t=1}^{n} z_t \tilde{x}_t^T I\left(q_{1t} + q_{2t}^T\hat{\psi} > 0\right).$$

For G_ψ, we can estimate it using a standard Nadaraya-Watson kernel estimator,

$$\hat{G}_\psi = -\frac{1}{nb}\sum_{t=1}^{n} z_t \hat{\delta}^T \tilde{x}_t q_{2t}^T \phi\left(\frac{q_{1t}+q_{2t}^T\hat{\psi}}{b}\right),$$

where $\phi(.)$ is the second-order kernel function and b is the bandwidth.

Let $\hat{\Omega} = \frac{1}{n}\sum_{t=1}^{n} g_t(\hat{\theta})g_t^T(\hat{\theta})$. As $n \to \infty$, $\hat{G}$ and $\hat{\Omega}$ converge in probability respectively to G and Ω following the uniform law of large number, the consistency of the Nadaraya-Watson estimator and the kernel density estimator for α mixing data.

[2] We provide a heuristic example in section S7 to explain the smoothness of the GMM and compare the limiting behaviours among the least square estimator, the smoothed least square estimator, and the GMM estimator.

S1.3. Test for Linearity

Similar to section 5 in the paper, we consider the null hypothesis of $\delta = 0$. Let

$$SupWald = \sup_{\psi \in \Theta_\psi} Wald(\psi), \tag{1.1.3}$$

where $\hat{\beta}(\psi)$ and $\hat{\delta}(\psi)$ are the the GMM estimates of β and δ, repectively,

$$Wald(\psi) = n\left(R\left[\hat{\beta}^T(\psi), \hat{\delta}^T(\psi)\right]^T\right)^T \left[R(\hat{G}(\psi)^T \tilde{\Omega}^{-1}\hat{G}(\psi))^{-1}R^T\right]^{-1} R\left[\hat{\beta}^T(\psi), \hat{\delta}^T(\psi)\right]^T,$$

$$\hat{G}(\psi) = \left[\hat{G}_\beta, \hat{G}_\delta(\psi)\right],$$

$$R = [0_{l\times k}, I_{l\times l}],$$

and $\tilde{\Omega}$ is the estimate of Ω under the null of linearity.

Theorem 1.2. *Suppose that* $\inf_{\psi \in \Theta_\psi} |G(\psi)^T \Omega^{-1} G(\psi)|$ *is positive, with Assumptions 4.1(a)-(d), and 4.1(f) hold in the paper, under the null hypothesis, we have*

$$SupWald \xrightarrow{d} \sup_{\psi \in \Theta_\psi} V^T \Omega^{-1/2} G(\psi)^T \left(G(\psi)^T \Omega^{-1} G(\psi)\right)^{-1} R^T \left(R\left(G(\psi)^T \Omega^{-1} G(\psi)\right)^{-1}\right) R^T)^{-1}$$

$$\times R\left(G(\psi)^T \Omega^{-1} G(\psi)\right)^{-1} G(\psi) \Omega^{-1/2} V, \tag{1.2.1}$$

where $V \sim N(0, I_l)$ *and* I_l *is an* l *by* l *identity matrix.*

Similar to the smoothed GMM estimator discuss in the paper, the p-values can be simulated following the same bootstrap.

Proof. Under the null of $\delta_0 = 0$, we have

$$\hat{\delta}^s(\psi) = R\left(\hat{G}^s(\psi)^T \tilde{\Omega}(\psi)^{-1} \hat{G}^s(\psi)\right)^{-1} \hat{G}^s(\psi)^T \tilde{\Omega}(\psi)^{-1} g_n^s(\theta_n). \tag{1.2.2}$$

By applying Lemma 3, we can show $\sup_{\psi \in \Theta_\psi} \|\hat{G}_\delta^s(\psi) - G_\delta(\psi)\| = o_p(1)$. Thus, applying continuous mapping theorem, for each $\psi \in \Theta_\psi$, we have

$$\sqrt{n}\hat{\delta}^s(\psi) \Longrightarrow R\left(G(\psi)^T \Omega^{-1} G(\psi)\right)^{-1} G(\psi)^T \Omega^{-1/2} V, \tag{1.2.3}$$

where R and V are defined in (1.2) and concludes our proof.

Q.E.D.

S2. LEMMAS

Lemma 2.1. *Under Assumption 4.1, we have*

$$\sup_{\psi \in \Theta_\psi} \left\| \frac{1}{n} \sum_{t=1}^{n} z_t \tilde{x}_t{}^T I(q_{1t} + q_{2t}^T \psi > 0) - E[z_t \tilde{x}_t{}^T I(q_{1t} + q_{2t}^T \psi > 0)] \right\| \xrightarrow{p} 0.$$

$$\sup_{\psi \in \Theta_\psi} \left\| \frac{1}{n} \sum_{t=1}^{n} x_t \tilde{x}_t{}^T I(q_{1t} + q_{2t}^T \psi > 0) - E[x_t \tilde{x}_t{}^T I(q_{1t} + q_{2t}^T \psi > 0)] \right\| \xrightarrow{p} 0.$$

PROOF. Under Assumptions 4.1, $E||z_t\tilde{x}_t^T||$ and $E||x_t\tilde{x}_t^T||$ are bounded. Then, the proof is straightforward by applying Lemma 1 of Seo and Linton (2007).

Q.E.D.

LEMMA 2.2. *Under Assumptions 4.1 (a)-(b) and (f), there is a $C<\infty$ such that for any $\psi_1,\psi_2\in\Theta_\psi$, we have*

$$\|E\left(X_t\left(I(\psi_1)-I(\psi_2)\right)\right)\|\le C\left\|\psi_1-\psi_2\right\|,$$
$$\|E\left(X_t\varepsilon_t\left(I(\psi_1)-I(\psi_2)\right)\right)\|\le C\left\|\psi_1-\psi_2\right\|,$$

where $I(\psi)=I\left(q_{1t}+q_{2t}^T\psi>0\right)$.

PROOF. Note that, for any random variable w, we have

$$\frac{\partial E\left(wI(q_{1t}+q_{2t}^T\psi>0)\right)}{\partial\psi_i}=E\left(wq_{2it}\left|v_t(\psi)=0\right.\right)f_{v_t(\psi)}(0),$$

where v_t defines in Assumption 4.1 (g).

Thus, applying the first-order Taylor approximation, we have,

$$\|E\left(X_t\left(I(\psi_1)-I(\psi_2)\right)\right)\|\le\left\|E(X_tq_{2t}^T|v_t(\psi_2)=0)\right\|f_{v_t(\psi_2)}(0)\left\|\psi_1-\psi_2\right\|+O(1)$$
$$\|E\left(X_t\varepsilon_t\left(I(\psi_1)-I(\psi_2)\right)\right)\|\le\left\|E(\ X_t\varepsilon_tq_{2t}^T|v_t(\psi_2)=0)\right\|f_{v_t(\psi_2)}(0)\left\|\psi_1-\psi_2\right\|+O(1).$$

Applying Assumptions 4.1 (b) and (f), we can show that there exists a C such that $\|E(X_tq_{2t}|v_t(\psi_2)=0)\|f_{v_t(\psi_2)}(0)<C<\infty$ and $\|E(\ X_t\varepsilon_tq_{2t}^T|v_t(\psi_2)=0)\|f_{v_t(\psi_2)}(0)<C<\infty$. This completes the proof of the Lemma.

Q.E.D.

LEMMA 2.3. *Under Assumptions 4.1-4.2, we have*

$$\underset{\theta\in\Theta}{sup}\left\|g_n^s(\theta)-g_n(\theta)\right\|\overset{p}{\to}0, \tag{2.0.1}$$

and hence,

$$\underset{\theta\in\Theta}{sup}\left\|g_n^s(\theta)-E\left(g_t(\theta)\right)\right\|\overset{p}{\to}0. \tag{2.0.2}$$

PROOF. Note that, by Hölder's inequality, we can prove (2.0.1) by showing

$$\underset{\theta\in\Theta}{sup}\|g_n^s(\theta)-g_n(\theta)\|=\underset{\theta\in\Theta}{sup}\left\|\frac{1}{n}\sum_{t=1}^{n}z_t\delta^T\tilde{x}_t\left(K(\frac{q_{1t}+q_{2t}^T}{h_n})-I\left(q_{1t}+q_{2t}^T\psi>0\right)\right)\right\|$$
$$\le\left(\underset{\theta\in\Theta}{sup}-\frac{1}{n}\sum_{t=1}^{n}\left\|z_t\delta^T\tilde{x}_t\right\|^a\right)^{1/a}\left(\underset{\theta\in\Theta}{sup}\frac{1}{n}\sum_{t=1}^{n}|K(\frac{q_{1t}+q_{2t}^T\psi}{h_n})-I\left(q_{1t}+q_{2t}^T\psi>0\right)|^b\right)^{1/b},$$

where the first term is bounded by Assumptions 4.1 (a) and (b), and the second term to zero almost surely by Lemma 4 of Horowitz (1992), provided that, for any $\eta>0$, $\frac{1}{n}\sum_{t=1}^{n}I\left(|q_{1t}+q_{2t}^T\psi|<\eta\right)$ converges to $Pr\left(|q_{1t}+q_{2t}^T\psi|<\eta\right)$, almost surely uniformly over $\psi\in\Theta_\psi$, which follows Lemma 2.1 since

$$I\left(|q_{1t}+q_{2t}^T\psi|<\eta\right)=I\left(q_{1t}+q_{2t}^T\psi<\eta\right)-I\left(q_{1t}+q_{2t}^T\psi\le-\eta\right).$$

Then, under Assumptions 4.1-4.2 and applying Lemma 1 we concludes the proof of (2.0.2).

Q.E.D.

LEMMA 2.4. *Under Assumption 6.1, we have*

$$\underset{\theta\in\Theta}{sup}\left\|g_n^{sfd}(\theta)-g_n^{fd}(\theta)\right\|\xrightarrow{p}0,$$

where $g_n^{fd}(\theta)=\frac{1}{n}\sum_{i=1}^n g_i^{fd}(\theta)$.

PROOF. For $t=t_0,\dots,T$, by triangular inequality, we have

$$\underset{\theta\in\Theta}{sup}\left\|g_n^{sfd}(\theta)-g_n^{fd}(\theta)\right\|\le\underset{\theta\in\Theta}{sup}\left\|\begin{bmatrix}\frac{1}{n}\sum_{i=1}^n z_{it_0}\delta^T\tilde{x}_{it_0}\left(I(q_{i,1t_0}+q_{i,2t_0}^T\psi>0)-K(\frac{q_{i,1t_0}+q_{i,2t_0}^T\psi}{h_n})\right)\\ \cdots\\ \frac{1}{n}\sum_{i=1}^n z_{iT}\delta^T\tilde{x}_{iT}\left(I(q_{i,1T}+q_{i,2T}^T\psi>0)-K(\frac{q_{i,1T}+q_{i,2T}^T\psi}{h_n})\right)\end{bmatrix}\right\|$$
$$+\underset{\theta\in\Theta}{sup}\left\|\begin{bmatrix}\frac{1}{n}\sum_{i=1}^n z_{it_0}\delta^T\tilde{x}_{it_0-1}\left(I(q_{i,1t_0-1}+q_{i,2t_0-1}^T\psi>0)-K(\frac{q_{i,1t_0-1}+q_{i,2t_0-1}^T\psi}{h_n})\right)\\ \cdots\\ \frac{1}{n}\sum_{i=1}^n z_{iT}\delta^T\tilde{x}_{iT-1}\left(I(q_{i,1T-1}+q_{i,2T-1}^T\psi>0)-K(\frac{q_{i,1T-1}+q_{i,2T-1}^T\psi}{h_n})\right)\end{bmatrix}\right\|.$$

Then, by applying Hölder's inequality and following the proof of Lemma 2.3[3], we conclude the proof.

Q.E.D.

S3. MONTE CARLO SIMULATION

In this section, we investigate the finite sample performance of the smoothed GMM estimator. We use the following structure to carry out the simulations:

$$\begin{aligned} y_t &= I(q_{1t}+q_{2t}\le 0)+e_t, \\ e_t &= 0.1\varepsilon_t+k_1v_{q_{1t}}+k_2v_{q_{2t}}, \\ q_{1t} &= 0.5q_{1t-1}+v_{q_{1t}}, \\ q_{2t} &= 0.5q_{2t-1}+v_{q_{2t}}, \end{aligned} \tag{3.0.1}$$

where $v_{q_{1t}}$, $v_{q_{2t}}$ and ε_t are independently normally distributed with mean zero and variance one.

We let q_{1t} and q_{2t} follow an $AR(1)$ process, $I(.)$ is the indication function and $\psi_0=1$. The degree of endogeneity of the threshold variable is controlled by k_1 and k_2. We use q_{1t-1} and q_{2t-1} as the instrument for q_{1t} and q_{2t} respectively.

Clearly, this data generation process (DGP) is a simpler version of the general model, $y_t=x_t^T\beta+\delta^T\tilde{x}_tI(q_{1t}+q_{2t}^T\psi>0)+e_t$, with $\beta=0$, $\delta=1$ and $x_t=\tilde{x}_t=1$ for all $t=1,2,...$ We estimate the model with the smoothed least squares (LS) method of Seo and Linton (2007) and the smoothed GMM estimator. For both the smoothed LS and the smoothed GMM, we use the same kernel function and the bandwidth choice with the simulations

[3]The proof here is easier than Lemma 2.3 since we can directly apply Lemma 4 of Horowitz (1992) by Assumption 6.1 (a).

reported in Seo and Linton (2007). We use 2000 replications with sample sizes n = 100, 300 and 500 respectively. To investigate the endogeneity in threshold variable, we vary k_1 and k_2 with values 0, 0.3 & 0.5. All simulations are executed in Matlab.

S3.1. Bias, MSE, and Standard Deviation

For each simulation, we report the MSE, average Bias, and the standard deviation of the threshold estimates. Tables A1 - A9 report the simulation results. Specifically, Table 1 reports results with both exogenous q_1 and q_2. Tables 2 reports the results with endogenous q_1 and q_2.

Note that the DGP is designed with a fixed threshold effect. Hence, the smoothed GMM estimator converges at the normal root-n rate, whereas the convergence rate of the smoothed LS estimator is $\sqrt{\frac{n}{h}}$. In Table A1, for the linear threshold estimate ψ, we observe the smoothed LS estimator does converge at a faster rate than the smoothed GMM estimator, which confirms the super-consistency of the smoothed LS estimator.

Table A9 reports the results with both endogenous q_1 and q_2, where $k_1 = k_2 = 0.5$. We observe a large bias in the smoothed LS estimator, and the bias cannot shrink as sample size increases. This is consistent with our expectation since the smoothed LS estimator has an asymptotic bias with endogenous threshold variables. In contrast, all MSEs of our proposed smoothed GMM estimator decreases to zero as the sample size increase, confirming the consistency.

S3.2. Coverage Probability

We also explore the coverage probability with the data generation process (3.0.1). Following Seo and Shin (2016), we choose the bandwidth by the Silverman's rule of thumb multiplied by h, and report the results for $h = (0.5, 1, 1.5)$. Table A3 reports the empirical coverage probabilities of the 95% confidence intervals for the smoothed GMM estimator with both exogenous q_1 and q_2, exogenous q_2 and endogenous q_1, exogenous q_1 and endogenous q_2 respectively. As n increases, we observe the coverage steadily improves for both exogenous and endogenous cases. We also observe the coverage increases as h increases. Interestingly, unlike the bias and MSE, the coverage for ψ seems to be better than those of β and δ, especially for the case when q_1 is endogenous.

S3.3. More Monte Carlo Results

This subsection provides additional simulation results of the above DGP. Table A4 and A6 show the results of small and large endogenous q_1 only. Table A5 and A7 show the results of small and large endogenous q_2 only. The rest of the tables present results of both endogenous q_1 and q_2 with varying degrees.

Table A1. Finite Sample Performance of the Smoothed Least Squares Estimator and the Smoothed GMM Estimator. k_1=k_2=0 (Exogenous Case)

MSE

	Smoothed LS			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ
100	0.0065	0.0009	0.0021	0.0413	0.0017	0.0055
300	0.0014	0.0002	0.0005	0.0120	0.0004	0.0015
500	0.0007	0.0001	0.0003	0.0058	0.0003	0.0008

Bias

	Smoothed LS			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ
100	0.0052	0.0128	-0.0251	0.0102	0.0032	-0.0063
300	0.0010	0.0061	-0.0119	0.0044	0.0013	-0.0020
500	0.0006	0.0035	-0.0071	0.0030	0.0002	-0.0004

Standard Error

	Smoothed LS			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ
100	0.0802	0.0268	0.0379	0.2030	0.0406	0.0739
300	0.0375	0.0136	0.0195	0.1096	0.0207	0.0383
500	0.0259	0.0101	0.0142	0.0764	0.0159	0.0289

Note: This table reports the simulation results of the smoothed LS estimator and the smoothed GMM estimator for the DGP defined by equation (3.0.1) with exogenous q_{1t} and q_{2t}. The first column shows the sample size. The second to the fourth columns report the results of the smoothed LS estimator for ψ, β & δ respectively. The fifth to the seventh columns present the results of the smoothed GMM estimator.

S4. FINITE SAMPLE PERFORMANCE OF THE TEST FOR LINEARITY

To assess the finite sample performance of the linearity test, we use a simple model,

$$\begin{aligned} y_t &= bI(q_{1t} + q_{2t} \leq 0) + \varepsilon_t, \\ q_{1t} &= 0.5q_{1t-1} + v_{q1t}, \\ q_{2t} &= 0.5q_{2t-1} + v_{q2t}, \end{aligned}$$

where v_{q1t}, v_{q2t} and ε_t are independently normally distributed with mean zero and variance one.

The simulations are done for five sample sizes, $n = 50$, $n = 100$, $n = 200$, $n = 300$, $n = 500$, and five threshold effects, $b = 0$, $b = 0.2$, $b = 0.5$, $b = 0.8$, $b = 1$. We report the results in Table S4 and S4. The replication number is 2000. Throughout the analysis, we use a significance level of 5%. As expected, for both testings, size is approaching to 5% as sample size increases. Power is increasing in b, and increasing in n.

Table A2. Finite Sample Performance of the Smoothed Least Square Estimator, the GMM Estimator, and the Smoothed GMM Estimator. k_1=0.5, k_2 = 0.5 (Endogenous q_1 and q_2)

MSE

	Smoothed LS			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ
100	2.3009	0.3519	1.4086	0.6175	0.0560	0.2080
300	1.8697	0.3469	1.3855	0.2094	0.0135	0.0445
500	1.5386	0.3500	1.4012	0.1353	0.0082	0.0275

Bias

	Smoothed LS			Smoothed GMM		
n	ψ	β	δ		ψ	β
δ						
100	-1.0058	0.5833	-1.1698	-0.0585	0.0592	-0.1167
300	-1.0648	0.5847	-1.1693	0.0554	0.0093	-0.0203
500	-1.0177	0.5890	-1.1789	0.0474	0.0011	-0.0041

Standard Error

	Smoothed LS			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ
100	1.1357	0.1079	0.2008	0.7838	0.2291	0.4410
300	0.8581	0.0710	0.1353	0.4543	0.1160	0.2100
500	0.7093	0.0558	0.1070	0.3649	0.0903	0.1659

Note: This table reports the simulation results of the smoothed LS estimator and the smoothed GMM estimator for the DGP defined by equation (3.0.1) with endogenous q_{1t} and q_{2t}. The first column shows the sample size. The second to the fourth columns report the results of the smoothed LS estimator for ψ, β & δ respectively. The fifth to the seventh columns present the results of the smoothed GMM estimator.

S5. SYSTEM GMM IN THE THRESHOLD MODEL

It is well-documented that, in many empirical problems, for example, the democracy-growth nexus as we discuss more in Section 7 of the main paper, we need to tackle the weak instrument problem in the presence of a highly persistent regressor of a FD estimator. In this section, following Blundell and Bond (1998), we propose to use a system GMM to alleviate the weak instrument issue.

S5.1. The Problem of Weak Instruments and A System GMM Estimator

The most natural moment conditions for dynamic panels is proposed by Arellano and Bond (1991) and can be expressed as $E(y_{it-s}\Delta\varepsilon_{it}) = 0$, for $t = 3, ..., T$ and $s \geq 2$. Thus,

Table A3. Coverage Frequency of the Smoothed GMM Estimator

k_1=k_2=0 (Exogenous Case)

	$h=0.5$			$h=1$			$h=1.5$		
n	ψ	β	δ	ψ	β	δ	ψ	β	δ
100	0.817	0.797	0.771	0.829	0.797	0.771	0.835	0.797	0.771
300	0.875	0.807	0.774	0.904	0.807	0.774	0.901	0.807	0.774
500	0.875	0.820	0.797	0.903	0.820	0.797	0.906	0.820	0.797

k_1=0.5, k_2=0 (Endogenous q_1)

	$h=0.5$			$h=1$			$h=1.5$		
n	ψ	β	δ	ψ	β	δ	ψ	β	δ
100	0.853	0.775	0.741	0.936	0.775	0.741	0.963	0.775	0.741
300	0.850	0.865	0.854	0.919	0.865	0.854	0.944	0.865	0.854
500	0.889	0.892	0.887	0.932	0.892	0.887	0.941	0.892	0.887

k_1=0, k_2=0.5 (Endogenous q_2)

	$h=0.5$			$h=1$			$h=1.5$		
n	ψ	β	δ	ψ	β	δ	ψ	β	δ
100	0.621	0.794	0.761	0.677	0.794	0.761	0.710	0.794	0.761
300	0.813	0.858	0.845	0.851	0.858	0.845	0.866	0.858	0.845
500	0.830	0.886	0.890	0.878	0.886	0.890	0.885	0.886	0.890

Note: This table reports the coverage probability results of the smoothed GMM estimator for the DGP defined by equation (3.0.1) with both exogenous and endogenous threshold variables, q_{1t}, q_{2t}. The first column shows the sample size. The second to the fourth columns report the results with the $h=0.5$. The fifth to the seventh columns present the results with the $h=1$. The last three columns report the results with $h=1.5$.

we can use following instrument matrix,

$$Z_i^{AB} = \begin{bmatrix} y_{i1} & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & y_{i1} & y_{i2} & \dots & 0 & \dots & 0 \\ . & . & . & \dots & . & \dots & . \\ 0 & 0 & 0 & \dots & y_{i1} & \dots & y_{iT-2} \end{bmatrix}, \tag{5.0.1}$$

to construct the sample moments (6.12) in the main paper.

However, as underscored in Blundell and Bond (1998), for a dynamic panel model with endogeneity, the instrument set (5.0.1) becomes less informative as the coefficient of y_{it-1} increases toward unity. Similarly, for the dynamic panels with threshold effect and endogeneity, we may expect the weak instrument problem also occurs when $\beta+\delta$ increases toward unity. To elaborate this point, we consider a heuristic model with $T=3$, and $x_{it}=\tilde{x}_{it}=y_{it-1}$ for all $i=1,..,N$ and $t=1,...,T$. Therefore, model (6.10) of the main paper becomes

$$y_{i2} = \beta_1 y_{i1} + \delta y_{i1} I(q_{i,12} + q_{i,22}^{\top}\psi > 0) + \eta_i + \varepsilon_{i2}, \tag{5.0.2}$$

for $t=2$.

Table A4. Finite Sample Performance of the Smoothed Least Square Estimator, the GMM Estimator, and the Smoothed GMM Estimator. k_1=0.3, $k_2 = 0$ (Endogenous q_1)

MSE

	Smoothed LS			GMM			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ	ψ	β	δ
100	0.1709	0.0274	0.1073	0.1286	0.0100	0.0323	0.1796	0.0107	0.0351
300	0.0488	0.0237	0.0939	0.0510	0.0030	0.0100	0.0510	0.0030	0.0100
500	0.0306	0.0228	0.0903	0.0294	0.0017	0.0058	0.0270	0.0017	0.0057

Bias

	Smoothed LS			GMM			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ	ψ	β	δ
100	0.3563	0.1603	-0.3213	0.0488	0.0091	-0.0180	0.1149	0.0217	-0.0430
300	0.2086	0.1522	-0.3045	0.0275	0.0045	-0.0095	0.0395	0.0074	-0.0150
500	0.1684	0.1499	-0.2994	0.0175	0.0010	-0.0021	0.0215	0.0027	-0.0060

Standard Error

	Smoothed LS			GMM			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ	ψ	β	δ
100	0.2096	0.0412	0.0638	0.3554	0.0997	0.1789	0.4080	0.1009	0.1823
300	0.0728	0.0228	0.0351	0.2241	0.0542	0.0995	0.2224	0.0540	0.0990
500	0.0476	0.0173	0.0261	0.1705	0.0416	0.0760	0.1631	0.0412	0.0750

Note: This table reports the simulation results of the smoothed LS estimator, the GMM estimator, and the smoothed GMM estimator for the DGP defined in Section S3 with endogenous q_{1t} and exogenous q_{2t}. The first column shows the sample size. The second to the fourth columns report the results of the smoothed LS estimator for ψ, β & δ respectively. The fifth to the seventh columns present the results of the GMM estimator. The last three columns show the results of the smoothed GMM estimator.

By deducting both sides by y_{i1}, we have

$$\Delta y_{i2} = \begin{cases} (\beta_1 - 1)y_{i1} + \eta_i + \varepsilon_{i2}, \; q_{i,12} + q_{i,22}^{\top}\psi \le 0, \\ (\beta_1 - 1 + \delta)y_{i1} + \eta_i + \varepsilon_{i2}, \; q_{i,12} + q_{i,22}^{\top}\psi > 0. \end{cases} \tag{5.0.3}$$

As β_1 approaches to unity, y_{i1} can be less informative to work as an instrument for Δy_{i2} in the low regime. Similarly, as $\beta_1 + \delta$ approaches to unity, y_{i1} is less informative in the high regime.

To deal with above weak instrument problem, following Blundell and Bond (1998), we exploit the additional moment condition, $E(\Delta y_{it-1}u_{it})$, where $u_{it} = \eta_i + \varepsilon_{it}$, and construct a instrument system that can be expressed as

$$Z_i^{sys} = \begin{bmatrix} Z_i^{AB} & 0 & 0 & \dots & 0 \\ 0 & \Delta y_{i2} & 0 & \dots & 0 \\ 0 & 0 & \Delta y_{i3} & \dots & 0 \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & \Delta y_{iT-1} \end{bmatrix}, \tag{5.0.4}$$

Table A5. Finite Sample Performance of the Smoothed Least Square Estimator, the GMM Estimator, and the Smoothed GMM Estimator. k_1=0, k_2 = 0.3 (Endogenous q_2)

MSE

	Smoothed LS			GMM			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ	ψ	β	δ
100	0.0738	0.0277	0.1068	0.1172	0.0093	0.0321	0.1299	0.0090	0.0315
300	0.0311	0.0233	0.0931	0.0523	0.0029	0.0095	0.0528	0.0029	0.0096
500	0.0213	0.0226	0.0902	0.0333	0.0017	0.0055	0.0317	0.0017	0.0055

Bias

	Smoothed LS			GMM			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ	ψ	β	δ
100	-0.2471	0.1615	-0.3206	-0.0900	0.0110	-0.0209	-0.0414	0.0236	-0.0453
300	-0.1695	0.1510	-0.3030	-0.0050	0.0009	-0.0025	0.0087	0.0038	-0.0083
500	-0.1415	0.1494	-0.2991	0.0080	-0.0003	0.0005	0.0120	0.0013	-0.0031

Standard Error

	Smoothed LS			GMM			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ	ψ	β	δ
100	0.1127	0.0406	0.0631	0.3304	0.0960	0.1779	0.3581	0.0921	0.1718
300	0.0482	0.0226	0.0351	0.2287	0.0540	0.0976	0.2297	0.0541	0.0979
500	0.0349	0.0179	0.0277	0.1824	0.0412	0.0743	0.1777	0.0412	0.0742

Note: This table reports the simulation results of the smoothed LS estimator, the GMM estimator, and the smoothed GMM estimator for the DGP defined in Section S3 with endogenous q_2 and exogenous q_1. The first column shows the sample size. The second to the fourth columns report the results of the smoothed LS estimator for ψ, β & δ respectively. The fifth to the seventh columns present the results of the GMM estimator. The last three columns show the results of the smoothed GMM estimator.

where Z_i^{AB} is defined in (5.0.1).

Using (5.0.4) to estimate model (6.10) in the main paper essentially extends the system GMM method to a nonlinear setting. Below, we carry out a small simulation and show the proposed smoothed system GMM estimator has a much better finite sample performance than the smoothed FD-GMM estimator in the presence of a persistent regressor in the dynamic panel context. To this end, we recommend using the system GMM if there is a persistent regressor.

S5.2. Monte Carlo study: System GMM in the Threshold Model

This subsection reports a Monte Carlo study that investigates the finite sample performance of the system GMM in the presence of a highly persistent regressor.

Table A6. Finite Sample Performance of the Smoothed Least Squares Estimator and the Smoothed GMM Estimator. k_1=0.5, $k_2 = 0$ (Endogenous q_1)

MSE

	Smoothed LS			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ
100	4.8371	0.4161	1.6623	0.3494	0.0252	0.0886
300	4.7505	0.3772	1.5041	0.1246	0.0069	0.0234
500	4.6371	0.3678	1.4688	0.0732	0.0040	0.0135

Bias

	Smoothed LS			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ
100	-1.4103	0.5804	-1.1642	0.1536	0.0225	-0.0501
300	-1.3636	0.5504	-1.1002	0.0845	0.0063	-0.0134
500	-1.3951	0.5432	-1.0857	0.0443	0.0021	-0.0034

Standard Error

	Smoothed LS			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ
100	1.6881	0.2815	0.5541	0.5710	0.1573	0.2934
300	1.7008	0.2726	0.5422	0.3428	0.0830	0.1524
500	1.6408	0.2698	0.5387	0.2670	0.0631	0.1163

Note: This table reports the simulation results of the smoothed LS estimator, the GMM estimator, and the smoothed GMM estimator for the DGP defined in Section S3 with endogenous q_{1t} and exogenous q_{2t}. The first column shows the sample size. The second to the fourth columns report the results of the smoothed LS estimator for ψ, β & δ respectively. The fifth to the seventh columns present the results of the smoothed GMM estimator.

S5.3. Design

We consider the following data generation processes (DGP),

$$y_{it} = \beta y_{it-1} + \delta y_{it-1} I(q_{1,it} + \psi_1 + \psi_2 q_{2,it} \leq 0) + 0.2\eta_i + v_{it}, \tag{5.0.5}$$

for $i = 1, .., N$ and $t = 1, ..., T$, where in each case $q_{1,it}$, $q_{2,it}$,, η_i, α_t, and v_{it} are drawn as mutually independent i.i.d. N(0,1) random variable.

Model (5.0.5) extends the DGP in Blundell and Bond (1998) to a dynamic panel model with a threshold effect. To investigate the sensitivity of the relevance of the instrument to the persistence, we vary β with values $(0.5, 0.95)$ and δ with values $(-0.1, -0.5)$. We use 599 replications with $N \in (100, 200)$ respectively. We fix $T = 10$.

Table A12-A13 report simulation results of estimations with a linear and a nonlinear DGP, respectively. We examine the mean and root mean square error (RMSE) for each estimator. First, from Table A12, we confirm the results of Blundell and Bond (1998) that system GMM performs better by overcoming the weak instrument problem in the

Table A7. Finite Sample Performance of the Smoothed Least Squares Estimator and the Smoothed GMM Estimator. k_1=0, k_2=0.5 (Endogenous q_2)

MSE

	Smoothed LS			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ
100	1.7200	0.0501	0.1895	0.3200	0.0286	0.1023
300	1.4364	0.0510	0.2006	0.1177	0.0072	0.0236
500	1.3417	0.0525	0.2076	0.0771	0.0041	0.0138

Bias

	Smoothed LS			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ
100	-1.2012	0.2153	-0.4263	-0.0821	0.0362	-0.0738
300	-1.0790	0.2234	-0.4453	0.0233	0.0048	-0.0115
500	-1.0193	0.2277	-0.4541	0.0202	0.0029	-0.0084

Standard Error

	Smoothed LS			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ
100	0.5266	0.0612	0.0879	0.5598	0.1652	0.3114
300	0.5218	0.0334	0.0479	0.3424	0.0845	0.1533
500	0.5504	0.0265	0.0380	0.2770	0.0643	0.1172

Note: This table reports the simulation results of the smoothed LS estimator, the GMM estimator, and the smoothed GMM estimator for the DGP defined in Section S3 with endogenous q_{2t} and exogenous q_{1t}. The first column shows the sample size. The second to the fourth columns report the results of the smoothed LS estimator for ψ, β & δ respectively. The fifth to the seventh columns present the results of the smoothed GMM estimator.

linear dynamic panel setup. Next, under the true model is a nonlinear model ($\delta \neq 0$), from Table A13, we uniformly observe FD-GMM has a much higher RMSE for β when the true $\beta = 0.95$ than the one when true $\beta = 0.5$. We find the system GMM significantly reduces the RMSE β regardless of the value of true β. Our results show that the system GMM has a good finite sample performance. Overall, our simulation results support the advantage of the system instrument in the linear model as addressed in Blundell and Bond (1998) can be also extended to the nonlinear model.

S6. ADDITIONAL RESULTS FOR THE DEMOCRACY-GROWTH NEXUS

S6.1. Linear Model with Interaction Term

As hinted in the paper, we posit the existence of a nonlinear impact of democracy on economic growth. To explore this hypothesis, an intuitive approach is to examine the interactions between democracy and the level of institutional quality. However, it is im-

Table A8. Finite Sample Performance of the Smoothed Least Square Estimator, the GMM Estimator, and the Smoothed GMM Estimator. k_1=0.3, k_2 = 0.5 (Endogenous q_1 and q_2)

MSE

	Smoothed LS			GMM			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ	ψ	β	δ
100	6.0640	0.1407	0.5587	0.3878	0.0438	0.1528	0.4800	0.0442	0.1569
300	6.5514	0.1437	0.5707	0.1396	0.0090	0.0304	0.1546	0.0089	0.0303
500	6.7359	0.1440	0.5737	0.1001	0.0057	0.0193	0.1011	0.0056	0.0192

Bias

	Smoothed LS			GMM			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ	ψ	β	δ
100	-2.2108	0.3708	-0.7425	-0.1267	0.0411	-0.0808	-0.0476	0.0632	-0.1259
300	-2.3208	0.3779	-0.7541	0.0101	0.0059	-0.0133	0.0365	0.0119	-0.0249
500	-2.3576	0.3788	-0.7567	0.0328	0.0027	-0.0062	0.0434	0.0063	-0.0134

Standard Error

	Smoothed LS			GMM			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ	ψ	β	δ
100	1.0848	0.0567	0.0857	0.6099	0.2052	0.3825	0.6914	0.2006	0.3756
300	1.0798	0.0302	0.0442	0.3736	0.0948	0.1738	0.3916	0.0938	0.1722
500	1.0854	0.0233	0.0337	0.3148	0.0752	0.1387	0.3151	0.0749	0.1380

Note: This table reports the simulation results of the smoothed LS estimator, the GMM estimator, and the smoothed GMM estimator for the DGP defined in Section S3 with endogenous q_1 and q_2. The first column shows the sample size. The second to the fourth columns report the results of the smoothed LS estimator for ψ, β & δ respectively. The fifth to the seventh columns present the results of the GMM estimator. The last three columns show the results of the smoothed GMM estimator.

portant to note that, compared to the threshold approach, the interaction term overlooks the potential heterogeneous impact of other variables, including democracy itself.

With this in mind, this subsection aims to present the estimates of the linear model with interaction terms. Table A14 provides the results from this analysis. We observe that the interaction term is significantly positive for the FD-GMM estimation, implying that the impact of democracy does indeed depend on the level of current institutional quality. However, this significant impact diminishes when we employ the system-GMM approach to estimate the model. This intriguing contradiction motivates us to utilize a threshold approach to further investigate the potential nonlinearity between democracy and growth.

S6.2. Full Results of Estimates Using Alternative Dichotomous Measures of Democracy

Table 3 in the main paper presents the linear FD-GMM and threshold smoothed-FD-GMM estimates of the impact of democracy on GDP per capita, using various alternative

Table A9. Finite Sample Performance of the Smoothed Least Square Estimator, the GMM Estimator, and the Smoothed GMM Estimator. k_1=0.5, $k_2 = 0.3$ (Endogenous q_1 and q_2)

MSE

	Smoothed LS			GMM			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ	ψ	β	δ
100	1.9676	0.3735	1.4915	0.3040	0.0326	0.1087	0.3947	0.0335	0.1130
300	1.8073	0.3575	1.4297	0.1375	0.0097	0.0334	0.1477	0.0095	0.0324
500	1.7583	0.3584	1.4332	0.1027	0.0057	0.0198	0.1004	0.0057	0.0197

Bias

	Smoothed LS			GMM			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ	ψ	β	δ
100	-1.2314	0.6023	-1.2058	-0.0370	0.0256	-0.0498	0.0466	0.0484	-0.0967
300	-1.1964	0.5906	-1.1817	0.0482	0.0026	-0.0032	0.0647	0.0099	-0.0178
500	-1.2072	0.5924	-1.1852	0.0419	0.0000	-0.0004	0.0452	0.0030	-0.0068

Standard Error

	Smoothed LS			GMM			Smoothed GMM		
n	ψ	β	δ	ψ	β	δ	ψ	β	δ
100	0.6720	0.1039	0.1939	0.5503	0.1788	0.3259	0.6267	0.1765	0.3221
300	0.6133	0.0934	0.1822	0.3677	0.0987	0.1828	0.3789	0.0969	0.1792
500	0.5488	0.0864	0.1692	0.3177	0.0756	0.1408	0.3136	0.0755	0.1404

Note: This table reports the simulation results of the smoothed LS estimator, the GMM estimator, and the smoothed GMM estimator for the DGP defined in Section S3 with endogenous q_1 and q_2. The first column shows the sample size. The second to the fourth columns report the results of the smoothed LS estimator for ψ, β & δ respectively. The fifth to the seventh columns present the results of the GMM estimator. The last three columns show the results of the smoothed GMM estimator.

Table A10. Rejection Probabilities of the Linearity Test for the GMM Estimator

	Sample Size				
	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 500$
$b = 0$	0.0955	0.077	0.0705	0.0645	0.0565
$b = 0.2$	0.088	0.1304	0.1959	0.2749	0.4313
$b = 0.5$	0.2699	0.5112	0.8446	0.9615	0.999
$b = 0.8$	0.5972	0.9115	0.9975	0.9995	0.9995
$b = 1$	0.8186	0.9945	0.9995	0.9995	0.9995

Note: This table presents the rejection rate of the linearity test for the GMM estimator. The first column gives the different settings of the sample splittings. With $b = 0$, there is no threshold effect. Higher value of b gives higher degree of the threshold effect.

Table A11. Rejection Probabilities of the Linearity Test for the Smoothed GMM Estimator

	Sample Size				
	$n = 50$	$n = 100$	$n = 200$	$n = 300$	$n = 500$
$b = 0$	0.0115	0.0055	0.0140	0.0225	0.0325
$b = 0.2$	0.5815	0.7780	0.9605	0.9935	0.9995
$b = 0.5$	0.9645	0.9940	0.9995	1.0000	1.0000
$b = 0.8$	0.9890	0.9995	1.0000	1.0000	1.0000
$b = 1$	0.9905	0.9985	1.0000	1.0000	1.0000

Note: This table presents the rejection rate of the linearity test for the smoothed GMM estimator. The first column gives the different settings of the sample splittings. With $b = 0$, there is no threshold effect. Higher value of b gives higher degree of the threshold effect.

Table A12. Linear Dynamic Panel Threshold Model Simulation Results, $\delta = 0$

			FD-GMM				System GMM			
δ	N	β	Mean β	RMSE β	Mean δ	RMSE δ	Mean β	RMSE β	Mean δ	RMSE δ
0	100	0.5	0.3887	0.1307			0.4314	0.0911		
		0.95	0.0402	0.9317			0.9573	0.0419		
	200	0.5	0.4463	0.0734			0.4712	0.0511		
		0.95	0.0890	0.8825			0.9724	0.0385		

Note: This table shows the estimation and testing results of the data generation process defined in 5.0.5 by using FD-GMM and system GMM approaches when $\delta = 0$. The first column is the true value ofδ. The second column is the cross-sectional sample size. The third column shows the true value ofβ. Mean β/δ is the average value of β/δ estimates. RMSE β/δ is the root-mean-square error of β/δ estimates.

Table A13. Dynamic Panel Threshold Model Simulation Results, $\delta \in (-0.1, -0.5)$

			FD-GMM				System GMM			
δ	N	β	Mean β	RMSE β	Mean δ	RMSE δ	Mean β	RMSE β	Mean δ	RMSE δ
-0.1	100	0.5	0.3569	0.1951	-0.0469	0.4446	0.3903	0.1614	-0.0470	0.4199
		0.95	0.2512	0.7328	-0.0386	0.1583	0.9281	0.1081	-0.0687	0.1929
	200	0.5	0.4168	0.1630	-0.0725	0.3708	0.4360	0.1542	-0.0693	0.3700
		0.95	0.4295	0.5623	-0.0559	0.1670	0.9767	0.1630	-0.0874	0.2037
-0.5	100	0.5	0.2977	0.3947	-0.1777	0.5653	0.4948	0.2066	-0.1744	0.5569
		0.95	0.6544	0.3898	-0.2625	0.3774	0.9305	0.1258	-0.3417	0.3300
	200	0.5	0.4953	0.3183	-0.4155	0.4406	0.4882	0.2121	-0.3910	0.4386
		0.95	0.6949	0.2850	-0.3504	0.2988	0.9698	0.0553	-0.3657	0.2894

Note: This table shows the estimation and testing results of the data generation process defined in 5.0.5 by using FD-GMM and system GMM approaches when the true model is a nonlinear model. The first column is the true value ofδ. The second column is the cross-sectional sample size. The third column shows the true value ofβ. Mean β/δ is the average value of β/δ estimates. RMSE β/δ is the root-mean-square error of β/δ estimates.

Table A14. Estimation of Linear Model with Interaction Term

	Linear FD-GMM Estimates				Linear System-GMM Estimates			
D_{it}	0.448	-0.132	0.898	1.313	1.757*	1.621*	0.955	-0.856
	(0.915)	(0.857)	(0.879)	(0.909)	(0.955)	(0.976)	(0.852)	(0.863)
Interaction	3.269**	4.571***	2.635*	2.974**	-0.215	-0.155	-0.223	2.058
	(1.414)	(1.390)	(1.397)	(1.434)	(1.289)	(1.335)	(1.280)	(1.274)
y_{it-1}	0.663***	0.501***	0.483***	0.471***	1.003***	0.841***	0.774***	0.732***
	(0.010)	(0.014)	(0.013)	(0.014)	(0.000)	(0.015)	(0.015)	(0.015)
y_{it-2}		0.223***	0.145***	0.160***		0.163***	0.190***	0.173***
		(0.014)	(0.015)	(0.017)		(0.015)	(0.017)	(0.017)
y_{it-3}			0.148***	0.152***			0.040***	0.167***
			(0.012)	(0.012)			(0.013)	(0.016)
y_{it-4}				0.013				-0.069***
				(0.011)				(0.012)
Observations	5748	5595	5442	5289	5748	5595	5442	5289
Countries	154	154	154	154	154	154	154	154

Note: $***$, $**$, and $*$ denote statistical significance at a 1%, 5%, and 10% level, respectively. This table reports the interaction estimates of the effect of democracy on GDP per capita, using ANRR-proposed dichotomous measures of democracy, D. Standard errors robust against heteroskedasticity at the country level are reported in parentheses. We control for a full set of country fixed effects. We use the lagged GDP per capita as instrumental variables.

measures of democracy with a single lag. This subsection provides results for all specifications up to three lags of GDP per capita for each alternative measure. Specifically, Table A15 to Table A18 report the estimation results when we use a dichotomous version of the Freedom House democracy index, Cheibub et al. (2009)'s measure of democracy, Boix et al. (2012)'s measure of democracy, and Polity IV as the democracy measure, denoted as D, in the dynamic threshold regression model.

Overall, except for the cases using Polity IV as the measure[4], we observe that most results remain consistent with those obtained with a single lag. This suggests that our results are robust against variations in the measure of democracy for most specifications.

S7. A HEURISTIC EXAMPLE TO ILLUSTRATE THE SMOOTHNESS OF THE GMM ESTIMATOR

To provide more intuition for the Theorem 4.2 in the paper and Theorem 1.1 in the supplement section S1, we use a simple example to explain the smoothness of the GMM and how the smoothness determines the asymptotic normality. Furthermore, this section also aims to provide some background on the different asymptotic forms of the least square estimator (LSE), the smoothed least square estimator (SLSE), the GMM estimator (GMM), and the smoothed GMM estimator (SGMM). The model considered is defined as follows,

$$y_i = I(q_{1i} + q_{2i}\psi_0 \leq 0) + \varepsilon_i,$$

where $q_{1i},\ q_{2i} \sim U[0,1]$, and $\varepsilon_i \sim N(0,\sigma^2)$.

Thus, we assume all threshold variables are exogenous, and the threshold effect is fixed.

[4]In fact, the insignificance of Polity IV aligns with ANRR's findings.

Table A15. Estimation of Linear and Threshold Model Using House of Freedom as a Measure of Decomcracy

	Linear	Threshold		Linear	Threshold		Linear	Threshold	
ψ_1		-4.2			-0.4			-2.6	
ψ_2		5			3.6			1.7	
		Low	High		Low	High		Low	High
D_{it}	1.251**	1.740***	5.610***	1.458**	1.345***	9.532	1.267**	0.956**	14.131**
	(0.509)	(0.557)	(1.256)	(0.596)	(0.461)	(6.610)	(0.534)	(0.434)	(6.726)
y_{it-1}	0.434***	0.431***	0.431***	0.375***	0.361***	0.791***	0.367***	0.350***	0.921***
	(0.015)	(0.015)	(0.015)	(0.017)	(0.016)	(0.151)	(0.018)	(0.019)	(0.892)
y_{it-2}				0.077***	0.079***	-0.354**	0.056***	0.048***	-0.982
				(0.016)	(0.012)	(0.152)	(0.020)	(0.019)	(1.077)
y_{it-3}							0.058***	0.069***	0.506
							(0.021)	(0.018)	(0.513)
Sup Wald		26.2			27.7			30.6	
Boot P		0.000***			0.000***			0.000***	
Observations	3533	1915	1618	3391	1342	2049	3249	2436	813
Countries	142	142		142	142		142	142	

Note: ***, **, and * denote statistical significance at a 1%, 5%, and 10% level, respectively. This table reports the linear FD-GMM and threshold smoothed-FD-GMM estimates of the effect of democracy on GDP per capita, using the House of Freedom as the dichotomous measure of democracy. The threshold variables used are the ANRR-proposed dichotomous measures of democracy, D, and the polity IV (multiplied by 0.1). Standard errors robust against heteroskedasticity at the country level are reported in parentheses. We control for a full set of country and year fixed effects. We use the jack-knifed average of democracy in a region × initial regime cell proposed by ANRR as the instrumental variables.

Table A16. Estimation of Linear and Threshold Model Using CGV as a Measure of Decomcracy

	Linear	Threshold		Linear	Threshold		Linear	Threshold	
ψ_1		-4			-1.1			-2.2	
ψ_2		5			1.3			1.3	
		Low	High		Low	High		Low	High
D_{it}	0.645**	0.964***	2.776***	1.051***	1.261***	0.797**	0.757**	0.517	0.645*
	(0.319)	(0.325)	(0.643)	(0.335)	(0.337)	(0.346)	(0.353)	(0.338)	(0.384)
y_{it-1}	0.442***	0.440***	0.440***	0.379***	0.358***	0.793***	0.361***	0.343***	0.354
	(0.016)	(0.010)	(0.010)	(0.017)	(0.015)	(0.156)	(0.018)	(0.018)	(0.703)
y_{it-2}				0.093***	0.095***	-0.341**	0.062***	0.054***	-0.425
				(0.016)	(0.007)	(0.155)	(0.020)	(0.018)	(0.953)
y_{it-3}							0.066***	0.073***	0.535
							(0.021)	(0.018)	(0.497)
Sup Wald		60.1			52.8			35	
Boot P		0.000***			0.000***			0.000***	
Observations	3533	1741	1792	3391	1427	1964	3249	2436	813
Countries	142	142		142	142		142	142	

Note: ***, **, and * denote statistical significance at a 1%, 5%, and 10% level, respectively. This table reports the linear FD-GMM and threshold smoothed-FD-GMM estimates of the effect of democracy on GDP per capita, using the CGV proposed by Cheibub et al. (2009) as the dichotomous measure of democracy. The threshold variables used are the ANRR-proposed dichotomous measures of democracy, D, and the polity IV (multiplied by 0.1). Standard errors robust against heteroskedasticity at the country level are reported in parentheses. We control for a full set of country and year fixed effects. We use the jack-knifed average of democracy in a region × initial regime cell proposed by ANRR as the instrumental variables.

Table A17. Estimation of Linear and Threshold Model Using BMR as a Measure of Decomcracy

	Linear	Threshold		Linear	Threshold		Linear	Threshold	
ψ_1		-4			-4.6			-2.6	
ψ_2		5			4			1.6	
		Low	High		Low	High		Low	High
D_{it}	0.615**	1.067***	2.624***	0.280	0.121	3.658***	0.562	0.956***	-16.949***
	(0.311)	(0.295)	(0.580)	(0.295)	(0.267)	(1.353)	(0.339)	(0.307)	(6.126)
y_{it-1}	0.439***	0.441***	0.442***	0.373***	0.360***	0.412	0.364***	0.330***	1.795
	(0.015)	(0.013)	(0.013)	(0.017)	(0.017)	(0.412)	(0.018)	(0.016)	(1.169)
y_{it-2}				0.074***	0.077***	0.024	0.055***	0.010	-2.084
				(0.015)	(0.014)	(0.414)	(0.020)	(0.019)	(1.558)
y_{it-3}							0.065***	0.095***	0.733
							(0.021)	(0.016)	(0.791)
Sup Wald		27.7			15			26.4	
Boot P		0.000***			0.1715			0.000***	
Observations	3533	1741	1792	3391	2287	1104	3391	2436	813
Countries	142	142		142	142		142	142	

Note: ***, **, and * denote statistical significance at a 1%, 5%, and 10% level, respectively. This table reports the linear FD-GMM and threshold smoothed-FD-GMM estimates of the effect of democracy on GDP per capita, using the BMR proposed by Boix et al. (2012) as the dichotomous measure of democracy. The threshold variables used are the ANRR-proposed dichotomous measures of democracy, D, and the polity IV (multiplied by 0.1). Standard errors robust against heteroskedasticity at the country level are reported in parentheses. We control for a full set of country and year fixed effects. We use the jack-knifed average of democracy in a region × initial regime cell proposed by ANRR as the instrumental variables.

Table A18. Estimation of Linear and Threshold Model Using Polity IV as a Measure of Decomcracy

	Linear	Threshold		Linear	Threshold		Linear	Threshold	
ψ_1		-1.2			-0.9			-0.9	
ψ_2		5			1.1			1.1	
		Low	High		Low	High		Low	High
D_{it}	0.066	-2.362***	0.295	0.457	-0.745**	1.344	0.401	-1.267	1.334
	(0.273)	(0.664)	(1.265)	(0.359)	(0.413)	(0.954)	(0.302)	(0.848)	(0.921)
y_{it-1}	0.464***	0.462***	0.457***	0.385***	0.390***	0.854***	0.387***	0.395***	0.821***
	(0.017)	(0.016)	(0.016)	(0.017)	(0.016)	(0.171)	(0.017)	(0.017)	(0.187)
y_{it-2}				0.083***	0.083***	-0.387**	0.082***	0.081***	-0.433*
				(0.015)	(0.009)	(0.171)	(0.015)	(0.010)	(0.257)
y_{it-3}							0.008	0.011	0.095
							(0.018)	(0.014)	(0.126)
Sup Wald		148.6			21			26.3	
Boot P		0.000***			0.0108**			0.000***	
Observations	3644	1520	2124	3502	1503	1999	3360	1415	1945
Countries	142	142		142	142		142	142	

Note: ***, **, and * denote statistical significance at a 1%, 5%, and 10% level, respectively. This table reports the linear FD-GMM and threshold smoothed-FD-GMM estimates of the effect of democracy on GDP per capita, using the polity IV as the measure of democracy. The threshold variables used are the ANRR-proposed dichotomous measures of democracy, D, and the polity IV (multiplied by 0.1). Standard errors robust against heteroskedasticity at the country level are reported in parentheses. We control for a full set of country and year fixed effects. We use the jack-knifed average of democracy in a region × initial regime cell proposed by ANRR as the instrumental variables.

S7.1. The LSE

As shown in Yu and Fan (2020), the LSE can be obtained as,

$$\widehat{\psi} = arg \min_{\psi \in \Theta_\psi} S_n(\psi), \tag{7.0.1}$$

where $S_n(\psi) = \frac{1}{n}\sum_{i=1}^{n}\left(I(\psi_0) + \varepsilon_i - I(\psi)\right)^2$ and $I(\psi) = I\left(q_{1i} + q_{2i}\psi \leq 0\right)$.

Assuming the knowledge of the consistency and the convergence rate, let $\psi = \psi_0 + \frac{v}{n}$. Following Yu and Phillips (2018), we can show the centered process as,

$$D_n^{LSE}(v) = S_n(\psi) - S_n(\psi_0) = n^{-1}\sum_{i=1}^{n}\left(I(\psi_0 + \frac{v}{n}) - I(\psi_0)\right)^2 + 2n^{-1}\sum_{i=1}^{n}\left(I(\psi_0 + \frac{v}{n}) - I(\psi_0)\right)\varepsilon_i.$$

This implies,

$$n(\psi - \psi_0) = \arg\min_v nD_n^{LSE}(v) = \begin{cases} \sum_{i=1}^{N_{1n}(|v|)} \bar{z}_{1i}, & \text{if } v \leq 0 \\ \sum_{i=1}^{N_{2n}(v)} \bar{z}_{2i}, & \text{if } v > 0 \end{cases},$$

where $N_{1n}(|v|) = \sum_{i=1}^{n} I\left(\frac{v}{n} \leq -\frac{q_{1i}+q_{2i}\psi_0}{q_{2i}} \leq 0\right)$, $N_{2n}(v) = \sum_{i=1}^{n} I\left(0 \leq -\frac{q_{1i}+q_{2i}\psi_0}{q_{2i}} \leq \frac{v}{n}\right)$, $\bar{z}_{1i} = 1 + 2\varepsilon_i$, and $\bar{z}_{2i} = 1 - 2\varepsilon_i$

Note that for any finite number v, $N_{2n}(v) \sim B(n, P_n(v))$ where $B(.,.)$ is a binomial process, $P_n(v) = F(0) - F(\frac{v}{n}) \approx f(0)\frac{v}{n}$, where $F(.)$ and $f(.)$ are CDF and PDF of (-$\frac{q_{1i}+q_{2i}\psi_0}{q_{2i}}$) respectively. Let $\lambda = nP_n(v)$. Hence, $\lambda \to f_z(0)v$.[5] As λ is a finite number, we have $N_{2n}(v) \to N_2(v)$. Similarly, we have $N_{1n}(|v|) \to N_1(|v|)$, where $N_1(|v|), N_2(v)$ are two independent Poisson process with intensity $f_z(0)$.

As a result,

$$n(\widehat{\psi} - \psi_0) \xrightarrow{d} \arg\min_v D^{LSE}(v), \tag{7.0.2}$$

where $D^{LSE}(v)$ is a compound Poisson process with the form,

$$D^{LSE}(v) = \begin{cases} \sum_{i=1}^{N_1(|v|)} z_{1i}, & \text{if } v \leq 0 \\ \sum_{i=1}^{N_2(v)} z_{2i}, & \text{if } v > 0 \end{cases},$$

where $z_{1i} = \lim_{\Delta\uparrow 0} \bar{z}_{1i} I\left(\Delta \leq -\frac{q_{1i}+q_{2i}\psi}{q_{2i}} \leq 0\right)$, and $z_{2i} = \lim_{\Delta\downarrow 0} \bar{z}_{2i} I\left(0 \leq -\frac{q_{1i}+q_{2i}\psi}{q_{2i}} \leq \Delta\right)$.

S7.2. The SLSE

Following Seo and Linton (2007), the SLSE can be obtain as,

$$\widehat{\psi}^{SLSE} = \arg\min_{\psi \in \Theta_\psi} S_n^{SLS}(\psi), \tag{7.0.3}$$

where $S_n^{SLS}(\psi) = \frac{1}{n}\sum_{i=1}^{n}(y_i - K(\psi, \sigma_n))^2$, $K(\psi, \sigma_n) = K(\frac{q_{1i}+q_{2i}\psi}{\sigma_n})$, $K(.)$ is a kernel function as defined in assumption 3 of Seo and Linton (2007), and σ_n is the bandwidth parameter.

[5]With "diminishing threshold "assumption, the convergence rate is slower than n. As a result, $\lambda \to \infty$, which implies we have infinitely many $-\frac{q_{1i}+q_{2i}\psi_0}{q_{2i}}$ in the local neighborhood of 0. Therefore, we can apply central limit theorem for a given v.

Note that, unlike the LSE, the objective function in this case is smoothed in ψ. Hence, we can apply the standard first order Taylor series to obtain the asymptotic normality.

By simple calculation, we have,

$$T_n(\psi,\sigma_n) = \frac{\partial S_n^{SLS}(\psi)}{\partial \psi} = -\frac{2}{n}\sum_{i=1}^{n} I(\psi_0)K'(\psi,\sigma_n)\frac{q_{2i}}{\sigma_n} + \frac{2}{n}\sum_{i=1}^{n} K(\psi,\sigma_n)K'(\psi,\sigma_n)\frac{q_{2i}}{\sigma_n} \tag{7.0.4}$$

$$-\frac{2}{n}\sum_{i=1}^{n} K'(\psi,\sigma_n)\frac{q_{2i}}{\sigma_n}\varepsilon_i = A(\psi) + B(\psi) + C(\psi),$$

where $K'(\psi,.) = \frac{\partial K(.,.)}{\partial \psi}$.

First, by assmuption 3(b) of Seo and Linton (2007), we can show,

$$\sigma_n^{-h}A(\psi_0) \xrightarrow{p} \sigma_n^{-h}E\left(I(\psi_0)K'(\psi_0,\sigma_n)\frac{q_{2i}}{\sigma_n}\right) = O(1),$$

where h defines h^{th} order kernel.

This implies, as long as $\sqrt{n\sigma_n}\sigma_n^{-h} \to 0$, $\sqrt{n\sigma_n}A(\psi_0) = o_p(1)$. Similarly, we can show $\sqrt{n\sigma_n}B(\psi_0) \xrightarrow{p} 0$.

Next, similar to the proof of lemma 3 of Seo and Linton (2007), we have,

$$\sqrt{n\sigma_n}C(\psi_0) \xrightarrow{d} N\left(0, V^{\psi}\right),$$

where $V^{\psi} = 4Var\left(K'(\psi_0,\sigma_n)q_{2i}\varepsilon_i\right)$.

Hence, we have,

$$\sqrt{n\sigma_n}T_n\left(\psi_0,\sigma_n\right) \xrightarrow{d} N\left(0, V^{\psi}\right).$$

Then, by the first order Taylor series,

$$T_n\left(\widehat{\psi}^{SLSE},\sigma_n\right) = T_n(\psi_0,\sigma_n) + Q_n\left(\tilde{\psi},\sigma_n\right)\left(\widehat{\psi}^{SLSE} - \psi_0\right) = 0,$$

where $Q_n(\psi) = \frac{\partial T_n(.,.)}{\partial \psi}$, and $\tilde{\psi}$ is between $\widehat{\psi}^{SLSE}$ and ψ_0.

As $\sigma_n Q_n\left(\tilde{\psi},\sigma_n\right) \xrightarrow{p} Q$, we have

$$\sqrt{n\sigma_n^{-1}}\left(\widehat{\psi}^{SLSE} - \psi_0\right) \xrightarrow{d} N\left(0, Q^{-1}V^{\psi}Q^{-1}\right),$$

where $Q = K'(0)E\left(q_{2i}^2|z_i = 0\right)f_z(0)$, $z_i = q_{1i} + q_{2i}\psi_0$, and $f_z(.)$ is the density of z_i.

S7.3. The GMM

We propose to use use the moment condition, $E(q_{2i}\varepsilon_i) = 0$, for all $i = 1, ..., n$. Therefore, the GMM estimator can be obtained as,

$$\widehat{\psi}^{GMM} = \arg\min_{\psi\in\Theta_\psi} S_n^{GMM}(\psi), \tag{7.0.5}$$

where $S_n^{GMM} = \left(\frac{1}{n}\sum_{i=1}^{n} q_{2i}\left(I(\psi_0) + \varepsilon_i - I(\psi)\right)\right)^2$.

Similar to the LSE, the objective function is non-smooth in ψ. Now, assuming the knowledge of the consistency and the converge rate [6], let $\psi = \psi_0 + \frac{v}{n^{1/2}}$. Hence, the centered process can be shown as,

$$D_n^{GMM}(v) = S_n^{GMM}(\psi) - S_n^{GMM}(\psi_0) = n^{-2}\left[\sum_{i=1}^n q_{2i}\left(I(\psi_0) - I(\psi_0 + \tfrac{v}{n^{1/2}})\right)\right]^2 \\ +2n^{-2}\sum_{i=1}^n q_{2i}\left(I(\psi_0) - I(\psi_0 + \tfrac{v}{n^{1/2}})\right)\sum_{i=1}^n q_{2i}\varepsilon_i.$$

Note that, by comparing D_n^{LSE} with D_n^{GMM} , it is obvious that the second term is quite different. For the D_n^{LSE}, the sum of error cannot be isolated from v. As a result, we cannot directly apply the central limit theorem (CLT) [7]. On the contrary, for the D_n^{GMM}, the CLT can be applied to $\sum_{i=1}^n q_{2i}\varepsilon_i$ as long as the multiplier is bounded. The reason comes from the nature of the sample averaging condition.

This implies,

$$n^{1/2}(\psi - \psi_0) = \arg\min_v \; nD_n^{GMM}(v) = \arg\min_v \; n^{-1}\left[\sum_{i=1}^n q_{2i}\left(I(\psi_0) - I(\psi_0 + \tfrac{v}{n^{1/2}})\right)\right]^2 \\ +2n^{-1/2}\sum_{i=1}^n q_{2i}\left(I(\psi_0) - I(\psi_0 + \tfrac{v}{n^{1/2}})\right)\tfrac{1}{n^{1/2}}\sum_{i=1}^n q_{2i}\varepsilon_i = \arg\min_v A^{GMM}(v) + B^{GMM}(v).$$

Then, by the Glivenko-Cantelli theorem, for any v,

$$A^{GMM}(v) \xrightarrow{p} n\left[E\left(q_{2i}\left(I(\psi_0) - I(\psi_0 + \frac{v}{n^{1/2}})\right)\right)\right]^2 = G_\psi(\psi_0)^2 v^2,$$

where $G_\psi(\psi_0) = \frac{dE(q_{2i}I(\psi))}{d\psi}|_{\psi=\psi_0}$.

Similarly, we can show that,

$$n^{-1/2}\sum_{i=1}^n q_{2i}\left(I(\psi_0) - I(\psi_0 + \frac{v}{n^{1/2}})\right) \xrightarrow{p} n^{1/2}E\left(q_{2i}\left(I(\psi_0) - I(\psi_0 + \frac{v}{n^{1/2}})\right)\right) = G_\psi(\psi_0)v.$$

Hence, by applying the CLT and the continous mapping theorem,

$$B^{GMM}(v) \xrightarrow{d} G_\psi(\psi_0)vN(0, \Omega),$$

where $\Omega = Var(q_{2i}\varepsilon_i)$.

This follows that,

$$n^{1/2}\left(\widehat{\psi}^{GMM} - \psi\right) \xrightarrow{d} \hat{v} = \underset{v}{argmin}\; G_\psi(\psi_0)^2 v^2 + G_\psi(\psi_0)vW,$$

where $W \sim N(0, \Omega)$.

Obviously, $\hat{v} = -W/G_\psi(\psi_0)$. This provides the asymptotic normality,

$$n^{1/2}\left(\widehat{\psi}^{GMM} - \psi_0\right) \xrightarrow{d} N\left(0, \left(G_\psi(\psi_0)\Omega^{-1}G_\psi(\psi_0)\right)^{-1}\right).$$

S7.4. The Smoothed GMM

Again, we consider the moment condition $E(q_{2i}\varepsilon)$ for all $i = 1, ..., n$. The smoothed GMM estimator can be obtained as,

$$\widehat{\psi}^{SGMM} = arg\min_{\psi\in\Theta_\psi} S_n^{SGMM}(\psi), \tag{7.0.6}$$

[6]The example is designed with a fixed threshold effect. Hence, the theoretical convergence rate of threshold estimator is $\sqrt{n}$.

[7]With diminishing threshold framework, the functional central limit theorem can be applied to D_n^{LSE}, which leads to a limiting distribution formed by a two-sided Brownian motion (Hansen (2000)). Yu and Phillips (2018) explains on how compound Poisson process can be approximated by two-sided Brownian motion.

where $S_n^{SGMM} = \left[\frac{1}{n}\sum_{i=1}^n q_{2i}\left(I(\psi_0) + \varepsilon_i - K(\psi)\right)\right]^2$, and $K(\psi) = K(\frac{q_{1i}+q_{2i}\psi}{h_n})$.

In contrast to the objective function of GMM method, S_n^{SGMM} is differentiable over ψ. Therefore, we can apply standard Taylor expansion to derive the limiting results by imposing some assumptions on the kernel. Note that S_n^{SGMM} is a quantity of the squared of sample averages. However, S_n^{SLSE} is a quantity of an average over squared term.

By simple calculation, we have

$$T_n^s(\widehat{\psi}^s, h_n) = -\frac{2}{n}\sum_{i=1}^n q_{2i}\left[I(\psi) - K(\psi_0)\right]\frac{1}{n}\sum_{i=1}^n \frac{q_{2i}^2}{h_n}K'(\widehat{\psi}^s) \tag{7.0.7}$$
$$-\frac{2}{n}\sum_{i=1}^n q_{2i}\left[K(\psi_0) - K(\widehat{\psi}^s)\right]\frac{1}{n}\sum_{i=1}^n \frac{q_{2i}^2}{h_n}K'(\widehat{\psi}^s)$$
$$-\frac{2}{n}\sum_{i=1}^n q_{2i}\varepsilon_i\frac{1}{n}\sum_{i=1}^n \frac{q_{2i}^2}{h_n}K'(\widehat{\psi}^s) = -2A^s(\widehat{\psi}^s) - 2B^s(\widehat{\psi}^s) - 2C^s(\widehat{\psi}^s) = 0.$$

Comparing $C^s(.)$ of (7.0.7) to $C(.)$ of (7.0.4), the nature of the sample averaging in the GMM allows us to apply CLT to $\frac{1}{\sqrt{n}}\sum_{t=1}^n q_{2i}\varepsilon_i$ of $C^s(.)$ as $\frac{1}{n}\sum_{i=1}^n \frac{q_{2i}^2}{h_n}K'(\widehat{\psi}^s) = O_p(1)$. As a result, $Var\left(\sqrt{n}C^s(.)\right) = O_p(1)$. In contrast, $Var\left(\sqrt{n}C(\psi)\right) = O_p(h_n^{-1})$ since $K'(\psi,\sigma_n)$ of $C(\psi)$ cannot be isolated from applying the central limit theorem. In addition, the feature of squared of the sample averaging of the GMM objective function makes the first order Taylor expansion use up to the first order of $K(.)$ in smoothed GMM and implies $\frac{\partial T_n^s(\widehat{\psi}^s,h_n)}{\partial\psi} = O_p(1)$. However, the first order Taylor expansion of the smoothed least square objective function use up to the second order of $K(.)$, which follows $\frac{\partial T_n(\widehat{\psi}^{SLS},h_n)}{\partial\psi} = O_p(h_n^{-1})$.

Let $g_i(\psi) = q_{2i}\left(I(\psi_0) + \varepsilon_i - K(\psi)\right)$ and $G_i(\psi) = -\frac{q_{2i}^2}{h_n}K'(\psi)$, we can rewrite $T_n^s(\widehat{\psi}^s, h_n) = 0$ as $\frac{1}{n}\sum_{i=1}^n g_i(\widehat{\psi}^s)\frac{1}{n}\sum_{i=1}^n G_i(\widehat{\psi}^s) = 0$. Applying Taylor expansion and by consistency, we have

$$\sqrt{n}\left(\widehat{\psi}^s - \psi_0\right) = \left(\frac{1}{n}\sum_{i=1}^n G_i(\widehat{\psi}^s)\right)^{-2}\frac{1}{n}\sum_{i=1}^n G_i(\widehat{\psi}^s)\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[q_{2i}\varepsilon_i + q_{2i}\left(I(\psi_0) - K(\psi_0)\right)\right] + o(1). \tag{7.0.8}$$

As shown in the proof of Theorem 4, we can show $\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[q_{2i}\varepsilon_i + q_{2i}\left(I(\psi_0) - K(\psi_0)\right)\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^n q_{2i}\varepsilon_i + o(1) \xrightarrow{d} N(0,\Omega)$ and $\frac{1}{n}\sum_{i=1}^n G_i(\widehat{\psi}^s) \xrightarrow{p} E\left(q_{2i}^2|v_i = 0\right)f_v(0)$, where $v_i = q_{1i} + q_{2i}\psi_0$. Hence, by continuous mapping theorem, we have

$$\sqrt{n}\left(\widehat{\psi}^s - \psi_0\right) \xrightarrow{d} N\left(0, Q_s^{-1}\Omega Q_s^{-1}\right). \tag{7.0.9}$$

REFERENCES

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

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies 58*(2), 277–297.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143.

Boix, C., M. Miller, and S. Rosato (2012). A complete data set of political regimes, 1800–2007. *Comparative Political Studies 46*(12), 1523–1554.

Cheibub, J. A., J. Gandhi, and J. R. Vreeland (2009). Democracy and dictatorship revisited. *Public Choice 143*(1–2), 67–101.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica 68*, 575–603.

Horowitz, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica 60*(3), 505–531.

Lee, S., Y. Liao, M. H. Seo, and Y. Shin (2021). Factor-driven two-regime regression. *The Annals of Statistics 49*(3).

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

Seo, M. H. and O. Linton (2007). A smoothed least squares estimator for threshold regression models. *Journal of Econometrics 141*(2), 704–735.

Seo, M. H. and Y. Shin (2016). Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics 195*(2), 169–186.

Vaart, A. W. v. d. and J. A. Wellner (2000). *Weak convergence and empirical processes: with applications to statistics*. Springer.

Yu, P. and X. Fan (2020). Threshold regression with a threshold boundary. *Journal of Business Economic Statistics*, 1–19.

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