A GMM Estimator for Linear Index Threshold Model

Chaoyi Chen*

September 1, 2019

Abstract

This paper investigates the linear index threshold regression model with endogeneity. We propose a two-step GMM estimation method to estimate the model, which allows both the threshold variable and regressors to be endogenous. We show the consistency of the GMM estimator and derive the asymptotic distribution of the GMM estimator for weakly dependent data. We suggest a test of the exogeneity null hypothesis for both the threshold and the slope regressors. Monte Carlo simulations are used to assess the finite sample performance of our proposed estimator. Finally, we present an empirical application investigating the threshold effect of a linear index between external debt and public debt on economic growth for developing countries.

JEL Classification: C13 C32 O40

Keywords: Endogenous threshold effects and regressors, GMM, Index model, Public debt, Threshold Regression

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

1 Introduction

Parametric threshold regression models are widely used to characterize nonlinearities in economic relationships. There are many applications of threshold regression model in both time series and cross-sectional scenario. Examples include the pricing asymmetry of oil prices and the nonlinear effect of public debt to GDP ratio on the per capita GDP growth. Threshold models allow us to identify the unknown threshold variable and draw inferences. It has been well established that the estimator of the threshold parameter is super-consistent, while the slope regressors estimators converge at the standard square-n rate. However, there are different approaches to obtain the asymptotic distribution of the threshold parameter. Firstly, this can be derived using a "fixed threshold effect" assumption. Chan (1993) establishes that the threshold parameter estimator converges to a functional of a compound Poisson process. Yet, in that case statistical inference is impossible to implement in practice due to the presence of nuisance parameters in the joint distribution of the covariates. Secondly, using a "diminishing threshold effect" assumption, introduced by Hansen (2000), the limiting distribution involves two independent Brownian motions and is available through simulations. In that case, inference can be carried out fairly easily.

However, both Chan (1993) and Hansen (2000) rely on the crucial exogeneity assumption for both the slope regressors and the threshold variable. Recently, there is a growing interest in threshold models that allows for endogeneity. Using a two-stage least square method, Caner and Hansen (2004) allow for the slope regressors to be endogenous. In the spirit of the sample selection methodology of Heckman (1979), with a joint normality assumption, Kourtellos et al. (2016) allow for an endogenous threshold variable. Seo and Shin (2016) propose a two-step GMM estimator for the dynamic panel threshold model, which also allows for endogeneity. It is worth noticing that the GMM method allows for both fixed and small threshold effects and the rate of convergence for the GMM threshold estimator is not super-consistent. By relaxing the joint normality assumption of Kourtellos et al. (2016), Kourtellos et al. (2017) propose a two-step estimation method based on a nonparametric control function approach to correct for threshold endogeneity. The semiparametric threshold model separates the threshold effect into two parts, namely, an exogeneous threshold effect and an endogenous threshold effect. Therefore, with a "small threshold" effect, the convergence rate of the estimator of the threshold variable depends on diminishing rates of both these two effects.

Seo and Linton (2007) consider a more general model than Hansen (2000). In the spirit of Horowitz (1992), they propose a smoothed least square estimator by allowing the threshold to be a linear index of regressors. The linear index threshold regression model can capture the joint threshold effect between two possible threshold variables. Therefore, this model allows empirical researchers to investigate the threshold effect in a broader setting. Yu (2015) develops the limiting asymptotic results of the least square estimator for the linear index threshold model in both the fixed threshold and the diminishing threshold effect framework. However, both the smoothed least square estimator and the least square estimator rely on the assumption of exogeneity in both the slope regressors and the threshold variables, which may limit the usefulness of these models.

In this paper, we propose a two-step GMM linear index threshold estimator, which allows both threshold variables and the regressors to be endogenous. We also relax the fixed threshold effect assumption by allowing both for fixed and diminishing threshold effects. We develop the estimation strategy and the limiting results for weakly dependent data. The asymptotic distribution is similar to Seo and Shin (2016). They concentrate on the dynamic panel threshold model, whereas we focus on the linear index threshold model. Similar to Seo and Shin (2016), the convergence rate of the threshold estimators are $n^{\frac{1}{2}-\alpha}$ and not super-consistent, where α measures the diminishing rate of the threshold effect. The slope coefficients converge at the usual root-n rate. We further suggest a test of the linear index threshold effect and provide a Hausman type test for the exogeneity of the regressors. The finite performance of the proposed estimator are studied through Monte Carlo Simulations. We compare our estimator with the smoothed least square estimator of Seo and Linton (2007) and we report the average bias, mean square error and the standard deviation of the threshold estimator specifically. Finally, we investigate the threshold effect of a linear index model between external debt and public debt in economic growth for developing countries. We estimate the augmented Solow linear index threshold using both GMM method and Seo and Linton (2007). We find that after correcting for endogeneity, the joint threshold effect becomes insignificant.

The rest of the paper is organized as follows. In section 2, we introduce the linear index threshold model with endogeneity. In section 3, we propose the two-step GMM estimator of the linear index threshold model and in section 4 we derive the asymptotic results. Section 5 provides the inference for the threshold effect and the endogeneity in slope regressors. In section 6 we report the Monte Carlo results for the proposed estimators. Section 7 presents the empirical application, while section 8 concludes the paper. In the appendix we collect the proofs and we present additional evidence for the small sample performance of the proposed linearity test and some additional heuristic arguments for the smoothness of the GMM objective function that we adopt in our analysis.

2 The Model

Consider the following linear index threshold model suggested by Seo and Linton (2007)

$$y_t = x_t^T \beta + \delta^T \tilde{x}_t I(q_{1t} + q_{2t}^T \psi > 0) + \varepsilon_t$$
$$t = 1, ..., n \tag{1}$$

where y_t is the dependent variable, x_t is a $k \times 1$ vector and $\tilde{x}_t$ is an $l \times 1$ vector. Also $q_t = [q_{1t}^T, q_{2t}^T]^T$ is an $h \times 1$ threshold variable vector. Note that $x_t, \tilde{x}_t, q_t$ may have common variables. Many models in the previous literature can be viewed as a special case of this model. For example, for the case that $x_t = \tilde{x}_t$, q_{1t} is a constant and q_{2t} is a scalar, the model becomes the threshold model considered by Hansen (2000). If we further assume that x_t consists of the lagged y_t and $q_{2t} = y_{t-d}$, the model becomes the self-exciting threshold autoregressive (SETAR) model suggested by Tong and Lim (1980).

Similar to Seo and Shin (2016), we allow for both "fixed threshold effect" and the "diminishing threshold effect", of Hansen (2000). That is, we have

$$\delta = \delta_n = \delta_0 n^{-\alpha}, \ \alpha \in [0, 1/2) \tag{2}$$

Endogeneity is allowed in both the slope regressors ($E(x_t \varepsilon_t) \neq 0$) and the threshold variables ($E(q_t \varepsilon_t) \neq 0$). To fix the endogeneity problem, we need to find an $m \times 1$ vector of instrumental variables, z_t, for $t = 1, ..., n$, where $m \geq k+l+h-1$, satisfying the following orthogonality condition:

$$E(z_t \varepsilon_t) = 0 \tag{3}$$

for all $t = 1, ..., n$

3 Estimation Strategy

We consider the following moment condition:

$$E(g_t(\theta_n)) = E(z_t \varepsilon_t) = 0 \tag{4}$$

where θ_n is the true value with $\theta_n = [\beta_0^T, \delta_n^T, \psi_0^T]^T$, $\delta_n = \delta_0 n^{-\alpha}$, and

$$g_t(\theta) = z_t[y_t - x_t^T\beta - \delta^T\tilde{x}_t I(q_{1t} + q_{2t}^T\psi > 0)].$$

Naturally, the sample analogue to $E(g_t(\theta))$ is,

$$g_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} g_t(\theta). \tag{5}$$

Given that the general identification condition hold for $E(g_t(\theta_n))$, the GMM estimators can be obtained as

$$\hat{\theta}^{GMM} = \underset{\theta\epsilon\Theta}{argmin}\ Q_n(\theta), \tag{6}$$

where

$$Q_n(\theta) = g_n(\theta)^T W_n g_n(\theta) = [\frac{1}{n}\sum_{t=1}^{n} g_t(\theta)]^T W_n[\frac{1}{n}\sum_{t=1}^{n} g_t(\theta)] \tag{7}$$

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

For a given ψ, the model is linear in β and δ. Since $Q_n(\theta)$ is not continuous in ψ, it is more practical to use a grid search empirically.

For a given ψ and a weight matrix W_n ,

$$\left(\hat{\beta}_{(\psi)}^T, \hat{\delta}_{(\psi)}^T\right)^T = [\hat{G}(\psi)^T W_n \hat{G}(\psi)]^{-1}\hat{G}(\psi)^T W_n[-\frac{1}{n}\sum z_t y_t], \tag{8}$$

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

Then, the threshold index estimators can be obtained by

$$\hat{\psi}^{GMM} = \underset{\psi\in\Theta_\psi}{\text{argmin}}\ Q_n(\psi) = [\frac{1}{n}\sum_{t=1}^{n} g_t(\hat{\beta}_{(\psi)}, \hat{\delta}_{(\psi)}, \psi)]^T W_n[\frac{1}{n}\sum_{t=1}^{n} g_t(\hat{\beta}_{(\psi)}, \hat{\delta}_{(\psi)}, \psi)] \tag{9}$$

and

$$(\hat{\beta}^{GMM^T}, \hat{\delta}^{GMM^T})^T = \left(\hat{\beta}^T(\hat{\psi}), \hat{\delta}^T(\hat{\psi})\right)^T \tag{10}$$

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 $W_n = [\frac{1}{n}\sum_{t=1}^{n}(z_t z_t^T \hat{e}_t^2)]^{-1}$.

4 Asymptotic Results

In this section, we develop the asymptotic theory for the GMM estimator of the linear index threshold model. The regularity assumptions required for deriving the limiting results of the proposed estimator are as follows:

Assumption 1: $\{(X_t,\ z_t,\ q_t,\ \varepsilon_t)\}$ is a sequence of strictly stationary strong mixing random variables with mixing numbers α_s, s = 1,2,...... that satisfies α_s= o ($\mathrm{s}^{-\gamma/(\gamma-1)}$) as s $\to \infty$ for some $\gamma \geq 1$.

Assumption 2: For some $\eta > 1$, $\mathrm{E}||X_t X_t^T||^{\eta+\gamma} < \infty$, $\mathrm{E}||z_t \varepsilon_t||^{\eta+\gamma} < \infty$, $\mathrm{E}||z_t X_t^T||^{\eta+\gamma} < \infty$. $E[z_t X_t^T X_t z_t^T|q_t] > 0$ a.s.

Assumption 3: $\{(\varepsilon_t, \mathcal{F}_{n,t})\}_{t=1}^{n}$ is a martingale difference sequence with $E(\varepsilon_t^2|\mathcal{F}_{n,t-1}) < \infty$, where $\mathcal{F}_{n,t}$ is the smallest sigma-field generated from $\{(X_s^T, z_{s+1}^T, q_s^T, \varepsilon_s) : 1 \leq s < t \leq n\}$, and $Var(\mathrm{n}^{-1/2}\sum_{t=1}^{n} z_t \varepsilon_t)$ is a positive definite matrix.

Assumption 4: The true values of β and ψ are fixed at β_0 and ψ_0 . The true δ depends on n such that $\delta_n = \delta_0 n^{-\alpha}$ for some $\alpha \in [0, \frac{1}{2})$ and $\delta_0 \neq 0$. $\theta_n = [\beta_0^T, \delta_n^T, \psi_0^T]^T$ is an interior point of $\Theta = \Theta_\beta \times \Theta_\delta \times \Theta_\psi$, which is a compact set. $\Omega = E(g_t(\theta_n) g_t(\theta_n)^T)$ is finite and positive definite.

Assumption 5: $E(q_{2t}q_{2t}^T)$ is positive definite.

Assumption 6: For all $\psi \in \Theta_\psi$, the linear index of the threshold variables, $v_t(\psi) = q_{1t} + q_{2t}^T\psi$, has a continuous and bounded density, $f_{v_t(\psi)}(.)$, such that $f_{v_t(\psi)}(0) > 0$; $E\left(z_t\delta_0^T\tilde{x}_t q_{2t}^T|v_t(\psi) = 0\right)$ is continuous at ψ_0

Define:

$$G_\beta = -E(z_t x_t^T), \tag{11}$$

$$G_\delta(\psi) = -E\{z_t\tilde{x}_t^T I(q_{1t} + q_{2t}^T\psi > 0))\}, \tag{12}$$

$$G_\psi(\psi) = -E\left(z_t\delta_0^T\tilde{x}_t q_{2t}^T|v_t(\psi) = 0\right) f_{v_t(\psi)}(0), \tag{13}$$

where G_β is an $m \times k$ matrix, $G_\delta(\psi)$ is an $m \times l$ matrix, $G_\psi(\psi)$ is an $m \times (h-1)$ matrix and $f_{v_t}(0)$ is the density for v_t at $v_t = 0$.

Assumption 7: $G = (G_\beta, G_\delta(\psi_0), G_\psi(\psi_0))$, then G is a full column rank matrix.

Assumption 1 gives standard conditions on the stochastic process. We can apply the generic uniform law of large numbers of Andrews (1987) to prove the consistency of our estimator. Assumptions 2-4 are regularity assumptions of the generalized method moment method. We assume ε_t is the martingale difference sequence. We allow both for fixed threshold effect and for diminishing threshold effect. If $\delta_0 = 0$ (no threshold effect), ψ can not be identified. Assumption 5 corresponds to assumption 1(d) in Seo and Linton (2007) and assumption 6 in Yu (2015). This assumption is required for the asymptotic uniqueness of the GMM estimator. Assumption 6 is a smoothness assumption on the distributions of the threshold variables and their linear index and the conditional moments, which is standard in threshold models. Assumption 7 is the GMM full rank condition .

Theorem 1: Under assumptions 1-7, as n$\rightarrow \infty$, we have

$$\hat{\theta}^{GMM} \xrightarrow{p} \theta_n. \tag{14}$$

Theorem 2: Under assumptions 1-7, as n$\rightarrow \infty$,

$$\begin{bmatrix} \sqrt{n} & 0 & 0 \\ 0 & \sqrt{n} & 0 \\ 0 & 0 & n^{\frac{1}{2}-\alpha} \end{bmatrix} \begin{bmatrix} \hat{\beta} - \beta_0 \\ \hat{\delta} - \delta_n \\ \hat{\psi} - \psi_0 \end{bmatrix} \xrightarrow{d} N\left(0, \left(G^T \Omega^{-1} G\right)^{-1}\right), \tag{15}$$

where Ω and G are defined in assumption 4 and assumption 7.

The convergence rate for the estimator of the slope parameter is standard root-n. The convergence rate for the threshold variables 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 [1].

G_β and G_δ can be estimated as

$$\hat{G}_\beta = -\frac{1}{n}\sum_{t=1}^{n} z_t x_t^T, \tag{16}$$

$$\hat{G}_\delta = -\frac{1}{n}\sum_{t=1}^{n} z_t \tilde{x}_t^T I(q_{1t} + q_{2t}^T \hat{\psi} > 0). \tag{17}$$

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

$$\hat{G}_\psi = -\frac{1}{nh}\sum_{t=1}^{n} z_t \hat{\delta}^T \tilde{x}_t q_{2t}^T K(\frac{q_{1t} + q_{2t}^T \hat{\psi}}{b}), \tag{18}$$

where $K(.)$ 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 (Robinson (1983), Robinson (1986)).

5 Testing

5.1 Test for Linearity

In equation (2), the threshold effect disappears under the null hypothesis, $\delta_n = 0$. However, due to the presence of unidentified parameters under the null, the natural way to test for nonlinearity is the *Sup*-Wald test, which is formed as follows,

[1] We provide a heuristic example in the appendix to explain the smoothness of the GMM and to compare the limiting behaviors among the least square estimator, the smoothed least square estimator, and the GMM estimator.

$$SupWald = \sup_{\psi \in \Theta_\psi} Wald(\psi), \tag{19}$$

where

$$Wald(\psi) = n(R[\hat{\beta}^T_{(\psi)}, \hat{\delta}^T_{(\psi)}]^T)^T [R(\hat{G}(\psi)^T \hat{\Omega}^{-1} \hat{G}(\psi))^{-1} R^T]^{-1} R[\hat{\beta}^T_{(\psi)}, \hat{\delta}^T_{(\psi)}]^T, \tag{20}$$

$$\hat{G}(\psi) = [\hat{G}_\beta, \hat{G}_\delta(\psi)], \tag{21}$$

$$R = [0_{l\times k}, I_{l\times l}]. \tag{22}$$

Theorem 3: Suppose that $\inf_{\psi \in \Theta_\psi} |G(\psi)^T \Omega^{-1} G(\psi)|$ is positive, with assumptions 1,2,and 6 hold, under the null, we have

$$\begin{aligned} SupWald \xrightarrow{d} & \sup_{\psi \in \Theta_\psi} V^T \Omega^{-1/2} G(\psi)^T (G(\psi)^T \Omega^{-1} G(\psi))^{-1} R^T (R(G(\psi)^T \Omega^{-1} G(\psi))^{-1}) R^T)^{-1} \\ & \times R(G(\psi)^T \Omega^{-1} G(\psi))^{-1} G(\psi) \Omega^{-1/2} V, \end{aligned} \tag{23}$$

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

Thus, the asymptotic distribution of $SupWald$ is the supremum of the "chi-squar" process and depends upon the covariance function. However, the critical value is non-tabulated in general. Following Hansen (1996), the asymptotic critical values and the p value can be generated by bootstrapping [2].

5.2 Test for Exogeneity

In this section, extending the new Hausman-type test suggested by Kapetanios (2010), we propse a Hausman test to test for the exogeneity of the slope regressors of the linear index threshold model.

Consider the following null hypothesis, for all t,

$$H_0 : E\big(\varepsilon_t | X_t\big) = 0 \tag{24}$$

[2]In the appendix, we provide a small simulation to assess the finite sample performance of the suggested bootstrapping test.

Given a consistent threshold estimate $\hat{\psi}$, let $\tilde{\theta}(\hat{\psi}) = [\tilde{\beta}(\hat{\psi})^T, \tilde{\delta}(\hat{\psi})^T]^T$ denote the slope estimator with moment condition $E[X_t\varepsilon_t|\hat{\psi}] = 0$, and $\hat{\theta}(\hat{\psi}) = [\hat{\beta}(\hat{\psi})^T, \hat{\delta}(\hat{\psi})^T]^T$ denote the slope estimator with moment condition $E[z_t\varepsilon_t|\hat{\psi}] = 0$. Evidently, with conditional homoskedasticity, if there is no endogeneity in regressors, both estimators are consistent and $\tilde{\theta}(\hat{\psi})$ is more efficient. However, if the slope regressors are endogenous, only $\hat{\theta}(\hat{\psi})$ is consistent.

Therefore, the test statistic is of the form,

$$H = (\hat{\theta}(\hat{\psi}) - \tilde{\theta}(\hat{\psi}))^T[Var(\hat{\theta}(\hat{\psi})) - Var(\tilde{\theta}(\hat{\psi}))]^+(\hat{\theta}(\hat{\psi}) - \tilde{\theta}(\hat{\psi})), \tag{25}$$

where " + " denotes the Moore-Penrose pseudoinverse.

Theorem 4: With assumptions 1-7 and the conditional homoskedasticity ($E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_\varepsilon^2$), under the null hypothesis,

$$H \xrightarrow{d} \chi^2_{k_1}, \tag{26}$$

where $k_1 = rank\bigg(Var\big((\hat{\theta}(\hat{\psi})\big) - Var\big((\tilde{\theta}(\hat{\psi})\big)\bigg)$.

As Kapetanios (2000) has shown, the asymptotic variances of the test statistic $\sqrt{n}(\hat{\theta} - \tilde{\theta})$ may be problematic due to the nature of the nonlinear model. Therefore, following Kapetanios (2010), the properties of the asymptotic tests can be improved using bootstrapping.

6 Monte Carlo Simulation

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

$$y_t = I(q_{1t} + q_{2t} \leq 0) + e_t, \tag{27}$$

$$e_t = 0.1\varepsilon_t + k_1 v_{q1t} + k_2 v_{q2t}, \tag{28}$$

$$q_{1t} = 0.5q_{1t-1} + v_{q1t}, \tag{29}$$

$$q_{2t} = 0.5q_{2t-1} + v_{q2t}, \tag{30}$$

where v_{q1t}, v_{q2t} 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 DGP is a simpler version of the general model, $y_t = x_t^T\beta + \delta^T\tilde{x}_t I(q_{1t} + q_{2t}^T\psi > 0) + e_t$, with $\beta = 0$, $\delta = 1$ and $x_t = \tilde{x}_t = 1$ for all $t = 1, 2, ...$ We estimate the model both with the GMM and the smoothed least square (LS) method of Seo and Linton (2007). For the smoothed LS, we use the same kernel function and the bandwidth choice with the simulations 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. For each simulation, we report the average MSE, Bias and the standard deviation of the threshold estimates. Tables 1 - 5 report the simulation results. Tables 2-3 reports the results with exogenous q_{2t} and endogenous q_{1t}. Finally, Tables 4-5 presents the results with exogenous q_{1t} and endogenous q_{2t}.

Table 1 shows the results with both exogenous threshold variables. For the linear threshold estimate ψ, the smoothed least square estimator achieves a better performance than the GMM estimator. This results from the super-consistency of the threshold estimate of the smoothed LS. Since the DGP is designed with a fixed threshold effect, the GMM estimator converges at the normal $\sqrt{n}$ rate, which implies a slower convergence speed than smoothed LS estimator.

Tables 2-3 report the results with exogenous q_{2t} and endogenous q_{1t}. Tables 4-5 presents the results with exogenous q_{1t} and endogenous q_{2t}. Therefore, for both cases, as expected the smoothed least square has an asymptotic bias. The average biases of the smoothed LS estimator are much larger than the GMM estimators for all cases. In addition, the stronger the endogeneity, the larger the average bias. For the GMM estimator, as the sample size increases, all MSEs of the GMM estimator decrease and converge to zero confirming the consistency of the GMM estimator.

Table 1: Simulation Performance of the GMM and the smoothed least square estimators, k_1=k_2=0 (exogenous case)

MSE

	GMM			Smoothed LS		
n	ψ	β	δ	ψ	β	δ
100	0.0279	0.0007	0.0019	0.0060	0.0009	0.0021
300	0.0070	0.0002	0.0006	0.0013	0.0002	0.0005
500	0.0030	0.0001	0.0003	0.0006	0.0001	0.0003

Bias

	GMM			Smoothed LS		
n	ψ	β	δ	ψ	β	δ
100	-0.0360	0.0100	-0.0195	0.0005	0.0132	-0.0260
300	-0.0053	0.0058	-0.0121	0.0009	0.0057	-0.0115
500	-0.0028	0.0045	-0.0092	0.0006	0.0038	-0.0074

Standard Error

	GMM			Smoothed LS		
n	ψ	β	δ	ψ	β	δ
100	0.1631	0.0255	0.0389	0.0776	0.0276	0.0382
300	0.0835	0.0127	0.0206	0.0363	0.0138	0.0197
500	0.0550	0.0089	0.0142	0.0252	0.0101	0.0144

This table reports the simulation results of the GMM estimator and the smoothed least square estimator for the DGP defined by equation (27) with exogenous threshold variables. The first column shows the sample size that the simulation used. The second to the fourth columns report the results of the GMM estimator for ψ, β & δ respectively. The fifth to the last column show the results of the smoothed LS estimator.

Table 2: Simulation Performance of the GMM and the smoothed least square estimators, k_1=0.3, k_2=0

MSE

	GMM			Smoothed LS		
n	ψ	β	δ	ψ	β	δ
100	0.0747	0.0281	0.1099	0.1605	0.0269	0.1058
300	0.0209	0.0242	0.0961	0.0501	0.0237	0.0935
500	0.0089	0.0237	0.0940	0.0296	0.0228	0.0904

Bias

	GMM			Smoothed LS		
n	ψ	β	δ	ψ	β	δ
100	0.0606	0.1604	-0.3228	0.3453	0.1586	-0.3192
300	0.0491	0.1538	-0.3079	0.2111	0.1520	-0.3038
500	0.0354	0.1530	-0.3054	0.1657	0.1501	-0.2995

Standard Error

	GMM			Smoothed LS		
n	ψ	β	δ	ψ	β	δ
100	0.2666	0.0487	0.0755	0.2032	0.0415	0.0630
300	0.1360	0.0242	0.0361	0.0741	0.0238	0.0355
500	0.0874	0.0182	0.0267	0.0467	0.0177	0.0268

This table reports the simulation results of the GMM estimator and the smoothed least square estimator for the DGP defined by equation (27) with small endogenous effect from q_{1t}. The first column shows the sample size that the simulation used. The second to the fourth columns report the results of the GMM estimator for ψ, β & δ respectively. The fifth to the last column show the results of the smoothed LS estimator.

Table 3: Simulation Performance of the GMM and the smoothed least square estimators, k_1=0.5, k_2=0

MSE

	GMM			Smoothed LS		
n	ψ	β	δ	ψ	β	δ
100	0.2606	0.0753	0.2971	4.9297	0.4247	1.7088
300	0.1087	0.0631	0.2505	4.7556	0.3820	1.5278
500	0.0759	0.0623	0.2491	4.6673	0.3666	1.4681

Bias

	GMM			Smoothed LS		
n	ψ	β	δ	ψ	β	δ
100	0.1510	0.2588	-0.5192	-1.4815	0.5910	-1.1884
300	0.1882	0.2487	-0.4977	-1.3859	0.5549	-1.1114
500	0.1503	0.2480	-0.4975	-1.3964	0.5422	-1.0854

Standard Error

	GMM			Smoothed LS		
n	ψ	β	δ	ψ	β	δ
100	0.4877	0.0914	0.1659	1.6541	0.2747	0.5447
300	0.2707	0.0360	0.0526	1.6841	0.2722	0.5410
500	0.2310	0.0277	0.0395	1.6488	0.2695	0.5386

This table reports the simulation results of the GMM estimator and the smoothed least square estimator for the DGP defined by equation (27) with large endogenous effect from q_{1t}. The first column shows the sample size that the simulation used. The second to the fourth columns report the results of the GMM estimator for ψ, β & δ respectively. The fifth to the last column show the results of the smoothed LS estimator.

Table 4: Simulation Performance of the GMM and the smoothed least square estimators, k_1=0, k_2=0.3

MSE

	GMM			Smoothed LS		
n	ψ	β	δ	ψ	β	δ
100	0.0577	0.0283	0.1093	0.0739	0.0278	0.1071
300	0.0121	0.0243	0.0965	0.0324	0.0237	0.0939
500	0.0060	0.0235	0.0933	0.0212	0.0226	0.0897

Bias

	GMM			Smoothed LS		
n	ψ	β	δ	ψ	β	δ
100	-0.1024	0.1622	-0.3240	-0.2476	0.1615	-0.3210
300	-0.0554	0.1539	-0.3086	-0.1728	0.1522	-0.3044
500	-0.0359	0.1521	-0.3043	-0.1417	0.1492	-0.2984

Standard Error

	GMM			Smoothed LS		
n	ψ	β	δ	ψ	β	δ
100	0.2173	0.0449	0.0661	0.1122	0.0411	0.0636
300	0.0951	0.0238	0.0351	0.0504	0.0229	0.0348
500	0.0685	0.0188	0.0272	0.0342	0.0179	0.0268

This table reports the simulation results of the GMM estimator and the smoothed least square estimator for the DGP defined by equation (27) with small endogenous effect from q_{2t}. The first column shows the sample size that the simulation used. The second to the fourth columns report the results of the GMM estimator for ψ, β & δ respectively. The fifth to the last column show the results of the smoothed LS estimator.

Table 5: Simulation Performance of the GMM and the smoothed least square estimators, k_1=0, k_2=0.5

MSE

	GMM			Smoothed LS		
n	ψ	β	δ	ψ	β	δ
100	0.1783	0.0660	0.2564	1.7022	0.0496	0.1871
300	0.0666	0.0632	0.2509	1.4350	0.0518	0.2024
500	0.0373	0.0630	0.2505	1.3642	0.0529	0.2090

Bias

	GMM			Smoothed LS		
n	ψ	β	δ	ψ	β	δ
100	-0.2466	0.2484	-0.4966	-1.1965	0.2148	-0.4241
300	-0.1669	0.2487	-0.4981	-1.0766	0.2250	-0.4471
500	-0.1195	0.2496	-0.4989	-1.0327	0.2286	-0.4556

Standard Error

	GMM			Smoothed LS		
n	ψ	β	δ	ψ	β	δ
100	0.3429	0.0658	0.0988	0.5204	0.0587	0.0852
300	0.1969	0.0360	0.0526	0.5255	0.0343	0.0495
500	0.1519	0.0277	0.0401	0.5458	0.0255	0.0372

This table reports the simulation results of the GMM estimator and the smoothed least square estimator for the DGP defined by equation (27) with large endogenous effect from q_{2t}. The first column shows the sample size that the simulation used. The second to the fourth columns report the results of the GMM estimator for ψ, β & δ respectively. The fifth to the last column show the results of the smoothed LS estimator.

7 Empirical Application

For many countries, especially certain advanced economies, public debt has been steadily increasing over the past decades, and there is a growing concern about its impact on long-term growth. Therefore, one of the most active areas of research recently has been to test whether debt has a nonlinear effect on growth. To investigate the potential threshold effect of public debt on growth, many researchers have carried out empirical studies to examine its magnitude of this effect and estimate the level beyond which debt will be detrimental to growth (threshold level of debt). By using Hansen's (2000) threshold regression model, Cecchetti et al. (2011), Caner et al. (2011) and Afonso & Jalles (2013) find that the public debt will have an adverse effect on economic growth when the public debt to GDP ratio exceeds 85%, 77%, and 59% respectively. By correcting for endogeneity in both slope regressors and the threshold variable with a structural threshold regression model of Kourtellos et al. (2016), Kourtellos et al. (2013) fail to find the significant threshold effect for the public debt.

However, the above findings ignore country heterogeneity. Moreover, all results in the literature estimate the threshold effect by assuming the nonlinearity exists only in public debt. For developing countries, it is natural to expect a threshold effect from external debt. For example, Patillo et al. (2002) show that there is a U shape relationship between external debt and growth in developing countries. In contrast to Patillo et al. (2002), Schclarek (2004) fails to detect any nonlinearity in foreign debt on growth for developing countries.

One of the methodological problems in the past literature is that the model only allows for one threshold variable. Furthermore, most research relies on the homogeneity assumptions in both slope regressors and the threshold variable, which is highly dubious. It may be useful to conjecture that nonlinearity of growth in developing countries could originate from the joint linear threshold effect between both public debt and external debt. As such, we apply the linear index threshold model to investigate this issue. We examine the following linear index threshold Solow growth model:

$$g_t = x_t^T \beta + \delta^T x_t I(d_{1t} + d_{2t}\psi_1 + \psi_2 \leq 0) + \varepsilon_t, \tag{31}$$

where g_t is the growth rate, d_{1t} is the demeaned public debt to GDP ratio, d_{2t} is the demeaned external debt to GDP ratio, x is the Solow controlling set including constant & five Solow variables, namely, initial income per capita, schoolings, investment, population growth, and openness. It also includes public debt to GDP ratio and external debt to GDP ratio. A detailed data resource description of all variables is given in Table 7. We also account for time fixed effects. We observe that, according to the heavily indebted poor countries (HIPC) initiative, 33 out of 37 HIPC in our dataset are from the Sub-Saharan African area. Therefore, we also include the regional effects with the Latin-American dummy and the Sub-Saharan dummy.

We employ an averaged ten-year period panel data covering 54 developing countries in 1980-1989, 1990- 1999, 2000-2009 and 2010-2016. The growth rate of real per capita GDP is from PWT 9.0. The public debt and external debt to GDP ratio are from the IMF Historical Public Debt Database and the data bank of the world bank. In this paper, all variables are instrumented by their lagged values. We estimate the model using both smoothed least square method of Seo and Linton (2007) and our proposed GMM method. We test the nonlinearity by using the sup-wald statistic. As suggested by Hansen (2000), we use the bootstrap method to test for the existence of the threshold effect.

We present the results in Table 6. The smoothed least square estimate shows the presence of the significant threshold effect at 1% level with the bootstrap P value equaling 0.0001. It is worth noting that, with all else being equal, higher external debt leads to higher growth if the country is in the low debt regime and lower growth if the country is in the high debt regime. The finding supports for the inverted-U relationship of the external debt with growth. Furthermore, the positive effect of the external debt on growth in low debt regime is insignificant while the adverse impact in high debt regime is significant at 1% level.

Surprisingly, after correcting the endogeneity in both slope regressors and the threshold variables, the nonlinearity result of the GMM method becomes insignificant with the bootstrap P value equals 0.2727. Therefore, our finding suggests there is little evidence of nonlinearity in the effects of debt on growth and any finding to the contrary may be the result that the effects of possible self-selection or endogeneity by various countries is ignored in how they behave towards their debt obligations. The linear index threshold effect in external debt is found to be endogenous and the main reason of the heterogeneity in the debt-growth relationship is not the level of public debt and/or external debt.

Table 6: Estimation and Testing Results of the Linear Index Threshold Solow Growth Model

Method	GMM-Index		Smoothed LS		Linear-GMM	Linear-LS
ψ_1	0.57		0.90			
ψ_2	0.24		0.12			
	Low	High	Low	High		
Constant	-0.0713	-0.1609	-0.0135	0.0608	-0.1639	0.0695
	(0.8059)	(0.2534)	(0.8844)	(0.5344)	(0.1763)	(0.1019)
Initial income	-0.0009	-0.0044	-0.0017	-0.0013	-0.0036	-0.0033
	(0.8742)	(0.4905)	(0.8234)	(0.8532)	(0.4571)	(0.2108)
Schooling	0.0027	0.0013	0.0002	-0.0039	0.0021	0.001
	(0.7179)	(0.8526)	(0.9835)	(0.6108)	(0.7109)	(0.7669)
Investment	0.0101	-0.0092^{**}	-0.001	0.0044	-0.0013	-0.0011
	(0.3219)	(0.0188)	(0.9294)	(0.4331)	(0.7442)	(0.7541)
Population	-0.0233	-0.0757	-0.016	0.00253	-0.0732	0.0153
	(0.8385)	(0.2019)	(0.5920)	(0.5281)	(0.1434)	(0.2630)
Public debt	0.0110	0.1320	-0.0300^{**}	-0.0231^{*}	-0.0006	-0.0068
	(0.6240)	(0.1587)	(0.0191)	(0.0998)	(0.9159)	(0.1172)
External debt	0.0265^{**}	-0.0354	0.0056	-0.0899^{***}	-0.0076	-0.0033
	(0.0452)	(0.2132)	(0.7034)	(0.0046)	(0.6059)	(0.7117)
Openness	0.0085	0.0123^{**}	-0.0028	0.0166^{***}	0.0075^{*}	0.0090^{***}
	(0.2322)	(0.0463)	(0.7577)	(0.0068)	(0.0665)	(0.0050)
SupWald	29.6561		80.2709			
SupWald Boot P value	0.2727		0.0001^{***}			
Observations	216		216		216	216

This table presents the estimation of the smoothed least square threshold index model of Seo and Linton (2007) and the GMM threshold index model. The first column shows the slope regressors. The second and third column give the results of the GMM method. The fourth and the fifth column report the the results of the smoothed least square method. The last two columns report the GMM and LS results that ignores the presence of a threshold. All variables are instrumented by the lag values. Time dummies and regional dummies are included but not reported. "***" denotes significantly different from zero at the 1% level, "**" denotes significantly different from zero at the 5% level, and "*" denotes significantly different from zero at the 10% level.

8 Conclusion

In this paper, we propose a GMM estimator for the linear index threshold model. The GMM estimator allows for the endogeneity of the threshold variable as well as the slope regressors. We show the consistency of the GMM estimator and derive the limiting distribution. We study the finite sample performance of the proposed estimator through Monte Carlo simulation. We compare the performance of the GMM estimator with the smoothed least square estimator of Seo and Linton (2007) under both exogenous and endogenous threshold variable design. The simulation results are consistent with the theory. We use the linear index threshold model to investigate the threshold effect of the linear index combined by the public debt and the external debt on the economic growth in developing countries. The nonlinearity testing result of the GMM estimator shows the threshold effect is insignificant.

Table 7: Data Resource

Variables	Description
Initial income	Logarithm of real per capita GDP in chain series at 1980, 1990, 2000, 2010 Lagged values correspond to 1975, 1985, 1995 & 2005 Source: PWT 9.0
Population growth rates	Logarithm of average population growth rates plus 0.05 for the periods 1980-1989, 1990-1999, 2000-2009 & 2010-2016 Lagged values correspond to 1975-1979, 1985-1989, 1995-1999 & 2005-2009 Source: PWT 9.0
Investment	Logarithm of average ratios over each period of investment to real GDP per capita for the periods 19801989, 1990-1999, 2000-2009 & 2010-2016 Lagged values correspond to 1975-1979, 19851989, 19951999 & 2005-2009. Source: PWT 9.0
Schooling	Logarithm of average years of male secondary plus tetriary school attainment (25+) in 1980, 1990, 1999 & 2010. Lagged values correspond to 1975, 1985, 1995 & 2005 Source: Barro and Lee (2010).
Openness	Average ratios of exports plus imports to real GDP per capita 1980-1989, 1990-1999, 2000-2009 and 2010-2016, Lagged values correspond to 1975-1979, 1985- 1989, 1995-1999 and 2005-2009 Source: PWT 9.0
Public Debt	Logarithm of public debt to GDP for the periods 1980-1989, 1990-1999, 2000-2009 and 2010-2016, Lagged values correspond to 1975-1979, 1985- 1989, 1995-1999 and 2005-2009 Source: IMF Historical Public Debt Dataset
External Debt	Logarithm of external debt to GDP for the periods 1980-1989, 1990-1999, 2000-2009 and 2010-2016, Lagged values correspond to 1975-1979, 1985- 1989, 1995-1999 and 2005-2009 Source: World bank

References

[1] Afonso, A., Jalles, J, (2013), Growth and productivity: The role of government debt. *International Review of Economics and Finance*, Vol. 25, pp. 384-407.

[2] Andrews, D.W.K, (1987), Consistency in nonlinear econometric models: a generic uniform law of large numbers.*Econometrica*, Vol.55, pp.817858.

[3] Andrews, D.W.K., (1994), Empirical Process Methods in Econometrics, *Handbook of Econometrics*, Vol. 4, R.F. Eagle and D.L. McFadden, eds., New York: Elsevier Science B.V., Ch. 37, 2247

[4] Caner, M., Hansen, B, (2004), Instrumental variable estimation of a threshold model, *Econometric Theory*, Vol. 20, pp. 813-843.

[5] Caner, M, Grennes, T., Koehler-Geib, F, (2011), Finding the Tipping Point When Sovereign Debt Turns Bad. In: Braga, C. (Ed.) (2011) *Sovereign debt and the financial crisis: will this time be different?*, Washington DC.

[6] Cecchetti, S., Mohanty, M., Zampolli, F, (2011), The Real Effects of Debt. *Bank For International Settlements Working Papers*, No. 352.

[7] Chan, K. S, (1993), Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model, *Annals of Statistics*, Vol. 21, pp. 520-533.

[8] De Jong, R., Woutersen, T., 2004. Dynamic time series binary choice, *Econometric Theory*, Vol.27, pp., 673-702.

[9] Hansen, B.E., (1999), Threshold effects in non-dynamic panels: estimation, testing and inference, *Journal of Econometrics*, Vol. 93 (2), pp. 345368.

[10] Hansen, B.E., (2000), Sample splitting and threshold estimation, *Econometrica*, Vol. 68, pp. 575-603.

[11] Hausman, J. A., (1978), Specification Tests in Econometrics, *Econometrica*, Vol. 46, pp. 12511271.

[12] Heckman, J.J, (1979), Sample Selection Bias as a Specification Error, *Econometrica*, Vol. 47, pp. 153-161.

[13] Horowitz, J.L, (1992), A smoothed maximum score estimator for the binary response model. *Econometrica*, Vol. 60, pp.505-531.

[14] Kapetanios, G., (2010), Testing for exogeneity in threshold models, *Econometric Theory*, Vol.26, pp. 231-259.

[15] Kourtellos, A., Stengos, T., Tan, C.M, (2013), The effect of public debt on growth in multiple regimes, *Journal of Macroeconomics*, Vol. 38, pp. 35-43.

[16] Kourtellos, A., Stengos, T., Tan, C.M, (2016), Structural Threshold Regression, *Econometric Theory*, Vol.32, pp. 827-860.

[17] Kourtellos, A., T. Stengos, Y. Sun, (2017), Endogeneity in Semiparametric Threshold Regression, *Working paper*, University of Cyprus and University of Guelph.

[18] Newey, W., and D. McFadden, (1994), Large Sample Estimation and Hypothesis Testing, in R.. Engle, and D. McFadden, ed.: *Handbook of Econometrics*, Vol. 4 . pp. 2113-2241 (Springer: New York).

[19] Pattillo, C, H. Poirson, L. Ricca, (2002), External Debt and Growth, *IMF Working Paper*, No. 02/69.

[20] Robinson, P.M, (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*. Vol.4(3), pp.185-207.

[21] Robinson, P. M, (1986) On the consistency and finite sample properties of nonparametric kernel time series regression, autoregression, and density estimators. *Annals of the Institute of Statistical Mathematics*. Vol.38, pp.539-549.

[22] Seo, M. H., O. Linton, (2007), A smoothed least squares estimator for threshold regression models, *Journal of Econometrics*, Vol.141, pp. 704-735.

[23] Seo, M. H., Y. Shin, (2016), Dynamic Panels with Threshold Effect and Endogeneity, *Journal of Econometrics*, Vol.195, pp. 169-186.

[24] Van der Vaart, A.W. and J. A.Wellner, (1996), Weak Convergence and Empirical Processes: With Applications to Statistics (New York: Springer).

[25] White, H, (1984), Asymptotic Theory for Econometricians (Academic Press: London).

[26] Yu, P, (2015), Threshold Regression With A Threshold Boundary, Mimeo, Hong Kong University.

[27] Yu, P, and P. C.b. Phillips, (2018), Threshold Regression Asymptotics: From the Compound Poisson Process to Two-sided Brownian Motion, *Economics Letter*, Vol.172, pp. 123-126.

Appendices

Throughout the proof, let $||.||$ denote the Euclidean norm. The integral is taken over $(-\infty, \infty)$ unless specified otherwise. All limits are taken as $n \to \infty$. $\xrightarrow{a.s}$, $\xrightarrow{p}$, and $\xrightarrow{d}$ denote almost sure convergence, convergence in probability, and convergence in distribution respectively. $\bigwedge$ and $\bigvee$ denote the minimum and maximum operators.

By definition, we have

$$\begin{aligned} g(\theta) = E(g_t(\theta)) &= E[z_t(y_t - x_t^T\beta - \delta^T\tilde{x}_t I(q_{1t} + q_{2t}^T\psi > 0))] \\ &= E[z_t y_t - z_t x_t^T\beta - z_t\delta^T\tilde{x}_t I(q_{1t} + q_{2t}^T\psi > 0)], \end{aligned} \tag{32}$$

and the sample analogue is

$$g_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} g_t(\theta) = \frac{1}{n}\sum_{t=1}^{n}[z_t y_t - z_t x_t^T\beta - z_t\delta^T\tilde{x}_t I(q_{1t} + q_{2t}^T\psi > 0)].$$

A Proof of Lemma

Lemma 1: Under assumptions 1 and 2, it can be shown that:

$$\underset{\psi\in\Theta_\psi}{Sup}\,||\frac{1}{n}\sum_{t=1}^{n} z_t\tilde{x}_t{}^T I(q_{1t} + q_{2t}^T\psi > 0) - E[z_t\tilde{x}_t{}^T I(q_{1t} + q_{2t}^T\psi > 0)]|| \xrightarrow{p} 0. \tag{33}$$

$$\underset{\psi\in\Theta_\psi}{Sup}\,||\frac{1}{n}\sum_{t=1}^{n} x_t\tilde{x}_t{}^T I(q_{1t} + q_{2t}^T\psi > 0) - E[x_t\tilde{x}_t{}^T I(q_{1t} + q_{2t}^T\psi > 0)]|| \xrightarrow{p} 0. \tag{34}$$

Proof. Under assumptions 1 and 2, $E||z_t\tilde{x}_t^T||$ and $E||x_t\tilde{x}_t^T||$ are bounded. Then, the proof is straightforward by applying lemma 1 of Seo & Linton (2007). ■

Lemma 2: Under assumptions 1, 2, and 6, there is a $C < \infty$ such that for any $\psi_1, \psi_2 \in \Theta_\psi$, we have

$$\|E\Big(X_t\Big(I(\psi_1) - I(\psi_2)\Big)\Big)\| \leq C||\psi_1 - \psi_2||,$$
$$\|E\Big(X_t\varepsilon_t\Big(I(\psi_1) - I(\psi_2)\Big)\Big)\| \leq C||\psi_1 - \psi_2||, \tag{35}$$

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

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

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

where v_t defines in Assumption 7.

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

$$\|E\Big(X_t\big(I(\psi_1) - I(\psi_2)\big)\Big)\| \leq \|E(X_tq_{2t}^T|v_t(\psi_2) = 0)\|f_{v_t(\psi_2)}(0)||\psi_1 - \psi_2|| + O(1)$$
$$\|E\Big(X_t\varepsilon_t\|\big(I(\psi_1) - I(\psi_2)\big)\Big) \leq \|E(\ X_t\varepsilon_tq_{2t}^T|v_t(\psi_2) = 0)\|f_{v_t(\psi_2)}(0)||\psi_1 - \psi_2|| + O(1). \tag{36}$$

Appying assumptions 2 and 6, 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. ■

B Proof of Theorem 1:

First, under assumptions 1, 2, and 3, by applying lemma 1, we have

$$\underset{\beta\in\Theta_\beta,\delta\in\Theta_\delta,\psi\in\Theta_\psi,}{Sup} ||\frac{1}{n}\sum_{t=1}^{n}(z_ty_t - z_tx_t^T\beta - z_t\delta^T\tilde{x}_tI(q_{1t} + q_{2t}^T\psi > 0)) - E[(z_ty_t - z_tx_t^T\beta - z_t\delta^T\tilde{x}_tI(q_{1t} + q_{2t}^T\psi > 0)]|| \xrightarrow{p} 0. \tag{37}$$

That is

$$\underset{\theta\in\Theta}{Sup}||g_n(\theta) - E(g_t(\theta))|| \xrightarrow{p} 0. \tag{38}$$

Evidently, $E(g_t(\theta))$ is continuous in θ.

Next, we show that $E(g_t(\theta)) = 0$ iff $\theta = \theta_n$.

Applying simple calculations gives

$$G_\beta = -E(z_t x_t^T), \tag{39}$$

$$G_\delta(\psi) = -E(z_t\{\tilde{x}_t I(q_{1t} + q_{2t}^T\psi > 0)\}^T), \tag{40}$$

$$G_\psi(\psi) = -E\left(z_t\delta^T\tilde{x}_t q_{2t}^T | v_t(\psi) = 0\right) f_{v_t(\psi)}(0), \tag{41}$$

Now, suppose $\beta = \beta_0$, $\delta = \delta_n$ but $\psi \neq \psi_0$, we have

$$E(g_t(\theta)) = E(g_t(\beta_0, \delta_n, \psi)) = E(g_t(\beta_0, \delta_n, \psi)) - E(g_t(\beta_0, \delta_n, \psi_0))$$

$$= E\{z_t\tilde{x}_t^T[I(q_{1t} + q_{2t}^T\psi_0 > 0) - I(q_{1t} + q_{2t}^T\psi > 0)]\}\delta_n = [G_\delta(\psi) - G_\delta(\psi_0)]\delta_n. \tag{42}$$

Let $A = \{-q_{2t}^T\psi < q_{1t} < -q_{2t}^T\psi_0\} \bigcup \{-q_{2t}^T\psi_0 < q_{1t} < -q_{2t}^T\psi\}$.

Under assumptions 5 and 6, the set A has a positive probability. Therefore, we have

$$E[I(q_{1t} + q_{2t}^T\psi_0 > 0) - I(q_{1t} + q_{2t}^T\psi > 0)|A] \neq 0. \tag{43}$$

Under equation (43), assumption 2, and 4, we obtain

$$E[z_t\tilde{x}_t^T(I(q_{1t} + q_{2t}^T\psi_0 > 0) - I(q_{1t} + q_{2t}^T\psi > 0))|A]\delta_n \neq 0, \tag{44}$$

which implies the unconditional expectation

$$E[z_t\tilde{x}_t^T(I(q_{1t} + q_{2t}^T\psi_0 > 0) - I(q_{1t} + q_{2t}^T\psi > 0))]\delta_n \neq 0. \tag{45}$$

If $\beta \neq \beta_0$ or $\delta \neq \delta_n$ but $\psi = \psi_0$, we have

$$E(g_t(\theta)) = E(g_t(\beta, \delta_n, \psi_0)) = E(g_t(\beta, \delta_n, \psi_0)) - E(g_t(\beta_0, \delta_n, \psi_0))$$
$$= -E(z_t x_t^T)(\beta - \beta_0) = G_\beta(\beta - \beta_0) \neq 0, \tag{46}$$

and

$$E(g_t(\theta)) = E(g_t(\beta_0, \delta, \psi_0)) = E(g_t(\beta_0, \delta, \psi_0)) - E(g_t(\beta_0, \delta_n, \psi_0))$$
$$= -E(z_t\{\tilde{x}_t I(q_{1t} + q_{2t}^T\psi_0 > 0)\}^T)(\delta - \delta_n) = G_\delta(\psi_0)(\delta - \delta_n) \neq 0 \tag{47}$$

where the inequality follows assumption 7.

If $\beta \neq \beta_0$ or $\delta \neq \delta_n$ and $\psi \neq \psi_0$, with almost same arguments, we have

$$E(g_t(\theta)) = E(g_t(\beta, \delta, \psi)) = E(g_t(\beta, \delta, \psi)) - \underbrace{E(g_t(\beta_0, \delta_n, \psi_0))}_{E(g_t(\theta_n))=0}$$
$$= [G_\delta(\psi) - G_\delta(\psi_0)]\delta_n + G_\beta(\beta - \beta_0) + G_\delta(\psi)(\delta - \delta_n) \neq 0. \tag{48}$$

Hence, we obtain $E(g_t(\theta)) = 0$ if and only if $\theta = \theta_n$.

Therefore, $Q(\theta) = E(g_t(\theta))^T W E(g_t(\theta))$ has a unique minimum at $\theta = \theta_n$, where W is a positive definite matrix.

Last, we show $Q_n(\theta)$ converges uniformly in probability to $Q(\theta)$.

$$\underset{\theta\in\Theta}{Sup}|Q_n(\theta) - Q(\theta)| = \underset{\theta\in\Theta}{Sup}|g_n(\theta)^T W g_n(\theta) - g(\theta)^T W g(\theta)|$$
$$= \underset{\theta\in\Theta}{Sup}|(g_n(\theta) - g(\theta))^T W (g_n(\theta) - g(\theta)) + 2(g_n(\theta) - g(\theta))^T W g(\theta)|$$
$$\leq \underset{\theta\in\Theta}{Sup}\{||g_n(\theta) - g(\theta)||^2 \; ||W|| + 2||g_n(\theta) - g(\theta)|| \; ||W|| \; ||g(\theta)||\}. \tag{49}$$

Applying equation (38), we have $\underset{\theta\in\Theta}{Sup}|Q_n(\theta) - Q(\theta)| \xrightarrow{p} 0$, which completes our proof by following theorem 2.1 of Newey and McFadden (1994).

C Proof of Theorem 2

To derive the asymptotic normality with nonsmooth objective function, we follow theorm 7.1 of Newey and McFadden (1994).

First, by central limit theorem (CLT), we have $\sqrt{n}g_n(\theta_n) \xrightarrow{d} N(0, \Omega)$, where $\Omega = E(g_t(\theta_n)g_t(\theta_n)^T)$.

Now, let k_n be a $k+l+h-1$ dimensional diagonal matrix whose first $k+l$ diagonals are ones and the other element is n^α, $W_n \xrightarrow{p} W = \Omega^{-1}$, and

$$D_n = k_n^{-1} G^T W_n g_n(\theta_n), \tag{50}$$

$$H = k_n^{-1} G^T W G k_n^{-1}, \tag{51}$$

$$R(\theta) = (\frac{Q_n(\theta) - Q_n(\theta_n) - Q(\theta) - D_n^T(\theta - \theta_n)}{||\theta - \theta_n||}). \tag{52}$$

Next, we show the stochastic differentiability condition hold.

That is, for any $\gamma_n \to 0$, we have

$$\underset{||\theta-\theta_n|| \leq \gamma_n}{Sup} |\frac{\sqrt{n}R(\theta)}{1+\sqrt{n}||\theta-\theta_n||}| = o_p(1). \tag{53}$$

Define

$$\varepsilon_n(\theta) = \frac{g_n(\theta) - g_n(\theta_n) - g(\theta)}{1+\sqrt{n}||\theta-\theta_n||} \tag{54}$$

.

For $\gamma_n \to 0$ and $U = \{||\theta - \theta_n|| \leq \gamma_n\}$, $\underset{\theta \in U}{Sup}\{\sqrt{n}||\varepsilon_n(\theta)||\} \xrightarrow{p} o_p(1)$ 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 the assumption 4. Therefore, we only need to check the stochastic equicontinuity of the empirical process $\frac{1}{\sqrt{n}}\sum_{t=1}^n[z_t\tilde{x}_t I(q_{1t} + q_{2t}^T\psi > 0) - E\{z_t\tilde{x}_t I(q_{1t} + q_{2t}^T\psi > 0)\}]$. Let $F = (||z_t\tilde{x}_t||) \underset{||\psi-\psi_0|| \leq \gamma_n}{sup} I(q_{1t} > -q_{2t}^T\psi \bigwedge -q_{2t}^T\psi_0)$ be the envelope function. Since the indicator functions of half intervals constitute a *type I* class or a Vapnik Chervonenkis (VC) class, by assumptions

1, 2, and 4, the stochastic equicontinuity follows the Theorem 1 of Andrews (1994) and the Theorem 2.14.1 of Van der Vaart and Wellner (1996). Evidently, $\varepsilon_n(\theta_n) = 0$.

Following proof of theorem 7.2 of Newy & MaFadden (1994), we decompose $|\frac{\sqrt{n}R(\theta)}{1+\sqrt{n}||\theta-\theta_n||}|$ into 5 terms,

$$|\frac{\sqrt{n}R(\theta)}{1+\sqrt{n}||\theta-\theta_n||}| \leq \sum_{j=1}^{5} r_{nj}(\theta), \tag{55}$$

where

$$\begin{aligned} r_{n1}(\theta) &= \frac{\sqrt{n}(2\sqrt{n}||\theta-\theta_n|| + ||\theta-\theta_n||^2) \left|\varepsilon_n(\theta)^T W_n \varepsilon_n(\theta)\right|}{||\theta-\theta_n||(1+\sqrt{n}||\theta-\theta_n||)} \\ &= \frac{(2n+\sqrt{n}||\theta-\theta_n||)|\varepsilon_n(\theta)^T W_n \varepsilon_n(\theta)|}{(1+\sqrt{n}||\theta-\theta_n||)}, \end{aligned} \tag{56}$$

$$r_{n2}(\theta) = \frac{\sqrt{n}|[g(\theta) - Gk_n^{-1}(\theta-\theta_n)]^T W_n g_n(\theta_n)|}{||\theta-\theta_n||(1+\sqrt{n}||\theta-\theta_n||)}, \tag{57}$$

$$r_{n3}(\theta) = \frac{n|[g(\theta) + g_n(\theta_n)]^T W_n \varepsilon_n(\theta)|}{(1+\sqrt{n}||\theta-\theta_n||)}, \tag{58}$$

$$r_{n4}(\theta) = \frac{\sqrt{n}\left|g(\theta)^T W_n \varepsilon_n(\theta)\right|}{||\theta-\theta_n||}, \tag{59}$$

$$r_{n5}(\theta) = \frac{\sqrt{n}\left|g(\theta)^T [W_n - W] g(\theta)\right|}{||\theta-\theta_n||(1+\sqrt{n}||\theta-\theta_n||)}. \tag{60}$$

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

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

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

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

$$\underset{\theta\in U}{Sup}\{\frac{||g(\theta)-Gk_n^{-1}(\theta-\theta_n)||}{||\theta-\theta_n||(1+\sqrt{n}||\theta-\theta_n||)}\} \leq \underset{\theta\in U}{Sup}\{\frac{||g(\theta)-Gk_n^{-1}(\theta-\theta_n)||}{||\theta-\theta_n||}\}$$
$$= \underset{\theta\in U}{Sup}\{\frac{||g(\theta)-g(\theta_n)-Gk_n^{-1}(\theta-\theta_n)||}{||\theta-\theta_n||}\} = o(1).$$

Therefore, by Cauchy-Schwarz inequality, we have

$$\underset{\theta\in U}{Sup}\{r_{n2}(\theta)\} = \underset{\theta\in U}{Sup}\{\frac{\left|[g(\theta)-Gk_n^{-1}(\theta-\theta_n)]^T W_n\sqrt{n}g_n(\theta_n)\right|}{||\theta-\theta_n||(1+\sqrt{n}||\theta-\theta_n||)}\} = o_p(1), \tag{61}$$

$$\underset{\theta\in U}{Sup}\{r_{n3}(\theta)\} \leq \underset{\theta\in U}{Sup}\{\frac{\sqrt{n}||g(\theta)+g_n(\theta_n)||\ ||W_n||\ ||\sqrt{n}\varepsilon_n(\theta)||}{(1+\sqrt{n}||\theta-\theta_n||)}\} = o_p(1).$$

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

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

To sum up, we obtain

$$\underset{\theta\in U}{Sup}\{|\frac{\sqrt{n}R(\theta)}{1+\sqrt{n}||\theta-\theta_n||}|\} \leq \sum_{j=1}^{5}\underset{\theta\in U}{Sup}\{r_{nj}(\theta)\} \leq \sum_{j=1}^{5} o_p(1) = o_p(1). \tag{62}$$

Next, we show $k_n^{-1}(\hat{\theta} - \theta_n) = O_p(n^{-1/2})$, where $\hat{\theta} = \underset{\theta \in \Theta}{argmin}\ Q_n(\theta)$

By Taylor expansion, we have

$$Q(\theta) = Q(\theta_n) + (\theta - \theta_n)^T H(\theta - \theta_n) + o(||\theta - \theta_n||^2), \tag{63}$$

where $Q(\theta)$ achevies minimum at $\theta = \theta_n$, and H is positive definite.

This implies we can find a constant $C > 0$ such that

$$(\theta - \theta_n)^T H(\theta - \theta_n) + o(||\theta - \theta_n||^2) \geq C||\theta - \theta_n||^2 \geq \lambda_{min}(H)||\theta - \theta_n||^2 \tag{64}$$

where $\lambda_{min}(H)$ is the smallest eigenvalue of H.

Therefore,

$$Q(\theta) - Q(\theta_n) \geq C||\theta - \theta_n||^2. \tag{65}$$

Since $Q_n(\hat{\theta}) \leq \underset{\theta \in \Theta}{Sup} Q_n(\theta) + o_p(n^{-1})$, we have

$$0 \geq Q_n(\hat{\theta}) - Q_n(\theta_n) - o_p(n^{-1}) = Q(\hat{\theta}) - Q(\theta_n) + D_n^T(\hat{\theta} - \theta_n) + ||\hat{\theta} - \theta_n||R(\hat{\theta}) - o_p(n^{-1}). \tag{66}$$

For $\theta \in U$, we have $|R(\theta)| = (1 + \sqrt{n}||\theta - \theta_n||)o_p(n^{-1/2})$.

Therefore,

$$0 \geq C||\hat{\theta} - \theta_n||^2 + ||G^T W_n g_n(\theta_n)||\ ||k_n^{-1}(\hat{\theta} - \theta_n)|| - ||\hat{\theta} - \theta_n||(1 + \sqrt{n}||\hat{\theta} - \theta_n||)o_p(n^{-1/2}) - o_{p}.(n^{-1}) \tag{67}$$

By the fact that $G^T W_n g_n(\theta_n) \xrightarrow{p} O_p(n^{-1/2})$, we have

$$\begin{aligned} 0 &\geq C||\hat{\theta} - \theta_n||^2 + O_p(n^{-1/2})\ ||k_n^{-1}(\hat{\theta} - \theta_n)|| - ||\hat{\theta} - \theta_n||o_p(n^{-1/2}) - ||\hat{\theta} - \theta_n||^2 o_p(1) - o_p(n^{-1}) \\ &\geq [C + o_p(1)]||\hat{\theta} - \theta_n||^2 + O_p(n^{-1/2})||k_n^{-1}(\hat{\theta} - \theta_n)|| - o_p(n^{-1}). \end{aligned} \tag{68}$$

Since $C + o_p(1)$ is bounded away from zero, we have

$$||\hat{\theta} - \theta_n||^2 + O_p(n^{-1/2})||k_n^{-1}(\hat{\theta} - \theta_n)|| \leq o_p(n^{-1}). \tag{69}$$

Hence,

$$[||k_n^{-1}(\hat{\theta} - \theta_n)|| + O_p(n^{-1/2})]^2 \leq ||(\hat{\theta} - \theta_n)||^2 + O_p(n^{-1/2})||k_n^{-1}(\hat{\theta} - \theta_n)|| + O_p(n^{-1}) \leq O_p(n^{-1}). \tag{70}$$

Taking square root for both sides yields,

$$\Big| ||k_n^{-1}(\hat{\theta} - \theta_n)|| + O_p(n^{-1/2}) \Big| = O_p(n^{-1/2}). \tag{71}$$

Therefore, by the triangle inequality, we have

$$||k_n^{-1}(\hat{\theta} - \theta_n)|| \leq \Big| ||k_n^{-1}(\hat{\theta} - \theta_n)|| + O_p(n^{-1/2}) \Big| + |-O_p(n^{-1/2})| = O_p(n^{-1/2}), \tag{72}$$

which completes the proof that $k_n^{-1}(\hat{\theta} - \theta_n)$ is $\sqrt{n}$ consistent.

Next, let

$$\tilde{\theta} = \theta_n - [k_n^{-1} G^T W G k_n^{-1}]^{-1}(k_n^{-1} G^T W_n) g_n(\theta_n). \tag{73}$$

Therefore, with $W = \Omega^{-1}$ and $W_n \xrightarrow{p} W$, we have

$$\sqrt{n}[k_n^{-1}(\tilde{\theta} - \theta_n)] = -[G^T W G]^{-1}(G^T W_n)\sqrt{n} g_n(\theta_n) \xrightarrow{d} N(0, [G^T \Omega^{-1} G]^{-1}). \tag{74}$$

Since

$$Q_n(\theta) - Q_n(\theta_n) \approx 2D_n^T(\theta - \theta_n) + (\theta - \theta_n)^T H(\theta - \theta_n), \tag{75}$$

we have

$$Q_n(\theta) - Q_n(\theta_n) + \underbrace{Q(\theta_n)}_{=0} = (\theta - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\theta - \theta_n) + 2(\theta - \theta_n) k_n^{-1} (G^T W_n) g_n(\theta_n) + o_p(n^{-1}). \tag{76}$$

By the definition of $\tilde{\theta}$, we have

$$-(G^T W G)[k_n^{-1}(\tilde{\theta} - \theta_n)] = (G^T W_n) g_n(\theta_n). \tag{77}$$

Therefore,

$$Q_n(\theta) - Q_n(\theta_n) = (\theta - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\theta - \theta_n) - 2(\theta - \theta_n) k_n^{-1} G^T W G k_n^{-1} (\tilde{\theta} - \theta_n) + o_p(n^{-1}). \tag{78}$$

Similarly, we can get

$$\begin{aligned} Q_n(\tilde{\theta}) - Q_n(\theta_n) &= (\tilde{\theta} - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\tilde{\theta} - \theta_n) + 2(\tilde{\theta} - \theta_n)^T k_n^{-1} (G^T W_n) g_n(\theta_n) + o_p(n^{-1}) \\ &= (\tilde{\theta} - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\tilde{\theta} - \theta_n) - 2(\tilde{\theta} - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\tilde{\theta} - \theta_n) + o_p(n^{-1}) \\ &= -(\tilde{\theta} - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\tilde{\theta} - \theta_n) + o_p(n^{-1}). \end{aligned} \tag{79}$$

Since $\tilde{\theta} \in \Theta$, we have

$$Q_n(\tilde{\theta}) - Q_n(\hat{\theta}) = Q_n(\tilde{\theta}) - Q_n(\theta_n) - (Q(\hat{\theta}) - Q_n(\theta_n)) = o_p(n^{-1}). \tag{80}$$

Therefore,

$$-(\tilde{\theta} - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\tilde{\theta} - \theta_n) - (\hat{\theta} - \theta_n)^T k_n^{-1} G^T W G k_n^{-1} (\hat{\theta} - \theta_n) + 2(\hat{\theta} - \theta_n) k_n^{-1} G^T W G k_n^{-1} (\tilde{\theta} - \theta_n) = o_p(n^{-1}), \tag{81}$$

which implies

$$-(k_n^{-1}(\hat{\theta} - \theta_n) - k_n^{-1}(\tilde{\theta} - \theta_n))^T (k_n^{-1} G^T W G k_n^{-1})(k_n^{-1}(\hat{\theta} - \theta_n) - k_n^{-1}(\tilde{\theta} - \theta_n)) = o_p(n^{-1}). \tag{82}$$

Since $G^T W G$ is positive definite, we can find a constant $C \geq 0$ such that

$$-C||k_n^{-1}(\hat{\theta} - \theta_n) - k_n^{-1}(\tilde{\theta} - \theta_n)||^2 = o_p(n^{-1}). \tag{83}$$

Therefore,

$$||k_n^{-1}(\hat{\theta} - \theta_n) - k_n^{-1}(\tilde{\theta} - \theta_n)|| = o_p(n^{-1/2}). \tag{84}$$

Hence,

$$\sqrt{n} k_n^{-1} ||\hat{\theta} - \tilde{\theta}|| \xrightarrow{p} 0. \tag{85}$$

Following

$$\sqrt{n} k_n^{-1}(\tilde{\theta} - \theta_n) \xrightarrow{d} N(0, (G^T \Omega^{-1} G)^{-1}), \tag{86}$$

we have

$$\sqrt{n} k_n^{-1}(\hat{\theta} - \theta_n) \xrightarrow{d} N(0, (G^T \Omega^{-1} G)^{-1}), \tag{87}$$

D Proof of Theorem 3:

For a fixed $\psi \in \Theta_\psi$,

$$\begin{pmatrix} \hat{\beta}(\psi) - \beta_0 \\ \hat{\delta}(\psi) - \delta_n \end{pmatrix} = \Big(\hat{G}(\psi)^T \hat{\Omega}(\hat{\psi})^{-1} \hat{G}(\psi)\Big)^{-1} \hat{G}(\psi)^T \hat{\Omega}(\hat{\psi})^{-1} \Big(g_n(\theta_n) + \frac{1}{n} \sum_{t=1}^{n} z_t \delta_n^T \tilde{x}_t (I(\psi_0) - I(\psi))\Big) \tag{88}$$

Under the null, $\delta_n = 0$, we have

$$\hat{\delta}(\psi) = R\Big(\hat{G}(\psi)^T \hat{\Omega}(\hat{\psi})^{-1} \hat{G}(\psi)\Big)^{-1} \hat{G}(\psi)^T \hat{\Omega}(\hat{\psi})^{-1} g_n(\theta_n), \tag{89}$$

where R is defined in (23).

First, by applying lemma 1, it is straightforward to show that $\hat{G}(\psi) \xrightarrow{p} G(\psi)$ uniformly in $\psi \in \Theta_\psi$.

Next, we show that $\hat{\Omega}(\hat{\psi}) \xrightarrow{p} \Omega$.

Simple calculation shows

$$\hat{\varepsilon}_t = \varepsilon_t - \left(\hat{\beta} - \beta_0\right)^T x_t - \delta_n^T \tilde{x}_t \left(I(\hat{\psi}) - I(\psi_0)\right) - \left(\hat{\delta} - \delta_n\right)^T \tilde{x}_t I(\hat{\psi}). \tag{90}$$

Hence,

$$\begin{aligned}
& \hat{\Omega}(\hat{\psi}) - \frac{1}{n}\sum_{t=1}^{n} z_t z_t^T \varepsilon_t^2 \\
= \quad & -\frac{2}{n}\sum_{t=1}^{n} z_t z_t^T \varepsilon_t x_t^T \left(\hat{\beta} - \beta_0\right) \\
& -\frac{2}{n}\sum_{t=1}^{n} z_t z_t^T \varepsilon_t \tilde{x}_t^T I(\hat{\psi}) \left(\hat{\delta} - \delta_n\right) \\
& -\frac{2}{n}\sum_{t=1}^{n} z_t z_t^T \varepsilon_t \delta_n^T \tilde{x}_t \left(I(\hat{\psi}) - I(\psi_0)\right) \\
& +\frac{1}{n}\sum_{t=1}^{n} z_t z_t^T \left(\hat{\beta} - \beta_0\right)^T x_t x_t^T \left(\hat{\beta} - \beta_0\right) \\
& +\frac{1}{n}\sum_{t=1}^{n} z_t z_t^T \delta_n^T \tilde{x}_t \tilde{x}_t^T \delta_n \left(I(\hat{\psi}) - I(\psi_0)\right) \\
& +\frac{1}{n}\sum_{t=1}^{n} z_t z_t^T \left(\hat{\delta} - \delta_n\right)^T \tilde{x}_t \tilde{x}_t^T \left(\hat{\delta} - \delta_n\right) I(\hat{\psi}).
\end{aligned} \tag{91}$$

For the first term, we have

$$\frac{1}{n}|| \sum_{t=1}^{n} z_t z_t^T \varepsilon_t x_t^T \left(\hat{\beta} - \beta_0\right) || \leq \frac{1}{n}\sum_{t=1}^{n} ||z_t||^2 \; ||\varepsilon_t|| \; ||x_t|| \; ||\hat{\beta} - \beta_0|| \xrightarrow{p} 0, \tag{92}$$

because the boundedness assumption and the consistency of $\hat{\beta}$.

Similarly, we can show

$$
\begin{aligned}
\frac{1}{n}||\sum_{t=1}^{n} z_t z_t^T \varepsilon_t \tilde{x}_t^T I(\hat{\psi})\Big(\hat{\delta}-\delta_n\Big)|| &\le \frac{1}{n}\sum_{t=1}^{n} ||z_t||^2\ ||\varepsilon_t||\ ||\tilde{x}_t||\ ||\hat{\delta}-\delta_n|| \xrightarrow{p} 0 \\
\frac{1}{n}||\sum_{t=1}^{n} z_t z_t^T \Big(\hat{\beta}-\beta_0\Big)^T x_t x_t^T \Big(\hat{\beta}-\beta_0\Big)|| &\le \frac{1}{n}\sum_{t=1}^{n} ||z_t||^2\ ||x_t||^2\ ||\hat{\beta}-\beta_0||^2 \xrightarrow{p} 0 \\
\frac{1}{n}||\sum_{t=1}^{n} z_t z_t^T \Big(\hat{\delta}-\delta_n\Big)^T \tilde{x}_t \tilde{x}_t^T \Big(\hat{\delta}-\delta_n\Big) I(\hat{\psi})|| &\le \frac{1}{n}\sum_{t=1}^{n} ||z_t||^2\ ||\tilde{x}_t||\ ||\hat{\delta}-\delta_n||^2 \xrightarrow{p} 0 \qquad (93)
\end{aligned}
$$

Next, under assumptions 1, 2, 3 and applying lemma 1, we obtain

$$
||\frac{1}{n}\sum_{t=1}^{n} z_t z_t^T \varepsilon_t \delta_n^T \tilde{x}_t \Big(I(\hat{\psi})-I(\psi_0)\Big) - E\Big(z_t z_t^T \varepsilon_t \delta_n^T \tilde{x}_t \Big(I(\hat{\psi})-I(\psi_0)\Big)\Big)|| \xrightarrow{p} 0.
$$

By applying lemma 2, we have,

$$
E\Big(||z_t z_t^T \varepsilon_t \delta_n^T \tilde{x}_t \Big(I(\hat{\psi})-I(\psi_0)\Big)||\Big) \le C||\hat{\psi}-\psi_0|| \xrightarrow{p} 0, \qquad (94)
$$

where $C < \infty$ and $||\hat{\psi}-\psi_0|| \xrightarrow{p} 0$.

Similarly, we can show,

$$
\begin{aligned}
&\frac{1}{n}||\sum_{t=1}^{n} z_t z_t^T \delta_n^T \tilde{x}_t \tilde{x}_t^T \delta_n \Big(I(\hat{\psi})-I(\psi_0)\Big) - E\Big(z_t z_t^T \delta_n^T \tilde{x}_t \tilde{x}_t^T \delta_n \Big(I(\hat{\psi})-I(\psi_0)\Big)\Big)|| \xrightarrow{p} 0 \\
&E\Big(||z_t z_t^T \delta_n^T \tilde{x}_t \tilde{x}_t^T \delta_n \Big(I(\hat{\psi})-I(\psi_0)\Big)||\Big) \xrightarrow{p} 0 \qquad (95)
\end{aligned}
$$

Due to $\frac{1}{n}\sum_{t=1}^{n} z_t z_t^T \varepsilon_t^2 \xrightarrow{a.s} \Omega$, to sum up, we have $\hat{\Omega}(\hat{\psi}) \xrightarrow{p} \Omega$.

Then, applying the continuous mapping theorem on equation (89), we have,

$$
\sqrt{n}\hat{\delta}(\psi) \xrightarrow{d} R(G(\psi)^T \Omega^{-1} G(\psi))^{-1} G(\psi)^T \Omega^{-1/2} V, \qquad (96)
$$

where R and V are defined in (23). This completes the proof.

E Proof of Theorem 4

Let $\theta_0 = [\beta_0^T, \delta_n^T]^T$. By theorem 2, we have

$$\sqrt{n}(\hat{\theta}(\hat{\psi}) - \theta_0) - \sqrt{n}(\hat{\theta}(\psi_0) - \theta_0) = o_p(1), \tag{97}$$

$$\sqrt{n}(\bar{\theta}(\hat{\psi}) - \theta_0) - \sqrt{n}(\bar{\theta}(\psi_0) - \theta_0) = o_p(1). \tag{98}$$

For the purposes of this theorem, we assume knowledge of ψ_0. Therefore, we can show that

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}(\psi_0) - \theta_0) \\ \sqrt{n}(\tilde{\theta}(\psi_0) - \theta_0) \end{pmatrix} = \begin{pmatrix} \left(\hat{G}(\psi_0)^T \hat{\Omega}(\psi_0)^{-1} \hat{G}(\psi_0)\right)^{-1} \hat{G}(\psi_0)^T \hat{\Omega}(\psi_0)^{-1} & 0 \\ 0 & \left(\tilde{G}(\psi_0)^T \tilde{\Omega}(\psi_0)^{-1} \tilde{G}(\psi_0)\right)^{-1} \tilde{G}(\psi_0)^T \tilde{\Omega}(\psi_0)^{-1} \end{pmatrix}$$

$$\times \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n z_t \varepsilon_t \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n x_t \varepsilon_t \end{pmatrix},$$

where

$$\tilde{G}(\psi_0) = [-\frac{1}{n} \sum_{t=1}^n x_t x_t^T, -\frac{1}{n} \sum_{t=1}^n x_t \tilde{x}_t^T I(\psi_0)],$$

$$\tilde{\Omega}(\psi_0) = \frac{1}{n} \sum_{i=1}^n x_t x_t^2 \tilde{\varepsilon}_t,$$

$$\tilde{\varepsilon}_t = y_t - \tilde{\theta}(\psi_0)^T (x_t, \tilde{x}^T I(\psi_0)).$$

Following the proof of of lemma 1 of Kapetanios (2000), we have

$$\begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n z_t \varepsilon_t \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n x_t \varepsilon_t \end{pmatrix} \xrightarrow{d} N\left(0, \begin{matrix} \Omega & \Omega_{zx} \\ \Omega_{xy} & \Omega_{xx} \end{matrix}\right), \tag{99}$$

where $\Omega = E(z_t z_t^T \varepsilon_t^2)$, $\Omega_{zx} = E(z_t x_t^T \varepsilon_t^2)$, $\Omega_{xy} = E(x_t z_t^T \varepsilon_t^2)$, and $\Omega_{xx} = E(x_t x_t^T \varepsilon_t^2)$.

Next, similar to the proof of $\hat{\Omega}(\hat{\psi}) \xrightarrow{p} \Omega$, it is straightforward to show $\hat{\Omega}(\psi_0) \xrightarrow{p} \Omega$ and $\tilde{\Omega}(\psi_0) \xrightarrow{p} \Omega_{xx}$. By lemma 1, we have $\hat{G}(\psi_0) \xrightarrow{p} G(\psi_0)$ and $\tilde{G}(\psi_0) \xrightarrow{p} G_{xx}(\psi_0)$, where $G_{xx}(\psi_0) = [-E(x_t x_t^T), -E(x_t \tilde{x}_t^T I(\psi_0))]$.

Therefore, under the null hypothesis of no endogeneity in regressors, we have

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}(\psi_0) - \theta_0) \\ \sqrt{n}(\tilde{\theta}(\psi_0) - \theta_0) \end{pmatrix} \xrightarrow{d} N\Bigg(0, \begin{pmatrix} \Upsilon_{zx} & \Upsilon_{zx} G(\psi_0)^T \Omega^{-1} \Omega_{zx} \Omega_{xx}^{-1} G_{xx}(\psi_0) \Upsilon_{xx} \\ \Upsilon_{xx} G_{xx}(\psi_0)^T \Omega_{xx}^{-1} \Omega_{zx} \Omega^{-1} G(\psi_0) \Upsilon_{zx} & \Upsilon_{xx} \end{pmatrix}\Bigg)$$

where $\Upsilon_{zx} = \Big(G(\psi_0)^T \Omega^{-1} G(\psi_0)\Big)^{-1}$ and $\Upsilon_{xx} = \Big(G_{xx}(\psi_0)^T \Omega_{xx}^{-1} G_{xx}(\psi_0)\Big)^{-1}$

This implies

$$\sqrt{n}\Big(\hat{\theta}(\psi_0) - \tilde{\theta}(\psi_0)\Big) \xrightarrow{d} N\Big(0, V\Big),$$

where

$$\begin{aligned} V = \Upsilon_{zx} + \Upsilon_{xx} - \Upsilon_{zx} G(\psi_0)^T \Omega^{-1} \Omega_{zx} \Omega_{xx}^{-1} G_{xx}(\psi_0) \Upsilon_{xx} - \Upsilon_{xx} G_{xx}(\psi_0)^T \Omega_{xx}^{-1} \Omega_{zx} \Omega^{-1} G(\psi_0) \Upsilon_{zx} \\ = Var(\hat{\theta}(\psi_0)) + Var(\tilde{\theta}(\psi_0)) - 2Cov(\hat{\theta}(\psi_0), \tilde{\theta}(\psi_0)). \end{aligned}$$

Evidently, with conditional homoskedasticity, $\Upsilon_{xx} = \sigma_\varepsilon^2 G_{xx}(\psi_0)^{-1}$ and $\Upsilon_{zx} = \sigma_\varepsilon^2 \Big(G(\psi_0)_{xx}^T(\psi_0)^{-1} G(\psi_0)\Big)^{-1}$, which implies $\tilde{\theta}(\psi_0)$ is more efficient. Hence, following Hausman (1978), $V = Var(\hat{\theta}(\psi_0)) - Var(\tilde{\theta}(\psi_0))$, which completes our proof.

F Finite Sample Performance of the Test for Linearity

To assess the finite sample performance of the linearity test, we use a model similar to (27),

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

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 8. The replication number is 2000. Throughout the analysis, we use a significance level of 5%. As expected, size is approaching to 5% as sample size increases. Power is increasing in b, and increasing in n.

Table 8: 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

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.

G A Heuristic Example to Illustrate the Smoothness of the GMM Estimator

To provide more intuition for the Theorem 2, 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), and the GMM estimator (GMM). 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 $\varepsilon_i \sim N(0,\sigma^2)$.

Hence, in this example, we assume all threshold variables are exogenous, and the threshold effect is fixed.

G.1 The LSE

As shown in Yu (2015), the LSE can be obtained as,

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

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

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,

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

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 $N_{1n}(|v|) = \sum_{i=1}^{n}(I(\frac{v}{n} \leq -\frac{q_{1i}+q_{2i}\psi_0}{q_{2i}} \leq 0),\ N_{2n}(v) = \sum_{i=1}^{n}(I(0 \leq -\frac{q_{1i}+q_{2i}\psi_0}{q_{2i}} \leq \frac{v}{n}),\ \bar{z}_{1i} = 1 + 2\varepsilon_i$, 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, $\lambda \to f_z(0)v$. As $n \to \infty$, $P_n(v) \to 0$, which implies $N_{2n}(v) \to N_2(v)$. 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,

$$n(\widehat{\psi} - \psi_0) \xrightarrow{d} \arg\min_v D^{LSE}(v), \tag{100}$$

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

G.2 The SLSE

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

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

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.

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,

$$\begin{aligned} T_n(\psi, \sigma_n) = \frac{\partial S_n^{SLS}(\psi)}{\partial \psi} &= -\frac{2}{n}\sum_{i=1}^{n} I(\psi_0)K'(\psi, \sigma_n)\frac{q_{2i}}{\sigma_n} + \frac{2}{n}\sum_{i=1}^{n} K(\psi, \sigma_n)K'(\psi, \sigma_n)\frac{q_{2i}}{\sigma_n} - \frac{2}{n}\sum_{i=1}^{n} K'(\psi, \sigma_n)\frac{q_{2i}}{\sigma_n}\varepsilon_i \\ &= A(\psi) + B(\psi) + C(\psi), \end{aligned}$$

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

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

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

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,

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

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

Hence, we have,

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

Then, by the first order Taylor series,

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

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

As a result, this provides the asymptotic normality,

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

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

G.3 The GMM

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

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

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

Note that, similar to the LSE, the objective function is non-smooth in ψ. Now, assuming the knowledge of the consistency and the converge rate [3], let $\psi = \psi_0 + \frac{v}{n^{1/2}}$. Hence, the centered

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

process can be shown as,

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

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) [4]. 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,

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

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

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

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

Similarly, we can show that,

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

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

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

[4] 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 Philips (2018) explains on how compound Poisson process can be approximated by two-sided Brownian motion.

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

This follows that,

$$n^{1/2}(\widehat{\psi}^{GMM} - \psi) \xrightarrow{d} \hat{v} = \arg\min_{v}[2G_{\psi}(\psi_0)^2 v^2 + 2G_{\psi}(\psi_0)vW],$$

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

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

$$n^{1/2}(\widehat{\psi}^{GMM} - \psi_0) \sim N\Big(0, \big(G_{\psi}(\psi_0)\mathit{\Omega}^{-1}G_{\psi}(\psi_0)\big)^{-1}\Big).$$