JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. (2020)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12574

ORIGINAL ARTICLE

THRESHOLD MODEL WITH A TIME-VARYING THRESHOLD BASED ON FOURIER APPROXIMATION

LIXIONG YANG,[a*] CHINGNUN LEE[b] AND I-PO CHEN[b]

[a] *School of Management, Lanzhou University, Lanzhou, China*
[b] *Institute of Economics, National Sun Yat-sen University, Kaohsiung, Taiwan*

Classical threshold models assume that threshold values are constant and stable, which appears overly restrictive and unrealistic. In this article, we extend Hansen's (2000) constant threshold regression model by allowing for a time-varying threshold which is approximated by a Fourier function. Least-square estimation of regression slopes and the time-varying threshold is proposed, and test statistics for the existence of threshold effect and threshold constancy are constructed. We also develop the asymptotic distribution theory for the time-varying threshold estimator. Through Monte Carlo simulations, we show that the proposed estimation and testing procedures work reasonably well in finite samples, and there is little efficiency loss by the allowance for Fourier approximation in the estimation procedure even when there is no time-varying feature in the threshold. On the contrary, the slope estimates are seriously biased when the threshold is time-varying but being treated as a constant. The model is illustrated with an empirical application to a nonlinear Taylor rule for the United States.

Received 19 June 2019; Revised 18 August 2020; Accepted 23 November 2020

Keywords: Threshold model; time-varying threshold; Fourier approximation; estimation; testing.

MOS subject classification: 62F03, 62F12, 62M10.

JEL. C13; C51; C52.

1. INTRODUCTION

Threshold models have been widely applied in economics. The classical threshold models, which specify that individual observations can be divided into subclasses/subsamples based on the value of an observed variable, assume that threshold values are constant. Under this assumption, the econometric theory of threshold models has been fully investigated by the literature such as Hansen (1996, 2000). However, varying-coefficient models have attracted considerable attention in the past two decades, which is particularly true for both cross-sectional and time series varying-coefficient models (Feng *et al.*, 2017).

Clearly, it may be restrictive to assume that threshold values are time-invariant during the whole sampling period. In fact, economic models are unlikely to have constant parameters over time and threshold models with time-dependent thresholds are called for. The intuition of the time-varying threshold is that 'usually high/small values of an economic variable may sometimes be best thought of in relative terms' (Dueker *et al.*, 2013). For example, public debt may be considered low or high not in absolute terms but relative to relevant macroeconomic variables that shape the state of the economy (Yang and Su, 2018). Furthermore, a constant threshold setting is questioned by the literature in other contexts. For instance, in the target-zone literature, Bessec (2003) demonstrates that a constant threshold may be inappropriate in modeling a band of inaction, because the monetary authorities may modify the band at various times, resulting into a better definition of the band of inaction in the time-varying threshold model; in the classical return-to-schooling literature it is reasonable to believe that the threshold levels of education for men and women should be different (Yu and Fan, 2020), leading to a threshold depending on a dummy variable.

*Correspondence to: Lixiong Yang, School of Management, Lanzhou University, Lanzhou 730000, China. Email: ylx@lzu.edu.cn

Time-dependence or instability in the threshold so far in the literature may be implemented through either a change-point mechanism or a function of candidate variables. Bessec (2003) proposes a Self-Exciting Threshold Autoregressive (SETAR) model with time-varying thresholds using ‘a one-time break where the change point is not estimated but corresponds to the time of the change in the official margins’, and finds that the data cannot reject the time-varying threshold specification. Yu and Fan (2020) propose a threshold regression with a threshold boundary, in which the threshold is modeled as a function of a dummy variable or other observable variables. Dueker *et al.* (2013) present a model in which the threshold varies as a function of some observable variables, and demonstrate that models with constant thresholds have been outperformed by models with a time-varying threshold both in terms of in-sample fit and out-sample forecast accuracy.

A drawback in Bessec’s (2003) approach is that the dates and number of breaks are usually unknown in most empirical analysis. It is now well-known that it is difficult to precisely estimate the number and magnitudes of multiple breaks especially when the breaks are of opposite sign (Prodan, 2008). Meanwhile, the approach proposed by Dueker *et al.* (2013) and Yu and Fan (2020) requires that the variables affecting threshold are given or predetermined, but, in most applications, it is difficult to explore the factors which affect the threshold value in advance.

To complement the literature, we propose threshold models with time-varying thresholds which are approximated by a Fourier function. As pointed out by Enders and Lee (2012a,b), the Fourier form approximation has several advantages. First, it works reasonably well for types of time-varying features often observed in economic analysis. Second, the Fourier function with a single-frequency component can be a reasonable approximation for a time-varying threshold of an unknown form even if the time-varying threshold itself is not periodic. Third, the Fourier approximation involves only the choice of the appropriate component in the Fourier function and hence avoids the complication of selecting break dates, the number of breaks and the form of breaks. Therefore, an important advantage of our model is its simplicity in empirical applications.

The remainder of this article is organized as follows: Section 2 introduces the threshold model with a time-varying threshold, describes least-square estimation of the model parameters, and constructs tests for the threshold effect and threshold constancy. In this section, we also establish the asymptotic theory for the estimator of the time-varying threshold. Sections 3 presents Monte Carlo experiments that assess the finite-sample properties of the estimation procedure and the test statistics. Section 4 provides an application and Section 5 concludes. Throughout this article, the notation $\|\cdot\|$ stands for the Euclidean norm, that is, $\|\mathbf{A}\| = \sqrt{\mathrm{tr}(\mathbf{A}'\mathbf{A})}$ for a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\det(\mathbf{A})$ and $\mathbf{A}'$ denote the determinant of a square matrix $\mathbf{A}$ and its transpose respectively.

2. THRESHOLD MODEL WITH A TIME-VARYING THRESHOLD

The model we consider in the article is given by

$$y_t = \begin{cases} \boldsymbol{\beta}_1'\mathbf{x}_t + e_t, \text{ if } q_t \le \gamma_t \\ \boldsymbol{\beta}_2'\mathbf{x}_t + e_t, \text{ if } q_t > \gamma_t \end{cases}, \; t = 1, 2, \dots, T, \tag{1}$$

where $\mathbf{x}_t$ is a m-dimensional vector of exogenous regressors, q_t is the threshold variable and is used to split the sample into two subgroups. The random variable e_t is the regression disturbance, and γ_t is the time-varying threshold.

If γ_t is an absolutely integrable function of time, then it can be approximated to any desired level of accuracy by the sufficiently long Fourier series:

$$\gamma_t = \gamma_0 + \sum_{k=1}^{n} \gamma_{1,k} \sin\left(\frac{2\pi kt}{T}\right) + \sum_{k=1}^{n} \gamma_{2,k} \cos\left(\frac{2\pi kt}{T}\right); n < \frac{T}{2}, \tag{2}$$

where n represents the number of frequencies, k represents a particular frequency, and T is the number of observations.

DOI: 10.1111/jtsa.12574

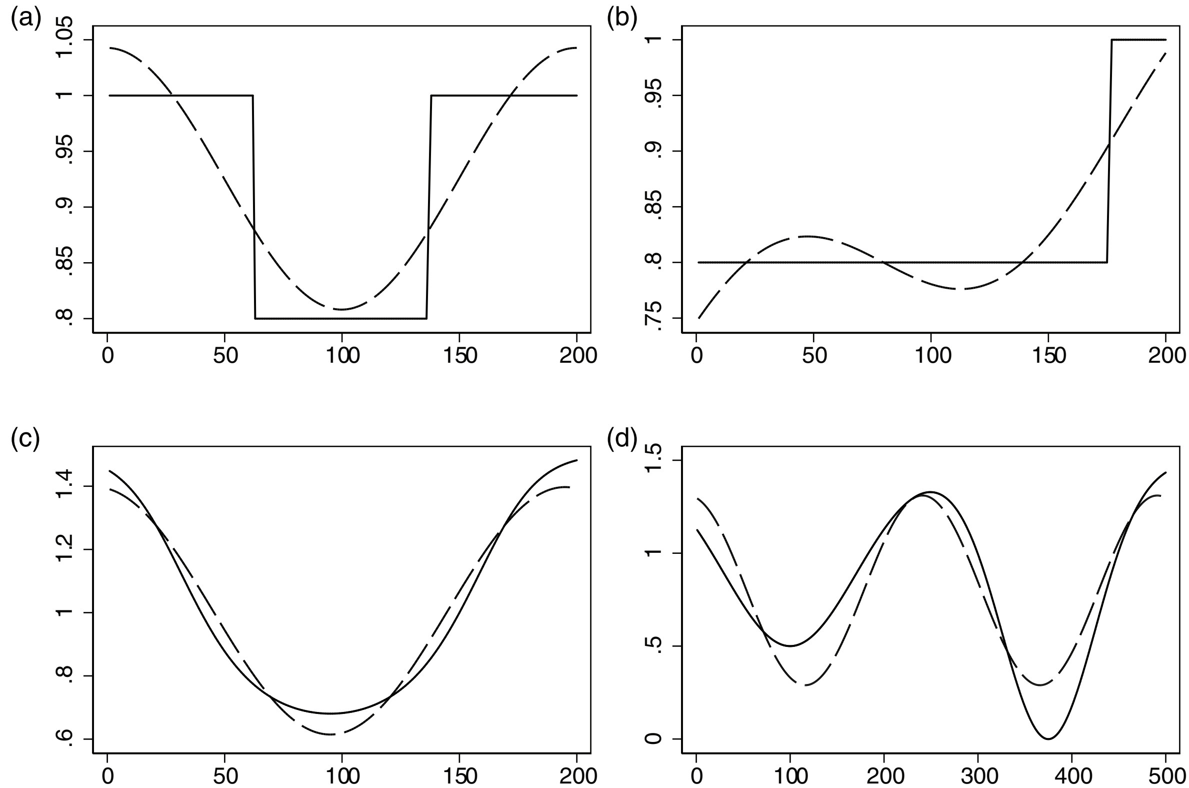

FIGURE 1. Fourier approximations. (a) $\gamma_t = 0.8$ if $62 \le t \le 137$, else $\gamma_t = 1$; (b) $\gamma_t = 0.8$ if $t \le 176$, else $\gamma_t = 1$; (c) $\gamma_t = -1/[1 + \exp(0.0005(t - 0.75T)(t - 0.2T))] + 1.5$; (d) $\gamma_t = 1.5[1 - \exp(-0.002(t - 0.75T)^2)) - \exp(-0.001(t - 0.2T)^2]$. The solid line denotes the series containing time-varying features, and the dashed line denotes the Fourier approximation for the time-varying series γ_t

As is illustrated by the literature (e.g. Becker *et al.*, 2006), beginning with $n = 1$, it is always possible to improve the approximation accuracy by using additional frequencies; the fit of γ_t will be perfect when $n = \frac{T}{2}$. However, there are several reasons why it is not suitable to use a large n as discussed by Enders and Lee (2012a). First, using many frequency components can lead to an over-fitting problem. Second, n should be small because it allows us to deal with a gradual change in threshold, as the higher frequencies may be associated with stochastic parameter variability (Becker *et al.*, 2006). Furthermore, Becker *et al.* (2006) and Enders and Lee (2012a) illustrate that the essential characteristics of a time series containing time-varying features can often be captured using a single-frequency component of a Fourier approximation, as can be seen in Figure 1. Therefore, in this article we focus on the following single-frequency approximation:

$$\gamma_t(\boldsymbol{\gamma}) = \gamma_0 + \gamma_1 \sin(2\pi kt/T) + \gamma_2 \cos(2\pi kt/T), \tag{3}$$

where $\boldsymbol{\gamma} = [\gamma_0, \gamma_1, \gamma_2, k]'$, γ_0 is a constant, γ_1 and γ_2 measure the amplitude and displacement of sinusoidal components and k represents the frequency selected for the approximation.[1] It is important to use the cumulative frequencies in the threshold if our main interest is to obtain a precise time-varying threshold parameter. The

[1] The conventional threshold model assumes that the threshold is constant. Instead, we treat the threshold as a time-varying parameter. Thus, our methods are appropriate when the threshold is not constant, or when the researcher wishes to investigate the robustness of this assumption.

DOI: 10.1111/jtsa.12574

proposed method can be easily extended to the case of cumulative frequencies, in which, instead of choosing an optimal value of frequencies, we can choose the number of cumulative frequencies n for $n \leq n_{max} = 5$.[2]

It is worthy noting that, although the proposed time-varying threshold model can be used to deal with the cases where the actual change is discrete, it is designed to work best when the actual change is gradual, because the Fourier approximation seems to be less effective when the actual change is discrete, as can be seen from Figure 1. The gradual threshold change might be expected to be seen in economic variables with a gradually evolving trend, which is appropriate in many applications. For example, the reference (threshold) for changing monetary policy may evolve gradually since such a reference often reflects a long-run equilibrium which tends to change gradually.

2.1. Estimation of Regression Slopes and Time-varying Threshold

For ease of manipulation, we express the model defined in (1) and (3) in a more compacted form. Define $\boldsymbol{\beta} = [\boldsymbol{\beta}_2', \boldsymbol{\beta}_1' - \boldsymbol{\beta}_2']'$, $\mathbf{x}_t(\gamma_t(\boldsymbol{\gamma})) = [\mathbf{x}_t', \mathbf{x}_t'\{q_t \leq \gamma_t(\boldsymbol{\gamma})\}]'$, where $\{.\}$ is the indicator function, then the model defined in (1) and (3) can be rewritten as:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t(\gamma_t(\boldsymbol{\gamma})) + e_t. \tag{4}$$

For any fixed $\gamma_t(\boldsymbol{\gamma})$, the regression slopes in (4) can be estimated by OLS, and we obtain

$$y_t = \hat{\boldsymbol{\beta}}'(\boldsymbol{\gamma})\mathbf{x}_t(\gamma_t(\boldsymbol{\gamma})) + \hat{e}_t(\gamma_t(\boldsymbol{\gamma})), \tag{5}$$

where $\hat{e}_t(\gamma_t(\boldsymbol{\gamma}))$ is the regression error, and $\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$ is given by

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = \left[\sum_{t=1}^{T}\mathbf{x}_t(\gamma_t(\boldsymbol{\gamma}))\mathbf{x}_t'(\gamma_t(\boldsymbol{\gamma}))\right]^{-1}\left[\sum_{t=1}^{T}\mathbf{x}_t(\gamma_t(\boldsymbol{\gamma}))y_t\right]. \tag{6}$$

The sum of squared errors is given by

$$SSR_T(\boldsymbol{\gamma}) = \sum_{t=1}^{T}\hat{e}_t^2(\gamma_t(\boldsymbol{\gamma})) = \sum_{t=1}^{T}(y_t - \hat{\boldsymbol{\beta}}'(\boldsymbol{\gamma})\mathbf{x}_t(\gamma_t(\boldsymbol{\gamma})))^2. \tag{7}$$

Since $\gamma_t(\boldsymbol{\gamma}) = \gamma_0 + \gamma_1\sin(2\pi kt/T) + \gamma_2\cos(2\pi kt/T)$, the estimation of the time-varying threshold is as follows. We define the profiled LS estimator of parameters in $\gamma_t(\boldsymbol{\gamma})$ as

$$(\hat{k}, \hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2) = \underset{((\gamma_0,\gamma_1,\gamma_2),k)\in\Gamma\times\Gamma_k}{\arg\min} SSR_T(\boldsymbol{\gamma}), \tag{8}$$

where $\Gamma = \Gamma_0 \times \Gamma_1 \times \Gamma_2$, in which Γ_i is assumed to a bounded set such that $\Gamma_i = [\underline{\gamma}_i, \overline{\gamma}_i] \subset \mathbb{R}$ for $i = 0, 1, 2$, and $\Gamma_k = \{1, 2, 3, 4, 5\}$ which is following Becker *et al.* (2006). That is, we assume the integer frequency in this article.[3]

It is computationally convenient to use a combination of concentration and grid search, as is typical in the literature. For fixed $k \in \Gamma_k$, we first solve

$$(\hat{\gamma}_0(k), \hat{\gamma}_1(k), \hat{\gamma}_2(k)) = \underset{(\gamma_0,\gamma_1,\gamma_2)\in\Gamma}{\arg\min} SSR_T(\boldsymbol{\gamma}). \tag{9}$$

[2] We thank the anonymous referee to raise this point to us.
[3] In applications, as illustrated by Becker *et al.* (2006), a better approximation of breaks can be achieved by employing a fractional frequency.

DOI: 10.1111/jtsa.12574

Denote $\hat{\gamma}_t(k) = \hat{\gamma}_0(k) + \hat{\gamma}_1(k)\sin(2\pi kt/T) + \hat{\gamma}_2(k)\cos(2\pi kt/T)$ for $k \in \Gamma_k$, and $SSR_T^*(k) = \sum_{t=1}^{T}(y_t - \hat{\boldsymbol{\beta}}'(\hat{\gamma}_t(k))\mathbf{x}_t(\hat{\gamma}_t(k)))^2$. Accordingly, the optimum frequency value can be given as

$$\hat{k} = \arg\min_{k\in\Gamma_k} SSR_T^*(k) \tag{10}$$

After $\hat{k}$ is found, the estimates $\hat{\gamma}_0, \hat{\gamma}_1$, and $\hat{\gamma}_2$ are obtained by equation (9). Then the time-varying threshold is estimated as $\hat{\gamma}_t = \gamma_t(\hat{\boldsymbol{\gamma}}) = \hat{\gamma}_0 + \hat{\gamma}_1\sin(2\pi\hat{k}t/T) + \hat{\gamma}_2\cos(2\pi\hat{k}t/T)$.

2.2. Asymptotic Properties for the Estimates of Time-varying Threshold

Define the moment functionals $E(\mathbf{x}_t\mathbf{x}_t') = M$, $E(\mathbf{x}_t\mathbf{x}_t'e_t^2) = N$, $E(\mathbf{x}_t\mathbf{x}_t'\{q_t \leq \gamma_t\}) = M(\gamma_t)$, $E(\mathbf{x}_t\mathbf{x}_t'e_t^2\{q_t \leq \gamma_t\}) = N(\gamma_t)$, $E(\mathbf{x}_t\mathbf{x}_t'|q_t) = \mathbf{D}(q_t)$, and $E(\mathbf{x}_t\mathbf{x}_t'e_t^2|q_t) = \mathbf{V}(q_t)$. Note that this definition implies that $\mathbf{D}(q_t)$ and $\mathbf{V}(q_t)$ are function of q_t.

Let $f(q)$ be the density function of the threshold variable q_t, and denote $\gamma_\tau = \gamma_0 + \gamma_1\sin(2\pi k\tau) + \gamma_2\cos(2\pi k\tau)$, $\mathbf{D}_\tau = \mathbf{D}(\gamma_\tau^0)$, $\mathbf{V}_\tau = \mathbf{V}(\gamma_\tau^0)$ and $f_\tau = f(\gamma_\tau^0)$, where γ_τ^0 is the true value of γ_τ as $\gamma_\tau^0 = \gamma_0^0 + \gamma_1^0\sin(2\pi k^0\tau) + \gamma_2^0\cos(2\pi k^0\tau)$, in which $\gamma_0^0, \gamma_1^0, \gamma_2^0$, and k^0 are the true values of $\gamma_0, \gamma_1, \gamma_2$, and k respectively.

To obtain the asymptotic properties of the estimator of the time-varying threshold, $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2, \hat{k})$, we make the following assumptions.

Assumption 1. We assume that

1. $(\mathbf{x}_t, q_t, e_t)$ is strictly stationary, ergodic and ρ-mixing, with ρ-mixing coefficients satisfying $\sum_{m=1}^{\infty}\rho_m^{1/2} < \infty$.
2. There exists a filtration $\mathcal{F}_t, t = 1, \ldots, T$, such that $E(e_t|\mathcal{F}_{t-1}) = 0$.
3. $E\|\mathbf{x}_t\|^4 < \infty$ and $E\|\mathbf{x}_t e_t\|^4 < \infty$.
4. For all $\gamma_i \in \Gamma_i$ $(i = 0, 1, 2)$ and t, $E(\|\mathbf{x}_t\|^4|q_t = \gamma_t(\gamma_i)) \leq C$ and $E(\|\mathbf{x}_t e_t\|^4|q_t = \gamma_t(\gamma_i)) \leq C$ for some $C < \infty$, and $0 < f(\gamma_t(\gamma_i)) \leq \bar{f} < \infty$.
5. $f(\gamma_\tau(\gamma_i))$, $\mathbf{D}(\gamma_\tau(\gamma_i))$ and $\mathbf{V}(\gamma_\tau(\gamma_i))$ are continuous at $\gamma_i = \gamma_i^0$ for $i = 0, 1, 2$.
6. $\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2 = \boldsymbol{\delta}_T = \mathbf{c}T^{-\alpha}$ with $\mathbf{c} \neq \mathbf{0}$ and $0 < \alpha < 1/2$, where $\mathbf{c}$ is a $m \times 1$ vector of constant.
7. $\mathbf{c}'\mathbf{D}_\tau\mathbf{c} > 0$, $\mathbf{c}'\mathbf{V}_\tau\mathbf{c} > 0$, and $f_\tau > 0$.
8. $\det(\mathbf{M}) > \det(\mathbf{M}(\gamma_t(\gamma_i))) > 0$ and $\det(\mathbf{N}) > \det(\mathbf{N}(\gamma_t(\gamma_i))) > 0$ for all $\gamma_i \in \Gamma_i$ and t.

Assumptions 1.1–1.8 are natural extensions of Assumptions 1.1–1.8 in Hansen (2000), which are very conventional in the literature of threshold models. Similar assumptions are employed in Hansen (2000), Caner and Hansen (2001), Gonzalo and Pitarakis (2006, 2012) and Chen (2015). Assumption 1.1 supposes time series to be stationary as in Hansen (2000), and controls the degree of time series dependence by the $\rho-$mixing assumption. Assumption 1.2 imposes that the model defined in (1) is correctly specified in terms of the conditional mean. Assumptions 1.3 and 1.4 restrict unconditional and conditional moment bounds to be finite. Assumption 1.5 requires the threshold variable to have a continuous distribution. Assumption 1.6 is well-known as the small threshold effect assumption, which is very conventional in the literature of threshold models. Assumptions 1.7 and 1.8 are full rank conditions needed to have nondegenerate asymptotic distributions.

The following theorem establishes the consistency of the time-varying threshold estimator, $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2, \hat{k})$.

Theorem 1. Under Assumption 1, we have

$$\begin{aligned} &\hat{k} \xrightarrow{p} k^0, \text{ and} \\ &\hat{\gamma}_i \xrightarrow{p} \gamma_i^0, \quad \text{for } i = 0, 1, 2. \end{aligned} \tag{11}$$

Proof of Theorem 1. See the Appendix. ∎

DOI: 10.1111/jtsa.12574

Theorem 1 shows that the time-varying threshold estimators $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2, \hat{k})$ are consistent. Hence $\hat{\gamma}_t = \hat{\gamma}_0 + \hat{\gamma}_1 \sin(2\pi\hat{k}t/T) + \hat{\gamma}_2 \cos(2\pi\hat{k}t/T)$ is a consistent estimator of the true time-varying threshold $\gamma_t^0 = \gamma_0^0 + \gamma_1^0 \sin(2\pi k^0 t/T) + \gamma_2^0 \cos(2\pi k^0 t/T)$. The consistency of the threshold estimator implies that the observations can be correctly classified into subsets, and hence the true model parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ in (1) can be consistently estimated. On the contrary, if the true model is accompanied with a time-varying threshold, then any constant threshold estimate would lead to misclassification of observations, and thus the true model parameters cannot be retrieved from the estimation method based on the constant threshold model.

We next establish the convergence rate and asymptotic distribution of the time-varying threshold estimator $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2)$.[4]

Theorem 2. Under Assumption 1, we have

$$T^{1-2\alpha}(\hat{\gamma}_i - \gamma_i^0) \xrightarrow{d} \varpi_i \cdot \underset{-\infty<r<\infty}{\arg\max} \left[-\frac{1}{2}|r| + \mathbb{W}(r)\right], \quad \text{for } i = 0, 1, 2,$$

where $\mathbb{W}(r)$ is a two-sided Brownian motion that is defined in Hansen (2000, p. 580),[5] and $\varpi_i = \lambda_i/\mu_i^2$, in which

$$\lambda_0 = \mathbf{c}' \left(\int_0^1 \mathbf{V}_\tau f_\tau \mathrm{d}\tau\right) \mathbf{c}, \mu_0 = \mathbf{c}' \left(\int_0^1 \mathbf{D}_\tau f_\tau \mathrm{d}\tau\right) \mathbf{c},$$
$$\lambda_1 = \mathbf{c}' \left(\int_0^1 |\sin(2\pi k^0 \tau)| \mathbf{V}_\tau f_\tau \mathrm{d}\tau\right) \mathbf{c}, \mu_1 = \mathbf{c}' \left(\int_0^1 |\sin(2\pi k^0 \tau)| \mathbf{D}_\tau f_\tau \mathrm{d}\tau\right) \mathbf{c},$$
$$\lambda_2 = \mathbf{c}' \left(\int_0^1 |\cos(2\pi k^0 \tau)| \mathbf{V}_\tau f_\tau \mathrm{d}\tau\right) \mathbf{c}, \mu_2 = \mathbf{c}' \left(\int_0^1 |\cos(2\pi k^0 \tau)| \mathbf{D}_\tau f_\tau \mathrm{d}\tau\right) \mathbf{c}.$$

Proof of Theorem 2. See the Appendix. ■

Theorem 2 is a generalization to the asymptotic result of Hansen's (2000) constant threshold model.[6] Theorem 2 shows that the convergence rate of $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2)$ is $T^{1-2\alpha}$, which is decreasing in α. This convergence rate of $\hat{\gamma}_0$ is equal to the rate found for the constant threshold model in Hansen (2000). Hence, the inclusion of a time-varying threshold does not affect the convergence rate of the constant threshold parameter. Moreover, the limiting distribution of $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2)$ has the same form as that found for the constant threshold model in Hansen (2000), but the scale factors ϖ_i, for $i = 0, 1, 2$, are different. In our model, the scale factors depend on the Fourier components.

---

[4] It is noted that it is impossible to find the convergence rate and the asymptotic distribution of $\hat{k}$. From the consistency result in Theorem 1 and definition of convergence in probability, for $\forall \varepsilon > 0$ we have

$$P(|\hat{k} - k^0| < \varepsilon) \to 1.$$

Since we assume that the frequency is an integer, the distance between $\hat{k}$ and k^0 is also an integer (i.e. $|\hat{k} - k^0| \in \mathbb{Z}$). Based on the above observation, if we choose arbitrary small ε, say, $\varepsilon = 0.5$, then

$$P(|\hat{k} - k^0| < 0.5) \to 1; \text{ therefore we obtain } P(|\hat{k} - k^0| = 0) \to 1.$$

This result implies that even $|\hat{k} - k^0|$ is multiplied by some function of T, say v_T, $v_T|\hat{k} - k^0|$ still converges to zero in probability, that is $v_T|\hat{k} - k^0| = o_p(1)$. Hence, we cannot obtain the convergence rate and asymptotic distribution of $\hat{k}$.

[5] A two-sided Brownian motion $\mathbb{W}(r)$ on the real line is defined as $\mathbb{W}(r) = \begin{cases} W_1(-r), & r < 0 \\ 0, & r = 0 \\ W_2(r), & r > 0 \end{cases}$, where $W_1(r)$ and $W_2(r)$ are independent standard Brownian motions on $[0, \infty)$.

[6] In the constant threshold model, we have $\gamma_1^0 = \gamma_2^0 = 0$ and $\gamma_t^0 = \gamma_0^0$, that is, γ_t^0 does not depend on t in this case; therefore, γ_τ, $\mathbf{D}_\tau$, $\mathbf{V}_\tau$ and f_τ do not depend on τ. Denote $\mathbf{D} = \mathbf{D}(\gamma_0^0)$, $\mathbf{V} = \mathbf{V}(\gamma_0^0)$, and $f = f(\gamma_0^0)$, so we obtain $\lambda_0 = \mathbf{c}'(\mathbf{V}f \int_0^1 \mathrm{d}\tau)\mathbf{c} = \mathbf{c}'\mathbf{V}\mathbf{c}f$, $\mu_0 = \mathbf{c}'(\mathbf{D}f \int_0^1 \mathrm{d}\tau)\mathbf{c} = \mathbf{c}'\mathbf{D}\mathbf{c}f$, leading to the scale factor ϖ_0 becoming $\varpi_0 = \mathbf{c}'\mathbf{V}\mathbf{c}/[(\mathbf{c}'\mathbf{D}\mathbf{c})^2 f]$, which is the result of Hansen's (2000) Theorem 1.

DOI: 10.1111/jtsa.12574

2.3. Test Statistics

Testing for the existence of the threshold effect and threshold constancy is important in applications. The test hypothesis of no threshold effect in (1) can be represented as

$$H_0^1 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2, \text{ v.s. } H_1^1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2.$$

Under H_0^1, the model defined in (1) and (3) shrinks to a linear regression model:

$$y_t = \boldsymbol{\beta}_1' \mathbf{x}_t + e_t. \tag{12}$$

The threshold γ_t is not identified (and therefore k, γ_0, γ_1 and γ_2 are not identified), and hence the null distributions of the test statistics are non-standard due to the well-known Davies' problem and can be typically explored by taking the supremum of all possible values of unidentified parameters (e.g. Davies, 1987; Hansen, 1996).

This regression model (12) can be estimated by OLS, yielding estimate $\tilde{\boldsymbol{\beta}}_1$, residuals $\tilde{e}_t$ and sum of squared errors $S_0 = \sum_{t=1}^T \tilde{e}_t^2$. To test the null hypothesis H_0^1, it is natural to compare the sum of squared errors, S_0, with that of the time-varying threshold model, $SSR_T(\boldsymbol{\gamma})$. The sup-test statistic then is defined as

$$F_1 = \sup_{((\gamma_0,\gamma_1,\gamma_2),k)\in\boldsymbol{\Gamma}\times\Gamma_k} \frac{S_0 - SSR_T(\boldsymbol{\gamma})}{SSR_T(\boldsymbol{\gamma})/T - m}. \tag{13}$$

We reject the null hypothesis that linear model is appropriate if F_1 is large.

If the null hypothesis of no threshold effect in (1), H_0^1, is rejected, one can further examine whether or not this threshold is constant. This hypothesis is equivalent to a restriction that the two coefficients of time-varying threshold function are zeros. Consider the null hypothesis

$$H_0^2 : \gamma_1 = \gamma_2 = 0, \text{ v.s. } H_1^2 : \gamma_1 \neq 0 \text{ or } \gamma_2 \neq 0.$$

Under H_0^2, the model defined in (1) and (3) shrinks to a constant threshold model which has been investigated in Hansen (2000):

$$y_t = \begin{cases} \boldsymbol{\beta}_1' \mathbf{x}_t + e_t, & \text{if } q_t \leq \gamma_0 \\ \boldsymbol{\beta}_2' \mathbf{x}_t + e_t, & \text{if } q_t > \gamma_0 \end{cases}. \tag{14}$$

The frequency k is not identified and hence the test again suffers the Davies' problem.

To test the null hypothesis H_0^2, it is also natural to compare the sum of squared errors of the constant threshold model with that of the time-varying threshold model. Define the sum of squared errors of the constant threshold model as $SSR_2(\hat{\gamma}_0)$, in which $\hat{\gamma}_0$ is the estimated constant threshold value. For fixed frequency k, estimate the model in (1) and denote the sum of squared errors as $SSR_1(k)$. Thus consider the following test statistic:

$$F_2 = \sup_{k\in\Gamma_k} \frac{SSR_2(\hat{\gamma}_0) - SSR_1(k)}{SSR_1(k)/T - m}. \tag{15}$$

The null hypothesis of constant threshold, $H_0^2 : \gamma_1 = \gamma_2 = 0$, is rejected in favor of the time-varying threshold if F_2 is large.

Because of the above mentioned Davies' problem, the test statistics F_1 and F_2 have non-standard distributions. Many authors have examined the Davies' problem, for example, Andrews and Ploberger (1994) and Hansen (1996). In general, the construction of critical values for these distributions of test statistics is non-trivial (Andrews and Ploberger, 1994) and bootstrapping becomes a realistic alternative. Following the literature (e.g.

DOI: 10.1111/jtsa.12574

Mammen, 1993; Porter and Yu, 2015), we propose to construct the p-values of the test statistics F_1 and F_2 based on a two-point wild bootstrap procedure, which enables the wild bootstrap to remain consistent even in the presence of heteroscedasticity or model misspecification as illustrated by Kline and Santos (2012). Specifically, the following procedure is used in the test statistics F_1 and F_2.

Step 1: For $t = 1, 2, \dots, T$ and $j = 1, 2$, generate two-point wild bootstrap residual $e^*_{jt} = \hat{e}_{jt}(1 - \sqrt{5})/2$ with probability $(1 + \sqrt{5})/2\sqrt{5}$, and $e^*_{jt} = \hat{e}_{ji}(1 + \sqrt{5})/2$ with probability $(\sqrt{5} - 1)/2\sqrt{5}$, where $\hat{e}_{1t}$ are the OLS residual from model (12), and $\hat{e}_{2t}$ is the residuals of the constant threshold model (14).
Step 2: Generate $y^*_{1t} = \hat{\boldsymbol{\beta}}'\mathbf{x}_t + e^*_{1t}$ under the null H^1_0. Under the null H^2_0, we generate $y^*_{2t} = \hat{\boldsymbol{\beta}}'(\hat{\gamma}_0)\mathbf{x}_t + e^*_{2t}$.
Step 3: Compute the sup-test F^b_1 from $\{y^*_{1t}, \mathbf{x}_t\}^T_{t=1}$, and F^b_2 from $\{y^*_{2t}, \mathbf{x}_t, q_t\}^T_{t=1}$.
Step 4: Repeat Steps 1–3 B times. The empirical p-value can be obtained by calculating the frequency of simulated F^b_1 (F^b_2) that exceeds the observed F_1 (F_2) when the number of B is sufficiently large.

3. SIMULATION STUDIES

The goal of this Monte Carlo study is to investigate the finite sample properties of the estimation procedure and test statistics proposed in Section 2.

To evaluate the performance of the proposed estimation procedure, we consider the following data generating process (DGP):

$$\text{DGP1: } y_t = \begin{cases} \beta_{11} + \beta_{12}x_{1t} + e_t, & q_t \le \gamma_t \\ \beta_{21} + \beta_{22}x_{2t} + e_t, & q_t > \gamma_t \end{cases}, \tag{16}$$

where $x_{1t} = x_t\{q_t \le \gamma_t\}$, $x_{2t} = x_t\{q_t > \gamma_t\}$. $x_t \sim \mathrm{N}(0, 2^2)$, q_t and e_t follow $i.i.d.\mathrm{N}(0, 1)$.

With regard to γ_t in (16), we first consider the Fourier-form threshold setting and the constant threshold setting given by:

$$\text{Case 1(a): } \gamma_t = \gamma_0 + \gamma_1 \sin(2\pi kt/T) + \gamma_2 \cos(2\pi kt/T); \tag{17}$$

$$\text{Case 1(b): } \gamma_t \equiv 0. \tag{18}$$

Furthermore, to investigate the performance of the proposed model based on Fourier approximation, we also consider the following threshold settings to match the patterns of threshold change in Figure 1:

$$\text{Case 2(a): } \gamma_t = \begin{cases} 1\ , t < \frac{T}{3} \\ 0.8, \text{if } \frac{T}{3} \le t \le \frac{2T}{3} \\ 1\ , t > \frac{2T}{3} \end{cases}; \tag{19}$$

$$\text{Case 2(b): } \gamma_t = \begin{cases} 0.8, \text{if } t \le 0.8T \\ 1\ , \text{if } t > 0.8T \end{cases}; \tag{20}$$

$$\text{Case 2(c): } \gamma_t = -1/[1 + \exp(0.0005(t - 0.75T)(t - 0.2T))] + 1.5; \tag{21}$$

$$\text{Case 2(d): } \gamma_t = 1.5[1 - \exp(-0.002(t - 0.75T)^2)) - \exp(-0.001(t - 0.2T)^2]. \tag{22}$$

DOI: 10.1111/jtsa.12574

Finally, to compare the performance of the proposed procedure with that of the comparable method using the SETAR model and dummy variables proposed by Bessec (2003), we consider the following data generating process (DGP):

$$\text{DGP2: } y_t = \begin{cases} \alpha_1 + \rho_1 y_{t-1} + e_t, \; y_{t-1} \le \gamma_t \\ \alpha_2 + \rho_2 y_{t-1} + e_t, \; y_{t-1} > \gamma_t \end{cases}, \tag{23}$$

in which $\gamma_t = \begin{cases} 1 \text{ , if } t \le 0.8T \\ 1.5, \text{if } t > 0.8T \end{cases}$, e_t follows $i.i.d.$N(0, 1).

3.1. The Estimation Procedure

We conduct Monte Carlo experiments to examine the finite performance of the estimation procedure proposed in Section 2.1.

In the first experiment we consider two cases in DGP1: Case 1(a) with a Fourier-form threshold change and Case 1(b) with a constant threshold. For all simulations in this experiment, the values of the parameters are set at $\beta_{11} = \beta_{21} = 10$, $\beta_{12} = 10$, $\beta_{22} = 20$, $\gamma_0 = 0.5$, $\gamma_1 = 1$, and $\gamma_2 = 1$. We vary the parameter k to assess the model's performance to the magnitude of the Fourier frequency. We set $k = \{1, 2, 3, 4, 5\}$ and run experiments on a range of sample sizes ($T = 50, 100, 200$). The number of replications is 1000.

We first consider the case where the time-varying threshold is given by Case 1(a), while the time-varying threshold is overlooked in the estimation procedure (and hence the estimation procedure proposed by Hansen (2000) is used). Figure 2 reports the kernel densities for the regression slope estimates. It is easily seen that the estimated slopes in sub-samples cannot converge to the desired values by ignoring potentially time-varying features in the threshold.

We next evaluate the performance of the estimation procedure proposed in Section 2.1. In doing so, we consider two cases: Case 1(a) with a Fourier-form threshold change and Case 1(b) with a constant threshold. Figure 3 presents the kernel densities for the regression slope estimators for selected samples and the frequency $k = 1$. As predicted by Theorem 1, the consistency of the time-varying threshold estimator implies that the observations can be correctly classified into subsets, and hence the true model slope parameters β_{12} and β_{22} can be consistently estimated. As can be seen from Figure 3, the parameter estimation procedure works well in the finite samples.

Table I presents the summary statistics (i.e. mean, and standard deviation) for the least-square parameter estimates. Simulations show that the discrete frequency k can be almost estimated accurately,[7] and hence its standard deviation is not computed and reported. For all the parameters, the mean of each parameter comes close to its true value, and the accuracy of the model improves as T increases, regardless of the frequency k, which is also consistent with the results in Theorems 1 and 2. Moreover, lower panel of Table I show that there is little efficiency loss by allowance for Fourier approximation in the estimation procedure even when there is no time-varying feature in the threshold.[8]

In the second experiment, we evaluate the performance of the proposed estimation for different patterns of threshold change given by Case 2(a)–2(d). That is, the true threshold change is given by Case 2(a)–2(d), while approximated by a Fourier function. In these simulations, we set $\beta_{11} = \beta_{12} = 1$, and $\beta_{21} = \beta_{22} = 2$. The simulation results are reported in Table II, in which we report the summary statistics (i.e. mean and standard deviation) for the estimates based on the proposed estimation procedure. For each case of threshold change, the mean of each parameter is close to its true value,[9] and the standard deviation becomes smaller as the sample size increases, indicating that the estimation based on Fourier approximation works well in different patterns of threshold change.

[7] See footnote 4.
[8] In our unreported simulation results, we find that there is not much difference between the standard errors of the estimates based on the Hansen's (2000) estimation procedure and the estimation procedure proposed in Section 2 when the true DGP contains constant threshold.
[9] In these simulations, the threshold coefficients are not reported as we cannot compare the estimates with their true values, simply because the Fourier-form threshold is only an approximation of the actual threshold.

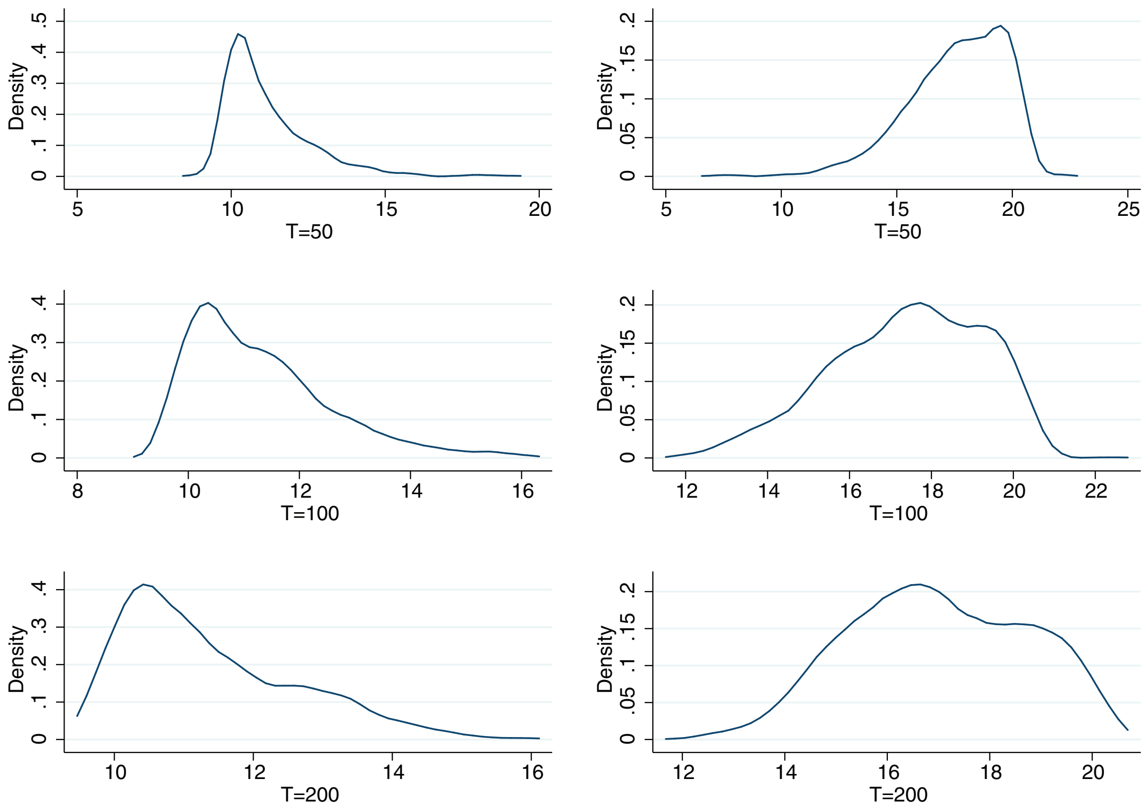

FIGURE 2. Monte Carlo experiment generated sample distributions of $\beta_{12} = 10$ (left panel) and $\beta_{22} = 20$ (right panel) for $k = 1$. DGP is the threshold model with a time-varying threshold given by (16) and (17), but parameters are estimated using the Hansen's (2000) approach

In the third experiment, we compare the performance of the proposed procedure with the method using dummy variables investigated by Bessec (2003) who assumes a one-time break in threshold where the change point is not estimated but given in advance. The true DGP is the SETAR model with a dummy variable in the threshold. In these simulations, the values of the parameters are set at $\alpha_1 = 0.5, \rho_1 = -0.3, \alpha_2 = 1$, and $\rho_2 = 0.3$. Table III reports the simulation results. As before we report the summary statistics (i.e. mean and standard deviation) for the estimates based on the proposed estimation procedure and the dummy procedure. As can be seen in Table III, when we employ the estimation procedure based on a Fourier approximation to estimate the SETAR model, the mean of each parameter is close to its true value, and the accuracy of the estimates improves as the sample size increases. Meanwhile, when we employ the dummy method assuming the break point being known, the mean of each parameter becomes close to its true value particularly when the sample size is larger than 500. Overall, the two methods give fairly similar performance. Comparing with the dummy method which assumes the change point being exogenous, the proposed procedure has the advantage of operating conveniently, especially when the change point is unknown.

3.2. Test Statistics

We conduct Monte Carlo experiments to examine the size and power properties of the test statistic for the existence of threshold, F_1, and for threshold constancy, F_2, proposed in Section 2.3. To examine the finite sample performance of the statistic F_1 (the linear model against the time-varying threshold model), we set $\beta_{11} = \beta_{21} = 1$,

DOI: 10.1111/jtsa.12574

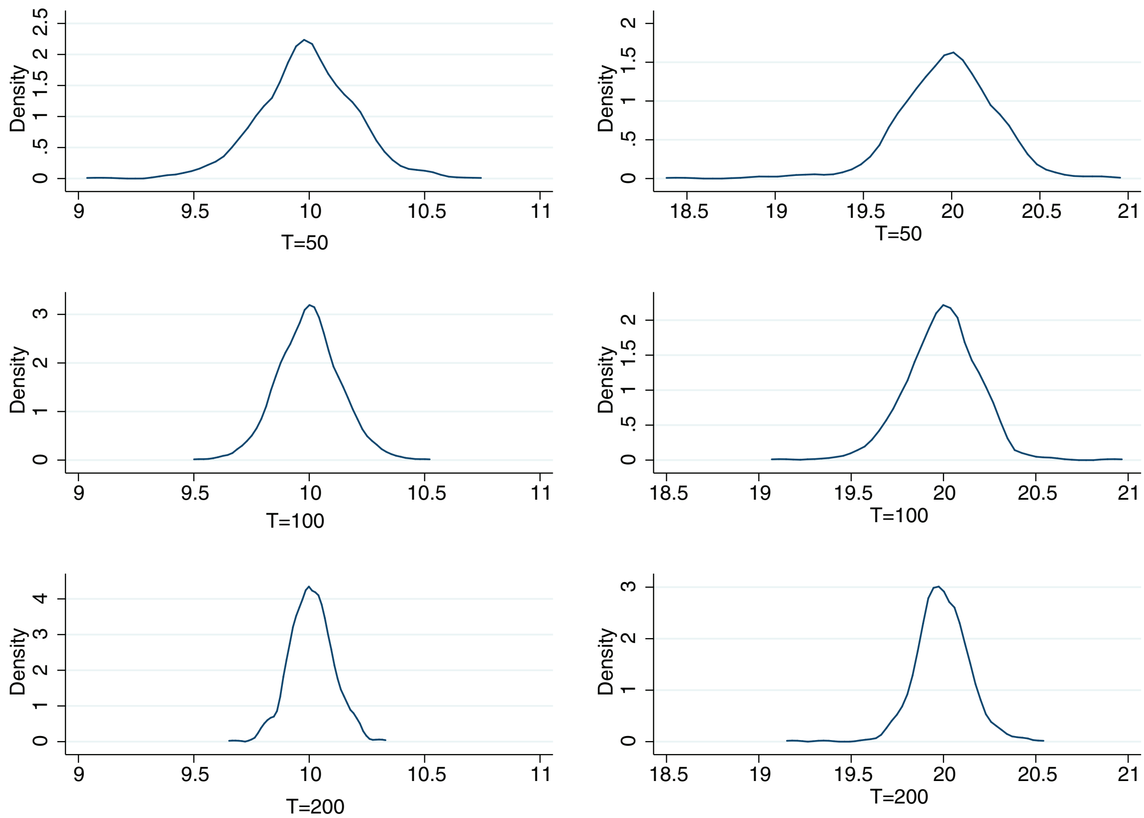

FIGURE 3. Monte Carlo experiment generated sample distributions of $\beta_{12} = 10$ (left panel) and $\beta_{22} = 20$ (right panel) for $k = 1$. DGP is the threshold model with a time-varying threshold given by (16) and (17), and parameters are estimated using the estimation procedure proposed in Section 2.1

$\beta_{12} = 1$, and set γ_t as in Case 1(a), Case 2(a)-2(d). Thus, the rejection frequencies under the DGP given by (16) with $\beta_{22} = 1$ and $\beta_{22} = \{1.5, 2\}$ are the size and power of the test statistic F_1 respectively.

Similarly, to evaluate the performance of the test for threshold constancy, F_2 (the constant threshold against the time-varying threshold model), we set $\beta_{11} = \beta_{21} = 1, \beta_{12} = 1$, and $\beta_{22} = \{1.5, 2\}$. Then the rejection frequencies under the DGP in (16) with $\{\gamma_t \equiv 0\}$ as in Case 1(b) and $\{\gamma_t\}$ as in Case 1(a), Case 2(a)-2(d) are the size and power of the test statistic F_2 respectively.[10]

Table IV reports size and power properties of the test statistics F_1 and F_2. The results show that the empirical size is close to the 5% significance level in all cases, implying the test statistics F_1 and F_2 are correctly sized. Meanwhile, the probability of rejection increases as the sample size gets larger, and the power performance is generally satisfactory in most patterns of threshold change, especially when the sample size T is not less than 200. Overall, the proposed tests work well when sample sizes are moderate.

4. EMPIRICAL APPLICATION

We apply our approach to investigate the Taylor rule that establishes interest rate rules in conducting monetary policy, which is considered as one of the most influential tools and has attracted much attention among academics and policymakers in past decades. In this section, we revisit the Taylor rule by employing the proposed time-varying

[10] Following Porter and Yu (2015), to save simulation time, the number of replications is set as 100, and B is set as 200 in the bootstrap method of Section 2.2.

TABLE I. Simulation results for selected sample sizes and frequencies

Case 1(a):		Time-varying threshold						Coefficients			
		$\gamma_0 = 0.5$		$\gamma_1 = 1$		$\gamma_2 = 1$		$\beta_{12} = 10$		$\beta_{22} = 20$	
T	$k(\hat{k})$	Mean	S.d	Mean	S.d	Mean	S.d	Mean	S.d	Mean	S.d
50	1	0.471	0.253	0.923	0.216	0.964	0.213	9.890	0.204	19.896	0.863
	2	0.474	0.192	0.932	0.212	0.970	0.202	9.995	0.197	20.028	0.608
	3	0.490	0.196	0.919	0.211	0.974	0.222	10.022	0.208	19.926	0.933
	4	0.461	0.186	0.883	0.212	0.942	0.212	9.999	0.191	19.944	0.713
	5	0.483	0.231	0.894	0.253	0.935	0.265	10.008	0.197	19.880	0.822
100	1	0.497	0.111	0.972	0.129	0.981	0.124	9.999	0.134	19.985	0.244
	2	0.485	0.129	0.962	0.139	0.971	0.133	10.007	0.142	19.965	0.512
	3	0.497	0.118	0.975	0.137	0.989	0.134	10.003	0.134	19.983	0.235
	4	0.509	0.112	0.985	0.131	0.984	0.131	9.999	0.135	19.982	0.206
	5	0.504	0.109	0.980	0.126	0.996	0.124	9.992	0.127	19.997	0.194
200	1	0.500	0.071	0.999	0.075	0.998	0.069	10.009	0.092	19.986	0.201
	2	0.507	0.055	1.003	0.068	1.003	0.067	9.994	0.098	19.986	0.119
	3	0.504	0.066	0.997	0.078	1.000	0.083	9.998	0.098	19.974	0.382
	4	0.498	0.057	0.997	0.072	1.001	0.073	10.008	0.097	19.986	0.135
	5	0.504	0.060	0.997	0.065	0.996	0.072	10.009	0.096	19.992	0.143
Case 1(b):		**Constant threshold**						**Coefficients**			
		$\gamma_0 = 0.5$		$\gamma_1 = 0$		$\gamma_2 = 0$		$\beta_{12} = 10$		$\beta_{22} = 20$	
T	$k(\hat{k})$	Mean	S.d	Mean	S.d	Mean	S.d	Mean	S.d	Mean	S.d
50	1	0.481	0.231	−0.021	0.215	−0.004	0.213	9.990	0.214	19.999	0.822
100	1	0.499	0.118	−0.012	0.121	−0.019	0.124	9.992	0.131	19.995	0.241
200	1	0.501	0.061	0.005	0.055	0.003	0.072	10.006	0.096	19.995	0.133

Notes: S.d. denotes standard deviation. The simulations were written in the GAUSS programming language.

TABLE II. Simulations results for different patterns of threshold change

	$\beta_{11} = 1$		$\beta_{12} = 1$		$\beta_{21} = 2$		$\beta_{22} = 2$	
	Mean	S.d	Mean	S.d	Mean	S.d	Mean	S.d
(a) Threshold setting in Case 2(a)								
$T = 100$	0.985	0.168	0.984	0.172	2.022	0.165	2.016	0.161
$T = 200$	1.008	0.112	1.001	0.116	2.001	0.103	2.001	0.107
$T = 500$	1.019	0.070	1.017	0.067	1.984	0.066	1.991	0.065
$T = 1000$	1.017	0.049	1.012	0.052	1.997	0.046	1.996	0.049
(b) Threshold setting in Case 2(b)								
$T = 100$	0.978	0.179	0.998	0.175	2.026	0.167	2.012	0.163
$T = 200$	0.998	0.112	0.999	0.109	2.002	0.108	1.996	0.102
$T = 500$	1.011	0.050	1.013	0.065	1.993	0.067	1.997	0.066
$T = 1000$	1.021	0.046	1.021	0.045	1.991	0.046	1.988	0.047
(c) Threshold setting in Case 2(c)								
$T = 100$	0.986	0.183	0.997	0.179	2.011	0.178	2.001	0.173
$T = 200$	1.033	0.119	1.023	0.121	1.961	0.117	1.971	0.121
$T = 500$	1.028	0.062	1.033	0.071	1.975	0.071	1.977	0.069
$T = 1000$	1.063	0.051	1.058	0.053	1.961	0.052	1.972	0.046
(d) Threshold setting in Case 2(d)								
$T = 100$	0.944	0.339	0.986	0.424	2.018	0.138	1.996	0.132
$T = 200$	1.026	0.122	1.011	0.115	1.973	0.109	1.991	0.101
$T = 500$	1.062	0.072	1.054	0.075	1.943	0.093	1.946	0.097
$T = 1000$	1.068	0.042	1.067	0.044	1.988	0.065	1.987	0.065

DOI: 10.1111/jtsa.12574

TABLE III. Comparison of the proposed procedure and the method using dummy variables

	$\alpha_1 = 0.5$		$\rho_1 = -0.3$		$\alpha_2 = 1$		$\rho_2 = 0.3$	
	Mean	S.d	Mean	S.d	Mean	S.d	Mean	S.d
(a) The proposed estimation procedure								
$T = 100$	0.505	0.568	−0.365	0.543	1.058	0.687	0.271	0.308
$T = 200$	0.478	0.278	−0.361	0.303	1.033	0.461	0.283	0.198
$T = 500$	0.493	0.075	−0.312	0.104	1.014	0.203	0.293	0.095
$T = 1000$	0.500	0.045	−0.301	0.065	1.003	0.132	0.296	0.063
(b) The dummy method								
$T = 100$	0.558	0.802	−0.291	0.713	0.846	0.531	0.349	0.239
$T = 200$	0.485	0.197	−0.326	0.232	0.903	0.365	0.334	0.166
$T = 500$	0.493	0.067	−0.315	0.099	0.963	0.196	0.312	0.093
$T = 1000$	0.498	0.045	−0.306	0.067	0.968	0.132	0.311	0.062

threshold model and using U.S. quarterly data from 1955Q1 to 2019Q4. The data are available from the website of Federal Reserve Bank of St. Louis (http://www.stlouisfed.org/) and the Congressional Budget Office (https://www.cbo.gov/).

As a benchmark, we follow the literature (e.g. Taylor and Davradakis, 2006; Zhu *et al.*, 2019) to consider the linear Taylor rule:

$$i_t = \rho i_{t-1} + (1-\rho)[r^* + \pi_{t-1} + \delta_1(\pi_{t-1} - \pi^*) + \delta_2 x_{t-1}] + e_t, \tag{24}$$

in which i_t is the short-term nominal interest rate at time t, r^* is the long-run real interest rate, π_t is the inflation rate, π^* is the target inflation rate, x_t is the output gap in period t, e_t is the error term, and ρ controls the smoothness of the interest rate, δ_1 and δ_2 measure the sensitivities of the policy interest rate to inflation deviation and output gap respectively.

Since we can observe the variables i_t, π_t and y_t, we re-parameter the above model as

$$i_t = \theta_0 + \theta_1\pi_{t-1} + \theta_2 x_{t-1} + \theta_3 i_{t-1} + e_t, \tag{25}$$

in which $\theta_0 = (1-\rho)(r^* - \delta_1\pi^*)$, $\theta_1 = (1-\rho)(1+\delta_1)$, $\theta_2 = (1-\rho)\delta_2$ and $\theta_3 = \rho$. Clearly, we have $\delta_1 = \frac{\theta_1}{1-\theta_3} - 1$ and $\delta_2 = \frac{\theta_2}{1-\theta_3}$; hence, after obtaining the estimates of (25), we can rewrite (25) to a form in which inflation deviation and output gap as explanatory variables as in (24).

Following Taylor and Davradakis (2006) and Zhu *et al.* (2019), we use the following threshold model with a constant threshold to capture asymmetries and nonlinearities in Taylor rule, in which the interest rate would experience two regimes depending on the magnitude of the threshold variable:

$$i_t = \begin{cases} \alpha_0 + \alpha_1\pi_{t-1} + \alpha_2 x_{t-1} + \alpha_3 i_{t-1} + e_t, & q_t \le \gamma \\ \beta_0 + \beta_1\pi_{t-1} + \beta_2 x_{t-1} + \beta_3 i_{t-1} + e_t, & q_t > \gamma \end{cases}, \tag{26}$$

where q_t is the threshold variable. In this article, the unemployment rate is chosen as the threshold variable, because it provides an intuitive measure of real economic activity, as illustrated in Zhu *et al.* (2019).

To capture the time-varying threshold effect in the Taylor rule, we extend the constant threshold model to the proposed model with a time-varying threshold, given by

$$i_t = \begin{cases} \alpha_0 + \alpha_1\pi_{t-1} + \alpha_2 x_{t-1} + \alpha_3 i_{t-1} + e_t, & q_t \le \gamma_t \\ \beta_0 + \beta_1\pi_{t-1} + \beta_2 x_{t-1} + \beta_3 i_{t-1} + e_t, & q_t > \gamma_t \end{cases}, \tag{27}$$

DOI: 10.1111/jtsa.12574

TABLE IV. Size and power of test statistics in (13) and (15)

	Threshold effect	Sample size	Test for threshold effect		Test for threshold constancy	
			Size	Power	Size	Power
Case 1(a)	$\beta_{22} = 1.5$	$T = 100$	0.040	0.680	0.060	0.530
		$T = 200$	0.060	0.920	0.060	0.910
		$T = 500$	0.030	1.000	0.060	1.000
	$\beta_{22} = 2$	$T = 100$	0.040	1.000	0.020	0.990
		$T = 200$	0.060	1.000	0.030	1.000
		$T = 500$	0.030	1.000	0.030	1.000
Case 2(a)	$\beta_{22} = 1.5$	$T = 100$	0.030	0.800	0.070	0.260
		$T = 200$	0.060	1.000	0.060	0.410
		$T = 500$	0.050	1.000	0.060	0.760
	$\beta_{22} = 2$	$T = 100$	0.030	1.000	0.060	0.270
		$T = 200$	0.060	1.000	0.060	0.590
		$T = 500$	0.050	1.000	0.050	0.830
Case 2(b)	$\beta_{22} = 1.5$	$T = 100$	0.030	0.810	0.040	0.210
		$T = 200$	0.050	1.000	0.060	0.320
		$T = 500$	0.050	1.000	0.050	0.570
	$\beta_{22} = 2$	$T = 100$	0.030	1.000	0.030	0.270
		$T = 200$	0.050	1.000	0.060	0.440
		$T = 500$	0.050	1.000	0.050	0.720
Case 2(c)	$\beta_{22} = 1.5$	$T = 100$	0.040	0.900	0.060	0.380
		$T = 200$	0.060	1.000	0.070	0.680
		$T = 500$	0.050	1.000	0.060	1.000
	$\beta_{22} = 2$	$T = 100$	0.040	1.000	0.040	0.410
		$T = 200$	0.050	1.000	0.080	0.900
		$T = 500$	0.060	1.000	0.050	1.000
Case 2(d)	$\beta_{22} = 1.5$	$T = 100$	0.060	0.660	0.040	0.250
		$T = 200$	0.050	1.000	0.030	0.860
		$T = 500$	0.060	1.000	0.050	1.000
	$\beta_{22} = 2$	$T = 100$	0.060	0.990	0.030	0.390
		$T = 200$	0.050	1.000	0.020	0.990
		$T = 500$	0.060	1.000	0.060	1.000

in which $\gamma_t = \gamma_0 + \gamma_1 \sin(2\pi kt/T) + \gamma_2 \cos(2\pi kt/T)$. Here, a time-varying threshold may be reasonable, because policymakers may respond differently to the unemployment rate in different economic environments that shape their attitude and tolerance for unemployment, leading to the reference (threshold) for assessing the unemployment rate being time-varying. Accordingly, the same level of unemployment rate may be regarded as high under a certain period but only moderate under other periods. As shown in this article, suppose that the threshold is indeed time-varying but being treated as a constant, we would end up with biased estimates, while there is little efficiency loss by considering a time-varying threshold even when there is no time-varying feature in the threshold.

The empirical results are reported in Table V, in which confidence intervals and p-values are calculated with $B = 2000$ bootstrap replications.[11] It can be seen that there is great difference between the estimates of linear model and that of threshold models. Thus, we select the optimal model by employing three test statistics F_1, F_2

[11] In practice, we follow Hansen (1996, 2017) to propose the use of wild bootstrap confidence intervals. The bootstrap procedure goes as follows: (i) For $t = 1, 2, \dots, T$, generate two-point wild bootstrap residual $e_t^* = \hat{e}_t(1-\sqrt{5})/2$ with probability $(1+\sqrt{5})/2\sqrt{5}$, and $e_t^* = \hat{e}_i(1+\sqrt{5})/2$ with probability $(\sqrt{5}-1)/2\sqrt{5}$, where $\hat{e}_t$ are the residuals of the proposed time-varying threshold model. (ii) Set $y_t^* = \hat{\boldsymbol{\beta}}'(\hat{\boldsymbol{\gamma}})\mathbf{x}_t(\gamma_t(\hat{\boldsymbol{\gamma}}))+e_t^*$, where $(\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\gamma}}')$ are the LS estimates of the proposed model (1) using the original sample $\{y_t, \mathbf{x}_t, q_t\}_{t=1}^T$. (iii) Using the observations $\{y_t^*, \mathbf{x}_t, q_t\}_{t=1}^T$, estimate the proposed threshold regression model with a time-varying threshold, yielding the parameter estimates $(\hat{\boldsymbol{\beta}}^{*\prime}, \hat{\boldsymbol{\gamma}}^{*\prime})$. (iv) Repeat Steps (i)–(iii) B times, and obtain a sample of simulated coefficient estimates $(\hat{\boldsymbol{\beta}}^{*\prime}, \hat{\boldsymbol{\gamma}}^{*\prime})$. Create $1-a$ bootstrap confidence intervals for the estimates

DOI: 10.1111/jtsa.12574

TABLE V. Empirical results and 90% confidence intervals

	Linear model	constant threshold		time-varying threshold	
		regime 1($q_t \le \hat{\gamma}$)	regime 2($q_t > \hat{\gamma}$)	regime 1($q_t \le \hat{\gamma}_t$)	regime 2($q_t > \hat{\gamma}_t$)
Intercept	0.124	0.005	−0.149	−0.188	0.623
	[−0.032,0.280]	[−0.203,0.211]	[−5.811,5.512]	[−0.378,0.002]	[−0.233,1.482]
π_{t-1}	0.066	0.107	−0.048	0.170	−0.021
	[−0.012,0.144]	[0.041,0.173]	[−0.399,0.301]	[0.089,0.249]	[−0.136,0.093]
x_{t-1}	0.056	0.045	−0.079	0.021	−0.093
	[0.023,0.089]	[0.003,0.087]	[−1.052,0.893]	[−0.014,0.056]	[−0.244,0.057]
i_{t-1}	0.936	0.938	0.935	0.945	0.811
	[0.895,0.976]	[0.896,0.977]	[0.541,1.331]	[0.918,0.974]	[0.644,0.978]
Inflation deviation	0.022	0.730	−1.739	2.139	−1.112
	[−0.385,1.136]	[−1.662,3.122]	[−9.846,6.368]	[0.281,3.999]	[−2.001, − 0.223]
Output gap	0.873	0.732	−1.223	0.383	−0.493
	[−0.062, 1.665]	[−0.333, 1.798]	[−6.652, 4.206]	[−0.376, 1.141]	[−1.544, 0.558]
Testing		F^C	$F1$	$F2$	
	Statistics	21.931	88.932	61.789	
	p-value	0.661	0.075	0.069	

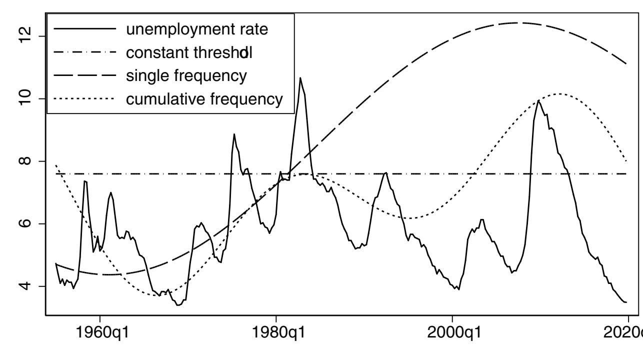

FIGURE 4. The estimated constant threshold and time-varying threshold

and F^C. The former two statistics are described in Section 2, while the test statistic F^C for the null hypothesis of the linear model against the constant threshold model is proposed by Hansen (1999). According to the F^C test statistic, the linear model cannot be rejected (p-value=0.661), while both F_1 and F_2 are with the p-value less than 10%, supporting the existence of a time-varying threshold. These testing results are reasonable, because ignoring the existed time-varying features in threshold can result in biased estimates, which may lead to further consequences including distorted testing results. Overall, these testing results indicate that the proposed time-varying threshold model is suitable in this application. As discussed by Dueker *et al.* (2010) and Zhu *et al.* (2019), the estimated time-varying threshold could be interpreted as the natural rate of unemployment, which is the foundation of a number of theoretical or empirical relationships such as Okun's Law. However, it is often difficult to directly estimate the natural rate of unemployment. In this case, the proposed time-varying threshold model may serve as a potential candidate for the natural rate. To obtain a more precise time-varying threshold level, we estimate the model with a cumulative frequency, in which the number of cumulative frequencies (n) is chosen by minimizing the residual sum of squares, leading to the estimate $\hat{n} = 2$.[12] The estimated time-varying threshold levels are given by Figure 4.

$(\hat{\boldsymbol{\beta}}^{*\prime}, \hat{\boldsymbol{\gamma}}^{*\prime})$ by the symmetric percentile method: the estimates plus and minus the $(1-a)$ quantile of the absolute centered bootstrap estimates. For example, the confidence interval of $\hat{\beta}_1$ is $\hat{\beta}_1 \pm q^*_{1-a}$, where q^*_{1-a} is the $1-a$ quantile of $|\hat{\beta}^*_1 - \hat{\beta}_1|$.

[12] In an unreported appendix, we show that the slope estimates based on the cumulative frequencies are similar to that based on the single frequency. Therefore, we do not report the empirical results based on the cumulative frequencies to save space.

DOI: 10.1111/jtsa.12574

According to the estimates based on the time-varying threshold model, the U.S. central band respond asymmetrically to inflation deviation and the output gap during periods of relatively good state ($q_t \le \hat{\gamma}_t$) and during periods of relatively bad state ($q_t > \hat{\gamma}_t$). The central bank responds positively (negatively) to inflation deviation and output gap during the good periods (bad periods); moreover, the response to inflation is clearly stronger during good periods than during bad periods, while the response to output gap is relatively weaker during good periods than during bad periods. The coefficients on inflation deviation in both regimes are both significant at the 5% level, while the coefficients on output gap are insignificant. Furthermore, the coefficients on i_{t-1} are 0.945 and 0.811, indicating that the short-term interest rates change more smoothly during good periods than bad periods. These empirical results cannot be obtained if the time-varying feature in threshold is overlooked.

5. CONCLUSION

Threshold models have been widely applied in economics. However it is restrictive to assume that the threshold values are stable or time-invariant. This article proposes a threshold model with a time-varying threshold, where the time-varying threshold is approximated by a Fourier function. A least-square based procedure is proposed to estimate the model parameters, and two statistics are constructed to test for the threshold effect and threshold constancy. The convergence rate and asymptotic distribution of the time-varying threshold estimator are also established, and Monte Carlo experiments are conducted to examine the finite properties of the estimation procedure and test statistics. We also provide the evidence supporting that the estimated slopes in sub-samples are biased when the threshold is time-varying but being treated as a constant. The model is illustrated with an empirical application to a nonlinear Taylor rule for the United States.

ACKNOWLEDGEMENTS

The authors thank the editor and two anonymous referees for very valuable comments and suggestions that result in a substantial improvement of this article. Remaining errors and omissions are our own. Lixiong Yang acknowledges the financial support from the National Natural Science Foundation of China (Grant No. 71803072).

DATA AVAILABILITY STATEMENT

The data analyzed in the empirical application are publicly available from the websites cited in the article. The simulated data are available on request from the authors. No newly collected data have been analyzed in this article.

APPENDIX : MATHEMATICAL PROOFS

This appendix provides the proofs of Theorems 1–2 in the article. To save space, we skip the details for some intermediary results and a more detailed version is available from the authors on request. For notational simplicity, we first clarify the following denotations.

1. The indicator function $1_{(q_t \le \gamma_t(\boldsymbol{\gamma}))} = \{q_t \le \gamma_t(\boldsymbol{\gamma})\}$, where $\boldsymbol{\gamma} = [\gamma_0, \gamma_1, \gamma_2, k]'$,
2. $\text{sgn}(x) = 1_{(x>0)} - 1_{(x\le 0)}$ (i.e. the sign function), hence $|x| = x1_{(x>0)} - x1_{(x\le 0)} = x(1_{(x>0)} - 1_{(x\le 0)}) = x\text{sgn}(x)$,
3. $\gamma_t(\gamma_0') = \gamma_t(\gamma_0', \gamma_1, \gamma_2, k)$, $\gamma_t(\gamma_1') = \gamma_t(\gamma_0, \gamma_1', \gamma_2, k)$ and $\gamma_t(\gamma_2') = \gamma_t(\gamma_0, \gamma_1, \gamma_2', k)$,
4. $\gamma_t^0(\gamma_0) = \gamma_t^0(\gamma_0, \gamma_1^0, \gamma_2^0, k^0)$, $\gamma_t^0(\gamma_1) = \gamma_t^0(\gamma_0^0, \gamma_1, \gamma_2^0, k^0)$ and $\gamma_t^0(\gamma_2) = \gamma_t^0(\gamma_0^0, \gamma_1^0, \gamma_2, k^0)$,
5. $d_t(\gamma_t^0(\gamma_i)) = \{q_t \le \gamma_t^0(\gamma_i)\}$, for $i = 0, 1, 2$,
6. $\Delta_t^0(\gamma_{i1}, \gamma_{i2}) = d_t(\gamma_t^0(\gamma_{i2})) - d_t(\gamma_t^0(\gamma_{i1}))$,
7. $a_T = T^{1-2\alpha}$, $\gamma_i = \gamma_i^0 + \omega_i/a_T$, $\Delta\gamma_i^0 = \gamma_i - \gamma_i^0$, $\Delta\gamma_t^0(\gamma_i) = \gamma_t^0(\gamma_i) - \gamma_t^0(\gamma_i^0)$, $\Delta\mathbf{M}(\gamma_t^0(\gamma_i)) = \mathbf{M}(\gamma_t^0(\gamma_i)) - \mathbf{M}(\gamma_t^0(\gamma_i^0))$,
8. $J_T(\gamma_t^0(\gamma_i)) = T^{-1/2} \sum_{t=1}^{T} \mathbf{c}'\mathbf{x}_t e_t d_t(\gamma_t^0(\gamma_i))$,

DOI: 10.1111/jtsa.12574

9. $G_T(\gamma_t^0(\omega_i)) = a_T T^{-1} \sum_{t=1}^T \mathbf{c}'\mathbf{x}_t\mathbf{x}_t'\mathbf{c}|d_t(\gamma_t^0(\gamma_i^0 + \omega_i/a_T)) - d_t(\gamma_t^0(\gamma_i^0))|$,
10. $G_T^m(\gamma_t^0(\omega_i)) = a_T T^{-1} \sum_{t=1}^T \mathbf{c}'\mathbf{x}_t\mathbf{x}_t'\mathbf{c}|d_t(\min\{\gamma_t^0(\gamma_i^0 + \omega_i/a_T), \gamma_t^0(\gamma_i^0)\}) - d_t(\gamma_t^0(\gamma_i^0))|$,
11. $u_{Tt}(\gamma_t^0(\omega_i)) = \mathbf{c}'\mathbf{x}_t e_t \Delta_t^0(\gamma_i^0, \gamma_i^0 + \omega_i/a_T)$,
12. $R_T(\gamma_t^0(\omega_i)) = \sqrt{a_T} T^{-1/2} \sum_{t=1}^T u_{Tt}(\gamma_t^0(\omega_i))$.

Before proving Theorem 1, we first prove the following Lemma, which is used to prove Theorem 1.

Lemma A1. If q_t is strictly stationary and ergodic, $E|\phi(q_t)|^2 < \infty$, q_t has a continuous density function $f(q)$ such that $\sup_{x\in\mathbb{R}} f(x) = \bar{f} < \infty$, $\gamma_t(\gamma_i)$ is a differentiable function, $\gamma_t'(\gamma_i) = d\gamma_t(\gamma_i)/d\gamma_i$ does not depend on γ_i, Γ_i is a compact set for $i = 0, 1, 2$, and Γ_k is a countably finite set (i.e. $\Gamma_k = \{1, 2, \ldots, K\}$ for some $K < \infty$ and $K \in \mathbb{N}$), then

$$\sup_{\boldsymbol{\gamma}\in\Gamma\times\Gamma_k} \left| \frac{1}{T}\sum_{t=1}^T \phi(q_t)\{q_t \le \gamma_t(\boldsymbol{\gamma})\} - E(\phi(q_t)\{q_t \le \gamma_t(\boldsymbol{\gamma})\}) \right| \xrightarrow{a.s.} 0 \tag{A1}$$

Proof of Lemma 1. This proof is similar to Lemma 1 in Hansen (1996). The proof is skipped here and is available from the authors on request. ■

Proof of Theorem 1. Define the moment $\mathbf{M}(\gamma_t) = E(\mathbf{x}_t\mathbf{x}_t'\{q_t \le \gamma_t\}) = \int_{-\infty}^{\gamma_t} E(\mathbf{x}_t\mathbf{x}_t'|q_t) f(q_t) \mathrm{d}q_t$. By Lemma 1, we have

$$T^{-1}\sum_{t=1}^T \mathbf{x}_t\mathbf{x}_t'\{q_t \le \gamma_t\} - T^{-1}\sum_{t=1}^T \mathbf{M}(\gamma_t) \xrightarrow{a.s.} 0, \tag{A2}$$

uniformly over $\boldsymbol{\gamma} \in \boldsymbol{\Gamma} \times \Gamma_k$.

Denote $\boldsymbol{\gamma} = [\gamma_0, \gamma_1, \gamma_2, k]'$, $\gamma_\tau = \gamma_0 + \gamma_1 \sin(2\pi k\tau) + \gamma_2 \cos(2\pi k\tau)$,

$$\overline{\mathbf{M}}(\boldsymbol{\gamma}) = \lim_{T\to\infty} T^{-1}\sum_{t=1}^T \mathbf{M}(\gamma_t) = \int_0^1 \mathbf{M}(\gamma_\tau)\mathrm{d}\tau, \tag{A3}$$

and

$$\begin{aligned}\overline{\mathbf{M}}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^0) &= \lim_{T\to\infty} T^{-1}\sum_{t=1}^T \mathbf{x}_t\mathbf{x}_t'\{q_t \le \gamma_t\}\{q_t \le \gamma_t^0\} \\ &= \lim_{T\to\infty} T^{-1}\sum_{t=1}^T \mathbf{x}_t\mathbf{x}_t'\{q_t \le \gamma_t \wedge \gamma_t^0\} \\ &= \int_0^1 \mathbf{M}(\gamma_\tau \wedge \gamma_\tau^0)\mathrm{d}\tau \end{aligned} \tag{A4}$$

in which $\gamma_t \wedge \gamma_t^0 = \min(\gamma_t, \gamma_t^0)$. We can show that

$$\begin{aligned} T^{2\alpha-1}\{SSR_T(\gamma_t) - SSR_T(\gamma_t^0)\} &= T^{-1}\mathbf{c}'\mathbf{X}_0'(\mathbf{I}_T - \mathbf{P}_\gamma)\mathbf{X}_0\mathbf{c} + O_p(T^{\alpha-\frac{1}{2}}) \\ &= \mathbf{c}'(T^{-1}\mathbf{X}_0'\mathbf{X}_0 - T^{-1}\mathbf{X}_0'\mathbf{P}_\gamma\mathbf{X}_0)\mathbf{c} + o_p(1), \end{aligned} \tag{A5}$$

in which $\mathbf{X}_{0[t]} = \mathbf{x}_t'\{q_t \le \gamma_t^0\}$, $\mathbf{P}_\gamma = \mathbf{X}_\gamma^*(\mathbf{X}_\gamma^{*\prime}\mathbf{X}_\gamma^*)^{-1}\mathbf{X}_\gamma^{*\prime}$, $\mathbf{X}_\gamma^* = [\mathbf{X}, \mathbf{X}_\gamma]$, $\mathbf{X}_{[t]} = \mathbf{x}_t'$ and $\mathbf{X}_{\gamma[t]} = \mathbf{x}_t'\{q_t \le \gamma_t\}$ for $t = 1, 2, \ldots, T$.

DOI: 10.1111/jtsa.12574

For the first term of (A5), we have

$$T^{-1}\mathbf{X}_0'\mathbf{X}_0 = T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\{q_t \le \gamma_t^0\} \xrightarrow{a.s.} \overline{\mathbf{M}}(\boldsymbol{\gamma}^0). \tag{A6}$$

The asymptotic distribution of the second term of (A5) is given by

$$\begin{aligned} T^{-1}\mathbf{X}_0'\mathbf{P}_\gamma\mathbf{X}_0 &= \left[\tfrac{1}{T}\mathbf{X}_0'\mathbf{X}\tfrac{1}{T}\mathbf{X}_0'\mathbf{X}_\gamma\right]\begin{bmatrix}\tfrac{1}{T}\mathbf{X}'\mathbf{X} & \tfrac{1}{T}\mathbf{X}'\mathbf{X}_\gamma\\ \tfrac{1}{T}\mathbf{X}_\gamma'\mathbf{X} & \tfrac{1}{T}\mathbf{X}_\gamma'\mathbf{X}_\gamma\end{bmatrix}^{-1}\begin{bmatrix}\tfrac{1}{T}\mathbf{X}_0'\mathbf{X}\\ \tfrac{1}{T}\mathbf{X}_0'\mathbf{X}_\gamma\end{bmatrix}\\ &\xrightarrow{a.s.}\left[\overline{\mathbf{M}}(\boldsymbol{\gamma}^0)\ \overline{\mathbf{M}}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0)\right]\begin{bmatrix}\mathbf{M} & \overline{\mathbf{M}}(\boldsymbol{\gamma})\\ \overline{\mathbf{M}}(\boldsymbol{\gamma}) & \overline{\mathbf{M}}(\boldsymbol{\gamma})\end{bmatrix}^{-1}\begin{bmatrix}\overline{\mathbf{M}}(\boldsymbol{\gamma}^0)\\ \overline{\mathbf{M}}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0)\end{bmatrix}. \end{aligned}$$

Furthermore, we have

$$\begin{aligned} &\left[\overline{\mathbf{M}}(\boldsymbol{\gamma}^0)\ \overline{\mathbf{M}}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0)\right]\begin{bmatrix}\mathbf{M} & \overline{\mathbf{M}}(\boldsymbol{\gamma})\\ \overline{\mathbf{M}}(\boldsymbol{\gamma}) & \overline{\mathbf{M}}(\boldsymbol{\gamma})\end{bmatrix}^{-1}\begin{bmatrix}\overline{\mathbf{M}}(\boldsymbol{\gamma}^0)\\ \overline{\mathbf{M}}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0)\end{bmatrix}\\ &= \left[\overline{\mathbf{M}}(\boldsymbol{\gamma}^0)\ \overline{\mathbf{M}}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0)\right]\begin{bmatrix}\mathbf{0}_{m\times m} & \mathbf{I}_m\\ \mathbf{I}_m & -\mathbf{I}_m\end{bmatrix}\begin{bmatrix}\mathbf{0}_{m\times m} & \mathbf{I}_m\\ \mathbf{I}_m & -\mathbf{I}_m\end{bmatrix}^{-1}\\ &\times\begin{bmatrix}\mathbf{M} & \overline{\mathbf{M}}(\boldsymbol{\gamma})\\ \overline{\mathbf{M}}(\boldsymbol{\gamma}) & \overline{\mathbf{M}}(\boldsymbol{\gamma})\end{bmatrix}^{-1}\\ &\times\begin{bmatrix}\mathbf{0}_{m\times m} & \mathbf{I}_m\\ \mathbf{I}_m & -\mathbf{I}_m\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{0}_{m\times m} & \mathbf{I}_m\\ \mathbf{I}_m & -\mathbf{I}_m\end{bmatrix}\begin{bmatrix}\overline{\mathbf{M}}(\boldsymbol{\gamma}^0)\\ \overline{\mathbf{M}}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0)\end{bmatrix}\\ &= \left[\overline{\mathbf{M}}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0)\ \overline{\mathbf{M}}(\boldsymbol{\gamma}^0)-\overline{\mathbf{M}}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0)\right]\begin{bmatrix}\overline{\mathbf{M}}(\boldsymbol{\gamma})^{-1} & \mathbf{0}_{m\times m}\\ \mathbf{0}_{m\times m} & (\mathbf{M}-\overline{\mathbf{M}}(\boldsymbol{\gamma}))^{-1}\end{bmatrix}\\ &\times\begin{bmatrix}\overline{\mathbf{M}}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0)\\ \overline{\mathbf{M}}(\boldsymbol{\gamma}^0)-\overline{\mathbf{M}}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0)\end{bmatrix}\\ &= \overline{\mathbf{M}}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0)\overline{\mathbf{M}}(\boldsymbol{\gamma})^{-1}\overline{\mathbf{M}}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0)+[\overline{\mathbf{M}}(\boldsymbol{\gamma}^0)-\overline{\mathbf{M}}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0)][\mathbf{M}-\overline{\mathbf{M}}(\boldsymbol{\gamma})]^{-1}\\ &\times[\overline{\mathbf{M}}(\boldsymbol{\gamma}^0)-\overline{\mathbf{M}}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0)]\\ &\equiv \mathbf{b}_2(\boldsymbol{\gamma},\boldsymbol{\gamma}^0). \end{aligned}\tag{A7}$$

From (A7), we obtain that $\mathbf{b}_2(\boldsymbol{\gamma}^0,\boldsymbol{\gamma}^0)=\overline{\mathbf{M}}(\boldsymbol{\gamma}^0)$ when $\boldsymbol{\gamma}=\boldsymbol{\gamma}^0$. Hence, we have

$$T^{-1}\mathbf{c}'\mathbf{X}_0'(\mathbf{I}_T-\mathbf{P}_\gamma)\mathbf{X}_0\mathbf{c}\xrightarrow{a.s.}\mathbf{c}'(\overline{\mathbf{M}}(\boldsymbol{\gamma}^0)-\mathbf{b}_2(\boldsymbol{\gamma},\boldsymbol{\gamma}^0))\mathbf{c}\equiv b(\boldsymbol{\gamma},\boldsymbol{\gamma}^0) \tag{A8}$$

It is easily seen that $b(\boldsymbol{\gamma}^0,\boldsymbol{\gamma}^0)=0$ when $\boldsymbol{\gamma}=\boldsymbol{\gamma}^0$. On the other hand, since $(\mathbf{I}_T-\mathbf{P}_\gamma)(\mathbf{I}_T-\mathbf{P}_\gamma)=(\mathbf{I}_T-\mathbf{P}_\gamma)$, we therefore obtain

$$\begin{aligned} T^{-1}\mathbf{c}'\mathbf{X}_0'(\mathbf{I}_T-\mathbf{P}_\gamma)\mathbf{X}_0\mathbf{c} &= T^{-1}\mathbf{c}'\mathbf{X}_0'(\mathbf{I}_T-\mathbf{P}_\gamma)(\mathbf{I}_T-\mathbf{P}_\gamma)\mathbf{X}_0\mathbf{c}\\ &= T^{-1}\mathbf{w}'\mathbf{w}\\ &= T^{-1}\sum_{t=1}^{T}w_t^2\ge 0. \end{aligned}\tag{A9}$$

DOI: 10.1111/jtsa.12574

Thus, we have

$$T^{2\alpha-1}\{SSR_T(\gamma_t) - SSR_T(\gamma_t^0)\} = T^{-1}\mathbf{c}'\mathbf{X}_0'(\mathbf{I}_T - \mathbf{P}_{\gamma})\mathbf{X}_0\mathbf{c} + O_p(T^{\alpha-\frac{1}{2}}) \\ \xrightarrow{p} b(\boldsymbol{\gamma}, \boldsymbol{\gamma}^0) \geq 0, \quad \text{(A10)}$$

where the equality holds if and only if $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0$. Since $\boldsymbol{\gamma}$ minimizes $SSR_T(\gamma_t) - SSR_T(\gamma_t^0)$, using Theorem 2.1 of Newey and McFadden (1994) we have $\hat{\boldsymbol{\gamma}} \rightarrow_p \boldsymbol{\gamma}^0$. This completes the proof of Theorem 1. ■

To prove Theorem 2, the following lemmas are needed.

Lemma A2. For any two random variables X and Y defined in the same probability space $(\Omega, \mathcal{F}, P)$:

(i). if $X(\omega) \leq Y(\omega)$, $\forall \omega \in \Omega$, then $E(X) \leq E(Y)$.
(ii). if $X(\omega) \leq Y(\omega')$, $\forall \omega, \omega' \in \Omega$, then $E(X) \leq E(Y)$.[13]

Lemma A3. Given that $f_n(x)$ and $f(x)$ are functions of x, if $f_n(x)$ and $f(x)$ are monotonic on a compact set $\mathbf{X}$, $f(x)$ is continuous function and $f_n(x)$ converges pointwise to $f(x)$ in probability, then $f_n(x)$ converges uniformly to $f(x)$ in probability.

Lemma A4. There is a $C_1 < \infty$ such that for $\gamma_{i1}, \gamma_{i2} \in \Gamma_i$, $i = 0, 1, 2$, and $r \leq 4$,

$$E\{\|\mathbf{x}_t\|^r |d_t(\gamma_t(\gamma_{i2})) - d_t(\gamma_t(\gamma_{i1}))|\} \leq C_1 |\gamma_{i2} - \gamma_{i1}|, \quad \text{(A11)}$$

$$E\{\|\mathbf{x}_t e_t\|^r |d_t(\gamma_t(\gamma_{i2})) - d_t(\gamma_t(\gamma_{i1}))|\} \leq C_1 |\gamma_{i2} - \gamma_{i1}|. \quad \text{(A12)}$$

Lemma A5. There is a $K_1 < \infty$ such that for $\gamma_{i1}, \gamma_{i2} \in \Gamma_i$, $i = 0, 1, 2$,

$$E\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[[\mathbf{c}'\mathbf{x}_t]^2|\Delta_t(\gamma_{i1}, \gamma_{i2})| - E([\mathbf{c}'\mathbf{x}_t]^2|\Delta_t(\gamma_{i1}, \gamma_{i2})|)]\right|^2 \leq K_1 |\gamma_{i2} - \gamma_{i1}|, \quad \text{(A13)}$$

and

$$E\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[[\mathbf{c}'\mathbf{x}_t e_t]^2|\Delta_t(\gamma_{i1}, \gamma_{i2})| - E([\mathbf{c}'\mathbf{x}_t e_t]^2|\Delta_t(\gamma_{i1}, \gamma_{i2})|)]\right|^2 \leq K_1 |\gamma_{i2} - \gamma_{i1}|. \quad \text{(A14)}$$

Lemma A6. There are finite constants K_1 and K_2 such that for all γ_{i1}, $\epsilon > 0$, $\eta > 0$, and $\delta \geq T^{-1}$, if $\sqrt{T} \geq K_2/\eta$, then

$$P(\sup_{\gamma_{i1} \leq \gamma_i \leq \gamma_{i1}+\delta} |J_T(\gamma_t^0(\gamma_i)) - J_T(\gamma_t^0(\gamma_{i1}))| > \eta) \leq \frac{K_1 \delta^2}{\eta^4}. \quad \text{(A15)}$$

Lemma A7. Under Assumption 1, we have

$$T^{\alpha}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) = o_p(1), \;\; T^{\alpha}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T) = o_p(1) \;\text{ or }\; T^{\alpha}\hat{\boldsymbol{\delta}} p \rightarrow \mathbf{c}. \quad \text{(A16)}$$

Lemma A8. Under Assumption 1, on a compact set Ψ_i, we have the following uniform convergence results

$$G_T(\gamma_t^0(\omega_i)) \xrightarrow{p} |\omega_i|\mu_i, \;\; \text{for } i = 0, 1, 2, \quad \text{(A17)}$$

[13] For example, let $Y = X + 1$, which is Case (i) in Lemma 1; let $X \sim \mathcal{U}(0, 1)$ and $Y \sim \mathcal{U}(1, 2)$, which is Case (ii), where $\mathcal{U}(a, b)$ is the continuous uniform distribution.

$$G_T^m(\gamma_t^0(\omega_i))p\rightarrow -\omega_i\mu_i 1_{(\omega_i\leq 0)}-\omega_i\mu_i^m, \text{ for } i=0,1,2, \tag{A18}$$

$$V_T(\gamma_t^0(\omega_i))p\rightarrow|\omega_i|\lambda_i, \text{ for } i=0,1,2, \tag{A19}$$

where

$$\begin{aligned}
\mu_0 &= \mathbf{c}'\left(\int_0^1 \mathbf{D}_\tau f_\tau \mathrm{d}\tau\right)\mathbf{c} \qquad, \lambda_0 = \mathbf{c}'\left(\int_0^1 \mathbf{V}_\tau f_\tau \mathrm{d}\tau\right)\mathbf{c},\\
\mu_1 &= \mathbf{c}'\left(\int_0^1 |\sin(2\pi k^0\tau)|\mathbf{D}_\tau f_\tau d\tau\right)\mathbf{c}, \lambda_1 = \mathbf{c}'\left(\int_0^1 |\sin(2\pi k^0\tau)|\mathbf{V}_\tau f_\tau \mathrm{d}\tau\right)\mathbf{c},\\
\mu_2 &= \mathbf{c}'\left(\int_0^1 |\cos(2\pi k^0\tau)|\mathbf{D}_\tau f_\tau \mathrm{d}\tau\right)\mathbf{c}, \lambda_2 = \mathbf{c}'\left(\int_0^1 |\cos(2\pi k^0\tau)|\mathbf{V}_\tau f_\tau \mathrm{d}\tau\right)\mathbf{c},
\end{aligned}$$

$$\begin{aligned}
\mu_0^m &= 0,\\
\mu_1^m &= \mathbf{c}'\left(\int_0^1 \min\{\sin(2\pi k^0\tau),0\}\mathbf{D}_\tau f_\tau \mathrm{d}\tau\right)\mathbf{c},\\
\mu_1^m &= \mathbf{c}'\left(\int_0^1 \min\{\cos(2\pi k^0\tau),0\}\mathbf{D}_\tau f_\tau \mathrm{d}\tau\right)\mathbf{c},
\end{aligned}$$

in which $\mathbf{D}_\tau = \mathbf{D}(\gamma_\tau^0)$, $\mathbf{V}_\tau = \mathbf{V}(\gamma_\tau^0)$, $f_\tau = f(\gamma_\tau^0)$ and $\gamma_\tau^0 = \gamma_0^0 + \gamma_1^0\sin(2\pi k^0\tau) + \gamma_2^0\cos(2\pi k^0\tau)$.

Lemma A9. Under Assumption 1, on any given compact set Ψ_i, we have

$$R_T(\gamma_t^0(\omega_i)) \Rightarrow \sqrt{\lambda_i}\mathbb{W}(\omega_i), \text{ for } i=0,1,2, \tag{A20}$$

where $\mathbb{W}(\omega_i)$ is a two-sided Brownian motion.

Proof of Lemmas 2–9. The proofs of Lemmas 2–9 are skipped here and are available from the authors on request. ■

Proof of Theorem 2. We first derive the convergence rate of the threshold estimators $a_T(\hat{\gamma}_i - \gamma_i^0) = O_p(1)$ for $i=0,1,2$. To prove this, we need to prove that, for some $\overline{\upsilon}_i > 0$ we have

$$\lim_{T\to\infty} P\left(|\hat{\gamma}_i - \gamma_i^0| \leq \frac{\overline{\upsilon}_i}{a_T}\right) = 1, \text{ for } i=0,1,2. \tag{A21}$$

For any $B>0$, define $V_B = \{(k,\gamma_0,\gamma_1,\gamma_2) : |\gamma_j - \gamma_j^0| \leq B, j=0,1,2, |k-k^0| \leq B\}$. Then when the sample size T is large enough, we have $\overline{\upsilon}_i/a_T < B$. By Theorem 1, we have $\hat{\boldsymbol{\gamma}}\rightarrow_p\boldsymbol{\gamma}^0$, and hence $\lim_{T\to\infty}P(\hat{\boldsymbol{\gamma}} \in V_B) = 1$. Therefore, we only need to examine the limiting behavior in V_B.

Define a subset of $V_B : V_{iB}(\overline{\upsilon}_i) = \{(k,\gamma_0,\gamma_1,\gamma_2) : |\gamma_j - \gamma_j^0| \leq B, j=0,1,2, |k-k^0| \leq B, \overline{\upsilon}_i/a_T < |\gamma_i - \gamma_i^0|\}$. To prove $\lim_{T\to\infty}P(|\hat{\gamma}_i - \gamma_i^0| < \overline{\upsilon}_i/a_T) = 1$, for $i=0,1,2$, we just need to prove $\lim_{T\to\infty}P(\hat{\boldsymbol{\gamma}} \in V_{iB}(\overline{\upsilon})) = 0$.

Let $\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}}$ be the estimation of $y_t = \boldsymbol{\beta}_2'\mathbf{x}_t + \boldsymbol{\delta}_T'\mathbf{x}_t 1(\gamma_t) + e_t$. Denote $S_T^*(\gamma_t) = S_T(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}}, \gamma_t)$ and $S_T^*(\gamma_t^0) = S_T(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}}, \gamma_t^0)$. From the estimation procedure of $\hat{\boldsymbol{\gamma}}$, we have $S_T^*(\hat{\gamma}_t) \leq S_T^*(\gamma_t^0)$. Thus it suffices to prove that for any $\boldsymbol{\gamma} \in V_{iB}(\overline{\upsilon}_i)$,

$$\lim_{T\to\infty} P(S_T^*(\gamma_t) - S_T^*(\gamma_t^0) > 0) = 1. \tag{A22}$$

DOI: 10.1111/jtsa.12574

We consider the case of $\gamma_i \gtrless \gamma_i^0$. In this case, the equation (A22) is equivalent to prove

$$\lim_{T\to\infty} P\left(\frac{S_T^*(\gamma_t) - S_T^*(\gamma_t^0)}{a_T(\gamma_i - \gamma_i^0)} \gtrless 0\right) = 1, \quad \text{for } i = 0, 1, 2. \tag{A23}$$

Since the true model can be rewritten as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}_2 + \mathbf{X}_0\boldsymbol{\delta}_T + \mathbf{e}$, thus we have

$$\begin{aligned}\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_2 - \mathbf{X}_\gamma\hat{\boldsymbol{\delta}} &= \mathbf{X}\boldsymbol{\beta}_2 + \mathbf{X}_0\boldsymbol{\delta}_T + \mathbf{e} - \mathbf{X}\hat{\boldsymbol{\beta}}_2 - \mathbf{X}_\gamma\hat{\boldsymbol{\delta}} \\ &= \mathbf{e} - \mathbf{X}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) - \mathbf{X}_0(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T) - \Delta\mathbf{X}_\gamma\hat{\boldsymbol{\delta}},\end{aligned}$$

where $\Delta\mathbf{X}_\gamma = \mathbf{X}_\gamma - \mathbf{X}_0$. Therefore,

$$\begin{aligned}&S_T^*(\gamma_t) - S_T^*(\gamma_t^0) \\ &\quad = [\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_2 - \mathbf{X}_\gamma\hat{\boldsymbol{\delta}}]'[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_2 - \mathbf{X}_\gamma\hat{\boldsymbol{\delta}}] - [\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_2 - \mathbf{X}_0\hat{\boldsymbol{\delta}}]'[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_2 - \mathbf{X}_0\hat{\boldsymbol{\delta}}] \\ &\quad = \hat{\boldsymbol{\delta}}'\Delta\mathbf{X}_\gamma'\Delta\mathbf{X}_\gamma\hat{\boldsymbol{\delta}} - 2\hat{\boldsymbol{\delta}}'\Delta\mathbf{X}_\gamma'\mathbf{e} + 2\hat{\boldsymbol{\delta}}'\Delta\mathbf{X}_\gamma'\mathbf{X}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) + 2\hat{\boldsymbol{\delta}}'\Delta\mathbf{X}_\gamma'\mathbf{X}_0(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T) \\ &\quad = S_1 - S_2 - S_3 + S_4 + S_5 + S_6,\end{aligned} \tag{A24}$$

where

$$\begin{aligned}S_1 &= \boldsymbol{\delta}_T'\Delta\mathbf{X}_\gamma'\Delta\mathbf{X}_\gamma\boldsymbol{\delta}_T, \\ S_2 &= 2\boldsymbol{\delta}_T'\Delta\mathbf{X}_\gamma'\mathbf{e}, \\ S_3 &= 2(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T)'\Delta\mathbf{X}_\gamma'\mathbf{e}, \\ S_4 &= 2\hat{\boldsymbol{\delta}}'\Delta\mathbf{X}_\gamma'\mathbf{X}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2), \\ S_5 &= 2\hat{\boldsymbol{\delta}}'\Delta\mathbf{X}_\gamma'\mathbf{X}_0(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T), \\ S_6 &= (\boldsymbol{\delta}_T + \hat{\boldsymbol{\delta}})'\Delta\mathbf{X}_\gamma'\Delta\mathbf{X}_\gamma(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T).\end{aligned}$$

Using a first-order Taylor series expansion and by Assumption 6 ($\boldsymbol{\delta}_T = \mathbf{c}T^{-\alpha}$) and Lemma 7, we have $\hat{\mathbf{c}} = \hat{\boldsymbol{\delta}}T^{\alpha}$, $\hat{\boldsymbol{\delta}} = \hat{\mathbf{c}}T^{-\alpha}$. Thus, we can show

$$\begin{aligned}\frac{S_1}{a_T(\gamma_i - \gamma_i^0)} &= \frac{\boldsymbol{\delta}_T'\Delta\mathbf{X}_\gamma'\Delta\mathbf{X}_\gamma\boldsymbol{\delta}_T}{a_T(\gamma_i - \gamma_i^0)} = \mathbf{c}'\frac{1}{T}\sum_{t=1}^{T}\left|\frac{\mathbf{x}_t\mathbf{x}_t'(\{q_t \le \gamma_t\} - \{q_t \le \gamma_t^0\})}{\gamma_i - \gamma_i^0}\right|\mathbf{c}\,\mathrm{sgn}(\gamma_i - \gamma_i^0) \\ &= O_p(1) \gtrless 0, \text{ if } \gamma_i \gtrless \gamma_i^0,\end{aligned} \tag{A25}$$

$$\begin{aligned}\frac{S_2}{a_T(\gamma_i - \gamma_i^0)} &= \frac{2\boldsymbol{\delta}_T'\Delta\mathbf{X}_\gamma'\mathbf{e}}{a_T(\gamma_i - \gamma_i^0)} = \sqrt{\frac{a_T}{\gamma_i - \gamma_i^0}}\frac{\mathrm{sgn}(\gamma_i - \gamma_i^0)}{\sqrt{T(\gamma_i - \gamma_i^0)}}\mathbf{c}'\Delta\mathbf{X}_\gamma'\mathbf{e} \\ &= O_p(\overline{v}_i^{-1/2})O_p(1),\end{aligned} \tag{A26}$$

$$\frac{S_3}{a_T(\gamma_i - \gamma_i^0)} = \frac{2(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T)'\Delta\mathbf{X}_\gamma'\mathbf{e}}{a_T(\gamma_i - \gamma_i^0)} = T^{\alpha}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T)'\sqrt{\frac{a_T}{\gamma_i - \gamma_i^0}}\frac{\mathrm{sgn}(\gamma_i - \gamma_i^0)}{\sqrt{T(\gamma_i - \gamma_i^0)}}\Delta\mathbf{X}_\gamma'\mathbf{e}$$

DOI: 10.1111/jtsa.12574

$$= o_p(1)O_p(\overline{v}_i^{-1/2})O_p(1) = o_p(1), \tag{A27}$$

$$\begin{aligned}\frac{S_4}{a_T(\gamma_i - \gamma_i^0)} &= \frac{2\hat{\boldsymbol{\delta}}'\Delta\mathbf{X}_\gamma'\mathbf{X}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)}{a_T(\gamma_i - \gamma_i^0)} = \hat{\boldsymbol{\delta}}'T^\alpha\frac{2}{T(\gamma_i - \gamma_i^0)}\Delta\mathbf{X}_\gamma'\mathbf{X}T^\alpha(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \\ &= O_p(1)O_p(1)o_p(1) = o_p(1),\end{aligned} \tag{A28}$$

$$\begin{aligned}\frac{S_5}{a_T(\gamma_i - \gamma_i^0)} &= \frac{2\hat{\boldsymbol{\delta}}'\Delta\mathbf{X}_\gamma'\mathbf{X}_0(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T)}{a_T(\gamma_i - \gamma_i^0)} = \hat{\boldsymbol{\delta}}'T^\alpha\frac{2}{T(\gamma_i - \gamma_i^0)}\Delta\mathbf{X}_\gamma'\mathbf{X}_0T^\alpha(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T) \\ &= O_p(1)O_p(1)o_p(1) = o_p(1),\end{aligned} \tag{A29}$$

$$\begin{aligned}\frac{S_6}{a_T(\gamma_i - \gamma_i^0)} &= \frac{(\boldsymbol{\delta}_T + \hat{\boldsymbol{\delta}})'\Delta\mathbf{X}_\gamma'\Delta\mathbf{X}_\gamma(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T)}{a_T(\gamma_i - \gamma_i^0)} \\ &= [\mathbf{c} + \hat{\boldsymbol{\delta}}T^\alpha]'\frac{1}{T(\gamma_i - \gamma_i^0)}\Delta\mathbf{X}_\gamma'\mathbf{X}_\gamma T^\alpha(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T) \\ &= O_p(1)O_p(1)o_p(1) = o_p(1).\end{aligned} \tag{A30}$$

Hence, it is possible to find a $\overline{v}_i < \infty$ and when $T \to \infty$ such that

$$\left|\frac{S_1}{a_T(\gamma_i - \gamma_i^0)}\right| > \left|\sum_{k=2}^{6}\frac{S_k}{a_T(\gamma_i - \gamma_i^0)}\right| \tag{A31}$$

holds with probability one. By (A31) we can obtain (A23). Thus we obtain the convergence rate.

We next establish the asymptotic distribution:

$$a_T(\hat{\gamma}_i - \gamma_i^0)d{\rightarrow}\underset{r\in\mathbb{R}}{\arg\max}\,Q_i(r), \quad \text{for } i = 0, 1, 2. \tag{A32}$$

Since the threshold parameters are consistent with convergence rate $a_T = T^{1-2\alpha}$. Thus we can study their asymptotic behavior in the neighborhood of the true thresholds, $\hat{\gamma}_i = \gamma_i^0 + \hat{\omega}_i/a_T$, for $i = 0, 1, 2$.

By the definition of the threshold estimator, we have

$$\begin{aligned}a_T(\hat{\gamma}_i - \gamma_i^0) \equiv \hat{\omega}_i &= \underset{\omega_i\in\Psi_i}{\arg\min}\,S_T^{*0}\left(\gamma_i^0 + \frac{\omega_i}{a_T}\right) - S_T^{*0}(\gamma_i^0) \\ &= \underset{\omega_i\in\Psi_i}{\arg\max} - \left\{S_T^{*0}\left(\gamma_i^0 + \frac{\omega_i}{a_T}\right) - S_T^{*0}(\gamma_i^0)\right\} \\ &\equiv \underset{\omega_i\in\Psi_i}{\arg\max}\,Q_T^0(\omega_i), \quad \text{for } i = 0, 1, 2,\end{aligned}$$

where $S_T^{*0}(\gamma_0) = S_T^*(\gamma_0, \gamma_1^0, \gamma_2^0, k^0)$, $S_T^{*0}(\gamma_1) = S_T^*(\gamma_0^0, \gamma_1, \gamma_2^0, k^0)$, $S_T^{*0}(\gamma_2) = S_T^*(\gamma_0^0, \gamma_1^0, \gamma_2, k^0)$ and $\Psi_i = [a_T(\underline{\gamma}_i - \gamma_i^0), a_T(\overline{\gamma}_i - \gamma_i^0)]$.

DOI: 10.1111/jtsa.12574

As above, we have $S_T^{*0}(\gamma_i^0+\omega_i/a_T)-S_T^{*0}(\gamma_i^0)=S_1-S_2-S_3+S_4+S_5+S_6$. We next derive the limiting behavior of each S_i respectively.

$$\begin{aligned} S_1 &= \boldsymbol{\delta}_T'\Delta\mathbf{X}_{\gamma_i}'\Delta\mathbf{X}_{\gamma_i}\boldsymbol{\delta}_T = \frac{1}{T^{2\alpha}}\mathbf{c}'\Delta\mathbf{X}_{\gamma_i}'\Delta\mathbf{X}_{\gamma_i}\mathbf{c} \\ &= \frac{a_T}{T}\sum_{t=1}^{T}\mathbf{c}'\mathbf{x}_t\mathbf{x}_t'\mathbf{c}|\Delta_t^0(\gamma_i^0,\gamma_i)| = G_T(\gamma_t^0(\omega_i)), \end{aligned} \tag{A33}$$

where $\Delta\mathbf{X}_{\gamma_i[t]}=[\mathbf{X}_{\gamma_i}-\mathbf{X}_0]_{[t]}=\mathbf{x}_t d_t(\gamma_t^0(\gamma_i))-\mathbf{x}_t d_t(\gamma_t^0(\gamma_i^0))=\mathbf{x}_t[d_t(\gamma_t^0(\gamma_i))-d_t(\gamma_t^0(\gamma_i^0))]$.

$$\begin{aligned} S_2 &= 2\boldsymbol{\delta}_T'\Delta\mathbf{X}_{\gamma_i}'\mathbf{e} = 2\frac{1}{T^{\alpha}}\mathbf{c}'\Delta\mathbf{X}_{\gamma_i}'\mathbf{e} \\ &= 2\sqrt{\frac{a_T}{T}}\sum_{t=1}^{T}\mathbf{c}'\mathbf{x}_t e_t\Delta_t^0(\gamma_i^0,\gamma_i) = 2R_T(\gamma_t^0(\omega_i)). \end{aligned} \tag{A34}$$

Note that

$$\begin{aligned} \left|\frac{a_T}{T}\mathbf{c}'\Delta\mathbf{X}_{\gamma_i}'\mathbf{X}_0\mathbf{c}\right| &= \left|\frac{a_T}{T}\sum_{t=1}^{T}\mathbf{c}'\mathbf{x}_t\mathbf{x}_t'\mathbf{c}[d_t(\min\{\gamma_t^0(\gamma_i),\gamma_t^0(\gamma_i^0)\})-d_t(\gamma_t^0(\gamma_i^0))]\right| \\ \left|\frac{a_T}{T}\mathbf{c}'\Delta\mathbf{X}_{\gamma_i}'\mathbf{X}_0\mathbf{c}\right| &\le \frac{a_T}{T}\sum_{t=1}^{T}\mathbf{c}'\mathbf{x}_t\mathbf{x}_t'\mathbf{c}|d_t(\min\{\gamma_t^0(\gamma_i),\gamma_t^0(\gamma_i^0)\})-d_t(\gamma_t^0(\gamma_i^0))| \\ &= G_T^m(\gamma_t^0(\omega_i)) = O_p(1), \end{aligned} \tag{A35}$$

$$\begin{aligned} \left|\frac{a_T}{T}\mathbf{c}'\Delta\mathbf{X}_{\gamma_i}'\mathbf{X}\mathbf{c}\right| &= \left|\frac{a_T}{T}\sum_{t=1}^{T}\mathbf{c}'\mathbf{x}_t\mathbf{x}_t'\mathbf{c}[d_t(\gamma_t^0(\gamma_i))-d_t(\gamma_t^0(\gamma_i^0))]\right| \\ &\le \frac{a_T}{T}\sum_{t=1}^{T}\mathbf{c}'\mathbf{x}_t\mathbf{x}_t'\mathbf{c}|d_t(\gamma_t^0(\gamma_i))-d_t(\gamma_t^0(\gamma_i^0))| \\ &= G_T(\gamma_t^0(\omega_i)) = O_p(1), \\ \frac{1}{T^{\alpha}}\mathbf{c}'\Delta\mathbf{X}_{\gamma_i}'\mathbf{e} &= R_T(\gamma_t^0(\omega_i)) = O_p(1), \\ \frac{a_T}{T}\mathbf{c}'\Delta\mathbf{X}_{\gamma_i}'\Delta\mathbf{X}_{\gamma_i}\mathbf{c} &= G_T(\gamma_t^0(\omega_i)) = O_p(1), \end{aligned} \tag{A36}$$

where $\mathbf{X}_{[t]}=\mathbf{x}_t$ and $\mathbf{X}_{0[t]}=\mathbf{x}_t d_t(\gamma_t^0(\gamma_i^0))$, and we have $\mathbf{c}=O(1)$ by Assumption 1, hence

$$\begin{aligned} &\frac{a_T}{T}\Delta\mathbf{X}_{\gamma_i}'\mathbf{X}_0 = O_p(1), \frac{a_T}{T}\Delta\mathbf{X}_{\gamma_i}'\mathbf{X} = O_p(1), \\ &\frac{1}{T^{\alpha}}\Delta\mathbf{X}_{\gamma_i}'\mathbf{e} = O_p(1), \frac{a_T}{T}\Delta\mathbf{X}_{\gamma_i}'\Delta\mathbf{X}_{\gamma_i} = O_p(1). \end{aligned} \tag{A37}$$

Next, we show that $|-S_3+S_4+S_5+S_6|=o_p(1)$.

$$|-S_3+S_4+S_5+S_6|$$

DOI: 10.1111/jtsa.12574

$$
\begin{aligned}
&\leq |S_3| + |S_4| + |S_5| + |S_6|,\\
&= 2|(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T)' \Delta \mathbf{X}'_{\gamma_i} \mathbf{e}| + 2|\hat{\boldsymbol{\delta}}' \Delta \mathbf{X}'_{\gamma_i} \mathbf{X}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)|\\
&\quad + 2|\hat{\boldsymbol{\delta}}' \Delta \mathbf{X}'_{\gamma_i} \mathbf{X}_0(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})| + |(\boldsymbol{\delta}_T + \hat{\boldsymbol{\delta}})' \Delta \mathbf{X}'_{\gamma_i} \Delta \mathbf{X}_{\gamma_i}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T)|,\\
&= 2\left| T^{\alpha}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T)' \frac{1}{T^{\alpha}} \Delta \mathbf{X}'_{\gamma_i} \mathbf{e} \right| + 2\left| T^{\alpha}\hat{\boldsymbol{\delta}}' \frac{a_T}{T} \Delta \mathbf{X}'_{\gamma_i} \mathbf{X} T^{\alpha}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \right|\\
&\quad + 2\left| T^{\alpha}\hat{\boldsymbol{\delta}}' \frac{a_T}{T} \Delta \mathbf{X}'_{\gamma_i} \mathbf{X}_0 T^{\alpha}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \right| + \left| (\mathbf{c} + T^{\alpha}\hat{\boldsymbol{\delta}})' \frac{a_T}{T} \Delta \mathbf{X}'_{\gamma_i} \Delta \mathbf{X}_{\gamma_i} T^{\alpha}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_T) \right|,\\
&= 2o_p(1)O_p(1) + 2O_p(1)O_p(1)o_p(1) + 2O_p(1)O_p(1)o_p(1)\\
&\quad + O_p(1)O_p(1)o_p(1),\\
&= o_p(1). \qquad \text{(A38)}
\end{aligned}
$$

Thus, by Lemmas 8 and 9, we have

$$
Q_T^0(\omega_i) \Rightarrow -|\omega_i|\mu_i + 2\sqrt{\lambda_i}\mathbb{W}(\omega_i), \quad \text{for } i = 0, 1, 2, \qquad \text{(A39)}
$$

Making the change of variable $\omega_i = r\lambda_i/\mu_i^2$ and setting $\varpi_i = 2\lambda_i/\mu_i$, we can show that

$$
\begin{aligned}
a_T(\hat{\gamma}_i - \gamma_i^0)d &\rightarrow \underset{-\infty<\omega_i<\infty}{\arg\max}[2\sqrt{\lambda_i}\mathbb{W}(\omega_i) - |\omega_i|\mu_i]\\
&= \frac{\lambda_i}{\mu_i^2} \underset{-\infty<r<\infty}{\arg\max} \left[2\frac{\lambda_i}{\mu_i}\mathbb{W}(r) - \frac{\lambda_i}{\mu_i}|r| \right]\\
&= \varpi_i \underset{-\infty<r<\infty}{\arg\max} \left[\mathbb{W}(r) - \frac{1}{2}|r| \right]. \qquad \text{(A40)}
\end{aligned}
$$

This completes the proof of Theorem 2. ■

REFERENCES

Andrews DWK, Ploberger W. 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* **62**:1383–1414.

Becker R, Enders W, Lee J. 2006. A stationarity test in the presence of an unknown number of smooth breaks. *Journal of Time Series Analysis* **27**:381–409.

Bessec M. 2003. The asymmetric exchange rate dynamics in the EMS: a time-varying threshold test. *European Review of Economics and Finance* **2**:3–40.

Caner M, Hansen BE. 2001. Threshold autoregression with a unit root. *Econometrica* **69**:1555–1596.

Chen H. 2015. Robust estimation and inference for threshold models with integrated regressors. *Econometric Theory* **31**:778–810.

Davies RB. 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74**:33–43.

Dueker M, Owyang M, Sola M. 2010. *A Time-Varying Threshold Star Model of Unemployment and the Natural Rate*. Working Papers 2010-029, Federal Reserve Bank of St. Louis. Available at SSRN: https://ssrn.com/abstract=1679954 or http://dx.doi.org/10.2139/ssrn.1679954

Dueker MJ, Psaradakis Z, Sola M. 2013. State-dependent threshold smooth transition autoregressive models. *Oxford Bulletin of Economics and Statistics* **75**:835–854.

Enders W, Lee J. 2012a. A unit root test using a fourier series to approximate smooth breaks. *Oxford Bulletin of Economics and Statistics* **74**:574–599.

Enders W, Lee J. 2012b. The flexible Fourier form and Dickey–Fuller type unit root tests. *Economics Letters* **117**:196–199.

Feng G, Gao J, Peng B. 2017. A varying-coefficient panel data model with fixed effects: theory and an application to US commercial banks. *Journal of Econometrics* **196**:68–82.

DOI: 10.1111/jtsa.12574

Gonzalo J, Pitarakis JY. 2006. Threshold effects in cointegrating relationships. *Oxford Bulletin of Economics and Statistics* **68**:813–833.

Gonzalo J, Pitarakis JY. 2012. Regime specific predictability in predictive regressions. *Journal of Business and Economic Statistics* **30**:229–241.

Hansen BE. 1996. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* **64**:413–430.

Hansen BE. 1999. Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics* **93**:345–368.

Hansen BE. 2000. Sample splitting and threshold estimation. *Econometrica* **68**:575–603.

Hansen BE. 2017. Regression kink with an unknown threshold. *Journal of Business and Economic Statistics* **35**:228–240.

Kline P, Santos A. 2012. A score based approach to wild bootstrap inference. *Journal of Econometric Methods* **1**:23–41.

Mammen E. 1993. Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics* **21**:255–285.

Newey WK, McFadden D. 1994. *Large sample estimation and hypothesis*. In *Handbook of Econometrics*, Engle RF, McFadden DL (eds.), Vol. IV Elsevier: New York, pp. 2113–2245.

Porter J, Yu P. 2015. Regression discontinuity designs with unknown discontinuity points: testing and estimation. *Journal of Econometrics* **189**:132–147.

Prodan R. 2008. Potential pitfalls in determining multiple structural changes with an application to purchasing power parity. *Journal of Business and Economic Statistics* **26**:50–65.

Taylor MP, Davradakis E. 2006. Interest rate setting and inflation targeting: evidence of a nonlinear Taylor rule for the United Kingdom. *Studies in Nonlinear Dynamics and Econometrics* **10**:1–18.

Yang L, Su J-J. 2018. Debt and growth: is there a constant tipping point? *Journal of International Money and Finance* **87**:133–143.

Yu P, Fan X. 2020. Threshold regression with a threshold boundary. *Journal of Business and Economic Statistic* Forthcoming. https://doi.org/10.1080/07350015.2020.1740712.

Zhu Y, Chen H, Lin M. 2019. Threshold models with time-varying threshold values and their application in estimating regime-sensitive Taylor rules. *Studies in Nonlinear Dynamics and Econometrics* **23**:1–17.

DOI: 10.1111/jtsa.12574