Econometric Theory, **31**, 2015, 778–810.
doi:10.1017/S0266466614000553

ROBUST ESTIMATION AND INFERENCE FOR THRESHOLD MODELS WITH INTEGRATED REGRESSORS

HAIQIANG CHEN
Xiamen University

This paper studies the robust estimation and inference of threshold models with integrated regressors. We derive the asymptotic distribution of the profiled least squares (LS) estimator under the diminishing threshold effect assumption that the size of the threshold effect converges to zero. Depending on how rapidly this sequence converges, the model may be identified or only weakly identified and asymptotic theorems are developed for both cases. As the convergence rate is unknown in practice, a model-selection procedure is applied to determine the model identification strength and to construct robust confidence intervals, which have the correct asymptotic size irrespective of the magnitude of the threshold effect. The model is then generalized to incorporate endogeneity and serial correlation in error terms, under which, we design a Cochrane–Orcutt feasible generalized least squares (FGLS) estimator which enjoys efficiency gains and robustness against different error specifications, including both $I(0)$ and $I(1)$ errors. Based on this FGLS estimator, we further develop a sup-Wald statistic to test for the existence of the threshold effect. Monte Carlo simulations show that our estimators and test statistics perform well.

1. INTRODUCTION

Threshold models, in which the model parameters switch when the value of a certain variable crosses a threshold, have received much attention over the past two decades. Given their ability to capture a very rich set of stylized facts of modern economics, such as multiple states, asymmetries, and cyclical effects, the use of threshold models has been advocated in many empirical applications. Examples include the analysis of asymmetries in persistence in US output growth (Potter, 1995), nonlinearities in unemployment rates (Hansen, 1997), and multiple states in cross-country growth regressions (Durlauf and Johnson, 1995; Hansen, 2000), among numerous others. The estimation and asymptotics for variants of

I would like to thank Professor Peter C.B. Phillips and the referees for helpful comments. Special thanks also go to Zongwu Cai, Jiti Gao, Wolfgang Härdle, Yongmiao Hong, and seminar participants at Cornell University and the 3rd XMU-Humboldt Workshop. The research was supported in part by the National Nature Science Foundation of China grants #11201390, #71201137, #71131008, and #71271179 as well as Deutsche Forschungsgemeinschaft (DFG) through the SFB649 "Economic Risk". Address correspondence to Haiqiang Chen, Xiamen University, Xiamen, Fujian 361005, China; e-mail: hc335@xmu.edu.cn.

 © Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

threshold models have been well established by Chan (1993), Hansen (1996, 2000), Gonzalo and Pitarakis (2002), and Seo and Linton (2007). More recently, Yu (2012) studies likelihood-based estimation and inference for a parametric discontinuous threshold regression model while Li and Ling (2012) study the least squares (LS) estimator of the multiple-regime threshold autoregressive (TAR) model.

However, most of the aforementioned studies maintain a restrictive assumption that the data should be either i.i.d. (independent and identically distributed) or stationary processes. Two exceptions are Caner and Hansen (2001) who study a threshold autoregressive model with a unit root and Gonzalo and Pitarakis (2006) who design a Wald statistic to test the threshold effect in a cointegrating regression. To the best of our knowledge, the asymptotic properties of threshold estimators, when all or some of the regressors are integrated processes, have not been established. In this paper, we investigate the robust estimation of threshold models with integrated regressors and establish its asymptotic properties allowing for different model identification strengths, endogenous regressors, and serially correlated error terms, including both I(0) and I(1) errors.

Specifically, we consider a threshold model of the following form:

$$y_t = \begin{cases} \alpha_1' x_t + u_t, & \text{if } q_t \le \gamma_0 \\ \alpha_2' x_t + u_t, & \text{if } q_t > \gamma_0 \end{cases}, \tag{1}$$

where x_t is a d_1-dimensional vector of integrated process of order 1 ($I(1)$) and y_t is the dependent variable.[1] The error term, which will be specified more precisely later, is denoted by u_t, and q_t is the threshold variable. The threshold value $\gamma_0 \in [\underline{\gamma}, \overline{\gamma}]$ is an unknown parameter pending estimation. If u_t is a stationary process, equation (1) can be regarded as a special case of nonlinear cointegration models (see Wang and Phillips, 2009; Gao, King, Lu, and Tjøstheim, 2009a, 2009b; Bierens and Martins, 2010; Choi and Saikkonen, 2010).[2]

Equation (1) can be rewritten as

$$y_t = \alpha' x_t + \delta' x_t I(q_t \le \gamma_0) + u_t, \tag{2}$$

where $\alpha = \alpha_2$ and $\delta = \alpha_1 - \alpha_2$. Here $I(q_t \le \gamma_0)$ is an indicator function taking the value one if $q_t \le \gamma_0$ and zero otherwise. The vector $(1, -\alpha)$ can be regarded as a benchmark long-run relationship between y_t and x_t and the term $\delta x_t I(q_t \le \gamma_0)$ captures the deviation from the linear equilibrium. In the literature, δ is often assumed to be a deterministic sequence converging to zero at some rate as the sample size n increases. This assumption is not only convenient for deriving the asymptotic distribution of threshold estimators as in Hansen (2000), but is also relevant in many empirical applications. For example, an important issue in the empirical finance literature is to investigate whether fundamental variables, such as the dividend-price ratio and the earning-price ratio, can predict asset returns. Linear prediction models have been extensively studied, but have failed to generate any unanimous conclusion (for more details, refer to Campbell and Yogo, 2006).[3] Recently, Gonzalo and Pitarakis (2012) document a regime-specific

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

predictability of S&P 500 index returns from the dividend ratio based on a threshold model. However, as is well known, stock returns commonly behave as martingale differences, while fundamental variables are highly persistent (integrated or nearly integrated processes). This imbalance on the order of integration implies that such predictive relationship should be very weak, which happens if and only if $\alpha = 0$ and δ is around zero in equation (2). For more discussions about the issue of balance and localizing coefficients to zero in predictive regressions, one could refer to Phillips and Lee (2013).

In this paper, we follow the literature by assuming the size of threshold effect converges to zero. Depending on how rapidly this sequence converges, the threshold value γ_0 may be identified or only weakly identified and we develop asymptotic theorems for both cases.[4] In the first case, we show that the profiled LS estimator is consistent and that its confidence intervals (CIs) can be constructed through inversion of certain standard test statistics; whereas in the second case, the profiled LS estimator is inconsistent and that its limiting distribution depends on some inestimable nuisance parameters. Because the standard method to construct CIs does not control the coverage probability, we take the supremum of quantiles for all possible values of nuisance parameters and then construct the least favorable CIs. These CIs have the correct asymptotic size under the weak identification case, but can be unnecessarily long when the model is identified. Motivated by Cheng (2008) and Shi and Phillips (2012), we then propose a model-selection procedure to choose the CIs. It can be shown that the CIs chosen by this method have approximately correct coverage probability irrespective of the magnitude of the threshold effect.

Furthermore, in a generalized model, we consider endogeneity and serial correlation, which are common in empirical studies with integrated regressors. Following Saikkonen (1991), we assume the error term to be an autoregressive (AR) process and use leads and lags of innovations as extra regressors to deal with endogeneity. We then design a Cochrane–Orcutt feasible generalized least squares (FGLS) estimator to estimate the model. It is well known that, in linear cointegration models, the FGLS estimator cannot improve the estimation efficiency, as Phillips and Park (1988) demonstrate by establishing their asymptotic equivalence. However, this equivalence does not hold when there exists a threshold effect and we analytically and numerically show that the FGLS estimation achieves more efficiency compared to the profiled LS estimation. The method to construct the robust CIs of the LS estimator can also be generalized for the FGLS estimator.

Another attraction of the FGLS estimator is its robustness with respect to different error specifications, including both I(0) and I(1) errors. This robustness allows us to design a sup-Wald statistic to test for the existence of threshold effects without any requirement on the stationarity of the regression error term and thus avoid the so-called joint hypothesis test problem in such models (see Balke and Fomby, 1997). The idea is similar to Perron and Yabu (2009), who consider testing for structural changes in the trend function of a time series using a quasi-FGLS procedure, without any requirement on whether the noise component is stationary

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

or integrated. In our case, we use the sup-Wald test statistic based on the FGLS estimator to test for the existence of a threshold effect and the critical values are generated from a bootstrap method. Monte Carlo simulations show that the FGLS estimator and the test statistic perform very well, even the error is an I(1) process.

The remainder of the paper is organized as follows. Section 2 presents the basic model. Section 3 considers the estimator and its asymptotics for the basic model. Section 4 generalizes the basic model to allow for serially correlated errors and model endogeneity. Monte Carlo simulations are discussed in Section 5. Section 6 concludes the paper. All proofs are in the Appendix.

Throughout the paper, $[ns]$ denotes the integer part of ns. The symbol '$\xrightarrow{p}$' represents convergence in probability, '$\Longrightarrow$' represents weak convergence, and $\xrightarrow{a.s.}$ represents almost sure convergence. All limits are defined as the sample size $n \to \infty$ unless otherwise stated. For a vector x, $||\cdot||$ signifies the euclidean norm, i.e., $||x|| = \sum x_i^2$. For a matrix A, $|A| = (tr(A'A))^{1/2}$ denotes the euclidean norm and A' denotes its transpose. $\int$ denotes integration from 0 to 1.

2. THE BASIC MODEL

The basic model we consider in the paper is given by

$$y_t = \begin{cases} \alpha_1' x_t + u_t, \text{ if } q_t \le \gamma_0 \\ \alpha_2' x_t + u_t, \text{ if } q_t > \gamma_0 \end{cases},$$

where x_t is a d_1-dimensional vector of unit root processes whose generating mechanism is given by

$$x_t = x_{t-1} + v_t, \ t = 1, 2, \ldots n,$$

and we set $x_0 = 0$ for convenience, although $x_0 = o_{a.s.}(\sqrt{n})$ is sufficient for the asymptotic results.

For the component time series u_t and v_t, we define the following partial sums as

$$U_n(s) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} u_t, \text{ and } V_n(s) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} v_t. \tag{3}$$

Before going further, we make the following assumptions.

Assumption 2.1. $\sup_{s\in[0,1]} ||(U_n(s), V_n(s)) - (B_u(s), \mathbf{B}_v(s))|| \xrightarrow{a.s.} 0$ as $n \to \infty$, where $(B_u(s), \mathbf{B}_v(s))$ is a vector of Brownian motions with a positive definite long-run covariance matrix.

Assumption 2.2. The error term u_t satisfies the following three conditions:

(a) $\mathbb{E}(u_t) = 0$, $\mathbb{E}|u_t|^4 < \infty$;

(b) For each n, there exists a filtration $\digamma_{n,t}$, $t = 1, \ldots, n$, such that $\{(u_t, \digamma_{n,t}) : t \ge 1\}$ is a stationary and ergodic martingale difference sequence with $\mathbb{E}(u_t^2 | \digamma_{n,t-1}) = \sigma_u^2 < \infty$ almost surely for all $t = 1, \ldots, n$;

(c) The term (q_t, x_t) is adapted to the filtration $\digamma_{n,t-1}$, $t = 1, \ldots, n$.

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

Assumption 2.3.

(a) The threshold parameter $\gamma_0 \in [\underline{\gamma}, \overline{\gamma}]$;

(b) The threshold variable q_t is strictly stationary, ergodic, and α-mixing with mixing coefficients α_m satisfying $\sum_{m=1}^{\infty} \alpha_m^{\frac{1}{2}-\frac{1}{r}} < \infty$ for $r > 2$;

(c) q_t has a continuous distribution $F(\cdot)$ and $f(\cdot)$ is the corresponding density function satisfying $0 < f(\gamma) \leq \overline{f} < \infty$ for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$.

Assumption 2.4. The d_1 dimensional $I(1)$ vector x_t is not cointegrated.

Assumption (2.1) considers a strong approximation for $(U_n(s), V_n(s))$, which is needed to derive the convergence rate of the threshold estimator. This assumption is stronger than the weak convergence result established by the multivariate invariance principle, but it is quite common in the literature (see Park and Phillips, 2001; Wang and Phillips, 2009; Shi and Phillips, 2012, among others). Sufficient conditions to derive the strong approximation are also well developed in the literature. For example, Park and Hahn (1999) establish conditions of strong approximations for general linear processes. Shao and Lu (1987) and Cai, Li, and Park (2009) provide conditions of strong approximations for an α-mixing process.

Assumption (2.2) supposes $\{(u_t, \mathcal{F}_{n,t}) : t \geq 1\}$ to be a martingale difference sequence and (q_t, x_t) is adapted to the filtration $\mathcal{F}_{n,t-1}$, $t = 1, \ldots, n$. Under this assumption, q_t and x_t become predetermined given the filtration $\mathcal{F}_{n,t-1}$. One natural example of $\mathcal{F}_t$ is the σ-field generated by the information set $\{(u_i, q_{i+1}, x_{i+1}) : 1 \leq i \leq t\}$. This assumption might be restrictive in linear cointegration models, but it is common in fully specified cointegrating regression models. It allows for arguments based on the martingale central limit theory to establish a weak convergence result for the empirical process $\sum_{t=1}^{[ns]} I_t(q_t \leq \gamma) u_t$ as in Caner and Hansen (2001).[5] Assumption (2.2) does not allow for serially correlated errors and model endogeneity. However, in Section 4, we relax this assumption by adopting the so-called "leads and lags" approach or dynamic OLS estimator proposed by Saikkonen (1991).

Assumption (2.3) is very conventional in the literature of threshold models. The threshold variable q_t is assumed to be strictly stationary and strong mixing for asymptotic purposes. The assumption excludes the case where q_t is a unit root process, under which one may need to use another technique such as the triangular array asymptotics proposed by Andrews and McDermott (1995). Assumption (2.4) is standard in the literature of regressions with integrated processes.

Following Caner and Hansen (2001), we define $W(s, \mu)$ as a two-parameter Brownian motion on $(s, \mu) \in [0, 1]^2$, which is a zero-mean Gaussian process with covariance given by

$$\mathbb{E}(W(s_1, \mu_1) W(s_2, \mu_2)) = (s_1 \wedge s_2)(\mu_1 \wedge \mu_2).$$

The two-parameter Brownian motion is a special tool to derive the limiting distribution in threshold models with integrated processes. Note that when $\mu = 1$, we have $W(s, 1) \equiv W(s)$, which is a Brownian motion on $s \in [0, 1]$.

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

Moreover, we can define the following stochastic integration with respect to $W(s,\mu)$ on the first argument (s) while holding the second argument (μ) constant as

$$\mathbf{J}_1(\mu) = \int \mathbf{B}_v(s)dW(s,\mu) = \operatorname*{plim}_{n\to\infty} \sum_{t=1}^{n} \left(\mathbf{B}_v\left(\frac{t-1}{n}\right)\left(W\left(\frac{t}{n},\mu\right) - W\left(\frac{t-1}{n},\mu\right)\right)\right), \tag{4}$$

where plim denotes convergence in probability and $\mathbf{J}_1(\mu)$ is a Gaussian process with an almost surely continuous sample path and the covariance kernel

$$\mathbb{E}(\mathbf{J}_1(\mu_1)\mathbf{J}_1(\mu_2)) = (\mu_1 \wedge \mu_2)\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds.$$

3. MAIN RESULTS

3.1. Profiled LS Estimation and Asymptotics

For ease of manipulation, we rewrite equation (2) in a more compacted form:

$$y_t = \theta' A_t(\gamma_0) + u_t, \tag{5}$$

where $A_t(\gamma_0) = (x_t', x_t'(\gamma_0))'$, $x_t(\gamma_0) = x_t I(q_t \le \gamma_0)$, and $\theta = (\alpha', \delta')'$.

For any fixed $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, the following model is estimated:

$$y_t = \widehat{\theta}(\gamma)' A_t(\gamma) + \widehat{u}_t(\gamma),$$

where $\widehat{\theta}(\gamma)$ is given by

$$\widehat{\theta}(\gamma) = \left[\sum_{t=1}^{n} A_t(\gamma)A_t(\gamma)'\right]^{-1}\left[\sum_{t=1}^{n} A_t(\gamma)y_t\right].$$

The sum of squared residuals is defined as

$$SSR_n(\gamma) = \sum_{t=1}^{n} \widehat{u}_t(\gamma)^2 = \sum_{t=1}^{n} \left(y_t - \widehat{\theta}(\gamma)' A_t(\gamma)\right)^2, \tag{6}$$

and we define the profiled LS estimator of γ_0 as the value that minimizes $SSR_n(\gamma)$, i.e.,

$$\widehat{\gamma}_n = \arg \min_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} SSR_n(\gamma).$$

The term $SSR_n(\gamma)$ is not differentiable due to the presence of the indicator functions; thus, we cannot write $\widehat{\gamma}_n$ in closed form from first-order conditions. Following Hansen (2000), we adopt a grid-searching method. Specifically, we

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

divide $[\underline{\gamma}, \overline{\gamma}]$ into N quantiles and let $\Gamma_N = \{q_1, q_2, \ldots q_N\}$. The estimator $\widehat{\gamma}_N = \arg\min_{\gamma \in \Gamma_N} SSR_n(\gamma)$ is a good approximation to $\widehat{\gamma}_n$ when N is large enough. The other parameters are then estimated by using the point estimate $\widehat{\gamma}_n$ via $\widehat{\theta} = \widehat{\theta}(\widehat{\gamma}_n)$ and $\widehat{\sigma}_u^2 = \frac{1}{n}\sum_{t=1}^{n} \widehat{u}_t(\widehat{\gamma}_n)^2$.

In what follows, we set $\delta = \delta_n = n^{-1/2-\tau}\delta_0$, where $-1/2 < \tau \leq 1/2$ and $\delta_0 \in R$ is a fixed parameter. Under this assumption, the size of the threshold effect converges to zero with the rate $n^{-1/2-\tau}$, where the value of τ determines the identification strength of the threshold value γ_0. We exclude the case with $\tau > 1/2$ since the nonlinear term is negligible asymptotically. In addition, when $\tau < -1/2$, the nonlinear term is explosive and is also excluded. We also exclude the case of $\tau = -1/2$ to focus on the limiting behavior of $\widehat{\gamma}_n$ when $\delta_n \to 0$.

The following theorem states the limiting results for $\widehat{\gamma}_n$.

THEOREM 3.1. *Under Assumptions* (2.1) − (2.4) *and* $\delta_n = n^{-1/2-\tau}\delta_0$, *the following limiting results hold:*

Case 1: if $-1/2 < \tau < 1/2$, *then*

$$n^{1-2\tau}|\widehat{\gamma}_n - \gamma_0| = O_p(1).$$

Furthermore,

$$n^{1-2\tau}\lambda(\widehat{\gamma} - \gamma_0) = r^* \Rightarrow \arg\max_{r \in (-\infty, \infty)} \left(\Lambda(r) - \frac{1}{2}|r|\right),$$

where

$$\lambda = \frac{\delta_0' \left(\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds\right)\delta_0 f_0}{\sigma_u^2},$$

with $f_0 = f(\gamma_0)$ *and* $\Lambda(r)$ *is a two-sided Brownian motion on the real line defined as:*

$$\Lambda(r) = \begin{cases} \Lambda_1(r) & \text{if } r > 0 \\ 0 & \text{if } r = 0 \\ \Lambda_2(-r) & \text{if } r < 0 \end{cases}. \tag{7}$$

The processes $\Lambda_1(r)$ *and* $\Lambda_2(r)$ *are two independent standard Brownian motions on* $[0, \infty)$.

Case 2: if $\tau = 1/2$, *then* $\widehat{\gamma}_n \Rightarrow \gamma(\gamma_0, \delta_0)$ *and* $\gamma(\gamma_0, \delta_0)$ *is a random variable that maximizes* $Q(\gamma, \gamma_0, \delta_0)$, *where*

$$Q(\gamma, \gamma_0, \delta_0) = \frac{1}{F(\gamma)(1 - F(\gamma))}\Gamma_1'(\gamma)\left(\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds\right)^{-1}\Gamma_1(\gamma), \tag{8}$$

with

$$\Gamma_1(\gamma) = \Gamma(\gamma) + [F(\gamma \wedge \gamma_0) - F(\gamma)F(\gamma_0)]\left(\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds\right)\delta_0, \tag{9}$$

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

and

$$\Gamma(\gamma) = \sigma_u \int \mathbf{B}_v(s) d\left[W(s, F(\gamma)) - F(\gamma) W(s)\right]. \tag{10}$$

Theorem 3.1 shows that the convergence results for $\widehat{\gamma}_n$ depend critically on the value of τ, which characterizes the convergence speed of δ_n.[6]

If $-1/2 < \tau < 1/2$, the threshold effect is identifiable and $\widehat{\gamma}_n$ is a consistent estimator. The rate of convergence is $n^{1-2\tau}$, which is decreasing in τ. The limiting distribution of $\widehat{\gamma}_n$ has the same form as that found for the stationary threshold model in Hansen (2000), but the scale factor λ is different. Note that f_0 is the density of q_t at γ_0. Intuitively, a larger f_0 implies more data points around γ_0, making $\widehat{\gamma}$ more accurate.

To generate the confidence interval of γ, we invert the following likelihood ratio statistic for the null hypothesis $\gamma = \gamma_0$, given by

$$LR_n(\gamma_0) = n \frac{SSR_n(\gamma_0) - SSR_n\left(\widehat{\gamma}_n\right)}{SSR_n\left(\widehat{\gamma}_n\right)}, \tag{11}$$

where $\widehat{\gamma}_n$ is the profiled LS estimator. Using a similar argument of Theorem 2 in Hansen (2000), we can show that $LR_n(\gamma_0) \Rightarrow \sup_{r\in(-\infty,\infty)}(2\Lambda(r) - |r|)$. Denote $q^I_{\gamma,1-a}$ as the $1-a$ quantile of the random variable $\sup_{r\in(-\infty,\infty)}(2\Lambda(r) - |r|)$ and $q^I_{\gamma,1-a}$ can be calculated by the formula $q^I_{\gamma,1-a} = -2\ln(1-\sqrt{1-a})$. Thus, the a-level confidence interval of γ in the case of identification can be expressed as

$$CI^I_{\gamma,n}(\alpha) = \left\{\gamma : LR_n(\gamma) \le q^I_{\gamma,1-a}\right\}. \tag{12}$$

If $\tau = 1/2$, the threshold effect is only weakly identified. The profiled LS estimator $\widehat{\gamma}_n$ converges to a random variable $\gamma(\gamma_0, \delta_0)$, reflecting the lack of information. Since γ_0 and δ_0 are not estimable, any statistical inference based on them is impossible. Following Cheng (2008) and Shi and Phillips (2012), we define the least favorable CI which is large enough for all possible γ_0 and δ_0. Denote $q^W_{\gamma,1-a}(\gamma_0, \delta_0)$ as the $1-a$ quantile of $|\gamma(\gamma_0, \delta_0) - \gamma_0|$ for each $\gamma_0 \in [\underline{\gamma}, \overline{\gamma}]$ and $\delta_0 \in R$. The a-level CI, given γ_0 and δ_0, is then defined as

$$CI^W_{\gamma,n}(1-a, \gamma_0, \delta_0) = \left\{\gamma : |\widehat{\gamma}_n - \gamma| \le q^W_{\gamma,1-a}(\gamma_0, \delta_0)\right\}. \tag{13}$$

Since γ_0 and δ_0 are two unknown variables, we define a robust quantile by taking the supremum of all possible γ_0 and δ_0. Let

$$q^W_{\gamma,1-a} = \sup_{\gamma_0\in[\underline{\gamma},\overline{\gamma}]} \sup_{\delta_0\in R} q^W_{\gamma,1-a}(\gamma_0, \delta_0). \tag{14}$$

The a-level least favorable CI in the case of weak identification is then defined as

$$CI^W_{\gamma,n}(a) = \{\gamma : |\widehat{\gamma}_n - \gamma| \le q^W_{\gamma,1-a}\}. \tag{15}$$

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

3.2. Robust Confidence Interval

In empirical studies, τ is unknown, raising the question of which CI should be used. In this subsection, based on a model selection procedure, we construct a robust CI which has approximately correct coverage probability irrespective of the value of τ.

For a fixed $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, let $X(\gamma) = (x_1(\gamma), x_2(\gamma), \ldots, x_n(\gamma))'$ and $X = (x_1, x_2, \ldots, x_n)'$. The Wald test statistic for testing $H_0 : \delta_n = 0$ can be defined as

$$T_n(\gamma) = \widehat{\delta}_n(\gamma)'(X'(\gamma)(I - P_n)X(\gamma))\widehat{\delta}_n(\gamma)/\widehat{\sigma}_u^2, \tag{16}$$

where P_n is the projection matrix of X, given by $P_n = X(X'X)^{-1}X'$. We define the sup-Wald test statistic as

$$T_n = \sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} T_n(\gamma).$$

The following theorem explores the limiting behaviors of the sup-Wald statistic under different model identification strengths.

THEOREM 3.2. *Under Assumptions* (2.1) − (2.4) *and* $\delta_n = n^{-1/2-\tau}\delta_0$, *where* δ_0 *is a nonzero constant, the following limiting results hold:*

(i) *if* $-1/2 < \tau < 1/2$, *then* $T_n \xrightarrow{p} \infty$.

(ii) *if* $\tau = 1/2$, *then*

$$T_n \Rightarrow T_1 = \sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} T_1(\gamma) = \sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \frac{1}{\sigma_u^2(F(\gamma)(1 - F(\gamma))} \Gamma_1(\gamma)' \left(\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds \right)^{-1} \Gamma_1(\gamma), \tag{17}$$

where $\Gamma_1(\gamma)$ *is defined by equation (9).*

Theorem 3.2 shows that $T_n \xrightarrow{p} \infty$ if $-1/2 < \tau < 1/2$ and $T_n < \infty$ if $\tau = 1/2$. This result enables us to develop the following model selection procedure. We define $\{\kappa_n : n \geq 1\}$ as a sequence of constants that diverge to infinity as $n \to \infty$. The parameter κ_n is referred to as a tuning parameter and we require the sequence κ_n to diverge to infinity at a rate slower than n^ν for any $\nu > 0$, i.e., $n^{-\nu}\kappa_n^{1/2} \to 0$. Suitable choices of κ_n include $d(\ln(n))^2$ where d is a positive constant, in accordance with BIC. The model selection procedure is designed to choose the model with the identified threshold effect if $T_n > \kappa_n$ and to choose the model with the weakly identified threshold effect otherwise. We use the CI chosen through this procedure as the final CI.

Specifically, for each confidence level a, define

$$CI_{\gamma,n}(a) = \left\{ \begin{array}{l} CI^I_{\gamma,n}(a), \text{ if } T_n > \kappa_n \\ CI^W_{\gamma,n}(a), \text{ if } T_n \leq \kappa_n \end{array} \right\}.$$

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

We focus on the smallest finite sample coverage probability of $CI_{\gamma,n}(a)$ over the whole parameter space, which can be approximated by the following asymptotic size

$$AsySZ_{\gamma}(a) = \lim_{n\to\infty} \inf_{\delta_0 \in R} \inf_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \Pr(\gamma \in CI_{\gamma,n}(a)). \tag{18}$$

The following theorem shows that the robust CI has the correct asymptotic size.

THEOREM 3.3. *Under Assumptions* (2.1) − (2.4)*, for any* $a \in (0,1)$*, we have* $AsySZ_{\gamma}(a) = a$.

4. THE GENERALIZED MODEL

In many economic applications of cointegration, error terms are serially correlated and correlated with regressors. Under these conditions, it is well known that the ordinary least squares (OLS) estimator contains a second-order bias in linear cointegration models. Several efficient estimators have been proposed, such as the fully modified (FM) OLS estimator of Phillips and Hansen (1990), the canonical cointegrating regression (CCR) estimator of Park (1992), and the dynamic ordinary least squares (DOLS) estimator proposed by Saikkonen (1991). In the following, we generalize the basic model in Section 2 to allow for serial correlation and model endogeneity.

More formally, we introduce the following assumptions.

Assumption 4.1. In equation (1), u_t can be decomposed as

$$u_t = \sum_{i=1}^{d_1} \sum_{j=-K}^{K} \beta_{ij} v_{i,t-j} + \eta_t = \beta' z_t + \eta_t, \tag{19}$$

$$\eta_t = \rho \eta_{t-1} + \varepsilon_t, \text{ with } \rho \in (-1, 1],$$

where $z_t = (v_{1,t-K}, \ldots, v_{d_1,t+K})$ and $\beta = (\beta_{1,-K}, \ldots, \beta_{d_1,K})$. The error term ε_t satisfies the following conditions:

(a) $\mathbb{E}(\varepsilon_t) = 0$, $\mathbb{E}|\varepsilon_t|^4 < \infty$;

(b) For each n, there exists a filtration $F_{n,t}$, $t = 1, \ldots, n$, such that $\{(\varepsilon_t, F_{n,t}) : t \geq 1\}$ is a stationary and ergodic martingale difference sequence with $\mathbb{E}(\varepsilon_t^2 | F_{n,t-1}) = \sigma^2 < \infty$ almost surely for all $t = 1, \ldots, n$;

(c) The term (q_t, x_t) is adapted to the filtration $F_{n,t-1}$, $t = 1, \ldots, n$.

Assumption 4.2. Let $\xi_n(s) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \varepsilon_t$ and $\sup_{s\in[0,1]} ||(\xi_n(s), V_n(s)) - (B_{\varepsilon}(s), \mathbf{B}_{v}(s))|| \xrightarrow{a.s.} 0$ as $n \to \infty$, where $(B_{\varepsilon}(s), \mathbf{B}_{v}(s))$ is a vector of Brownian motions with a positive definite long-run covariance matrix.

Under Assumption (4.1), the model endogeneity can be fully captured by $\beta' z_t$, where z_t is a vector of leads and lags of Δx_t. The parameter K can diverge to

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

infinity as the sample size increases. The idea of using leads and lags to deal with endogeneity in cointegration models was proposed by Saikkonen (1991). We assume β remains constant to focus on the threshold effect occurring in the cointegrating relationship. The extension allowing β to be regime-sensitive would be interesting and is left to future study. The term η_t is assumed to be *AR(1)* and ρ controls the stationarity of η_t. If $\rho = 1$, η_t is a unit root process and the model describes a structural spurious relationship,[7] while if $\rho < 1$, η_t is a stationary process and the model is a cointegrating relationship.

To estimate a regression with serial correlation, the Cochrane–Orcutt FGLS procedure is usually adopted. In linear cointegration models, as shown in Phillips and Park (1988), the FGLS estimator and the OLS estimator are equivalent in asymptotics. The Cochrane–Orcutt FGLS estimator also works for spurious regressions, as Phillips and Hodgson (1994) demonstrate by proving the asymptotic equivalence of the FGLS estimator to the OLS estimator when the error is an I(1) process. However, in the presence of threshold effects, there is no asymptotic equivalence between FGLS and OLS estimators. The following simple sketch may help to illustrate this difference.

For a linear cointegrating regression after transformation, $y_t - \rho y_{t-1} = \alpha'(x_t - \rho x_{t-1}) + (\eta_t - \rho \eta_{t-1})$, and

$$
\begin{aligned}
n(\widehat{\alpha}_{FGLS} - \alpha) &= \left(\sum_{t=1}^{n} (x_t - \rho x_{t-1})(x_t - \rho x_{t-1})' \right)^{-1} \\
&\quad \times \left(\sum_{t=1}^{n} (x_t - \rho x_{t-1})(\eta_t - \rho \eta_{t-1}) \right) \\
&\Rightarrow \left(\int \mathbf{B}_v(s) \mathbf{B}_v'(s) ds \right)^{-1} \int \mathbf{B}_v(s) dB_\eta(s),
\end{aligned}
$$

which is the same as the limiting result of the OLS estimator. However, for a cointegrating regression with a threshold effect after transformation, $y_t - \rho y_{t-1} = \alpha'(x_t - \rho x_{t-1}) + \delta_n'(x_t(\gamma) - \rho x_{t-1}(\gamma)) + (\eta_t - \rho \eta_{t-1}) = \alpha' \widetilde{x}_t + \delta_n' \widetilde{x}_t(\gamma) + (\eta_t - \rho \eta_{t-1})$, and ρ cannot be canceled in the limiting result because the denominator $\sum_{t=1}^{n} \widetilde{x}_t(\gamma) \widetilde{x}_t(\gamma)' \Rightarrow \big((1+\rho^2)F(\gamma) - 2\rho F_1(\gamma, \gamma)\big) \int \mathbf{B}_v(s) \mathbf{B}_v'(s) ds$, which depends on ρ, the marginal distribution function $F(\gamma) \equiv \Pr(q_t \le \gamma)$, and the joint distribution $F_1(\gamma, \gamma) \equiv \Pr(q_t \le \gamma, q_{t-1} \le \gamma)$, whereas the nominator $\sum_{t=1}^{n} \widetilde{x}_t(\gamma)(\eta_t - \rho \eta_{t-1}) \Rightarrow (1-\rho^2) \int \mathbf{B}_v(s) dB_\eta(s)$.

4.1. FGLS Estimator

To obtain a feasible GLS estimator, we first estimate the threshold value γ_0 through the profiled LS estimator without considering serial correlation and endogeneity. Then we estimate $\widehat{\rho}$ from the estimated error terms. Finally, we construct the Cochrane–Orcutt FGLS estimator based on $\widehat{\rho}$.

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

Specifically, we estimate γ_0 using

$$\widehat{\gamma}_n = \arg \min_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} SSR_n(\gamma),$$

where $SSR_n(\gamma)$ is the sum of squared residuals for the regression

$$y_t = \widehat{\alpha}' x_t + \widehat{\delta}_n' x_t I(q_t \le \gamma) + \widehat{\beta}' z_t + \widehat{\eta}_t.$$

By estimating the AR model

$$\widehat{\eta}_t(\widehat{\gamma}_n) = \widehat{\rho}\widehat{\eta}_{t-1}(\widehat{\gamma}_n) + \widehat{\varepsilon}_t,$$

the OLS estimator $\widehat{\rho}$ is estimated, where $\widehat{\eta}_t(\widehat{\gamma}_n) = y_t - \widehat{\alpha}'(\widehat{\gamma}_n)x_t - \widehat{\delta}_n'(\widehat{\gamma}_n)$ $x_t I(q_t \le \widehat{\gamma}_n) - \widehat{\beta}'(\widehat{\gamma}_n)z_t$.

The following theorem establishes the consistency and the convergence rate of $\widehat{\rho}$.

THEOREM 4.1. *Under Assumptions* (4.1), (4.2), (2.3), *and* (2.4), $\widehat{\rho} \to \rho$ *as* $n \to \infty$. *Furthermore,* $|\widehat{\rho} - \rho| = O_p(n^{-1/2})$ *if* $\rho < 1$, *and* $|\widehat{\rho} - 1| = O_p(n^{-1})$ *if* $\rho = 1$.

Theorem 4.1 shows that $\widehat{\rho}$ is consistent even when $\rho = 1$, i.e., the regression is a spurious relationship. The convergence rates are different due to the different convergence speeds of integrated and stationary processes. Moreover, we find that the limiting behavior of $\widehat{\rho}$ is not affected by the identification strength of the threshold effect. The intuition is as follows. If $\tau < 1/2$, γ_0 can be consistently estimated by $\widehat{\gamma}_n$. In each regime, if $\rho < 1$, the coefficients can be consistently estimated as well and thus it is obvious that $\widehat{\rho} \xrightarrow{p} \rho$. If $\rho = 1$, the coefficient estimators are not consistent; however, this inconsistency implies that the residual term $\widehat{\eta}_t(\widehat{\gamma}_n)$ is a unit root and that $\widehat{\rho} \xrightarrow{p} \rho = 1$. If $\tau = 1/2$, $\widehat{\gamma}_n$ is not consistent as shown in Theorem 3.1; however, the nonlinear term $\widehat{\delta}_n' x_t I(q_t \le \gamma)$ decays to zero so fast that it has no impact on the estimation of ρ asymptotically. Following Choi, Hu, and Ogaki (2008), the consistency of $\widehat{\rho}$ can also be obtained.

Based on the consistent estimator $\widehat{\rho}$, we construct the following Cochrane–Orcutt FGLS estimators. We first define $\widetilde{y}_t = y_t - \widehat{\rho} y_{t-1}$, and $\widetilde{z}_t$, $\widetilde{x}_t$, $\widetilde{\eta}_t$ in the same way. For each $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, define

$$\widetilde{x}_t(\gamma) = x_t I(q_t \le \gamma) - \widehat{\rho} x_{t-1} I(q_{t-1} \le \gamma).$$

Let $\widetilde{A}_{1t}(\gamma) = (\widetilde{x}_t', \widetilde{x}_t(\gamma)', \widetilde{z}_t')'$ and stack $\widetilde{x}_t$, $\widetilde{y}_t$, $\widetilde{z}_t$, $\widetilde{x}_t(\gamma)$, and $\widetilde{A}_{1t}(\gamma)$ to get the matrices: $\widetilde{X}$, $\widetilde{Y}$, $\widetilde{Z}$, $\widetilde{X}(\gamma)$, and $\widetilde{A}_1(\gamma)$.

After the transformation,

$$\widetilde{y}_t = \alpha' \widetilde{x}_t + \delta_n' \widetilde{x}_t(\gamma) + \beta' \widetilde{z}_t + \widetilde{\eta}_t = \widetilde{\theta}' \widetilde{A}_{1t}(\gamma) + \widetilde{\eta}_t. \tag{20}$$

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

Using equation (20), for each γ, we can define

$$\widehat{\widetilde{\theta}}(\gamma) = \left[\sum_{t=2}^{n} \widetilde{A}_{1t}(\gamma)\widetilde{A}_{1t}(\gamma)'\right]^{-1}\left[\sum_{t=2}^{n} \widetilde{A}_{1t}(\gamma)\widetilde{y}_t\right], \tag{21}$$

and the FGLS threshold estimator can be defined as

$$\widetilde{\gamma}_n = \arg\min_{\gamma\in[\underline{\gamma},\overline{\gamma}]} (\widetilde{SSR}_n(\gamma)),$$

where $\widetilde{SSR}_n(\gamma)$ is the sum of squared residuals given by

$$\widetilde{SSR}_n(\gamma) = \sum_{t=2}^{n}\left(\widetilde{y}_t(\gamma) - \widehat{\widetilde{\theta}}(\gamma)'\widetilde{A}_{1t}(\gamma)\right)^2. \tag{22}$$

The following theorem establishes the limiting results of the FGLS estimator.

THEOREM 4.2. *Under Assumptions* (4.1), (4.2), (2.3), *and* (2.4), *the following results hold:*

Case 1: if $-1/2 < \tau < 1/2$, *then* $n^{1-2\tau}|\widetilde{\gamma}_n - \gamma_0| = O_p(1)$. *Furthermore,*

$$n^{1-2\tau}\widetilde{\lambda}(\widetilde{\gamma}_n - \gamma_0) = r^* \Rightarrow \arg\max_{r\in(-\infty,\infty)} (\Lambda(r) - \frac{1}{2}|r|),$$

where

$$\widetilde{\lambda} = \frac{\left(1+\rho^2\right)\left(\delta_0'\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds\delta_0\right) f_0}{\sigma^2}, \tag{23}$$

and $\Lambda(r)$ *is defined by equation (7).*

Case 2: if $\tau = 1/2$, *then* $\widetilde{\gamma}_n \Rightarrow \widetilde{\gamma}(\gamma_0, \delta_0)$ *and* $\widetilde{\gamma}(\gamma_0, \delta_0)$ *is a random variable that maximizes* $\widetilde{Q}(\gamma, \gamma_0, \delta_0)$ *where*

$$\widetilde{Q}(\gamma, \gamma_0, \delta_0) = \widetilde{\Gamma}_1'(\gamma)\left(\widetilde{G}_{22}(\gamma) - \begin{pmatrix}\widetilde{G}_{21}(\gamma)\\ \widetilde{G}_{23}(\gamma)\end{pmatrix}'\begin{pmatrix}\widetilde{G}_{11}, \widetilde{G}_{13}\\ \widetilde{G}_{31}, \widetilde{G}_{33}\end{pmatrix}^{-1}\begin{pmatrix}\widetilde{G}_{12}(\gamma)\\ \widetilde{G}_{32}(\gamma)\end{pmatrix}\right)^{-1}\widetilde{\Gamma}_1(\gamma),$$

with

$$\widetilde{\Gamma}_1(\gamma) = \widetilde{\Gamma}(\gamma) + \left(\widetilde{G}_{22}(\gamma) - \begin{pmatrix}\widetilde{G}_{21}(\gamma)\\ \widetilde{G}_{23}(\gamma)\end{pmatrix}'\begin{pmatrix}\widetilde{G}_{11}, \widetilde{G}_{13}\\ \widetilde{G}_{31}, \widetilde{G}_{33}\end{pmatrix}^{-1}\begin{pmatrix}\widetilde{G}_{12}(\gamma)\\ \widetilde{G}_{32}(\gamma)\end{pmatrix}\right)\delta_0,$$

and

$$\widetilde{\Gamma}(\gamma) = \widetilde{\phi}_2(\gamma) - \begin{pmatrix}\widetilde{G}_{21}(\gamma)\\ \widetilde{G}_{23}(\gamma)\end{pmatrix}'\begin{pmatrix}\widetilde{G}_{11}, \widetilde{G}_{13}\\ \widetilde{G}_{31}, \widetilde{G}_{33}\end{pmatrix}^{-1}\begin{pmatrix}\widetilde{\phi}_1(\gamma)\\ \widetilde{\phi}_3(\gamma)\end{pmatrix}, \tag{24}$$

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

where the random matrix $\widetilde{G}(\gamma) = \begin{pmatrix} \widetilde{G}_{11}(\gamma), \widetilde{G}_{12}(\gamma), \widetilde{G}_{13}(\gamma) \\ \widetilde{G}_{21}(\gamma), \widetilde{G}_{22}(\gamma), \widetilde{G}_{23}(\gamma) \\ \widetilde{G}_{31}(\gamma), \widetilde{G}_{32}(\gamma), \widetilde{G}_{33}(\gamma) \end{pmatrix}$ *and the random vector* $\widetilde{\phi}(\gamma) = \begin{pmatrix} \widetilde{\phi}_1(\gamma) \\ \widetilde{\phi}_2(\gamma) \\ \widetilde{\phi}_3(\gamma) \end{pmatrix}$ *are specified in Lemma B.4 of the Appendix.*

Theorem 4.2 establishes the convergence results for the FGLS estimator $\widetilde{\gamma}_n$. If $-1/2 < \tau < 1/2$, the threshold value can be consistently estimated and its limiting distribution depends on the persistence parameter ρ. Note that $\widetilde{\lambda} = \left(1+\rho^2\right)\lambda$, implying that the FGLS estimator is more efficient than the profiled LS estimator when $\rho \neq 0$. The simulations in Section 5 demonstrate this result.

4.2. Generalized Sup-Wald Statistics

Testing for the existence of the threshold effect in a cointegration regression is challenging since it is a joint hypothesis problem (see Balke and Fomby, 1997). For example, when testing for the existence of threshold effects, the statistics based on error correction models (ECMs) need to assume the model is a cointegrating regression. Therefore, the rejection of the null hypothesis does not necessarily indicate that there is a threshold effect. It may mean that the regression is a spurious relationship. Thanks to the robustness of the FGLS estimator, we can design a generalized sup-Wald test statistic based on this estimator to test for the existence of threshold effects, without any requirement on the stationarity of the error term.

We consider to test the following hypothesis:

$$H_0 : \delta_n = 0 \quad v.s. \quad H_1 : \delta_n \neq 0.$$

Under the null, after transformation, the model is

$$\widetilde{y}_t = \alpha' \widetilde{x}_t + \beta' \widetilde{z}_t + \widetilde{\eta}_t, \tag{25}$$

while under the alternative, the transformed model is

$$\widetilde{y}_t = \alpha' \widetilde{x}_t + \delta_n' \widetilde{x}_t(\gamma_0) + \beta' \widetilde{z}_t + \widetilde{\eta}_t = \widetilde{\theta}' \widetilde{A}_{1t}(\gamma_0) + \widetilde{\eta}_t. \tag{26}$$

Let $\widetilde{V}_1 = (\widetilde{X}, \widetilde{Z})$, then a generalized Wald statistic can be defined as

$$\widetilde{T}_n(\gamma) = \widetilde{\delta}_n(\gamma)' (\widetilde{X}(\gamma)(I - \widetilde{P}(\gamma))\widetilde{X}(\gamma))\widetilde{\delta}_n(\gamma)/\widetilde{\sigma}^2,$$

where $\widetilde{P}(\gamma)$ is the projection matrix for $\widetilde{V}_1$ and $\widetilde{\sigma}^2 = \widetilde{SSR}_n(\gamma)/n$.

The generalized sup-Wald statistic is defined as

$$\widetilde{T}_n = \sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \widetilde{T}_n(\gamma).$$

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

THEOREM 4.3. *Under Assumptions* (4.1), (4.2), (2.3), *and* (2.4) *and* H_0 : $\delta_n = 0$, *the following limiting result holds:*

$$\widetilde{T}_n \Rightarrow \widetilde{T} = \sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \widetilde{T}(\gamma)$$
$$= \sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \frac{1}{\sigma^2} \widetilde{\Gamma}'(\gamma) \left(\widetilde{G}_{22}(\gamma) - \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix}' \begin{pmatrix} \widetilde{G}_{11}, \widetilde{G}_{13} \\ \widetilde{G}_{31}, \widetilde{G}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{G}_{12}(\gamma) \\ \widetilde{G}_{32}(\gamma) \end{pmatrix} \right)^{-1} \widetilde{\Gamma}(\gamma),$$

where $\widetilde{\Gamma}(\gamma)$ *is defined by equation (24).*

Theorem 4.3 establishes the limiting distribution of the generalized sup-Wald test statistic, which is nonstandard and we generate the critical values using a parametric bootstrap method. We first estimate $\widehat{\alpha}$ and $\widehat{\beta}$ using equation (25) under the null. Then, we obtain the residual terms $\{\widehat{\widetilde{\eta}}_t(\widetilde{\gamma}_n)\}_{t=2}^n$ using equation (26) under the alternative. We draw a random variable $\widetilde{\eta}_t^b$ from the sample $\{\widehat{\widetilde{\eta}}_t(\widetilde{\gamma}_n)\}_{t=2}^n$ for all $t = 2, \ldots, n$ and generate a new sequence $\{\widetilde{y}_t^b\}_{t=1}^n$ by $\widetilde{y}_t^b = \widehat{\alpha}'\widetilde{x}_t + \widehat{\beta}'\widetilde{z}_t + \widetilde{\eta}_t^b$. Define $y_1^b = y_1$ and $y_t^b = \widetilde{y}_t^b + \widehat{\rho} y_{t-1}^b$ for all $t = 2, \ldots, n$. Let $\widetilde{T}_n^b$ be the sup-Wald test calculated from the new data set $\{y_t^b, x_t, z_t, q_t\}_{t=2}^n$. Under the null, the distribution of $\widetilde{T}_n^b$ approximates the distribution of $\widetilde{T}_n$. The bootstrap p-value is obtained by calculating the frequency of simulated $\widetilde{T}_n^b$ that exceeds $\widetilde{T}_n$ when the number of simulations is sufficiently large. Following Hansen (1996), one can show that the generated p-value converges to the true size.

One can use the generalized sup-Wald statistic $\widetilde{T}_n$ to construct robust CIs for the threshold estimators following the procedures described in Section 3.2.

5. SIMULATIONS

This section demonstrates the finite sample performance of the estimators and test statistics through two simulation experiments.

Experiment 1. In this experiment, we examine the finite sample performance of the profiled LS estimator and the FGLS estimator under different model settings. The data generating process is given by

$$y_t = \alpha x_t + \delta_n x_t(q_t \le \gamma_0) + \beta v_t + u_t,$$

where $x_t = x_{t-1} + v_t$ and $u_t = \rho u_{t-1} + \varepsilon_t$. The threshold variable q_t is generated by an $AR(1)$ process: $q_t = 0.5q_{t-1} + e_t$. The innovation processes v_t, ε_t, and e_t follow i.i.d. $N(0, 1)$ and are independent of each other. The true threshold value is set as $\gamma_0 = 0$ and the coefficient α is set as 1. The parameter β is set as 0 or 0.5, which controls the correlation between x_t and u_t. To check the impact of the serial correlation of u_t on the estimation, we set ρ as either 0, 1, 0.95, or -0.95. Moreover, we set the size of the threshold effect as $\delta_n = 2n^{-1/2-\tau}$ with τ chosen as either 0 or 0.5, corresponding to the case with identification or the case with weak identification. The number of replications is $N = 1,000$. The sample size n is set as 100, 200, or 400. We let the number of lags $K = 5$ when construct the FGLS estimator.[8]

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

Table 1 reports the mean square errors (MSEs) of the profiled LS estimator and the FGLS estimator for the threshold value.

Specifically, Panel A of Table 1 displays the results when $\tau = 0$, i.e., the case with identification. From this panel, one can find that the MSE of the profiled LS estimator $\widehat{\gamma}$ decreases to zero as sample size increases when $\rho < 1$, but it does not when $\rho = 1$. On the other hand, the MSE of the FGLS estimator $\widetilde{\gamma}$ decays to zero regardless of whether $\rho < 1$ or $\rho = 1$. Of particular interest is that the FGLS estimator performs better than the profiled LS estimator when serial correlation and regressor endogeneity exist. Panel B reports the results when $\tau = 0.5$, i.e., the case with weak identification. From this panel, one can find that neither the MSE of profiled LS estimator or that of FGLS estimator converges to zero. However, the FGLS estimator has smaller MSE than the profiled LS estimator when serial correlation and regressor endogeneity exist. In summary, we conclude that the FGLS estimator does improve the estimation efficiency.

TABLE 1. The mean squared errors (MSEs) of threshold estimators

		$n = 100$		$n = 200$		$n = 400$	
Panel A: $\tau = 0$							
ρ	β	MSE($\widehat{\gamma}$)	MSE($\widetilde{\gamma}$)	MSE($\widehat{\gamma}$)	MSE($\widetilde{\gamma}$)	MSE($\widehat{\gamma}$)	MSE($\widetilde{\gamma}$)
0	0	.012	.019	.004	.0040	.0021	.0022
0	.5	.013	.009	.004	.0020	.0008	.0003
1	0	.191	.002	.221	.0056	.2030	.0002
1	.5	.314	.005	.197	.0019	.2340	.0002
.95	0	.129	.019	.097	.0010	.0320	.0001
.95	.5	.125	.002	.129	.0031	.0310	.0004
−.95	0	.093	.004	.043	.0002	.0270	.0002
−.95	.5	.171	.020	.095	.0031	.0230	.0003
Panel B: $\tau = 0.5$							
0	0	.365	.379	.485	.4720	.4540	.4550
0	.5	.519	.509	.518	.4880	.5100	.4540
1	0	1.17	.592	.654	.5270	.7170	.4080
1	.5	.435	.302	.460	.3890	.5440	.4290
.95	0	.586	.410	.572	.3800	.5220	.4040
.95	.5	.375	.359	.557	.4400	.4940	.3670
−.95	0	.800	.608	.422	.3640	.6060	.4120
−.95	.5	.742	.530	.532	.4200	.6220	.3560

Note: The model is $y_t = \alpha x_t + \delta_n x_t(q_t \le \gamma_0) + \beta v_t + u_t$ with $x_t = x_{t-1} + v_t$, $u_t = \rho u_{t-1} + \varepsilon_t$, and $q_t = 0.5q_{t-1} + e_t$, where v_t, ε_t, and e_t are i.i.d. $N(0, 1)$. The true threshold value γ_0 is set as 0 and the coefficient α is set as 1. The parameter ρ is set as either 0, 1, 0.95, or −0.95 and the parameter β is set as 0 or 0.5. The size of the threshold effect δ_n is set as $\delta_n = 2n^{-1/2-\tau}$, where τ is 0 or 0.5. MSE($\widehat{\gamma}$) is the MSE calculated for the profiled LS estimator of γ_0 and MSE($\widetilde{\gamma}$) is for the FGLS estimator. The sample size n takes either 100, 200, or 400. The replication number is 1,000.

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

Experiment 2. Through this experiment, we examine the performance of the test statistics and the model selection procedure. The data generating process is the same as in Simulation 1, but we consider the case of $\delta_n = 0$ to evaluate the size performance of the test statistics.

Table 2 reports the size performance of the sup-Wald statistic T_n and the generalized sup-Wald statistic $\widetilde{T}_n$, which are constructed using the profiled LS estimator and the FGLS estimator, respectively. Comparing Panel A with Panel B, one can see that $\widetilde{T}_n$ has better size performance than T_n when serial correlation or regressor endogeneity exists. In an unreported result with the sample size $n = 1,000$, the rejection rates of $\widetilde{T}_n$ are close to the size, but not for T_n.

Tables 3 and 4 report the power performance for T_n and $\widetilde{T}_n$, respectively. In both tables, Panel A shows the results for the case with identification and Panel B shows the results for the case with weak identification.

TABLE 2. Size performance

			$n = 100$			$n = 200$			$n = 400$		
ρ	δ_n	β	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: $\mathbf{T}_n$											
0	0	0	.078	.028	.005	.114	.056	.018	.098	.046	.008
0	0	.5	.120	.092	.034	.084	.050	.008	.076	.052	.010
1	0	0	.132	.082	.030	.064	.034	.014	.068	.044	.013
1	0	.5	.086	.052	.022	.134	.074	.022	.112	.076	.018
.95	0	0	.128	.073	.042	.178	.072	.030	.104	.046	.008
.95	0	.5	.134	.075	.020	.146	.064	.004	.132	.061	.016
−.95	0	0	.120	.072	.010	.106	.060	.018	.093	.042	.015
−.95	0	.5	.070	.034	.008	.118	.064	.006	.092	.039	.006
Panel B: $\widetilde{\mathbf{T}}_n$											
0	0	0	.110	.072	.030	.102	.062	.022	.098	.047	.010
0	0	.5	.144	.084	.028	.076	.044	.014	.088	.046	.008
1	0	0	.084	.058	.020	.112	.038	.018	.104	.054	.014
1	0	.5	.130	.062	.014	.082	.042	.014	.116	.055	.012
.95	0	0	.114	.062	.002	.072	.060	.008	.082	.042	.008
.95	0	.5	.128	.064	.032	.132	.076	.020	.094	.037	.008
−.95	0	0	.086	.036	.016	.142	.078	.024	.090	.044	.010
−.95	0	.5	.116	.066	.022	.089	.046	.016	.091	.058	.015

Note: The model is $y_t = \alpha x_t + \delta_n x_t(q_t \le \gamma_0) + \beta v_t + u_t$ with $x_t = x_{t-1} + v_t$, $u_t = \rho u_{t-1} + \varepsilon_t$, and $q_t = 0.5q_{t-1} + e_t$, where v_t, ε_t, and e_t are i.i.d. $N(0,1)$. The true threshold value γ_0 is set as 0 and the coefficient α is set as 1. The parameter ρ is set as either 0, 1, 0.95, or −0.95 and the parameter β is set as 0 or 0.5. The size of the threshold effect δ_n is set as 0 to check the size performance. The sample size n takes either 100, 200, or 400. The replication number is 1,000.

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

TABLE 3. Power performance of $\mathbf{T}_n$

		$n=100$				$n=200$				$n=400$			
Panel A: $\tau=0$													
ρ	β	10%	5%	1%	$>(\ln(n))^2$	10%	5%	1%	$>(\ln(n))^2$	10%	5%	1%	$>(\ln(n))^2$
0	0	1.00	.994	.984	.956	1.00	1.00	.998	.982	1.00	1.00	1.00	1.00
0	.5	.992	.980	.968	.904	1.00	1.00	.996	.974	1.00	1.00	1.00	.994
1	0	.478	.436	.360	.462	.854	.836	.816	.328	.742	.692	.586	.234
1	.5	.506	.460	.312	.486	.668	.614	.540	.346	.546	.488	.410	.266
.95	0	.796	.720	.564	.610	.856	.820	.732	.624	.910	.884	.820	.722
.95	.5	.830	.772	.654	.636	.790	.750	.708	.584	.942	.916	.868	.668
−.95	0	.956	.922	.884	.464	.962	.950	.912	.506	.994	.986	.958	.596
−.95	.5	.856	.802	.708	.444	.964	.956	.906	.502	.996	.984	.958	.648
Panel B: $\tau=0.5$													
0	0	.322	.224	.074	.024	.284	.222	.098	.004	.294	.196	.042	.002
0	.5	.258	.150	.030	.020	.254	.144	.078	.002	.166	.124	.054	.000
1	0	.164	.098	.050	.100	.098	.050	.016	.000	.386	.312	.160	.000
1	.5	.090	.043	.010	.030	.116	.050	.010	.030	.020	.012	.002	.012
.95	0	.070	.034	.010	.018	.082	.034	.016	.000	.074	.026	.004	.002
.95	.5	.322	.234	.072	.024	.078	.044	.002	.034	.094	.052	.010	.000
−.95	0	.296	.236	.146	.002	.318	.220	.124	.000	.092	.042	.020	.000
−.95	.5	.308	.198	.050	.004	.110	.050	.016	.000	.134	.064	.018	.000

Note: The model is $y_t=\alpha x_t+\delta_n x_t(q_t\le\gamma_0)+\beta v_t+u_t$ with $x_t=x_{t-1}+v_t$, $u_t=\rho u_{t-1}+\varepsilon_t$, and $q_t=0.5q_{t-1}+e_t$, where v_t, ε_t, and e_t are i.i.d. $N(0,1)$. The true threshold value γ_0 is set as 0 and the coefficient α is set as 1. The parameter ρ is set as either 0, 1, 0.95, or −0.95 and the parameter β is set as 0 or 0.5. The size of the threshold effect δ_n is set as $\delta_n=2n^{-1/2-\tau}$, where τ is 0 or 0.5. The sample size n takes either 100, 200, or 400. The replication number is 1,000.

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

TABLE 4. Power performance of $\widetilde{\mathbf{T}}_n$

		$n=100$				$n=200$				$n=400$			
Panel A: $\tau=0$													
ρ	β	10%	5%	1%	$>(\ln(n))^2$	10%	5%	1%	$>(\ln(n))^2$	10%	5%	1%	$>(\ln(n))^2$
0	0	1.00	.998	.988	.970	1.00	1.00	.998	.984	1.00	1.00	1.00	1.00
0	.5	.998	.994	.984	.952	1.00	1.00	1.00	.990	1.00	1.00	1.00	.998
1	0	.998	.996	.982	.958	1.00	1.00	1.00	.992	1.00	1.00	1.00	1.00
1	.5	1.00	.998	.992	.970	1.00	1.00	1.00	.988	1.00	1.00	1.00	.998
.95	0	.986	.986	.968	.968	1.00	1.00	1.00	.986	1.00	1.00	1.00	1.00
.95	.5	.996	.996	.994	.978	1.00	1.00	1.00	.992	1.00	1.00	1.00	.998
−.95	0	1.00	.998	.994	.994	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
−.95	.5	1.00	1.00	.998	.996	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Panel B: $\tau=0.5$													
0	0	.324	.206	.110	.034	.256	.198	.108	.004	.264	.166	.052	.000
0	.5	.206	.090	.030	.032	.250	.170	.066	.008	.272	.192	.108	.002
1	0	.278	.190	.078	.044	.280	.212	.126	.080	.248	.162	.052	.000
1	.5	.298	.240	.100	.048	.292	.174	.072	.014	.228	.146	.050	.002
.95	0	.296	.206	.118	.068	.428	.336	.194	.020	.264	.186	.068	.002
.95	.5	.268	.196	.084	.072	.308	.228	.068	.016	.292	.210	.082	.000
−.95	0	.550	.434	.182	.110	.442	.352	.268	.034	.546	.448	.302	.010
−.95	.5	.500	.392	.266	.088	.402	.292	.160	.030	.320	.204	.134	.010

Note: The model is $y_t=\alpha x_t+\delta_n x_t(q_t\le\gamma_0)+\beta v_t+u_t$ with $x_t=x_{t-1}+v_t$, $u_t=\rho u_{t-1}+\varepsilon_t$, and $q_t=0.5q_{t-1}+e_t$, where v_t, ε_t, and e_t are i.i.d. $N(0,1)$. The true threshold value γ_0 is set as 0 and the coefficient α is set as 1. The parameter ρ is set as either 0, 1, 0.95, or −0.95 and the parameter β is set as 0 or 0.5. The size of the threshold effect δ_n is set as $\delta_n=2n^{-1/2-\tau}$, where τ is 0 or 0.5. The sample size n takes either 100, 200, or 400. The replication number is 1,000.

When the threshold effect is identified, both T_n and $\widetilde{T}_n$ seem to be consistent since their power converges to one very quickly as the sample size increases. However, $\widetilde{T}_n$ performs better than T_n when there exists serial correlation or model endogeneity. Thus, we recommend to use the generalized sup-Wald statistic $\widetilde{T}_n$ for testing the existence of threshold effects in practice.

When the threshold effect is only weakly identified, both statistics seem to have low power even when the sample size is 400, which is consistent to the results of Theorem 3.2. In such case, we recommend to use the model selection procedure described in Section 3.2 to detect the weak identification at first. In our simulation, we choose $(\ln(n))^2$ as the tuning parameter κ_n and judge that the model is weakly identified if the test statistic is smaller than κ_n. From the columns corresponding to $(\ln(n))^2$ in Tables 3 and 4, one can find that the model selection procedure based on $\widetilde{T}_n$ works very well.

6. CONCLUSION

In the literature, statistical theory for threshold models with stationary explanatory variables has been well developed by Hansen (1996, 2000). However, in empirical macroeconomics and finance, many explanatory variables are nonstationary. This paper establishes the asymptotic theory for threshold models with nonstationary regressors under the diminishing threshold effect assumption. What is more important, we provide a method to construct robust confidence intervals for the threshold estimators, which have the correct asymptotic size irrespective of the magnitude of the threshold effect. The study can also be related to the literature of nonlinear cointegration as the model allows for a threshold effect in the cointegration relationship.

There are several directions open for further work. First, it may be interesting to develop a more general model with multiple threshold effects, each with a different identification strength. A sequential procedure can be applied to determine the number of regimes and their identification strengths. Second, the model can be extended to allow for stationary regressors. Moreover, there are many interesting applications. For example, one can use the model to study the regime-sensitive Taylor rule, where the dependent variable y_t could be the short-term interest rate, the explanatory variables x_t could be macroeconomic variables such as the inflation rate and the unemployment rate, and the threshold variable q_t could be the GDP growth rate. Another application is to model the regime-dependent predictability of the fundamental ratios, such as the dividend-price ratio and the earning-price ratio, to equity returns. The threshold variable could be a variable indicating the status of the economy. All these are left for future studies.

NOTES

1. The model could be extended to include an intercept term and stationary regressors.

2. Loosely speaking, if the response variable y_t is generated by a nonlinear transformation of integrated regressors x_t plus a stationary error term, then there exists a nonlinear cointegrating relationship between y_t and x_t.

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

3. Marmer (2008) shows that linear regression techniques actually may lead to spurious forecasts if there exist small departures from the linear structure, but improvements of the forecast accuracy are possible if the nonlinear components are properly considered.

4. See Elliott and Müller (2007) for weak identification in break-point models; Cheng (2008) and Shi and Phillips (2012) for weak identification in general nonlinear models.

5. Assumption (2.2) can be relaxed under certain circumstances. For instance, u_t could be generalized to follow a linear moving average process of finite order l. However, fully generalizing the model to allow for serially correlated errors would involve substantial added complexity (see Gonzalo and Pitarakis, 2006).

6. In an early version of this paper, we have also established the limiting distribution of the coefficient estimators $\widehat{\theta}\left(\widehat{\gamma}_n\right)$. We show that, if $-1/2 < \tau < 1/2$, the limiting distribution of the coefficients estimators is mixed normal, which makes conventional *t*-test and *chi-square* tests applicable, whereas if $\tau = 1/2$, the limiting result contains a bias term as a result of the inconsistency of $\widehat{\gamma}_n$.

7. Structural spurious regressions can be due to integrated measurement errors and missing integrated regressors (see Choi et al., 2008).

8. Other values for K, such as 10 and 15, are also applied with little change in the results.

REFERENCES

Andrews, D.W.K. & C.J. McDermott (1995) Nonlinear econometric models with deterministically trending variables. *Review of Economic Studies* 62(3), 343–360.

Balke, N. & T. Fomby (1997) Threshold cointegration. *International Economics Review* 8, 627–645.

Bierens, H. & L. Martins (2010) Time-varying cointegration. *Econometric Theory* 26, 1453–1490.

Cai, Z., Q. Li, & J.Y. Park (2009) Functional-coefficient models for nonstationary time series data. *Journal of Econometrics* 148, 101–113.

Campbell, J.Y. & M. Yogo (2006) Efficient tests of stock return predictability. *Journal of Financial Economics* 81, 27–60.

Caner, M. & B.E. Hansen (2001) Threshold autoregression with a unit root. *Econometrica* 69, 1555–1596.

Chan, K.S. (1993) Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics* 21, 520–533.

Chen, H. (2013) Robust Estimation and Inference for Threshold Models with Integrated Regressors. Working papers, Xiamen University. Available at SSRN: http://ssrn.com/abstract=2287442.

Cheng, X. (2008) Robust Confidence Intervals in Nonlinear Regression under Weak Identification. Manuscript, Department of Economics, Yale University.

Choi, C.Y., L. Hu, & M. Ogaki (2008) Robust estimation for structural spurious regressions and a Hausman-type cointegration test. *Journal of Econometrics* 142, 327–351.

Choi, I. & P. Saikkonen (2010) Test of nonlinear cointegration. *Econometric Theory* 26, 682–709.

Durlauf, S.N. & P.A. Johnson (1995) Multiple regimes and cross-country growth behavior. *Journal of Applied Econometrics* 10, 365–384.

Elliott, G. & U.K. Müller (2007) Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics* 141(2), 1196–1218.

Gao, J., M.L. King, Z. Lu, & D. Tjøstheim (2009a) Specification testing in nonstationary time series autoregression. *Annals of Statistics* 37, 3893–3928.

Gao, J., M.L. King, Z. Lu, & D. Tjøstheim (2009b) Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory* 25, 1869–1892.

Gonzalo, J. & J. Pitarakis (2002) Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics* 110, 319–352.

Gonzalo, J. & J.Y. Pitarakis (2006) Threshold effects in cointegrating regressions. *Oxford Bulletin of Economics and Statistics* 68, 813–833.

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

Gonzalo, J. & J.Y. Pitarakis (2012) Regime specific predictability in predictive regressions. *Journal of Business and Economic Statistics* 30(2), 229–241.
Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.
Hansen, B.E. (1997) Inference in TAR models. *Studies in Nonlinear Dynamics and Econometrics* 2, 1–14.
Hansen, B.E. (2000) Sample splitting and threshold estimation. *Econometrica* 68, 575–603.
Li, D. & S.Q. Ling (2012) On the least squares estimation of multiple-regime threshold autoregressive models. *Journal of Econometrics* 167, 240–253.
Marmer, V. (2008) Nonlinearity, nonstationarity, and spurious forecasts. *Journal of Econometrics* 142(1), 1–27.
Park, J.Y. (1992) Canonical cointegrating regressions. *Econometrica* 60, 119–143.
Park, J.Y. & S. Hahn (1999) Cointegration regression with time varying coefficients. *Econometric Theory* 15, 664–703.
Park, J.Y. & P.C.B. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrica* 69(1), 117–161.
Perron, P. & T. Yabu (2009) Testing for shifts in trend with an integrated or stationary noise component. *Journal of Business and Economics Statistics* 27(3), 369–396.
Phillips, P.C.B. & S.N. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 474–95.
Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99–125.
Phillips, P.C.B. & D.J. Hodgson (1994) Spurious regression and generalized least squares. *Econometric Theory* 10, 957–958.
Phillips, P.C.B. & J.H. Lee (2013) Predictive regression under varying degrees of persistence and robust long-horizon regression. *Journal of Econometrics* 177(2), 250–264.
Phillips, P.C.B. & J.Y. Park (1988) Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors. *Journal of the American Statistical Association* 83, 111–115.
Potter, S.M. (1995) A nonlinear approach to US GNP. *Journal of Applied Econometrics* 10, 109–125.
Saikkonen, P. (1991) Asymptotically efficient estimation of cointegration regressions. *Econometric Theory* 7, 1–20.
Seo, M. & O. Linton (2007) A smoothed least squares estimator for threshold regression models. *Journal of Econometrics* 141, 704–735.
Shao, Q. & C. Lu (1987) Strong approximation for partial sums of weakly dependent random variables. *Scientia Sinica* 15, 576–587.
Shi, X. & P.C.B. Phillips (2012) Nonlinear cointegration regression under weak identification. *Econometric Theory* 28(1), 1–39.
Wang, Q. & P.C.B. Phillips (2009) Structural nonparametric cointegrating regression. *Econometrica* 77(6), 1901–1948.
Yu, P. (2012) Likelihood estimation and inference in threshold regression. *Journal of Econometrics* 167, 274–294.

APPENDIX: Mathematical Proofs

In this appendix, Part A provides the proofs related to the basic model while Part B is for the generalized model. To save space, we skip the details for some intermediary results and a more detailed presentation can be found in Chen (2013).

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

Part A

LEMMA A.1. *Under Assumptions* (2.1) − (2.3), *for any* $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, *we have* $\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} I_t(q_t \le \gamma) u_t \Rightarrow \sigma_u W(s, F(\gamma))$.

LEMMA A.2. *Under Assumptions* (2.1) − (2.4), *for any* $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, *we have*

$$\frac{1}{n} \sum_{t=1}^{n} x_t I_t(\gamma) u_t \Rightarrow \sigma_u \int \mathbf{B}_v(s) dW(s, F(\gamma)).$$

LEMMA A.3. *Under Assumptions* (2.1) − (2.4), *for any* $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, *as* $n \to \infty$, *we have*

(i) $n^{-2} \sum_{t=1}^{n} A_t(\gamma) A_t'(\gamma) \xrightarrow{p} M(\gamma)$,

(ii) $n^{-1} \sum_{t=1}^{n} A_t(\gamma) u_t \Rightarrow \sigma_u \begin{pmatrix} \int \mathbf{B}_v(s) dW(s) \\ \int \mathbf{B}_v(s) dW(s, F(\gamma)) \end{pmatrix}$,

where $A_t(\gamma)$ *is defined in equation* (5) *and* $M(\gamma) = \begin{pmatrix} \int \mathbf{B}_v(s)\mathbf{B}_v'(s) ds & F(\gamma) \int \mathbf{B}_v(s)\mathbf{B}_v'(s) ds \\ F(\gamma) \int \mathbf{B}_v(s)\mathbf{B}_v'(s) ds & F(\gamma) \int \mathbf{B}_v(s)\mathbf{B}_v'(s) ds \end{pmatrix}$.

LEMMA A.4. *Under Assumptions* (2.1) − (2.4), *we have*

$$n\left(\widehat{\theta}(\gamma_0) - \theta\right) \Rightarrow \sigma_u M(\gamma_0)^{-1} \begin{pmatrix} \int \mathbf{B}_v(s) dW(s) \\ \int \mathbf{B}_v(s) dW(s, F(\gamma_0)) \end{pmatrix}.$$

For any $\gamma \neq \gamma_0$, *if* $-1/2 < \tau < 1/2$, *we have*

$$n^{\tau+1/2}(\widehat{\theta}(\gamma) - \theta) \Rightarrow M(\gamma)^{-1} \Pi(\gamma, \gamma_0, \delta_0);$$

if $\tau = 1/2$,

$$n\left(\widehat{\theta}(\gamma) - \theta\right) \Rightarrow \sigma_u M(\gamma)^{-1} \begin{pmatrix} \int \mathbf{B}_v(s) dW(s) \\ \int \mathbf{B}_v(s) dW(s, F(\gamma)) \end{pmatrix} + M(\gamma)^{-1} \Pi(\gamma, \gamma_0, \delta_0),$$

where $\Pi(\gamma, \gamma_0, \delta_0) = -\begin{pmatrix} (F(\gamma) - F(\gamma_0)) \int \mathbf{B}_v(s)\mathbf{B}_v'(s) ds \\ (F(\gamma) - F(\gamma_0 \wedge \gamma)) \int \mathbf{B}_v(s)\mathbf{B}_v'(s) ds \end{pmatrix} \delta_0$.

Proof of Lemmas A.1–A.4. The proofs are standard and are skipped here. ■

LEMMA A.5. *If* $-1/2 < \tau < 1/2$, *we have* $\widehat{\gamma}_n \xrightarrow{p} \gamma_0$.

Proof. To establish the consistency of $\widehat{\gamma}_n$, we need to prove $\Pr(|\widehat{\gamma}_n - \gamma_0| > \varepsilon) \to 0$ for any $\varepsilon > 0$. Denote $B(\epsilon) = \{\gamma : |\gamma - \gamma_0| > \varepsilon\}$ and $\overline{B}(\epsilon) = [\underline{\gamma}, \overline{\gamma}] \backslash B(\epsilon)$. By the definition of $\widehat{\gamma}_n$, we have

$$\begin{aligned} \Pr(|\widehat{\gamma}_n - \gamma_0| > \varepsilon) &= \Pr\left(\inf_{\gamma \in B(\epsilon)} SSR_n(\gamma) < \inf_{\gamma \in \overline{B}(\epsilon)} SSR_n(\gamma) \right) \\ &\le \Pr\left(\inf_{\gamma \in B(\epsilon)} SSR_n(\gamma) < SSR_n(\gamma_0) \right) \\ &= \Pr\left(\inf_{\gamma \in B(\epsilon)} n^{2\tau-1} (SSR_n(\gamma) - SSR_n(\gamma_0)) < 0 \right). \end{aligned}$$

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

Thus, to prove $\Pr(|\widetilde{\gamma}_n - \gamma_0| > \varepsilon) \to 0$, it suffices to show that $\inf_{\gamma \in B(\epsilon)} n^{2\tau-1}(SSR_n(\gamma) - SSR_n(\gamma_0)) > 0$ with probability 1.

Rewrite equation (2) as a matrix compacted form: $Y = X'\alpha + X(\gamma_0)'\delta_n + u$, where $Y, X, X(\gamma_0)$ and u stack $y_t, x_t, x_t I(q_t \le \gamma_0)$ and u_t, respectively. Denote $X^*(\gamma) = (X(\gamma), X - X(\gamma))$ and define its projection matrix $P^*_\gamma = X^*(\gamma)(X^*(\gamma)'X^*(\gamma))^{-1}X^*(\gamma)'$. By some simple algebra, we have

$$SSR_n(\gamma) = Y'(I - P^*_\gamma)Y = \delta_n' X(\gamma_0)'(I - P^*_\gamma)X(\gamma_0)\delta_n + 2\delta_n' X(\gamma_0)'(I - P^*_\gamma)u + u'(I - P^*_\gamma)u,$$

where the second equation uses the fact that $X'(I - P^*_\gamma) = 0$, since X is a linear combination of $(X(\gamma), X - X(\gamma))$.

When $\gamma = \gamma_0$, we have $SSR_n(\gamma_0) = u'(I - P^*_{\gamma_0})u$, and it follows that

$$\begin{aligned} n^{2\tau-1}(SSR_n(\gamma) - SSR_n(\gamma_0)) &= n^{-1+2\tau}\delta_n' X(\gamma_0)'(I - P^*_\gamma)X(\gamma_0)\delta_n \\ &\quad + n^{-1+2\tau}2\delta_n' X(\gamma_0)'(I - P^*_\gamma)u \\ &\quad + n^{-1+2\tau}(u'(I - P^*_\gamma)u - u'(I - P^*_{\gamma_0})u) \\ &\equiv S_1 + S_2 + S_3, \text{ say.} \end{aligned} \tag{A.1}$$

Next, we will show that $S_1 + S_2 + S_3$ uniformly converges to a function $b(\gamma)$ which is positive for any $\gamma \in B(\epsilon)$.

Given $-1/2 < \tau < 1/2$, by Lemma A.3, it can be shown that

$$\begin{aligned} S_2 = n^{-1/2+\tau}2\left(n^{1/2+\tau}\delta_n\right)'\frac{1}{n}X(\gamma_0)'(I - P^*_\gamma)u &= n^{-1/2+\tau}2\delta_0'\frac{1}{n}X(\gamma_0)'(I - P^*_\gamma)u \\ &= O_p(n^{-1/2+\tau}) \xrightarrow{p} 0, \end{aligned}$$

$$\begin{aligned} S_3 = n^{-1+2\tau}(u'(I - P^*_\gamma)u - u'(I - P^*_{\gamma_0})u) &= n^{-1+2\tau}(u'P^*_{\gamma_0}u - u'P^*_\gamma u) \\ &= O_p(n^{-1+2\tau}) \xrightarrow{p} 0. \end{aligned}$$

Using a similar argument of Lemma A.5 in Hansen (2000), we can show, for any $\gamma \ge \gamma_0$,

$$\begin{aligned} S_1 = n^{-1+2\tau}\delta_n' X'(\gamma_0)(I - P^*_\gamma)X(\gamma_0)\delta_n &\xrightarrow{p} (F(\gamma_0) \\ &\quad - F(\gamma_0)F(\gamma)^{-1}F(\gamma_0))\delta_0'\left(\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds\right)\delta_0 \equiv b_1(\gamma) \end{aligned}$$

uniformly. Since $(F(\gamma_0) - F(\gamma_0)F(\gamma)^{-1}F(\gamma_0)) \ge 0$ and $\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds$ is positive definite random matrix, $b_1(\gamma) \ge 0$ and the equality holds if and only if $\gamma = \gamma_0$.

For $\gamma \le \gamma_0$, we can show $S_1 \xrightarrow{p} (F(\gamma_0) - F(\gamma))\delta_0'\left(\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds\right)\delta_0 \equiv b_2(\gamma)$ uniformly, where $b_2(\gamma) \ge 0$ and the equality holds if and only if $\gamma = \gamma_0$. Define $b(\gamma) = b_1(\gamma)I(\gamma \ge \gamma_0) + b_2(\gamma)I(\gamma \le \gamma_0)$, and we have shown that $n^{-1+2\tau}(SSR_n(\gamma) - SSR_n(\gamma_0)) \xrightarrow{p} b(\gamma)$ uniformly for any $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and $b(\gamma)$ is strictly positive when $\gamma \in B(\epsilon)$. Thus, $\Pr(\inf_{\gamma \in B(\epsilon)} n^{2\tau-1}(SSR_n(\gamma) - SSR_n(\gamma_0)) > 0) \to 1$. ■

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

Proof of Theorem 3.1. We first prove the limiting results for the case with $-1/2 < \tau < 1/2$. Let $a_n = n^{1-2\tau}$. To prove $\widehat{\gamma}_n$ converge to γ_0 with rate a_n, we need to prove that $a_n|\widehat{\gamma}_n - \gamma_0| = O_p(1)$, or there exists a constant $\overline{v} > 0$, $\lim_{n\to\infty} \Pr(|\widehat{\gamma}_n - \gamma_0| \le \overline{v}/a_n) = 1$. For any $B > 0$, define $V_B = \{\gamma : |\gamma - \gamma_0| < B\}$. When n is large enough, we have $\overline{v}/a_n < B$. Since $\widehat{\gamma}_n \xrightarrow{p} \gamma_0$ according to Lemma A.5, $\Pr(\{\widehat{\gamma}_n \in V_B\}) \xrightarrow{p} 1$. Therefore, we only need to examine the limiting behavior of γ in V_B.

Define a subset $V_B(\overline{v}) = \{\gamma : \overline{v}/a_n < |\gamma - \gamma_0| < B\}$. Thus, $V_B(\overline{v}) \subset V_B$. To prove $\Pr(|\widehat{\gamma}_n - \gamma_0| \le \overline{v}/a_n) = 1$, we just need to prove $\Pr(\widehat{\gamma}_n \in V_B(\overline{v})) = 0$. Let $\widehat{\theta}$ and $\widehat{\delta}$ be the estimation of $\widehat{\theta}(\widehat{\gamma}_n)$ and $\widehat{\delta}(\widehat{\gamma}_n)$. Also, we denote $SSR_n^*(\gamma) = \sum_{t=1}^n (y_t - \widehat{\theta}' A_t(\gamma))^2$ and $SSR_n^*(\gamma_0) = \sum_{t=1}^n (y_t - \widehat{\theta}' A_t(\gamma_0))^2$. By the definition of $\widehat{\gamma}_n$, we have $SSR_n^*(\widehat{\gamma}_n) \le SSR_n^*(\gamma_0)$. Hence, it suffices to prove that for any $\gamma \in V_B(\overline{v})$, $SSR_n^*(\gamma) > SSR_n^*(\gamma^0)$ with probability 1.

We consider the case of $\gamma > \gamma^0$ first. Using an argument of symmetry, we can, without loss of generality, prove the result for the case of $\gamma < \gamma_0$. Note that

$$
\begin{aligned}
SSR_n^*(\gamma) - SSR_n^*(\gamma_0) &= \sum_{t=1}^n \left(y_t - \widehat{\theta}' A_t(\gamma)\right)^2 - \sum_{t=1}^n \left(y_t - \widehat{\theta}' A_t(\gamma_0)\right)^2 \\
&= \sum_{t=1}^n \delta_n' (x_t(\gamma) - x_t(\gamma_0))(x_t(\gamma) - x_t(\gamma_0))' \delta_n \\
&\quad - 2\delta_n' \sum_{t=1}^n (x_t(\gamma) - x_t(\gamma_0))u \\
&\quad + 2\widehat{\delta}' \sum_{t=1}^n (x_t(\gamma) - x_t(\gamma_0))(x_t(\gamma) - x_t(\gamma_0))'(\widehat{\theta} - \theta) \\
&\quad + 2\sum_{t=1}^n \left(\widehat{\delta} + \delta_n\right)' (x_t(\gamma) - x_t(\gamma_0))(x_t(\gamma) - x_t(\gamma_0))' \left(\widehat{\delta} - \delta_n\right) \\
&\quad + 2\left(\delta_n' - \widehat{\delta}'\right) \sum_{t=1}^n (x_t(\gamma) - x_t(\gamma_0))u \\
&\equiv R_1 - R_2 + R_3 + R_4 + R_5, \text{ say.}
\end{aligned} \tag{A.2}
$$

Next, we show that $\frac{R_1+R_2+R_3+R_4+R_5}{a_n(\gamma-\gamma_0)}$ converges to a positive random variable almost surely by studying the limiting behavior for each term specifically. First, note that

$$
\begin{aligned}
\frac{R_1}{a_n} &= \frac{1}{a_n} \sum_{t=1}^n \delta_n' (x_t(\gamma) - x_t(\gamma_0))(x_t(\gamma) - x_t(\gamma_0))' \delta_n \\
&= (F(\gamma) - F(\gamma_0))\delta_0' \int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds\,\delta_0 + o_p(1) \\
&= f(\gamma_0)(\gamma - \gamma_0)\delta_0' \left(\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds \right) \delta_0 + o_p(1).
\end{aligned}
$$

The last equation uses the first-order Taylor approximation of $F(\gamma)$ around γ_0.

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

Noting that $\overline{v}/a_n < |\gamma - \gamma_0| < B$ and $a_n = n^{1-2\tau}$ with $\tau < 1/2$, we have $\sqrt{\overline{v}} < \sqrt{a_n}\sqrt{(|\gamma - \gamma_0|)}$. Thus, there exists $k > 0$, such that

$$\frac{R_2}{a_n(\gamma - \gamma_0)} = \frac{2\delta_0' \frac{1}{n}\sum_{t=1}^{n}(x_t(\gamma) - x_t(\gamma_0))u}{\sqrt{a_n}(\gamma - \gamma_0)} = O_p\left(\frac{1}{\sqrt{a_n}\sqrt{(|\gamma - \gamma_0|)}}\right) \le k/\sqrt{\overline{v}}.$$

For any $B \to 0_+$, there exists $\overline{v} > 0$ and N, such that $k/\sqrt{\overline{v}} < f(\gamma_0)\delta_0' \int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds\delta_0$ with probability 1 and $\overline{v}/a_n < B$ when $n > N$. Therefore, for any $\gamma \in V_B(\overline{v})$, we have

$$\frac{R_1}{a_n(\gamma - \gamma_0)} - \frac{R_2}{a_n(\gamma - \gamma_0)} > 0. \tag{A.3}$$

Furthermore, from Lemma A.4, we have $n^{\tau+1/2}(\widehat{(\theta} - \theta) = O_p(\widehat{\gamma}_n - \gamma_0)$ and $n^{\tau+1/2}(\widehat{\delta}_n - \delta_n) = O_p(\widehat{\gamma}_n - \gamma_0)$, thus

$$\begin{aligned}\frac{R_3}{a_n(\gamma - \gamma_0)} &= \frac{2n^{\tau+1/2}\widehat{\delta}_n' n^{-2}\sum_{t=1}^{n}(x_t(\gamma) - x_t(\gamma_0))(x_t(\gamma) - x_t(\gamma_0))' n^{\tau+1/2}(\widehat{\theta} - \theta)}{(\gamma - \gamma_0)} \\ &= O_p(\widehat{\gamma}_n - \gamma_0) = o_p(1),\end{aligned} \tag{A.4}$$

$$\begin{aligned}&\frac{R_4}{a_n(\gamma - \gamma_0)} \\ &= \frac{2n^{\tau+1/2}\left(\widehat{\delta} + \delta_n\right)' n^{-2}\sum_{t=1}^{n}(x_t(\gamma) - x_t(\gamma_0))(x_t(\gamma) - x_t(\gamma_0))' n^{\tau+1/2}\left(\widehat{\delta} - \delta_n\right)}{(\gamma - \gamma_0)} \\ &= O_p(n^{\tau+1/2}\left(\widehat{\delta} - \delta_n\right) = O_p(\widehat{\gamma}_n - \gamma_0) = o_p(1),\end{aligned} \tag{A.5}$$

$$\begin{aligned}\frac{R_5}{a_n(\gamma - \gamma_0)} &= \frac{2n^{2\tau}\left(\delta_n' - \widehat{\delta}'\right) n^{-1}\sum_{t=1}^{n}(x_t(\gamma) - x_t(\gamma_0))u}{(\gamma - \gamma_0)} \\ &= O_p(n^{\tau+1/2}(\widehat{\delta} - \delta_n)n^{\tau-1/2}) = o_p(1).\end{aligned} \tag{A.6}$$

Combining the results of equations (A.3)–(A.6), we have $\frac{SSR_n^*(\gamma) - SSR_n^*(\gamma_0)}{a_n(\gamma - \gamma_0)} > 0$ with probability 1 for any $\gamma \in V_B(\overline{v})$ and $\gamma > \gamma^0$. Similarly, we can prove $SSR_n^*(\gamma) > SSR_n^*(\gamma^0)$ when $\gamma < \gamma^0$ and $\gamma \in V_B(\overline{v})$ with probability 1.

Next, we study the limiting distribution for the estimator $\widehat{\gamma}$. Given $-1/2 < \tau < 1/2$, $\widehat{\gamma}$ is a consistent estimator with convergence rate a_n, thus, we could focus on its asymptotic behavior in the neighborhood of the true thresholds. Let $\gamma = \gamma_0 + \frac{\omega}{a_n}$. By the definition of $\widehat{\gamma}_n$, we have

$$a_n(\widehat{\gamma}_n - \gamma_0) = \omega^* = \arg\min_{\omega \in (-\infty, \infty)}\left(SSR_n^*(\gamma_0 + \frac{\omega}{a_n}) - SSR_n^*(\gamma_0)\right).$$

By equation (A.2), we know $SSR_n^*(\gamma_0 + \frac{\omega}{a_n}) - SSR_n^*(\gamma_0) = R_1 - R_2 + R_3 + R_4 + R_5$, with γ replaced by $\gamma_0 + \frac{\omega}{a_n}$. Next, we turn to study the limiting behavior for each term.

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

We only provide the proof for the case where $\omega > 0$, as the proof for the case with $\omega \leq 0$ is analogous. Given $\omega > 0$, we have

$$\begin{aligned}
R_1 &= \delta_n' \sum_{t=1}^{n} \left(x_t\left(\gamma_0 + \frac{\omega}{a_n}\right) - x_t(\gamma_0) \right) \left(x_t\left(\gamma_0 + \frac{\omega}{a_n}\right) - x_t(\gamma_0) \right)' \delta_n \\
&= n^{1-2\tau} \delta_0' n^{-2} \sum_{t=1}^{n} \left(x_t\left(\gamma_0 + \frac{\omega}{a_n}\right) - x_t(\gamma_0) \right) \left(x_t\left(\gamma_0 + \frac{\omega}{a_n}\right) - x_t(\gamma_0) \right)' \delta_0 \\
&= n^{1-2\tau} \delta_0' \left(F\left(\gamma_0 + \frac{\omega}{a_n}\right) - F(\gamma_0) \right) \left(\int \mathbf{B}_v(s) \mathbf{B}_v'(s) ds \right) \delta_0 + o_p(1) \\
&\xrightarrow{p} f(\gamma_0) \omega \delta_0' \left(\int \mathbf{B}_v(s) \mathbf{B}_v'(s) ds \right) \delta_0. \qquad \textbf{(A.7)}
\end{aligned}$$

Note that the last equation uses the first-order Taylor expansion of $F(\gamma_0 + \frac{\omega}{a_n})$ around γ_0. For R_2, we have

$$\begin{aligned}
R_2 &= -2 \sum_{t=1}^{n} \delta_n' \left(x_t\left(\gamma_0 + \frac{\omega}{a_n}\right) - x_t(\gamma_0) \right) u_t \\
&= -2 n^{1/2-\tau} \delta_0' \frac{1}{n} \sum_{t=1}^{n} \left(x_t\left(\gamma_0 + \frac{\omega}{a_n}\right) - x_t(\gamma_0) \right) u_t \\
&\Rightarrow -2 \sqrt{a_n} \delta_0' \sigma_u \int \mathbf{B}_v(s) d \left(W\left(s, F\left(\gamma_0 + \frac{\omega}{a_n}\right)\right) - W(s, F(\gamma_0)) \right) \\
&= -2 \sqrt{a_n} \delta_0' \sigma_u \left(\mathbf{J}_1\left(F\left(\gamma_0 + \frac{\omega}{a_n}\right)\right) - \mathbf{J}_1(F(\gamma_0)) \right). \qquad \textbf{(A.8)}
\end{aligned}$$

Note that $\mathbf{J}_1(\mu) = \int \mathbf{B}_v(s) dW(s, \mu)$ is a mean-zero Gaussian process with an almost surely continuous sample path and with the covariance kernel $\mathbb{E}(\mathbf{J}_1(\mu_1)\mathbf{J}_1(\mu_2)) = (\mu_1 \wedge \mu_2) \int \mathbf{B}_v(s) \mathbf{B}_v'(s) ds$. Thus, $D(\omega) = \sqrt{a_n} \delta_0' \sigma_u \left(\mathbf{J}_1(F(\gamma_0 + \frac{\omega}{a_n})) - \mathbf{J}_1(F(\gamma_0)) \right)$ is also a Brownian motion with

$$\begin{aligned}
Var(D(\omega)) &= \sigma_u^2 a_n (F(\gamma_0 + \frac{\omega}{a_n}) - F(\gamma_0)) \delta_0' \int \mathbf{B}_v(s) \mathbf{B}_v'(s) ds \delta_0 \\
&= \sigma_u^2 f_0 \omega \int \mathbf{B}_v(s) \mathbf{B}_v'(s) ds + o_p(1).
\end{aligned}$$

By equations (A.4)–(A.6), we have $R_3 + R_4 + R_5 = \omega o_p(1) = o_p(1)$. Combining the convergence results of equations (A.7) and (A.8), we have

$$\begin{aligned}
SSR_n^*(\gamma) - SSR_n^*(\gamma_0) \Rightarrow\ & f_0 \omega \delta_0' \int \mathbf{B}_v(s) \mathbf{B}_v'(s) ds \delta_0 - 2D(\omega) \\
=\ & f_0 \omega \delta_0' \int \mathbf{B}_v(s) \mathbf{B}_v'(s) ds \delta_0 \\
& - 2\sigma_u \sqrt{f_0 \omega \delta_0' \int \mathbf{B}_v(s) \mathbf{B}_v'(s) ds \delta_0} \Lambda_1(\omega),
\end{aligned}$$

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

where $\Lambda_1(\omega)$ is a standard Brownian motion on $[0,\infty)$. Making the change of variables as follows

$$\omega = \frac{\sigma_u^2}{\delta_0'\left(\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds\right)\delta_0 f_0} r,$$

we have

$$SSR_n^*(\gamma) - SSR_n^*(\gamma_0) \Rightarrow 2\sigma_u^2\left(\frac{r}{2} - \Lambda_1(r)\right). \tag{A.9}$$

Using continuous mapping theorem, the asymptotic distribution of $\widehat{\gamma}$ can be expressed as

$$n^{1-2\tau}\lambda(\widehat{\gamma} - \gamma_0) = r^* \Rightarrow \arg\max_{r\in(0,\infty)} \left(\Lambda_1(r) - \frac{1}{2}|r|\right).$$

Lastly, we establish the limiting results for $\widehat{\gamma}_n$ when $\tau = 1/2$. By the definition of $\widehat{\gamma}_n$, we have $\widehat{\gamma}_n = \arg\min_{\gamma\in[\underline{\gamma},\overline{\gamma}]} SSR_n(\gamma) = \arg\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]}(SSR_n - SSR_n(\gamma))$, where SSR_n is defined as the sum of squared residuals by regressing y_t to x_t. After some standard algebra, we have $SSR_n - SSR_n(\gamma) = Y'(I-P_n)X(\gamma)(X'(\gamma)(I-P_n)X(\gamma))^{-1}X'(\gamma)(I-P_n)Y$, where $X(\gamma) = (x_1(\gamma), x_2(\gamma), \ldots, x_n(\gamma))'$ and $X = (x_1, x_2, \ldots, x_n)'$. $P_n = X(X'X)^{-1}X'$ is the projection matrix of X. Define $\Gamma_n(\gamma) = n^{-1}X'(\gamma)(I-P_n)Y$.

Given $\tau = 1/2$, we have $n\delta_n = \delta_0$. By Lemma A.3, it can be shown that

$$\Gamma_n(\gamma) \Rightarrow \sigma_u \int \mathbf{B}_v(s)d\left(W(s,F(\gamma)) - F(\gamma)W(s)\right) + \left(F(\gamma\wedge\gamma_0) - F(\gamma)F(\gamma_0)\right)\left(\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds\right)\delta_0 \equiv \Gamma_1(\gamma),$$

and $n^{-2}X'(\gamma)(I-P_n)X(\gamma) \Rightarrow F(\gamma)(1-F(\gamma))\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds$. By continuous mapping theorem, we can show that $(SSR_n - SSR_n(\gamma)) = \Gamma_n'(\gamma)(X'(\gamma)(I-P_n)X(\gamma)/n^2)^{-1}\Gamma_n(\gamma) \Rightarrow \frac{1}{F(\gamma)(1-F(\gamma))}\Gamma_1'(\gamma)\left(\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds\right)^{-1}\Gamma_1(\gamma)$. ■

Proof of Theorem 3.2. Note that

$$\begin{aligned} T_n(\gamma) &= \widehat{\delta}_n(\gamma)'(X'(\gamma)(I-P_n)X(\gamma))\widehat{\delta}_n(\gamma)/\widehat{\sigma}_u^2 \\ &= (I-P_n)Y'X(\gamma)(X'(\gamma)(I-P_n)X(\gamma))^{-1}X'(\gamma)(I-P_n)Y/\widehat{\sigma}_u^2 \\ &= \Gamma_n(\gamma)'\left(n^{-2}X(\gamma)'X(\gamma) - n^{-2}X(\gamma)'X(X'X)^{-1}XX(\gamma)\right)^{-1}\Gamma_n(\gamma)/\widehat{\sigma}_u^2. \end{aligned} \tag{A.10}$$

By Lemma A.3, we have $n^{-2}X(\gamma)'X(\gamma) - n^{-2}X(\gamma)'X(X'X)^{-1}XX(\gamma) \xrightarrow{p} \left(F(\gamma) - F(\gamma)^2\right)\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds$.

Next, we consider the limiting result for $\Gamma_n(\gamma)$ for different τ. When $-1/2 < \tau < 1/2$, we have

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

$$\begin{aligned}\Gamma_n(\gamma) &= \frac{1}{n}X'(\gamma)u - \frac{1}{n}X'(\gamma)X(X'X)^{-1}X'u + \frac{1}{n}X'(\gamma)X(\gamma_0)\delta_n \\ &\quad - \frac{1}{n}X'(\gamma)X(X'X)^{-1}X'X(\gamma_0)\delta_n \\ &\Rightarrow \sigma_u \int_0^1 \mathbf{B}_v(s)d\left(W(s,F(\gamma)) - F(\gamma)W(s)\right) + n^{1/2-\tau}\left(F(\gamma\wedge\gamma_0)\right. \\ &\quad \left. - F(\gamma)F(\gamma_0)\right)\left(\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds\right)\delta_0 \\ &= O_p(n^{1/2-\tau}) \xrightarrow{p} \infty.\end{aligned}$$

It follows that $T_n(\gamma) = O_p(n^{1-2\tau}) \xrightarrow{p} \infty$.

When $\tau = 1/2$, we have

$$\begin{aligned}\Gamma_n(\gamma) &= \frac{1}{n}X'(\gamma)u - \frac{1}{n}X'(\gamma)X(X'X)^{-1}X'u + \frac{1}{n}X'(\gamma)X(\gamma_0)\delta_n \\ &\quad - \frac{1}{n}X'(\gamma)X(X'X)^{-1}X'X(\gamma_0)\delta_n \\ &\Rightarrow \Gamma(\gamma) + \left(F(\gamma\wedge\gamma_0) - F(\gamma)F(\gamma_0)\right)\left(\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds\right)\delta_0 = \Gamma_1(\gamma).\end{aligned}$$

It follows that $T_n(\gamma) \Rightarrow \frac{1}{\sigma_u^2 F(\gamma)(1-F(\gamma))}\Gamma_1'(\gamma)\left(\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds\right)^{-1}\Gamma_1(\gamma)$. ■

Proof of Theorem 3.3. We prove the equality by showing the following two inequalities: $AsySZ_\gamma(a) \geq 1-a$ and $AsySZ_\gamma(a) \leq 1-a$ hold simultaneously. We first consider the proof of $AsySZ_\gamma(a) \geq 1-a$.

By the definition of $AsySZ_\gamma(a)$ as equation (18), we can find a parameter sequence (δ_n, γ_n) such that $AsySZ_\gamma(a) = \liminf_{n\to\infty} \Pr_{(\delta_n,\,\gamma_n)}(\gamma_n \in CI_{\gamma,n}(a))$. Let $\{b_n\}$ be a subsequence of $\{n\}$ such that $AsySZ_\gamma(a) = \lim_{n\to\infty} \Pr_{(\delta_{b_n},\,\gamma_{b_n})}(\gamma_{b_n} \in CI_{\gamma,b_n}(a))$. Denote $\delta_0^* = n\delta_n$ where $\delta_0^* \in R \cup \{-\infty, \infty\}$. Because the euclidean space is complete, we can always find a subsequence $\{c_n\}$ of $\{b_n\}$ such that $\left(c_n\delta_{c_n}, \gamma_{c_n}\right) \to (\delta_0^*, \gamma_0)$.

If $\tau = 1/2$, we have $\delta_0^* = nn^{-1/2-\tau}\delta_0 = \delta_0 \in R$. By Theorem 3.2, we have $T_n = O_p(n^{1-2\tau}) = O_p(1) < \kappa_n$ with probability one. Thus, $CI_{\gamma,n}(a) = CI^W_{\gamma,n}(a)$, and $AsySZ_\gamma(a) = \lim_{n\to\infty} \Pr_{(\delta_{c_n},\gamma_{c_n})}(\gamma_{c_n} \in CI_{\gamma,c_n}(a)) = \lim_{n\to\infty} \Pr_{(\delta_{c_n},\gamma_{c_n})}(|\widehat{\gamma}_{c_n} - \gamma_{c_n}| \leq \widehat{q}^W_{\gamma,1-a}) \geq \lim_{n\to\infty} \Pr_{(\delta_{c_n},\gamma_{c_n})}(|\widehat{\gamma}_{c_n} - \gamma_{c_n}| \leq \widehat{q}^W_{\gamma,1-a}(\gamma_0,\delta_0)) = 1-a$. The inequality uses the fact that $\widehat{q}^W_{\gamma,1-a} = \sup_{\gamma\in[\underline{\gamma},\overline{\gamma}]}\sup_{\delta\in R} q^W_{\gamma,1-a}(\gamma,\delta)$. The last equation uses the fact that $|\widehat{\gamma}_{c_n} - \gamma_{c_n}|$ converges to $|\widehat{\gamma}(\gamma_0,\delta_0) - \gamma_0|$ and $\widehat{q}^W_{\gamma,1-a}(\gamma_0,\delta_0)$ is defined as the $(1-a)$ quantile of the limiting distribution of $|\widehat{\gamma}(\gamma_0,\delta_0) - \gamma_0|$.

If $-1/2 < \tau < 1/2$, $\delta_0^* = nn^{-1/2-\tau}\delta_0 = n^{1/2-\tau}\delta_0 = \pm\infty$. Note that $n^{-\nu}\kappa_n^{1/2} \to 0$ for any $\nu > 0$, we have $T_n = O_p(n^{1-2\tau}) > \kappa_n$ with probability approaching one. Thus, $CI_{\gamma,c_n}(a) = CI^I_{\gamma,c_n}(a)$ and $AsySZ_\gamma(a) = \lim_{n\to\infty} \Pr_{(\delta_{c_n},\gamma_{c_n})}(\gamma_{c_n} \in CI_{\gamma,c_n}(a)) = \lim_{n\to\infty} \Pr(\gamma_{c_n} \in CI^I_{\gamma,c_n}(a)) = \lim_{n\to\infty} \Pr(LR(\gamma_{c_n}) \leq q^I_{1-\alpha})$. As $LR_n(\gamma_n) \Rightarrow \sup_{r\in(-\infty,\infty)}(2\Lambda(r) - |r|)$, $LR_{c_n}(\gamma_{c_n})$ converges to $\sup_{r\in(-\infty,\infty)}(2\Lambda(r) - |r|)$ and $q^I_{1-\alpha}$ is the $1-a$ quantile of $\sup_{r\in(-\infty,\infty)}(2\Lambda(r) - |r|)$. Thus, $AsySZ_\gamma(a) = \lim_{n\to\infty} \Pr(LR(\gamma_{c_n}) \leq q^I_{1-\alpha}) \geq 1-a$.

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

Next, we consider the other side $AsySZ_\gamma(a) \le 1-a$. Let $\delta_n = n^{-1/2}\delta_0$ and $\gamma_n = \gamma_0$ with $\delta_0 \in R/\{0\}$. By definition, we have $AsySZ_\gamma(a) \le \liminf_{n\to\infty} \Pr_{(\delta_0,\gamma_0)}(\gamma_0 \in CI_{\gamma,n}(a))$. When $\delta_n = n^{-1/2}\delta_0$, we have $T_n = O_p(n^{1-2\tau}) = O_p(n) > \kappa_n$ with probability approaching one. Thus, $CI_{\gamma,n}(a) = CI^I_{\gamma,n}(a)$, and $\liminf_{n\to\infty} \Pr_{(\delta_n,\gamma_0)}(\gamma_0 \in CI_{\gamma,n}(a)) = \liminf_{n\to\infty} \Pr(\gamma_0 \in CI^I_{\gamma,n}(a)) = \liminf_{n\to\infty} \Pr(LR_n(\gamma_0) \le q^I_{1-a}) = 1-a$ where the last equality holds because $LR_n(\gamma_0) \Rightarrow \sup_{r\in(-\infty,\infty)}(2\Lambda(r) - |r|)$. ■

Part B

Define the moment functionals for the stationary regressors z_t as

$$h(\gamma) = \mathbb{E}(z_t I(q_t \le \gamma)),\ h_1(\gamma) = \mathbb{E}(z_t I(q_{t-1} \le \gamma)),\ h_2(\gamma) = \mathbb{E}(z_{t-1} I(q_t \le \gamma)),$$
$$H = \mathbb{E}(z_t z_t'),\ H_1 = \mathbb{E}(z_t z_{t-1}').$$

LEMMA B.1. *Under Assumptions* (4.1), (4.2), (2.3), *and* (2.4), *for any* $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, *as* $n \to \infty$,

(a) $n^{-3/2}\sum_{t=1}^n z_t x_t'(\gamma) \xrightarrow{p} h(\gamma)\int \mathbf{B}_v'(s)ds$,

(b) $n^{-3/2}\sum_{t=1}^n z_t x_{t-1}'(\gamma) \xrightarrow{p} h_1(\gamma)\int \mathbf{B}_v'(s)ds$,

(c) $n^{-3/2}\sum_{t=1}^n z_{t-1} x_t'(\gamma) \xrightarrow{p} h_2(\gamma)\int' \mathbf{B}_v'(s)ds$,

(d) $n^{-1}\sum_{t=1}^n z_t z_t' \xrightarrow{p} H$,

(e) $n^{-1}\sum_{t=1}^n z_t z_{t-1}' \xrightarrow{p} H_1$.

LEMMA B.2. *Under Assumptions* (4.1), (4.2), (2.3), *and* (2.4), *for any* $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ *and* $\mu = F(\gamma)$ *as* $n \to \infty$,

(a) $n^{-1/2}\sum_{t=1}^n I(q_t \le \gamma)\varepsilon_t \Rightarrow \sigma W(s,\mu)$,

(b) $n^{-1/2}\sum_{t=1}^n I(q_{t-1} \le \gamma)\varepsilon_t \Rightarrow \sigma W_1(s,\mu)$,

(c) $n^{-1}\sum_{t=1}^n x_t(\gamma)'\varepsilon_t \Rightarrow \sigma\int \mathbf{B}_v'(s)dW(s,\mu)$,

(d) $n^{-1}\sum_{t=1}^n x_{t-1}(\gamma)'\varepsilon_t \Rightarrow \sigma\int \mathbf{B}_v'(s)dW_1(s,\mu)$,

(e) $n^{-1/2}\sum_{t=1}^n z_t\varepsilon_t \Rightarrow \sigma J_2$,

(f) $n^{-1/2}\sum_{t=1}^n z_{t-1}\varepsilon_t \Rightarrow \sigma J_3$,

where J_2 *and* J_3 *are Gaussian random variable with mean zero and variance* H. $W(s,\mu)$ *and* $W_1(s,\mu)$ *are two two-parameter Brownian motions.*

Proof of Lemmas B.1 and B.2. The proofs are standard and are skipped for space considerations. ■

LEMMA B.3. *If* $-1/2 < \tau < 1/2$, *under Assumptions* (4.1), (4.2), (2.3), *and* (2.4), *we have* $\widehat{\gamma}_n \xrightarrow{p} \gamma_0$ *for any* $\rho \in (-1, 1]$.

Proof. We conduct the proof by considering two cases according to the value of ρ. When $\rho < 1$, $u_t = \beta' z_t + \eta_t$ is stationary process. The proof is similar to Lemma A.5

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

and is skipped for brevity. When $\rho = 1$, $u_t = \beta' z_t + \eta_t$ is nonstationary since η_t is a unit root process. Using a similar argument of Lemma A.5, we only need to show that $n^{-1+2\tau}(SSR_n(\gamma) - SSR_n(\gamma_0))$ uniformly converges to a function $b(\gamma)$ which is positive when $\gamma \in B(\epsilon)$, where $B(\epsilon) = \{\gamma : |\gamma - \gamma_0| > \varepsilon\}$ for any $\varepsilon > 0$. Note that equation (A.1) still holds, i.e., $n^{-1+2\tau}(SSR_n(\gamma) - SSR_n(\gamma_0)) = S_1 + S_2 + S_3$. Next, we will show that both S_2 and S_3 converge to zero but S_1 uniformly converges a function $b(\gamma)$ which is positive when $\gamma \in B(\epsilon)$.

First, note that the stationary component $\beta' z_t$ of u_t is asymptotically negligible, implying that $S_3 = n^{-1+2\tau}(u'(I - P^*_\gamma)u - u'(I - P^*_{\gamma_0})u) = n^{-1+2\tau}(\eta' P^*_\gamma \eta - \eta' P^*_{\gamma_0}\eta) + o_p(1)$. Denote $B_\eta(s)$ as a Brownian motion such that $\frac{1}{\sqrt{n}}\eta_{[ns]} \xrightarrow{a.s.} B_\eta(s)$. By Lemma B.1, we have

$$n^{-2}\eta' X^*(\gamma)(X^*(\gamma)'X^*(\gamma))^{-1}X^*(\gamma)'\eta \xrightarrow{a.s.} \left(\int \mathbf{B}_\upsilon(s)B_\eta(s)ds\right)'\left(\int \mathbf{B}_\upsilon(s)\mathbf{B}'_\upsilon(s)ds\right)^{-1}\left(\int \mathbf{B}_\upsilon(s)B_\eta(s)ds\right),$$

which is unrelated to γ. Thus, $\eta' P^*_\gamma \eta - \eta' P^*_{\gamma_0}\eta = o_p(1)$ and it follows that $S_3 = n^{-1+2\tau}(\eta' P^*_\gamma \eta - \eta' P^*_{\gamma_0}\eta) + o_p(1) = o_p(1)$.

To prove S_2 converge to zero almost surely, we first consider the case where $\gamma > \gamma_0$. For the case where $\gamma \le \gamma_0$, the proof is similar and not repeated here. Given $\gamma > \gamma_0$, we have

$$n^{-2}X'(\gamma)X(\gamma_0) = n^{-2}X'(\gamma_0)X(\gamma_0) \xrightarrow{p} F(\gamma_0)\int \mathbf{B}_\upsilon(s)\mathbf{B}'_\upsilon(s)ds,$$

$$n^{-2}X(\gamma_0)'(X - X(\gamma)) = 0,$$

$$n^{-2}(X - X(\gamma_0))'X(\gamma) \xrightarrow{p} (F(\gamma) - F(\gamma_0))\int \mathbf{B}_\upsilon(s)\mathbf{B}'_\upsilon(s)ds.$$

It follows that

$$n^{-2}X^*(\gamma_0)'X^*(\gamma) \xrightarrow{p} \begin{pmatrix} F(\gamma_0)\int \mathbf{B}_\upsilon(s)\mathbf{B}'_\upsilon(s)ds, & 0 \\ (F(\gamma) - F(\gamma_0))\int \mathbf{B}_\upsilon(s)\mathbf{B}'_\upsilon(s)ds, & (1 - F(\gamma))\int \mathbf{B}_\upsilon(s)\mathbf{B}'_\upsilon(s)ds \end{pmatrix}.$$

Using the above convergence results, it can be shown that

$$n^{-1+2\tau}\delta'_n X(\gamma_0)'\eta = n^{1/2+\tau}\delta'_0 n^{-2} \times \sum_{t=1}^{n} x_t(\gamma_0)\eta \xrightarrow{p} n^{1/2+\tau}\delta'_0 \begin{pmatrix} F(\gamma_0)\int \mathbf{B}_\upsilon(s)B_\eta(s)ds \\ (1 - F(\gamma_0))\int \mathbf{B}_\upsilon(s)B_\eta(s)ds \end{pmatrix},$$

and

$$n^{-1+2\tau}\delta'_n X(\gamma_0)'P^*_\gamma \eta \xrightarrow{p} n^{1/2+\tau}\delta'_0 \begin{pmatrix} F(\gamma_0)\int \mathbf{B}_\upsilon(s)B_\eta(s)ds \\ (1 - F(\gamma_0))\int \mathbf{B}_\upsilon(s)B_\eta(s)ds \end{pmatrix}.$$

Thus, $S_2 = 2n^{-1+2\tau}\delta'_n X(\gamma_0)'(I - P^*_\gamma)u = 2n^{-1+2\tau}\delta'_n X(\gamma_0)'\eta - 2n^{-1+2\tau}\delta'_n X(\gamma_0)'P^*_\gamma \eta + o_p(1) = o_p(1)$.

Moreover, using a similar argument in Lemma A.5, we can show $S_1 \xrightarrow{p} b(\gamma)$ uniformly and $b(\gamma)$ is strictly positive when $\gamma \in B(\epsilon)$. Combining the convergence results for S_1, S_2, and S_3, we complete the proof. ■

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

Proof of Theorem 4.1. Under Assumption (4.1), the model can be rewritten as $y_t = \alpha' x_t + \delta_n' x_t(\gamma_0) + u_t = \theta_1' x_t^*(\widehat{\gamma}) + \eta_t^*$, where $x_t^*(\widehat{\gamma}) = (x_t(\widehat{\gamma})', x_t' - x_t'(\widehat{\gamma}), z_t)'$, $\theta_1 = (\alpha' + \delta_n', \alpha', \beta')'$, and $\eta_t^* = \eta_t + \theta_1'(x_t^*(\gamma_0) - x_t^*(\widehat{\gamma}))$. Since $\theta_1'(x_t^*(\gamma_0) - x_t^*(\widehat{\gamma})) = \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma}))$, we have $\eta_t^* = \eta_t + \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma}))$. The LS residual $\eta_t(\widehat{\gamma})$ can be expressed as $\widehat{\eta}_t(\widehat{\gamma}) = y_t - \widehat{\theta}_1(\widehat{\gamma})' x_t^*(\widehat{\gamma}) = \eta_t + \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})) + (\widehat{\theta}_1(\widehat{\gamma}) - \theta_1)' x_t^*(\widehat{\gamma})$. Next, we conduct the proof by considering two cases according to the value of τ.

Case 1: $-1/2 < \tau < 1/2$, by Lemma B.3, we have $\widehat{\gamma} \xrightarrow{p} \gamma_0$ and $\delta_n' = O_p(n^{-1/2-\tau}) = o_p(1)$. Thus $\delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})) = o_p(1)$. If $\rho = 1$, η_t is a unit root; thus, $\widehat{\eta}_t(\widehat{\gamma}) = \eta_t + \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})) + (\widehat{\theta}_1(\widehat{\gamma}) - \theta_1)' x_t^*(\widehat{\gamma})$ is also a unit root process and it can be shown that $\widehat{\rho} = 1 + O_p(\frac{1}{n})$. If $\rho < 1$, η_t is stationary. Since $\widehat{\gamma} \xrightarrow{p} \gamma_0$, it can be shown that $D_n(\widehat{\theta}_1(\widehat{\gamma}) - \theta_1) = O_p(n^{-1/2})$, where $D_n = diag\{n^{1/2} I_{d_1}, n^{1/2} I_{d_1}, I_{d_2}\}$ is a weighting matrix. It follows that $\widehat{\eta}_t(\widehat{\gamma}) = \eta_t + \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})) + (\widehat{\theta}_1(\widehat{\gamma}) - \theta_1)' x_t^*(\widehat{\gamma}) = \eta_t + o_p(1) + (\widehat{\theta}_1(\widehat{\gamma}) - \theta_1)' D_n D_n^{-1} x_t^*(\widehat{\gamma})' \xrightarrow{p} \eta_t$. As η_t is assumed to be $\eta_t = \rho \eta_{t-1} + \varepsilon_t$, it can be shown that $\widehat{\rho} = \rho + O_p(\frac{1}{\sqrt{n}})$.

Case 2: $\tau = 1/2$, $\widehat{\gamma}$ is inconsistent. Noting that $n\delta_n = \delta_0$, we have $\delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})) = O_p(\frac{1}{\sqrt{n}}) = o_p(1)$. If $\rho = 1$, the term $\widehat{\eta}_t(\widehat{\gamma}) = \eta_t + \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})) + (\widehat{\theta}_1(\widehat{\gamma}) - \theta_1)' x_t^*(\widehat{\gamma})$ is an I(1) process and it can be shown that $\widehat{\rho}(\widehat{\gamma}) = 1 + O_p(\frac{1}{n})$. If $\rho < 1$, we can show that $D_n(\widehat{\theta}_1(\widehat{\gamma}) - \theta_1) = O_p(n^{-1/2})$. Thus, the following convergence still holds: $\widehat{\eta}_t(\widehat{\gamma}) = \eta_t + \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})) + (\widehat{\theta}_1(\widehat{\gamma}) - \theta_1)' x_t^*(\widehat{\gamma}) \xrightarrow{p} \eta_t$, implying that $\widehat{\rho} = \rho + O_p(\frac{1}{\sqrt{n}})$. ■

LEMMA B.4. *For any $\rho \in (-1, 1]$, there exists a nonrandom weighting matrix $\widetilde{D}_n$ such that*

(i) $n^{-1} \widetilde{D}_n^{-1} \widetilde{A}_1(\gamma)' \widetilde{A}_1(\gamma) \widetilde{D}_n^{-1} \xrightarrow{p} \widetilde{G}(\gamma)$,

(ii) $n^{-1/2} \widetilde{D}_n^{-1} \widetilde{A}_1(\gamma)' \widetilde{\eta} \Rightarrow \widetilde{\phi}(\gamma)$,

where $\widetilde{G}(\gamma)$ and $\widetilde{\phi}(\gamma)$ are defined below.

Proof. When $\rho < 1$, define the weighting matrix $\widetilde{D}_n = diag\{n^{1/2} I_{d_1}, n^{1/2} I_{d_1}, I_{d_2}\}$,

$$G(\gamma) = \begin{pmatrix} \int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds & F(\gamma)\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds & 0 \\ F(\gamma)\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds & F(\gamma)\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds & \left(\int \mathbf{B}_v(s)ds\right) h(\gamma)' \\ 0 & h(\gamma)\left(\int \mathbf{B}_v'(s)ds\right) & H \end{pmatrix},$$

$$G_1(\gamma) = \begin{pmatrix} \int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds & F(\gamma)\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds & 0 \\ F(\gamma)\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds & F_1(\gamma,\gamma)\int \mathbf{B}_v(s)\mathbf{B}_v'(s)ds & \left(\int \mathbf{B}_v(s)ds\right) h_2'(\gamma) \\ 0 & h_1(\gamma)\left(\int \mathbf{B}_v'(s)ds\right) & H_1 \end{pmatrix},$$

and

$$\widetilde{\phi}(\gamma) = \begin{pmatrix} \sigma \int \mathbf{B}_v(s) d\widetilde{W}(s) \\ \sigma \int \mathbf{B}_v(s) d\widetilde{W}(s, F(\gamma)) \\ \sigma \widetilde{J} \end{pmatrix},$$

where $\widetilde{W}(s) = (1-\rho)W(s)$, $\widetilde{W}(s, F(\gamma)) = W(s, F(\gamma)) - \rho W_1(s, F(\gamma))$, and $\widetilde{J} = J_2 - \rho J_3$.

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press

By Lemmas B.1 and B.2 and Theorem 4.1, it can be shown that

$$\begin{aligned} n^{-1}\sum_{t=1}^{n}\widetilde{D}_n^{-1}\widetilde{A}_{1t}(\gamma)\widetilde{A}_{1t}(\gamma)'\widetilde{D}_n^{-1} &= n^{-1}\sum_{t=1}^{n}\widetilde{D}_n^{-1}A_{1t}(\gamma)A_{1t}(\gamma)'\widetilde{D}_n^{-1} \\ &\quad + n^{-1}\widehat{\rho}^2\sum_{t=1}^{n}\widetilde{D}_n^{-1}A_{1t-1}(\gamma)A_{1t-1}(\gamma)'\widetilde{D}_n^{-1} \\ &\quad - \widehat{\rho}\sum_{t=1}^{n}\widetilde{D}_n^{-1}A_{1t}(\gamma)A_{1t-1}(\gamma)'\widetilde{D}_n^{-1} \\ &\quad - \widehat{\rho}\sum_{t=1}^{n}\widetilde{D}_n^{-1}A_{1t-1}(\gamma)A_{1t}(\gamma)'\widetilde{D}_n^{-1} \\ &\xrightarrow{p} G(\gamma)+\rho^2 G(\gamma)-\rho\left(G_1(\gamma)+G_1'(\gamma)\right)\equiv\widetilde{G}(\gamma). \end{aligned}$$

By Lemmas B.1 and B.2 and Theorem 4.1, we can also show that

$$n^{-1/2}\sum_{t=1}^{n}\widetilde{D}_n^{-1}\widetilde{A}_{1t}(\gamma)\widetilde{\eta}_t = \begin{pmatrix} n^{-1}\sum_{t=1}^{n}\widetilde{x}_t\widetilde{\eta}_t \\ n^{-1}\sum_{t=1}^{n}D_n^{-1}\widetilde{x}_t(\gamma)\widetilde{\eta}_t \\ n^{-1/2}\sum_{t=1}^{n}\widetilde{z}_t\widetilde{\eta}_t \end{pmatrix} \Rightarrow \widetilde{\phi}(\gamma).$$

When $\rho = 1$, using a similar argument for the case of $\rho < 1$, we can find two random matrices $\widetilde{G}(\gamma)$ and $\widetilde{\phi}(\gamma)$ and establish the convergence results. ■

Proof of Theorem 4.2. The proof is similar to that of Theorem 3.1 and is skipped for brevity. ■

Proof of Theorem 4.3. The proof is similar to that of Theorem 3.2 and is skipped for brevity. ■

https://doi.org/10.1017/S0266466614000553 Published online by Cambridge University Press