Journal of
Risk and Financial Management

Article

Estimation and Inference for the Threshold Model with Hybrid Stochastic Local Unit Root Regressors

Chaoyi Chen [1,2,*,†] and Thanasis Stengos [3,†]

[1] Magyar Nemzeti Bank (Central Bank of Hungary), 1013 Budapest, Hungary
[2] MNB Institute, John von Neumann University, 6000 Kecskemét, Hungary
[3] Department of Economics and Finance, University of Guelph, Guelph, ON N1G 2W1, Canada; tstengos@uoguelph.ca
* Correspondence: chenc@mnb.hu
† These authors contributed equally to this work.

Abstract: In this paper, we study the estimation and inference of the threshold model with hybrid local stochastic unit root regressors. Our main contribution is to propose an estimator that generalizes the threshold model with various forms of nonstationary regressors and to obtain its limiting distribution theory. In particular, our proposed model generalizes the threshold model with unit root, local-to-unity, and stochastic unit root regressors. We provide the estimation strategy for the least squares estimator and derive the asymptotic results for the proposed estimator. Depending on the diminishing rate of the threshold effect, we find that the limiting distribution of the threshold estimator takes different forms. Monte Carlo simulations are used to assess our proposed estimator's finite sample performance, which is found to perform well.

Keywords: threshold model; hybrid local stochastic unit root

Citation: Chen, Chaoyi, and Thanasis Stengos. 2022. Estimation and Inference for the Threshold Model with Hybrid Stochastic Local Unit Root Regressors. *Journal of Risk and Financial Management* 15: 242. https://doi.org/10.3390/jrfm15060242

Academic Editor: Shigeyuki Hamori

Received: 12 May 2022
Accepted: 24 May 2022
Published: 28 May 2022

1. Introduction

Widely used to capture nonlinearities in economic relationships, the parametric threshold model has received much attention over the past two decades. Many empirical studies use threshold models in time series applications and examples include the pricing asymmetry of oil prices and the nonlinear effect of public debt-to-GDP ratio on the per capita GDP growth of a given country over time.

The statistical inference of the threshold regression model is well established with stationary regressors. Under the fixed threshold effect assumption, Chan (1993) develops a threshold estimator which converges to a function of a compound Poisson process. Hansen (2000) establishes the limiting results under the diminishing threshold effect assumption and shows that the limiting distribution of the threshold parameter estimator is in the form of two independent Brownian motions. Extending Hansen (2000), Caner and Hansen (2004) further allow for the endogenous slope regressors by using a two-step least squares estimation method. In the spirit of Heckman (1979), Kourtellos et al. (2016) propose using a parametric control function method to allow for the endogenous threshold variable under the joint normality assumption. Yu et al. (2021) correct and expand Kourtellos et al. (2016)'s method, assuming the endogeneity is known with a finite functional form. Kourtellos et al. (2021) further relax Yu et al. (2021)'s assumption on the known endogenous form and use a nonparametric approach to control for the unknown form of endogeneity. Chen et al. (2012) consider a threshold regression model with two separate exogenous threshold variables, while Chen et al. (2021) further allow two thresholds to be endogenous. Yang et al. (2021) extend Hansen (2000) by allowing for a time-varying threshold. Miao et al. (2020) study a panel threshold model with latent group structures. Seo and Linton (2007) explore a linear index threshold model that allows multiple stationary threshold variables to be a linear index of regressors, and Yu and Fan

(2020) provide the distributional theory of a least squares estimator for that model. In light of Seo and Shin (2016), Chen and Stengos (2021) employ a generalized method of moment approach to allow for the endogeneity in the linear index threshold model.

As for the threshold regression model with nonstationary regressors, surprisingly, the asymptotic properties have not been fully developed. Among the few, Caner and Hansen (2001) study a threshold autoregressive model with unit root regressors, and they derive the limiting distribution of Wald tests for a threshold, under the null of a linear model to be nonstandard and different from the stationary case. Gonzalo and Pitarakis (2012) extend the nonlinearity testing to the case of nearly unit root regressors, while Chen (2014) develops the estimation and inference of a threshold model with integrated regressors. One of the findings is that the model may be identified only weakly if the diminishing rate of the threshold effect is fast enough.

Nevertheless, the studies mentioned above restrict the regressors to either unit root or nearly unit root, though many series, particularly from empirical macroeconomics and finance, display various forms of dependence in different periods. Recent empirical work also sheds light on the advantages of linking multiple regimes of stationary and non-stationary behaviors and transition mechanisms between them (e.g., Phillips et al. 2011; Phillips and Yu 2011). Hence, a statistical inference of a threshold regression model that accommodates various forms of nonstationary regressors is needed.

Our paper aims to fill this gap in the literature. Our main contribution in this paper is to propose an estimator that generalizes the threshold model with various forms of nonstationary regressors and to obtain its limiting distribution theory. Following Lieberman and Phillips (2020), we propose a threshold regression model with hybrid local stochastic unit root (HLSUR) regressors. The proposed model generalizes the threshold model well, with unit root, local-to-unity, and stochastic unit root regressors. We provide a least squares estimation strategy and derive the limiting results of the threshold estimator. We also demonstrate the good finite sample performance of our proposed estimator via Monte Carlo simulations.

The remainder of the paper is organized as follows. In Section 2, we introduce the threshold model with HLSUR regressors and provide an estimation strategy, while in Section 3, we derive the limiting results. Section 4 provides a heuristic example to discuss the limiting results. In Section 5, we show the inference for the threshold effect. Section 6 reports the Monte Carlo results for the proposed estimators. Finally, Section 7 concludes the paper. Technical proofs are relegated to Appendix A.

To proceed, we adopt the following notation. The indicator function is denoted as $I(.)$. We use $\|.\|$ to denote the Euclidean norm. $[np]$ denotes the integer part of np. $\xrightarrow{p}$, $\xrightarrow{d}$, $\Longrightarrow$, and $\sim$ denote convergence in probability, convergence in distribution, weak convergence, and equivalent to, respectively. $\bigwedge$ and $\bigvee$ denote the minimum and maximum operators. Following Caner and Hansen (2001), we make $W(s,\mu)$ a two-parameter Brownian motion on $(s,\mu) \in [0,1]^2$.

2. The Model and Estimators

Consider the following threshold model,

$$y_t = \beta^T x_t + \delta_n^T x_t I_(q_t \leq \gamma_0) + \mu_t, \quad (1)$$

where x_t is a $d_x \times 1$ vector of hybrid stochastic unit root processes, for all $j = 1, \ldots, d_x$, with the form

$$x_{j,t} = \beta_{nt} x_{j,t-1} + v_{j,t}, \quad (2)$$
$$\beta_{nt} = e^{\frac{c_j}{n} + \frac{\alpha_j^T \varepsilon_{jt}}{\sqrt{n}}},$$

where c_j is a localizing constant, ε_{jt} is a $d_{\varepsilon_j} \times 1$ vector, and $x_{j1} = v_{j1}$. Note that the specification of β_{nt} allows for different types of dependence. For example, if all α_j and c_j are zero, we will have the standard unit root case.

Let $\varepsilon_t = [\varepsilon_{1t}^T, \ldots, \varepsilon_{d_xt}^T]^T$ and $v_t = [v_{1t}, \ldots, v_{d_xt}]^T$. We assume the vector $\eta_t = (\mu_t, v_t^T, \varepsilon_t^T)^T$ to be a strictly stationary martingale difference sequence, and the partial sums satisfy the following invariance principle

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[np]} \eta_t \Longrightarrow \mathbf{B}(p) = \left[B_\mu(p), \mathbf{B}_v^T(p), \mathbf{B}_\varepsilon^T(p)\right]^T, \tag{3}$$

where $\mathbf{B}_v(p) = [B_{v_1}(p), \ldots, B_{v_{d_x}}(p)]^T$, $\mathbf{B}_\varepsilon(p) = [B_{\varepsilon_1}^T(p), \ldots, B_{\varepsilon_{d_x}}^T(p)]^T$, and $\mathbf{B}(p)$ is a vector of Brownian motions with a long run positive definite covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_\mu^2 & 0 & 0 \\ 0 & \Sigma_v & 0 \\ 0 & 0 & \Sigma_\varepsilon \end{bmatrix}, \tag{4}$$

and $\sigma_\mu^2 > 0$, $\Sigma_v > 0$, and $\Sigma_\varepsilon > 0$.

We can rewrite model (1) in a stacked form

$$\mathbf{y} = X(\gamma)\theta + \mu, \tag{5}$$

where $\theta = \left[\beta^T, \delta_n^T\right]^T$, $\mathbf{y}$ stacks up y_t, μ stacks up μ_t, $X(\gamma)$ stacks up $X_t(\gamma)$, and $X_t(\gamma) = \left[x_t^T, x_t^T I(q_t \le \gamma)\right]^T$.

Hence, given a fixed $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, model (5) is linear in θ. We obtain

$$\widehat{\theta}(\gamma) = \left[X(\gamma)^T X(\gamma)\right]^{-1} X(\gamma)^T \mathbf{y}. \tag{6}$$

The least squares estimator of γ_0 is

$$\widehat{\gamma} = arg \min_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} SSR(\gamma), \tag{7}$$

where $SSR(\gamma) = \sum_{t=1}^n \left(y_t - \widehat{\theta}(\gamma)^T X_t(\gamma)\right)^2$.

Due to the non-smooth nature of the objective function, following Hansen (2000), we employ a grid-search method to estimate the model empirically.

3. Asymptotic Properties for the Estimators

We make the following assumptions to support model (1).

Assumption 1. *$\{q_t\}$ is a strictly stationary mixing sequence with the mixing coefficient of size $-r/(r-2)$ for some $r > 2$.*

Assumption 2. *Let $\mathcal{F}_{n,t}$ be the smallest sigma-field generated by $\{q_{s+1}, \eta_s^T : 1 \le s < t \le n\}$. $\{(\eta_t, \mathcal{F}_{n,t})\}_{t=1}^n$ is a strictly martingale difference sequence (MDS) with positive definite covariance $E(\eta_t \eta_t^T | \mathcal{F}_{n,t})$, whose partial sums satisfy the invariance principle, as in (3).*

Assumption 3. *The threshold variable q_t has a continuous distribution $F(.)$. $f(.)$ is the corresponding density function with $0 < f(\gamma) < \overline{f} < \infty$ for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$.*

Define

$$G_{\alpha_j, c_j}(s) = e^{sc_j + \alpha_j^T B_{\varepsilon_j}(s)} \left(\int_0^s e^{-pc_j - \alpha_j^T B_{\varepsilon_j}(p)} dB_{v_j}(p) \right), \tag{8}$$

and

$$G_{\alpha,c}(s) = \left[G_{\alpha_1,c_1}(s), \ldots, G_{\alpha_{d_x},c_{d_x}}(s)\right]^T. \tag{9}$$

Assumption 4. *$\int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds$ is positive definite.*

Assumption 1 ensures the threshold variable q_t is strictly stationary. Assumption 2 states that q_t is contemporaneously exogenous in model (1) and ensures the multivariate invariance principle for the partial sum of a martingale difference array. Assumption 3 is common in the threshold regression literature. This assumption assumes the threshold variable is continuous with positive density everywhere, ensuring they are dense near the true threshold level. Assumption 4 is a regular full rank condition.

Theorem 1. *Under Assumptions 1–4, with $\delta_n = n^{-1/2-\rho}\delta_0$ and $\rho \in (-1/2, 1/2)$, we have $\widehat{\gamma} - \gamma_0 = o_p(1)$.*

Theorem 2. *Under Assumptions 1–4, with $\delta_n = n^{-1/2-\rho}\delta_0$, the following limiting results hold:*

(i). *if $\rho \in (0, \frac{1}{2})$,*

$$n^{1-2\rho}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \lambda T, \tag{10}$$

where $\lambda = \frac{\sigma_\mu^2}{f(\gamma_0)\delta_0^T \int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds\delta_0}$, and $T = \underset{r\in(-\infty,+\infty)}{\text{argmax}}\ T(r)$ with $T(r)$ denotes a two-sided Brownian motion on the real line

$$T(r) = \begin{cases} W_1(r) - \frac{1}{2}r, & \text{if } r > 0, \\ 0, & \text{if } r = 0, \\ W_2(-r) - \frac{1}{2}|r|, & \text{if } r < 0, \end{cases} \tag{11}$$

and $W_1(r)$ and W_2 (r) are two independent standard Brownian motion processes defined on $[0, \infty)$;

(ii). *if $\rho = 0$,*

$$n(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \underset{v}{\text{argmin}}\ D(v), \tag{12}$$

where

$$D(v) = \begin{cases} \delta_0^T \sum_{t=1}^{N_1(v)} (x_t x_t^T \delta_0 - 2x_t\mu_t), & \text{if } v > 0, \\ \delta_0^T \sum_{t=1}^{N_2(|v|)} (x_t x_t^T \delta_0 + 2x_t\mu_t), & \text{if } v \le 0, \end{cases}$$

and $N_1(v)$ and $N_2(v)$ are two independent Poisson process with intensity $f(\gamma_0)$.

Theorems 1 and 2 show the threshold estimator's consistency and asymptotic distribution, respectively. We make the following remarks for Theorem 2.

Remark 1. *The convergence rate of $\widehat{\gamma}$ depends only on the diminishing rate of the threshold effect, ρ. It does not relate to both the localizing constant, c, and the coefficient of the stochastic term, α.*

Remark 2. *If the diminishing rate of the threshold effect is fast enough ($\rho \in (0, 1/2)$), the asymptotic distribution of $\widehat{\gamma}$ is in the form of the argmin of the drifted two-sided Brownian motion and is essentially in the same form of Hansen (2000) and Chen (2014), with a different scaling factor, λ. Intuitively, this is because we have infinite threshold variables in the local neighborhood of γ_0 asymptotically. Thus, we can apply the functional central limit theorem to derive the limiting results. We notice Chen (2014) also allows a "slowly" diminishing rate ($\rho \in (-1/2, 0)$). We exclude that case because, as we show in Lemma A4 of the appendix, the uniform law of large numbers and the functional central limit theorem can be applied if and only if $\frac{a_n}{n}$ goes to zero, where a_n is the convergence rate of the threshold estimator. This condition is well documented in the change point* *model. For example, Kejriwal and Perron (2008) assume the jump size is diminishing at a faster rate with the nonstationary regressors than with the stationary regressors.*

Remark 3. *If $\rho = 0$, only a finite number of q_t are involved asymptotically in the local neighborhood of the true threshold level. Under this case, that $P(N(v) \longrightarrow \infty) = 0$ for any finite v holds since the threshold variable is strictly stationarity. Therefore, we can extend Chan (1993)'s distributional theory to the nonstationary regressors. As a result, the limiting distribution of the threshold estimator is given as the argmin of the stochastic process $D(v)$. A similar argument can be found in the comment of proposition 2 of Bai (1997). Note that, if x_t is also strictly stationary, $D(v)$ reduces to a compound Poisson process, as shown in Theorem 2 of Chan (1993).* [1]

Remark 4. *If $\rho < 0$, the threshold estimator converges very fast, such that we have zero information around the local neighborhood of the true threshold level asymptotically. As a result, we cannot obtain a limiting distribution.*

We consider some extensions to Theorem 2. The following provides the limiting results of the threshold estimator for model (1) with some special HLSUR regressors.

Corollary 1. *Under Assumptions 1–4, and $c_j = 0$ for $j = 1, \ldots, d_x$, with $\rho \in (0, \frac{1}{2})$, we have $n^{1-2\rho}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \lambda' T$, where $\lambda' = \frac{\sigma_\mu^2}{f(\gamma_0)\delta_0^T \int_0^1 G_\alpha(s) G_\alpha^T(s) ds \delta_0}$, $G_\alpha(s) = \left[G_{\alpha_1}(s), \ldots, G_{\alpha_{d_x}}(s)\right]^T$, and $G_{\alpha_j}(s) = e^{\alpha_j^T B_{\varepsilon_j}(s)} \left(\int_0^s e^{-\alpha_j^T B_{\varepsilon_j}(p)} dB_{v_j}(p) \right)$.*

Corollary 2. *Under Assumptions 1–4, and $\alpha_j = 0$ for $j = 1, \ldots, d_x$, with $\rho \in (0, \frac{1}{2})$, we have $n^{1-2\rho}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \lambda'' T$, where $\lambda'' = \frac{\sigma_\mu^2}{f(\gamma_0)\delta_0^T \int_0^1 G_c(s) G_c^T(s) ds \delta_0}$, $G_c(s) = \left[G_{c_1}(s), \ldots, G_{c_{d_x}}(s)\right]^T$, and $G_{c_j}(s) = e^{sc_j}(\int_0^s e^{-pc_j} dB_{v_j}(p))$.*

Corollaries 1 and 2 show the limiting results of the threshold estimator in the case of a stochastic unit root and local-to-unity regressors, respectively. If we assume both $c_j = 0$ and $\alpha_j = 0$ for all $j = 1, \ldots, d_x$, the regressors are unit root processes. The limiting distribution further reduces to Theorem 3.1 of Chen (2014).

4. Heuristic Arguments and an Illustrative Example

In this section, we use a simple setup of model (1) to provide more intuitions on the limiting result in Theorem 2. Suppose x_t and α are of dimension one, $\delta_0 = 1$, and $\beta = 0$. In addition, we further assume $\sigma_\varepsilon^2 = \sigma_\mu^2 = \sigma_v^2 = 1$. Hence, model (1) becomes $y_t = n^{-1/2-\rho} x_t I(q_t \leq \gamma_0) + \mu_t$, where $E(G_{\alpha,c}(s)) = 0$ and $Var(G_{\alpha_c,c}(s)) = \frac{e^{2(c+\alpha^2)s} - 1}{2(c+\alpha^2)}$.

Assuming the knowledge of the consistency and the convergence rate, we examine the asymptotic behavior of $SSR(\gamma_0 + \frac{v}{a_n}) - SSR(\gamma_0)$, where $a_n = n^{1-2\rho}$. For any $v \in (0, \infty)$, we have,

$$SSR(\gamma_0 + \frac{v}{a_n}) - SSR_n(\gamma_0) = n^{-1-2\rho} \sum_{t=1}^{n} x_t^2 I(\frac{v}{a_n}) - 2n^{-1/2-\rho} \sum_{t=1}^{n} [x_t I(\frac{v}{a_n}) \mu_t]$$
$$= S_{n1}(v) - 2S_{n2}(v),$$

where $I(\frac{v}{a_n}) = I(\gamma_0 < q_t \leq \gamma_0 + \frac{v}{a_n})$.

Note that $S_{n1}(v) = \frac{1}{n} \sum_{t=1}^{n} (\frac{x_t}{\sqrt{n}})^2 a_n I(\frac{v}{a_n})$. Applying Lemma 1 of Hansen (2000), we have

$$E\left| \frac{a_n}{n} \sum_{t=1}^{n} \left[I(\frac{v}{a_n}) - E(I(\frac{v}{a_n})) \right] \right|^2 = \frac{a_n^2}{n} E\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[I(\frac{v}{a_n}) - E(I(\frac{v}{a_n})) \right] \right|^2 \leq \frac{a_n^2}{n} C \frac{v}{a_n} \longrightarrow 0,$$

where C is a constant. Thus, for $\rho \in (0, \frac{1}{2})$, we can verify, for any finite positive v, $\frac{a_n}{n}\sum_{t=1}^{n} I(\frac{v}{a_n}) \xrightarrow{p} f(\gamma_0)v$. Note that $\frac{1}{n}\sum_{t=1}^{n}(\frac{x_t}{\sqrt{n}})^2 \Longrightarrow \int G_{\alpha,c}(s)^2 ds$. Therefore, we have $S_{n1}(v) \Longrightarrow vf(\gamma_0)\int_0^1 G_{\alpha,c}(s)^2 ds$.

For $\rho = 0$, $E\left|\frac{a_n}{n}\sum_{t=1}^{n}\left(I(\frac{v}{a_n}) - E\left(I(\frac{v}{a_n})\right)\right)\right|^2 = O(1)$. This implies $\frac{a_n}{n}\sum_{t=1}^{n} I(\frac{v}{a_n}) \stackrel{p}{\nrightarrow} f(\gamma_0)v$, which is well explored in the case of the fixed threshold effect in a threshold regression model with stationary regressors. Following Chan (1993), we can show $\sum_{t=1}^{n} I(\frac{v}{a_n}) \sim Binomial(n, p(v)) \Longrightarrow N(v)$,[2] where $p(v) = f(\gamma_0)\frac{v}{a_n}$ and $N(v)$ is a Poisson distribution with intensity $f(\gamma_0)$ due to the fact that $np(v) = f(\gamma_0) < \infty$. As a result, $S_{n1}(v) \Longrightarrow \sum_{t=1}^{N(v)} x_t^2$.

Next, we can rewrite $S_{n2}(v)$ as $S_{n2}(v) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{x_t}{\sqrt{n}}\sqrt{a_n}I(\frac{v}{a_n})\mu_t$. For $\rho \in (0, \frac{1}{2})$, we have

$$E\left|\frac{a_n}{n}\sum_{t=1}^{n}\left[\mu_t^2 I(\frac{v}{a_n}) - E\left(\mu_t^2 I(\frac{v}{a_n})\right)\right]\right|^2 = \frac{a_n^2}{n}E\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\mu_t^2 I(\frac{v}{a_n}) - E\left(\mu_t^2 I(\frac{v}{a_n})\right)\right]\right|^2 \longrightarrow 0,$$

Thus, we can show $\frac{1}{n}\sum_{t=1}^{n} a_n I(\frac{v}{a_n})\mu_t^2 \xrightarrow{p} vf(\gamma_0)$ uniformly for all v. Then, we can apply the functional central limit theorem on $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sqrt{a_n}I(\frac{v}{a_n})\mu_t$. As $\frac{x_{t=[ns]}}{\sqrt{n}} \Longrightarrow G_{\alpha,c}(s)$, we have $S_{n2}(v) \Longrightarrow \int G_{\alpha,c}(s)dW(s, f(\gamma_0)v)$.

By contrast, for $\rho = 0$,

$$E\left|\frac{a_n}{n}\sum_{t=1}^{n}\left[\mu_t^2 I(\frac{v}{a_n}) - E\left(\mu_t^2 I(\frac{v}{a_n})\right)\right]\right|^2 = O(1).$$

Therefore, $\frac{1}{n}\sum_{t=1}^{n} a_n I(\frac{v}{a_n})\mu_t^2 \stackrel{p}{\nrightarrow} vf(\gamma_0)$. Following Chan (1993), we have $S_{n2}(v) \Longrightarrow \sum_{t=1}^{N(v)} x_t\mu_t$.

5. Test for the Linearity

The threshold effect of model (1) disappears if $\delta_0 = 0$. Thus, for a given $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, we can construct the Wald statistic to test the null hypothesis of $\delta_0 = 0$ as

$$W_n(\gamma) = \hat{\delta}^T\left[X_\gamma^T(I_n - P_x)X_\gamma\right]\hat{\delta}/\hat{\sigma}_\mu^2. \tag{13}$$

Following Hansen (1996), we define $SupW = \sup_{\gamma\in[\underline{\gamma},\overline{\gamma}]} W_n(\gamma)$.

Theorem 3. *Under Assumptions 1–4 and $\delta_0 \neq 0$, we have $W_n(\gamma) = O_p(n^{1/2-\rho})$.*

Corollary 3. *Under the null and Assumptions 1–4, we have*

$$SupW \Longrightarrow \sup_{\gamma\in[\underline{\gamma},\overline{\gamma}]}\left(\left\{\int_0^1 G_{\alpha,c}(s)dW(s, F(\gamma)) - F(\gamma)\int_0^1 G_{\alpha,c}(s)dW(s)\right\}^T\right. \tag{14}$$

$$\times\left\{\left[F(\gamma) - F(\gamma)^2\right]\int_0^1 G_{\alpha,c}(s)G_{\alpha,c}(s)^T ds\right\}^{-1} \tag{15}$$

$$\left.\times\left\{\int_0^1 G_{\alpha,c}(s)dW(s, F(\gamma)) - F(\gamma)\int_0^1 G_{\alpha,c}(s)dW(s)\right\}\right)/\sigma_\mu^2.$$

Theorem 3 shows $SupW \Longrightarrow \infty$ under alternative assumptions. Under the null hypothesis, the limiting random variable of $SupW$ depends on some unknown parameters.

Note that if $c = 0$ and $\alpha = 0$, $W_n(\gamma)$ is in the same form as Chen (2014). If $c < 0$ and $\alpha = 0$, Corrollary 3 reduces to Proposition 1 of Gonzalo and Pitarakis (2012).

6. Monte Carlo Studies

To evaluate the finite sample performance of our proposed estimator, we provide Monte Carlo simulation results in this section. We generate our data from following a simple structure:

$$\begin{aligned} y_t &= 0.2x_t + 2n^{-1/2-\rho}x_t I(q_t \leq 0.2) + \mu_t, \\ x_t &= \beta_{nt}x_{t-1} + v_t, \\ \beta_{nt} &= e^{\frac{c}{n}+\frac{\alpha\varepsilon_t}{\sqrt{n}}}, \end{aligned} \tag{16}$$

where q_t, μ_t, v_t, and ε_t are independently normally distributed with mean zero and variance one.

We vary $c \in (-0.5, 0, 0.5)$ and $\alpha \in (0, 0.07, 0.2)$. For each data-generation process, we replicate 2000 times. We consider sample sizes of $n = 300$, 500, and 1000. Tables 1–3 report the simulation results with $\rho = 0$, $1/4$, and $1/3$, respectively.

Overall, we see that the performance of the threshold estimator improves as the sample size increase for all cases. In addition, the finite sample performance slightly changes over various α and c. We observe more considerable bias and MSE as the threshold effect diminishes faster. These observations are consistent with our limiting results, suggesting that the convergence rate of the threshold estimator only relates to the diminishing rate of the threshold effect, ρ.

Table 1. Simulation results with $\rho = 0$.

Panel A: $\alpha = 0$									
	$c = -0.5$			$c = 0$			$c = 0.5$		
n	Bias	MSE	Std	Bias	MSE	Std	Bias	MSE	Std
300	−0.0367	0.0635	0.2494	−0.0270	0.0595	0.2425	−0.0204	0.0473	0.2166
500	−0.0179	0.0448	0.2109	−0.0175	0.0353	0.1870	−0.0119	0.0299	0.1727
1000	−0.0099	0.0170	0.1300	−0.0029	0.0129	0.1136	−0.0021	0.0095	0.0977
Panel B: $\alpha = 0.07$									
	$c = -0.5$			$c = 0$			$c = 0.5$		
n	Bias	MSE	Std	Bias	MSE	Std	Bias	MSE	Std
300	−0.0228	0.0532	0.2297	−0.0197	0.0493	0.2212	−0.0243	0.0399	0.1982
500	−0.0163	0.0292	0.1700	−0.0093	0.0235	0.1531	−0.0105	0.0180	0.1338
1000	−0.0059	0.0101	0.1004	−0.0053	0.0089	0.0942	−0.0037	0.0058	0.0758
Panel C: $\alpha = 0.2$									
	$c = -0.5$			$c = 0$			$c = 0.5$		
n	Bias	MSE	Std	Bias	MSE	Std	Bias	MSE	Std
300	−0.0141	0.0339	0.1835	−0.0077	0.0250	0.1580	−0.0167	0.0234	0.1519
500	−0.0119	0.0174	0.1316	−0.0094	0.0114	0.1062	−0.0047	0.0077	0.0879
1000	−0.0018	0.0039	0.0624	−0.0012	0.0023	0.0476	−0.0022	0.0017	0.0417

This table reports the simulation results with $\rho = 0$. The first column shows the sample size. The second to the fourth columns report the results with $c = -0.5$. The fifth to the seventh columns report the results with $c = 0$. The last three columns report the results with $c = 0.5$. Panel A shows the results with $\alpha = 0$. Panel B reports the results with $\alpha = 0.07$, and Panel C reports the results with $\alpha = 0.2$.

Table 2. Simulation results with $\rho = 1/4$.

Panel A: $\alpha = 0$									
	$c = -0.5$			$c = 0$			$c = 0.5$		
n	Bias	MSE	Std	Bias	MSE	Std	Bias	MSE	Std
300	−0.1519	0.2711	0.4981	−0.1304	0.2649	0.4980	−0.1448	0.2501	0.4788
500	−0.1380	0.2550	0.4859	−0.1201	0.2344	0.4691	−0.1109	0.2160	0.4514
1000	−0.1347	0.2494	0.4810	−0.1077	0.2241	0.4611	−0.1148	0.1951	0.4267
Panel B: $\alpha = 0.07$									
	$c = -0.5$			$c = 0$			$c = 0.5$		
n	Bias	MSE	Std	Bias	MSE	Std	Bias	MSE	Std
300	−0.1450	0.2502	0.4788	−0.1350	0.2311	0.4615	−0.1001	0.2066	0.4435
500	−0.1195	0.2264	0.4607	−0.1106	0.2149	0.4504	−0.1076	0.1894	0.4218
1000	−0.1100	0.1947	0.4274	−0.0640	0.1518	0.3845	−0.0675	0.1391	0.3669
Panel C: $\alpha = 0.2$									
	$c = -0.5$			$c = 0$			$c = 0.5$		
n	Bias	MSE	Std	Bias	MSE	Std	Bias	MSE	Std
300	−0.1095	0.2001	0.4338	−0.0841	0.1686	0.4021	−0.0723	0.1372	0.3633
500	−0.0644	0.1389	0.3671	−0.0619	0.1255	0.3489	−0.0452	0.0902	0.2970
1000	−0.0422	0.0779	0.2759	−0.0211	0.0526	0.2285	−0.0203	0.0433	0.2070

This table reports the simulation results with $\rho = 1/4$. The first column shows the sample size. The second to the fourth columns report the results with $c = -0.5$. The fifth to the seventh columns report the results with $c = 0$. The last three columns report the results with $c = 0.5$. Panel A shows results with $\alpha = 0$. Panel B reports results with $\alpha = 0.07$ and Panel C reports results with $\alpha = 0.2$.

Table 3. Simulation results with $\rho = 1/3$.

Panel A: $\alpha = 0$									
	$c = -0.5$			$c = 0$			$c = 0.5$		
n	Bias	MSE	Std	Bias	MSE	Std	Bias	MSE	Std
300	−0.1903	0.3193	0.5322	−0.1859	0.3105	0.5254	−0.1629	0.2868	0.5102
500	−0.2009	0.3091	0.5185	−0.1798	0.3019	0.5194	−0.1731	0.2886	0.5087
1000	−0.1615	0.2859	0.5099	−0.1524	0.2853	0.5121	−0.1553	0.2712	0.4972
Panel B: $\alpha = 0.07$									
	$c = -0.5$			$c = 0$			$c = 0.5$		
n	Bias	MSE	Std	Bias	MSE	Std	Bias	MSE	Std
300	−0.1549	0.2930	0.5188	−0.1527	0.2804	0.5071	−0.1541	0.2586	0.4848
500	−0.1483	0.2831	0.5111	−0.1551	0.2639	0.4899	−0.1233	0.2385	0.4727
1000	−0.1470	0.2549	0.4831	−0.1201	0.2324	0.4670	−0.1126	0.2162	0.4513
Panel C: $\alpha = 0.2$									
	$c = -0.5$			$c = 0$			$c = 0.5$		
n	Bias	MSE	Std	Bias	MSE	Std	Bias	MSE	Std
300	−0.1494	0.2624	0.4901	−0.1175	0.2288	0.4638	−0.1168	0.2038	0.4361
500	−0.1326	0.2229	0.4533	−0.0834	0.1782	0.4139	−0.0742	0.1512	0.3818
1000	−0.0786	0.1449	0.3725	−0.0633	0.1132	0.3305	−0.0468	0.0783	0.2759

This table reports the simulation results with $\rho = 1/3$. The first column shows the sample size. The second to the fourth columns report the results with $c = -0.5$. The fifth to the seventh columns report the results with $c = 0$. The last three columns report the results with $c = 0.5$. Panel A shows results with $\alpha = 0$. Panel B reports results with $\alpha = 0.07$ and Panel C reports results with $\alpha = 0.2$.

7. Concluding Remarks

The estimation and statistical theories of the threshold model with stationary or integrated regressors are well explored (e.g., Hansen 2000; Caner and Hansen 2001; Chen 2014). Yet, in empirical macroeconomics and finance, many explanatory variables follow either near unit root or stochastic unit root processes (e.g., Phillips et al. 2011; Phillips and Yu 2011; Lieberman and Phillips 2017, 2020). This paper proposes estimating a threshold regression model with HLSUR regressors that has both local-to-unity and stochastic unit root components. The estimation of the threshold parameter is based on a concentrated least squares method. We develop estimations and inferences for the threshold estimator under the diminishing threshold effect assumption, including a sup-Wald statistic to test for the existence of the threshold effect. Finally, we assess the performance of the proposed estimator using Monte Carlo simulations.

There is a wide range of directions open for future work. First, the current paper assumes the threshold variable is strictly stationary. It may be interesting to extend it to be nonstationary. Second, our model relies on exogenous regressors and the threshold variable. Future works can relax this assumption by applying a control function approach. Moreover, a more general model with nonstationary regressors and multiple threshold variables may be interesting. These extensions are left for future studies.

Author Contributions: These authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: Thanasis Stengos acknowledges the financial support from NSERC of Canada.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Lemma

Let $x_{t,\gamma} = x_t I(q_t \leq \gamma)$ and X_γ stack up $x_{t,\gamma}$. Let X stack up x_t, and let P_x be the projection matrix of x_t. Denote I_n as an $n \times n$ identity matrix.

Lemma A1. *Under Assumption 2, we have*

$$(i).\ n^{-1/2} x_{j,t=[ns]} \Longrightarrow G_{\alpha_j,c_j}(s). \tag{A1}$$

$$(ii).\ n^{-1/2} x_{t=[ns]} \Longrightarrow G_{\alpha,c}(s). \tag{A2}$$

Proof. The proof is given in Lemma 1 of Lieberman and Phillips (2020) with no presence of endogeneity in ε_t. □

Lemma A2. *Under Assumptions 1–4, uniformly for any $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, we have*

$$(i).\ n^{-3/2}\sum_{t=1}^{n} x_t \Longrightarrow \int_0^1 G_{\alpha,c}(s)ds, \tag{A3}$$

$$(ii).\ n^{-3/2}\sum_{t=1}^{n} x_{t,\gamma} \Longrightarrow F(\gamma)\int_0^1 G_{\alpha,c}(s)ds \tag{A4}$$

$$(iii).\ n^{-1/2}\sum_{t=1}^{[ns]} I(q_t \le \gamma)\mu_t \Longrightarrow \sigma_\mu W(s, F(\gamma)), \tag{A5}$$

$$(iv).\ n^{-1}\sum_{t=1}^{n} x_{t,\gamma}\mu_t \Longrightarrow \sigma_\mu \int_0^1 G_{\alpha,c}(s)dW(s, F(\gamma)), \tag{A6}$$

$$(v).\ n^{-2}\sum_{t=1}^{n} x_t x_t^T \Longrightarrow \int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds, \tag{A7}$$

$$(vi).\ n^{-2}\sum_{t=1}^{n} x_{t,\gamma}x_{t,\gamma}^T\mu_t \Longrightarrow F(\gamma)\int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds. \tag{A8}$$

$$(vii).\ n^{-2}\sum_{t=1}^{n} X_t(\gamma)X_t(\gamma)^T \Longrightarrow M(\gamma), \tag{A9}$$

where

$$M(\gamma) = \begin{bmatrix} \int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds & F(\gamma)\int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds \\ F(\gamma)\int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds & F(\gamma)\int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds \end{bmatrix}.$$

Proof. (i) is standard by applying Lemma A1. We show the proof of (ii), and the others can be proven in a similar way. Define $x_{nt} = n^{-1/2}x_t$. Closely following the proof of Theorm 3 of Caner and Hansen (2001) with our defined x_{nt} and $w_t = 1$, we can show, uniformly for $\gamma \in [\underline{\gamma}, \overline{\gamma}]$,

$$\sup_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \left| \frac{1}{n}\sum_{t=1}^{n} x_{nt}(I(q_t \le \gamma) - F(\gamma)) \right| \xrightarrow{p} 0.$$

Then, applying (i) and the continuous mapping theorem completes the proof of this Lemma. □

Lemma A3. *Under Assumptions 1–4, for any $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, we have*

$$n^{1/2+\rho}\left(\widehat{\theta}(\gamma) - \theta\right) \Longrightarrow M(\gamma)^{-1}A(\gamma, \gamma_0, \delta_0), \tag{A10}$$

where

$$A(\gamma, \gamma_0, \delta_0) = \begin{bmatrix} (F(\gamma_0) - F(\gamma))\int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds \\ (F(\gamma_0 \wedge \gamma) - F(\gamma))\int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds \end{bmatrix}\delta_0.$$

Proof. By definition, we have $\widehat{\theta}(\gamma) - \theta = \left(X(\gamma)^T X(\gamma)\right)^{-1} X(\gamma)^T[\mu + (X_{\gamma_0} - X_\gamma)\delta_n]$.

Hence, under Assumptions 1–4 and by Lemma A2, we can show

(i). $n^{1/2+\rho}\left(X(\gamma)^T X(\gamma)\right)^{-1} X(\gamma)^T \mu = o_p(1)$,

(ii). $n^{1/2+\rho}\left(X(\gamma)^T X(\gamma)\right)^{-1} X(\gamma)^T (X_{\gamma_0} - X_\gamma)\delta_n$

$$= \left(n^{-2}\sum_{t=1}^{n} X_t(\gamma)X_t(\gamma)^T\right)^{-1}\begin{bmatrix} n^{-2}\sum_{t=1}^{n} x_t\left(x_{t,\gamma_0}^T - x_{t,\gamma}^T\right)\delta_0 \\ n^{-2}\sum_{t=1}^{n} x_{t,\gamma}\left(x_{t,\gamma_0}^T - x_{t,\gamma}^T\right)\delta_0 \end{bmatrix} \Longrightarrow M(\gamma)^{-1}A(\gamma,\gamma_0,\delta_0),$$

where we use the fact that $I(q_t \le \gamma)I(q_t \le \gamma_0) = I(q_t \le \gamma \wedge \gamma_0)$.

This completes the proof of this Lemma. □

Let $a_n = n^{1-2\rho}$, $I(\frac{v}{a_n}) = I(\gamma_0 \le q_t \le \gamma_0 + \frac{v}{a_n})$, $u_t(v) = a_n I(\frac{v}{a_n}) - E(a_n I(\frac{v}{a_n}))$, and $v_{nt}(v) = \sqrt{a_n} e_t I(\frac{v}{a_n})$.

Lemma A4. *Under Assumptions 1–4 and $\rho \in (0, \frac{1}{2})$, for any $v \in [\underline{v}, \bar{v}]$, we have*

$$(i).\ \sup_{v\in[\underline{v},\bar{v}]} \left|\frac{1}{n}\sum_{t=1}^{n} u_t\right| \xrightarrow{p} 0, \tag{A11}$$

$$(ii).\ n^{-1/2}\sum_{t=1}^{n} v_{nt}(v) \Longrightarrow B(v), \tag{A12}$$

where $B(v)$ is a Brownian motion.

Proof. Note that, for each v, by Lemma 1 of Hansen (2000), we check the point convergence condition

$$E\left|\frac{1}{n}\sum_{t=1}^{n} a_n I(\frac{v}{a_n}) - E(a_n\frac{v}{a_n})\right|^2 = \frac{a_n^2}{n}E\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(I(\frac{v}{a_n}) - E(\frac{v}{a_n})\right)\right|^2 \le Cv\frac{a_n}{n} \longrightarrow 0,$$

where C is a finite number.

Then, we can apply Lemma 1 of Hansen (1996) to conclude the proof of (i). Next, given each v, by Lemma 1 of Hansen (2000), we can show

$$E\left|\frac{1}{n}\sum_{t=1}^{n} v_{nt}(v)^2 - E(\frac{1}{n}\sum_{t=1}^{n} v_{nt}(v)^2)\right|^2 \le C'v\frac{a_n}{n} \longrightarrow 0,$$

$$E\left|n^{-1/2}\max_{v\in[\underline{v},\bar{v}]} v_{nt}(v)\right|^4 \le C''v\frac{a_n}{n} \longrightarrow 0.$$

where C' and C'' are two finite numbers.

Then, we can use similar arguments in the proof of Lemma A.11 of Hansen (2000) to complete the proof of (ii). □

Appendix B. Proof of Theorem

Appendix B.1. Proof of Theorem 1

Define the projection matrix $P(\gamma) = X_\gamma^*(X_\gamma^{*T}X_\gamma^*)^{-1}X_\gamma^{*T}$, where $X_\gamma^* = [X_\gamma, X - X_\gamma]$. By simple calculation, we have

$$\begin{aligned} SSR(\gamma) &= \mathbf{y}^T(I_n - P(\gamma))\mathbf{y} \\ &= \delta_n^T X_{\gamma_0}^T(I_n - P(\gamma))X_{\gamma_0}\delta_n + 2\delta_n^T X_{\gamma_0}^T(I_n - P(\gamma))\mu + \mu^T(I_n - P(\gamma))\mu. \end{aligned} \tag{A13}$$

Hence, the centered process is as follows:

$$n^{2\rho-1}(SSR(\gamma)-SSR(\gamma_0))=n^{-2}\delta_0^T X_{\gamma_0}^T(I_n-P(\gamma))X_{\gamma_0}\delta_0+n^{\rho-3/2}2\delta_0^T X_{\gamma_0}^T(I_n-P(\gamma))\mu$$
$$+n^{2\rho-1}\Big[\mu^T(I_n-P(\gamma))\mu-\mu^T(I_n-P(\gamma_0))\mu\Big]=A_{n1}+A_{n2}+A_{n3}.$$

For all $\gamma \in (\gamma_0, \overline{\gamma}]$, closely following the same line of argument in the proof of Lemma A.5 of Chen (2014), and by applying Lemma A2, we can show that

$$(1).\ A_{n1}\Longrightarrow\Big(F(\gamma_0)-F(\gamma_0)F^{-1}(\gamma)F(\gamma_0)\Big)\delta_0^T\int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds,$$
$$(2).\ A_{n2}=O_p(n^{\rho-1/2}),$$
$$(3).\ A_{n3}=O_p(n^{2\rho-1}),$$

where we use the fact that $I(q_t \le \gamma_0)I(q_t \le \gamma) = I(q_t \le \gamma_0)$.

Therefore, we have

$$n^{2\rho-1}(SSR(\gamma)-SSR(\gamma_0))\Longrightarrow\Big(F(\gamma_0)-F(\gamma_0)F^{-1}(\gamma)F(\gamma_0)\Big)\delta_0^T\int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds.$$

Note that, for all $\gamma \in (\gamma_0, \overline{\gamma}]$, we have $F(\gamma) > F(\gamma_0)$. Thus, we have $n^{2\rho-1}S_{n1}\Big(SSR(\gamma) - SSR(\gamma_0)\Big) > 0$ uniformly for all $\gamma \in (\gamma_0, \overline{\gamma}]$. Similarly, we can show that $n^{2\rho-1}S_{n1}\Big(SSR(\gamma) - SSR(\gamma_0)\Big) < 0$ if $\gamma \in [\underline{\gamma}, \gamma_0)$, which completes the proof of this Theorem.

Appendix B.2. Proof of Theorem 2

First, closely following the proof of Lemma A.6 of Chen (2014), we can show $a_n(\widehat{\gamma} - \gamma_0) = \underset{v\in(-\infty,+\infty)}{argmin}\ Q_n(v) = O_p(1)$, where $Q_n(v) = SSR(\gamma_0 + \frac{v}{a_n}) - SSR(\gamma_0)$.

Next, note that

$$\mathbf{y}-X(\gamma)\widehat{\theta}=\mu+X(\beta-\widehat{\beta})+X_{\gamma_0}(\delta_n-\widehat{\delta}_n)-(X_\gamma-X_{\gamma_0})\widehat{\delta}_n.$$

Therefore, we can show

$$\begin{aligned}
&SSR(\gamma)-SSR(\gamma_0)\\
&=\widehat{\delta}_n^T(X_\gamma-X_{\gamma_0})^T(X_\gamma-X_{\gamma_0})\widehat{\delta}_n-2\big[\mu+X(\beta-\widehat{\beta})+X_{\gamma_0}(\delta_n-\widehat{\delta}_n)\big]^T\big[(X_\gamma-X_{\gamma_0})\widehat{\delta}_n\big]\\
&=\delta_n^T\sum_{t=1}^n x_t x_t^T|I(q_t\le\gamma)-I(q_t\le\gamma_0)|\delta_n\\
&-2\widehat{\delta}_n^T\sum_{t=1}^n \mu_t(x_{t,\gamma}-x_{t,\gamma_0})\\
&-2\widehat{\delta}_n^T\sum_{t=1}^n\big[(\beta-\widehat{\beta})^T x_t+(\delta_n-\widehat{\delta}_n)^T x_{t,\gamma_0}\big](x_{t,\gamma}-x_{t,\gamma_0})\\
&+\Big(\widehat{\delta}_n+\delta_n\Big)^T\sum_{t=1}^n x_t x_t^T|I(q_t\le\gamma)-I(q_t\le\gamma_0)|\Big(\widehat{\delta}_n-\delta_n\Big)\\
&=S_{n1}(\gamma)-2S_{n2}(\gamma)-2S_{n3}(\gamma)+S_{n4}(\gamma).
\end{aligned}$$

Now, we consider the limiting behavior of $S_{nj}(\gamma)$, for $j = 1,2,3,4$ with $\rho \in (0, \frac{1}{2})$. For any $v \in [\underline{v}, \overline{v}]$, a finite interval, by Lemmas A2–A4 and the Taylor expansion, we have

$$S_{n1}(\gamma_0 + \frac{v}{a_n}) =$$
$$n^{1-2\rho}\delta_0^T \left| F(\gamma_0 + \frac{v}{a_n}) - F(\gamma_0) \right| \int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds\delta_0 + o_p(1) \xrightarrow{p} |v| f_{\gamma_0}\delta_0^T \int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds\delta_0.$$

For S_{n2} and for $v > 0$, by Lemmas A2 and A3, we have

$$\begin{aligned}
&S_{n2}(\gamma_0 + \frac{v}{a_n}) \\
&= \delta_n^T \sum_{t=1}^{n} \mu_t x_t \left[I(q_t \le \gamma_0 + \frac{v}{a_n}) - I(q_t \le \gamma_0) \right] + (\widehat{\delta}_n - \delta_n)^T \sum_{t=1}^{n} \mu_t x_t \left[I(q_t \le \gamma_0 + \frac{v}{a_n}) - I(q_t \le \gamma_0) \right] \\
&\Longrightarrow n^{1/2-\rho}\delta_0^T \sigma_\mu \int_0^1 G_{\alpha,c}(s) d\left\{ W\left(s, F(\gamma_0 + \frac{v}{a_n})\right) - W(s, F(\gamma_0)) \right\} \left[1 + O_p(n^{-1/2}) \right].
\end{aligned}$$

Let $\mathbf{H}(\eta) = \int G_{\alpha,c}(s)dW(s,\eta)$. Note that $E[\mathbf{H}(\eta)] = 0$, $Var(\mathbf{H}(\eta)) = \eta \int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds$, and $E(\mathbf{H}(\eta_1)\mathbf{H}(\eta_2)) = (\eta_1 \wedge \eta_2) \int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds$. Thus, $\epsilon(v) = \sqrt{a_n}\delta_0^T \sigma_\mu [\mathbf{H}(\gamma_0 + \frac{v}{a_n}) - \mathbf{H}(\gamma_0)]$ is a Brownian motion with

$$Var(\epsilon(v)) = \sigma_\mu^2 f(\gamma_0) v \delta_0^T \int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds\delta_0 + o_p(1),$$

where we use the first-order Taylor expansion of $F(\gamma_0 + \frac{v}{a_n})$ around γ_0.

Next, we can show that, for all $v \in [\underline{v}, \overline{v}]$, $S_{n3}(v) = o_p(S_{n1}(v))$ and $S_{n4}(v) = o_p(S_{n1}(v))$. Thus, combining the above results together, we have

$$\begin{aligned}
&SSR(\gamma_0 + \frac{v}{a_n}) - SSR(\gamma_0) \\
&\Longrightarrow \begin{cases} v f(\gamma_0)\delta_0^T \int_0^1 G_{\alpha,c}G_{\alpha,c}^T(s)ds\delta_0 - 2\sigma_\mu \sqrt{f(\gamma_0) v \delta_0^T \int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds\delta_0} W_1(v), & \text{if } v > 0 \\ 0, & \text{if } v = 0 \\ |v| f(\gamma_0)\delta_0^T \int_0^1 G_{\alpha,c}G_{\alpha,c}^T(s)ds\delta_0 - 2\sigma_\mu \sqrt{f(\gamma_0)|v|\delta_0^T \int_0^1 G_{\alpha,c}(s)G_{\alpha,c}^T(s)ds\delta_0} W_2(-v), & \text{if } v < 0, \end{cases}
\end{aligned}$$

where $W_1(v)$ and $W_2(-v)$ are two independent standard Brownian motion process defined on $[0.\infty)$.

Following the proof of Theorem 1 of Hansen (2000) and Lemma A.7 in Chen (2014), by making the change of variables, we complete the proof of (i).

Now, we consider the limiting behavior with $\rho = 0$. Note that, for any positive finite v, $\sum_{t=1}^n I(\frac{v}{a_n}) \sim Binomial(n, p(v))$, where $p(v) \approx f(\gamma_0)\frac{v}{n}$ and $np(v)$ is finite. Hence, $\sum_{t=1}^n I(\frac{v}{a_n}) \Longrightarrow N_1(v)$ where $N(v)$ is a Poisson process with intensity $f(\gamma_0)$. Similarly, for any negative finite v, we have $\sum_{t=1}^n I(-\frac{v}{a_n}) \Longrightarrow N_2(|v|)$, which completes the proof of (ii).

Appendix B.3. Proof of Theorem 3

Note that, given $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, $\widehat{\delta} = [X_\gamma^T(I_n - P_x)X_\gamma]^{-1} X_\gamma^T(I_n - P_x)\mathbf{y}$. Hence, we have

$$\begin{aligned}
W_n(\gamma) &= \widehat{\delta}^T \left[X_\gamma^T(I_n - P_x)X_\gamma \right] \widehat{\delta}\widehat{\sigma}_\mu^{-2} \\
&= \mathbf{y}^T(I_n - P_x)X_\gamma \left[X_\gamma^T(I_n - P_x)X_\gamma \right]^{-1} X_\gamma^T(I_n - P_x)\mathbf{y},
\end{aligned}$$

where, by Lemma A2, we can show

$$(i).\ n^{-2}X_{\gamma}^{T}(I_n - P_x) \Longrightarrow \left[F(\gamma) - F(\gamma)^2\right]\int_0^1 G_{\alpha,c}(s)G_{\alpha,c}(s)^T ds,$$

$$(ii).\ n^{-1}X_{\gamma}^{T}(I_n - P_x)\mathbf{y} = n^{-1}\left[X_{\gamma}^{\mu} - X_{\gamma}^{T}P_x\mu + X_{\gamma}^{T}X_{\gamma_0}\delta_n - X_{\gamma}^{T}P_xX_{\gamma_0}\delta_n\right]$$

$$\Longrightarrow \int_0^1 G_{\alpha,c}(s)dW(s,F(\gamma)) - F(\gamma)\int_0^1 G_{\alpha,c}(s)dW(s) + n^{1-\rho}F(\gamma\bigwedge\gamma_0)\int_0^1 G_{\alpha,c}(s)G_{\alpha,c}(s)^T ds\delta_0$$

$$- n^{1-\rho}F(\gamma)F(\gamma_0)\int_0^1 G_{\alpha,c}(s)G_{\alpha,c}(s)^T ds\delta_0.$$

Thus, $W_n(\gamma) = O_p(n^{1-\rho})$. Under the null of $\delta_0 = 0$, we can show

$$W_n(\gamma) \Longrightarrow \left\{\int_0^1 G_{\alpha,c}(s)dW(s,F(\gamma)) - F(\gamma)\int_0^1 G_{\alpha,c}(s)dW(s)\right\}^T$$
$$\times\left\{\left[F(\gamma) - F(\gamma)^2\right]\int_0^1 G_{\alpha,c}(s)G_{\alpha,c}(s)^T ds\right\}^{-1}$$
$$\times\left\{\int_0^1 G_{\alpha,c}(s)dW(s,F(\gamma)) - F(\gamma)\int_0^1 G_{\alpha,c}(s)dW(s)\right\},$$

which completes the proof of this Theorem.

Notes

1. For more reference on how compound Poisson process can be approximated by two-sided Brownian motion, see Yu and Phillips (2018).
2. This is the limiting case when the finite q_t is involved in the local neighborhood of the true threshold level. If $\rho > 0$, we have infinite information, and if $\rho < 0$, $\sum_{t=1}^{n} I(\frac{v}{a_n}) \longrightarrow 0$, we have zero information provided in the local neighborhood of γ_0.

References

Bai, Jushan. 1997. Estimation of a Change Point in Multiple Regression Models. *Review of Economics And Statistics* 79: 551–63. [CrossRef]

Caner, Mehmet, and Bruce E. Hansen. 2001. Threshold Autoregression with a Unit Root. *Econometrica* 69: 1555–96. [CrossRef]

Caner, Mehmet, and Bruce E. Hansen. 2004. Instrumental Variable Estimation of a Threshold Model. *Econometric Theory* 20: 813–43. [CrossRef]

Chan, Kung-Sik. 1993. Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model. *The Annals of Statistics* 21: 520–33. [CrossRef]

Chen, Chaoyi, and Thanasis Stengos. 2021. The Democracy-Growth Nexus: A Smoothed GMM Linear Index Threshold Model Approach. *SSRN Electronic Journal*. [CrossRef]

Chen, Chaoyi, Thanasis Stengos, and Yiguo Sun. 2021. Endogeneity in Semiparametric Threshold Regression Model with Two Threshold Variables. *SSRN Electronic Journal*. [CrossRef]

Chen, Haiqiang. 2014. Robust Estimation and Inference for Threshold Models With Integrated Regressors. *Econometric Theory* 31: 778–810. [CrossRef]

Chen, Haiqiang, Terence Tai-Leung Chong, and Jushan Bai. 2012. Theory and Applications of TAR Model with Two Threshold Variables. *Econometric Reviews* 31: 142–70. [CrossRef]

Gonzalo, Jesús, and Jean-Yves Pitarakis. 2012. Regime-Specific Predictability in Predictive Regressions. *Journal of Business & Economic Statistics* 30: 229–41.

Hansen, Bruce E. 1996. Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis. *Econometrica* 64: 413–30. [CrossRef]

Hansen, Bruce E. 2000. Sample Splitting and Threshold Estimation. *Econometrica* 68: 575–603. [CrossRef]

Heckman, James J. 1979. Sample Selection Bias as a Specification Error. *Econometrica* 47: 153–61. [CrossRef]

Kejriwal, Mohitosh, and Pierre Perron. 2008. The limit distribution of the estimates in cointegrated regression models with multiple structural changes. *Journal of Econometrics* 146: 59–73. [CrossRef]

Kourtellos, Andros, Thanasis Stengos, and Chih Ming Tan. 2016. Structural Threshold Regression. *Econometric Theory* 32: 827–60. [CrossRef]

Kourtellos, Andros, Thanasis Stengos, and Yiguo Sun. 2021. Endogeneity in Semiparametric Threshold Regression. *Forthcoming to Econometric Theory*. [CrossRef]

Lieberman, Offer, and Peter C. B. Phillips. 2017. A multivariate stochastic unit root model with an application to derivative pricing. *Journal of Econometrics* 196: 99–110. [CrossRef]

Lieberman, Offer, and Peter C. B. Phillips. 2020. Hybrid stochastic local unit roots. *Journal of Econometrics* 215: 257–85. [CrossRef]

Miao, Ke, Liangjun Su, and Wendun Wang. 2020. Panel threshold regressions with Latent Group structures. *Journal of Econometrics* 214: 451–81. [CrossRef]

Phillips, Peter C. B., and Jun Yu. 2011. Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics* 2: 455–91. [CrossRef]

Phillips, Peter C. B., Yangru Wu, and Jun Yu. 2011. Explosive Behavior in the 1990s NASDAQ: When Did Exuberance Escalate Asset Values? *International Economic Review* 52: 201–26. [CrossRef]

Seo, Myung Hwan, and Oliver Linton. 2007. A smoothed least squares estimator for threshold regression models. *Journal of Econometrics* 141: 704–35. [CrossRef]

Seo, Myung Hwan, and Yongcheol Shin. 2016. Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics* 195: 169–86. [CrossRef]

Yang, Lixiong, Chingnun Lee, and I-Po Chen. 2021. Threshold model with a time-varying threshold based on Fourier approximation. *Journal of Time Series Analysis* 42: 406–30. [CrossRef]

Yu, Ping, and Peter C. B. Phillips. 2018. Threshold regression asymptotics: From the compound Poisson process to two-sided Brownian motion. *Economics Letters* 172: 123–26. [CrossRef]

Yu, Ping, and Xiaodong Fan. 2020. Threshold regression with a threshold boundary. *Journal Of Business & Economic Statistics* 39: 953–71.

Yu, Ping, Qin Liao, and Peter C. B. Phillips. 2021. *New Control Function Approaches in Threshold Regression with Endogeneity*. New York: Mimeo.