ISSN 2090-3359 (Print)
ISSN 2090-3367 (Online)

Advances in Decision Sciences

Volume 23
Issue 2
June 2019

Michael McAleer

Editor-in-Chief
University Chair Professor
Asia University, Taiwan

Published by Asia University, Taiwan

ADS@ASIAUNIVERSITY

ON BROWNIAN MOTION APPROXIMATION OF COMPOUND POISSON PROCESSES WITH APPLICATIONS TO THRESHOLD MODELS

Dong Li

Center for Statistical Science and Department of Industrial Engineering

Tsinghua University, China

Shiqing Ling

Department of Statistics, Hong Kong University of Science and Technology

Howell Tong

School of Mathematical Sciences, University of Electronic Science and Technology of China

and

Department of Statistics, London School of Economics and Political Science

Guangren Yang

Department of Statistics, School of Economics, Jinan University, China

Revised: June 2019

* Li's research is supported in part by the NSFC (No.11571348 and No.11771239). Ling's in part by Hong Kong Research Grants Commission (Grants GRF 16500915, 16307516, 16500117 and 16303118) and by Australian Research Council. Tong's in part by a research grant from the University of Electronic Science and Technology of China, Chengdu, China. Yang's research is supported in part by the NSFC (No. 11471086 and No. 11871173), the National Social Science Foundation of China grants 16BTJ032, the National Statistical Scientific Center 2015LD02 and the Fundamental Research Funds for the Central Universities of Jinan University Qimingxing Plan 15JNQM019. This paper was completed in February, 2018, and it is an independent research from Ping Yu and Peter C.B. Phillips (2018) "Threshold regression asymptotics: From the compound Poisson process to two-sided Brownian motion" in *Economics Letters*.

Abstract

Compound Poisson processes (CPP) constitute a fundamental class of stochastic processes and a basic building block for more complex jump-diffusion processes such as the Lévy processes. However, unlike those of a Brownian motion (BM), distributions of functionals, e.g. maxima, passage time, argmin and others, of a CPP are often intractable. The first objective of this paper is to propose a new approximation of a CPP by a BM so as to facilitate closed-form expressions in concrete cases. Specifically, we approximate, in some sense, a sequence of two-sided CPPs by a two-sided BM with drift. The second objective is to illustrate the above approximation in applications, such as the construction of confidence intervals of threshold parameters in threshold models, which include the threshold regression (also called two-phase regression or segmentation) and numerous threshold time series models. We conduct numerical simulations to assess the performance of the proposed approximation. We illustrate the use of our approach with a real data set.

Keywords: Brownian motion, compound Poisson process, TAR, TARMA, TCHARM, TDAR, TMA, threshold regression.

JEL: D31, D63

1. Introduction. In many statistical problems, especially those associated with change points, we often have to derive the distributions of functionals of a compound Poisson process (CPP). Examples include the estimation of the location of a discontinuity in density (e.g., Chernoff and Rubin (1956)), and the estimation of thresholds in threshold regression, also called two-phase regression or segmentation (e.g., Koul and Qian (2002), Seijo and Sen (2011) and Yu (2012, 2015)), threshold autoregressive (TAR) models (e.g., Chan (1993), Tsay (1989, 1998), Gonzalo and Pitarakis (2002) and Li and Ling (2012)), threshold double autoregressive (TDAR) models (e.g., Li, Ling and Zakoïan (2015) and Li, Ling and Zhang (2016)), conditionally heteroscedastic AR models with thresholds (T-CHARM) (e.g., Chan, et al. (2014)), threshold moving-average (TMA) models (e.g., Li, Ling and Li (2013)), threshold autoregressive moving-average (TARMA) models (e.g., Li, Li and Ling (2011)), and others. However, it is typically difficult to derive their distributions in closed-form.

To circumvent the difficulty, an important approach is to approximate the CPP by a Brownian motion (BM) in some sense because the situation is much more tractable in the latter. See, e.g., Stryhn (1996). The primary objective of this paper is to introduce a new approximation of

$$\mathcal{P}(z) = I(z<0)\sum_{k=1}^{N_1(-z)} \zeta_k^{(1)} + I(z\geq 0)\sum_{k=1}^{N_2(z)} \zeta_k^{(2)}, \qquad z\in\mathbb{R}, \tag{1.1}$$

where $\{N_1(z), z\geq 0\}$ and $\{N_2(z), z\geq 0\}$ are independent Poisson processes with rates λ_1 and λ_2, respectively, $\{\zeta_k^{(1)} : k\geq 1\}$ and $\{\zeta_k^{(2)} : k\geq 1\}$ are independent and identically distributed (i.i.d.) sequences with $E\zeta_1^{(i)} > 0$ for $i=1,2$, respectively, and mutually independent. $\{N_i(z)\}$ and $\{\zeta_k^{(j)}\}$ are also mutually independent. Throughout the paper, these assumptions are always supposed to hold.

This paper is organized as follows. We state the main results in Section 2. In Section 3, we describe some important applications in threshold models. In Section 4, we assess the efficacy of the theoretical results of approximation by numerical simulations. A real data set is also included. All proofs of Theorems are in the supplementary material.

2. Main results. Using the parameterizing technique (e.g. Kushner (1984) and Skorokhod, Hoppensteadt and Salehi (2002)), we introduce a new parameter $\gamma > 0$ to re-parameterize (1.1) to

$\{\mathcal{P}_\gamma(z) : z \in \mathbb{R}\}$ to

$$\mathcal{P}_\gamma(z) = I(z < 0) \sum_{k=1}^{N_1(|z|/\gamma)} \xi_k^{(1)} + I(z \geq 0) \sum_{k=1}^{N_2(z/\gamma)} \xi_k^{(2)}, \qquad z \in \mathbb{R}. \tag{2.1}$$

Denote

$$m_\gamma = s\text{-}\arg\min_{z \in \mathbb{R}} \mathcal{P}_\gamma(z),$$

where 's-argmin' stands for the smallest argmin.

Thus, we embed (1.1) into a sequence of $\{\mathcal{P}_\gamma(z) : z \in \mathbb{R}\}$. Our interests are m_γ and the limits of $\{\mathcal{P}_\gamma(z) : z \in \mathbb{R}\}$ as γ shrinks to zero under some suitable conditions. To this end, we first introduce two assumptions.

Assumption 1. *$E\xi_1^{(i)} = a_i\gamma + o(\gamma)$ and $E\{\xi_1^{(i)}\}^2 = b_i\gamma + o(\gamma)$ as $\gamma \to 0$ for some positive constants a_i and b_i, $i = 1, 2$.*

Assumption 2. *The rate of $N_i(\cdot)$ is $\lambda_i > 0$, $i = 1, 2$.*

Note that the choice of λ_i is critical. For example, if the rate of the component Poisson process is chosen as a function such that it tends to a deterministic continuous one, the CPP converges weakly to a CPP. See, e.g., Jacod and Shiryaev (2003). In our approach, λ_i's are fixed constants. The following theorem states that the sequence of stochastic processes $\{\mathcal{P}_\gamma(z) : z \in \mathbb{R}\}$ can be approximated by a two-sided BM with drift.

Theorem 1. *Let $\Longrightarrow$ stand for weak convergence. Let $\mathbb{D}(\mathbb{R})$ denote the space of functions defined on $\mathbb{R}$, which are right continuous and have left limits. Let the space be endowed with the Skorokhod topology. If Assumptions 1-2 hold, then, as $\gamma \to 0$,*

$$\mathcal{P}_\gamma(z) \Longrightarrow \mathbb{W}(z) \quad in\ \mathbb{D}(\mathbb{R}),$$

where

$$\mathbb{W}(z) = \begin{cases} \lambda_1 a_1 |z| - \sqrt{\lambda_1 b_1}\mathbb{B}_1(|z|), & \text{if } z \le 0, \\ \\ \lambda_2 a_2\, z - \sqrt{\lambda_2 b_2}\, \mathbb{B}_2(z), & \text{if } z > 0, \end{cases}$$

with $\mathbb{B}_1(z)$ *and* $\mathbb{B}_2(z)$ *being two independent standard Brownian motions on* $[0, \infty)$*. Further, let* $T := \arg\min_{z\in\mathbb{R}} \mathbb{W}(z)$*. Then* $m_\gamma \Longrightarrow T$*.*

In the literature, the density of T is readily available and has a closed form, which is given in the following theorem; see Proposition 1 in Stryhn (1996).

Theorem 2. *The probability density of* T *is given by*

$$f_T(x; a_i, b_i, \lambda_i) = \begin{cases} g(|x|;\ a_1\sqrt{\lambda_1/b_1},\ (a_2/b_2)\sqrt{\lambda_1 b_1}\,), & \text{for } x < 0, \\ \\ g(\,x;\ \ a_2\sqrt{\lambda_2/b_2},\ (a_1/b_1)\sqrt{\lambda_2 b_2}\,), & \text{for } x \ge 0, \end{cases}$$

where

$$g(x; \theta_1, \theta_2) = 2\theta_1(\theta_1 + 2\theta_2)\exp\{2\theta_2(\theta_1+\theta_2)x\}\Phi(-(\theta_1+2\theta_2)\sqrt{x}) - 2\theta_1^2\Phi(-\theta_1\sqrt{x}), \qquad x \ge 0,$$

and $\Phi(\cdot)$ *is the standard normal distribution.*

Corollary 1. *Suppose that* $\gamma = E\xi_1^{(1)} = E\xi_1^{(2)} > 0$*,* $\lambda_1 = \lambda_2 := \lambda$*, and* $E\{\xi_1^{(i)}\}^2 = b_i\gamma + o(\gamma)$ *as* $\gamma \to 0$ *for positive constants* b_i*,* $i = 1, 2$*. Then*

$$\lambda m_\gamma \Longrightarrow T_1 := \arg\min_{z\in\mathbb{R}} \mathbb{W}^*(z),$$

where

$$\mathbb{W}^*(z) = \begin{cases} |z| - \sqrt{b_1}\mathbb{B}_1(|z|), & \text{if } z \le 0, \\ \\ z - \sqrt{b_2}\,\mathbb{B}_2(z), & \text{if } z > 0, \end{cases}$$

and the density of T_1 *is*

$$f_{T_1}(x; b_1, b_2) = \begin{cases} g(|x|;\ 1/\sqrt{b_1},\ \sqrt{b_1}/b_2\ ), & \text{for } x < 0, \\ \\ g(\,x;\ \ 1/\sqrt{b_2},\ \sqrt{b_2}/b_1\ ), & \text{for } x \geq 0, \end{cases}$$

where $g(\cdot)$ *is as in Theorem 2.*

COROLLARY 2. *Suppose that* $\gamma = E\xi_1^{(1)} = E\xi_1^{(2)} > 0$, $\lambda_1 = \lambda_2 := \lambda$, *and* $E\{\xi_1^{(1)}\}^2 = E\{\xi_1^{(2)}\}^2 = b\gamma + o(\gamma)$ *as* $\gamma \to 0$ *for some positive constant* b. *Then*

$$\frac{4\lambda m_\gamma}{b} \Longrightarrow T_2,$$

where T_2 *has the density*

$$f_{T_2}(x) = \frac{3}{2}\Phi\big(-\frac{3}{2}\sqrt{|x|}\big)e^{|x|} - \frac{1}{2}\Phi\big(-\frac{1}{2}\sqrt{|x|}\big). \tag{2.2}$$

Thus, Theorem 2 includes the density (2.2) as a special case. Yao (1987) used this special case in his study of the approximation of the limiting distribution of the maximum likelihood estimate of a change-point problem. The distribution of T_2 has exponential tails; see Remark 1 in Yao (1987). Note that Theorem 1 includes Theorem 1 of Hansen (2000) as a special case. Figure 1 displays the density and the cumulative distribution function (CDF) of T_2. From Figure 1, we can see that T_2 is *symmetric*. Moreover, for our needs, it is easy to tabulate the quantiles of T_2. For any given level $\underline{\alpha} \in (0, 1)$, denote by $Q_{\underline{\alpha}}$ the $\underline{\alpha}$th quantile of T_2. Table 1 gives some commonly used quantiles.

TABLE 1
The quantiles of T_2.

$\underline{\alpha}$	0.5	0.95	0.975	0.995
$Q_{\underline{\alpha}}$	0	7.6873	11.0333	19.7665

3. Applications.

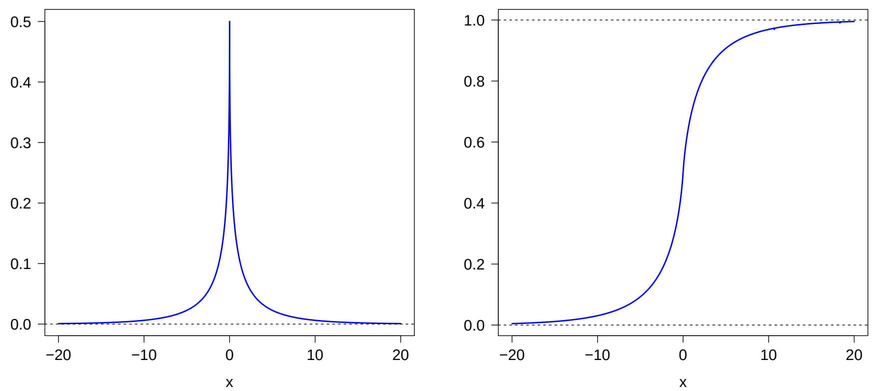

FIG 1. *The density (left panel) and the CDF (right panel) of T_2.*

3.1. *Threshold regression model.* To the best of our knowledge, the threshold regression model, also called the two-phase regression model or the segmentation model, can be dated back to Quandt (1958). Since then it has been widely used in economics and other areas. Asymptotics on statistical inference for such models have been considered; see, e.g., Hinkley (1969, 1971), Hansen (2000), Koul and Qian (2002), Seijo and Sen (2011) and Yu (2012, 2015).

We say $(\mathbf{x}', y, z)$ follows a threshold regression model if

$$y = \begin{cases} \mathbf{x}'\beta_1 + \sigma_1\varepsilon, & \text{if } z \le r, \\ \\ \mathbf{x}'\beta_2 + \sigma_2\varepsilon, & \text{if } z > r, \end{cases} \tag{3.1}$$

where y is a scalar dependent variable and $\mathbf{x} = (x_1, ..., x_p)'$ explanatory variables (or independent variables), z is called the threshold variable and r the threshold parameter, and ε is the error with zero mean and unit variance.

Suppose that $\{(\mathbf{x}_i', y_i, z_i)\}$ is a random sample of size n from model (3.1) with the true parameter $\theta_0 = (\beta_{10}', \beta_{20}', r_0)'$ and $(\sigma_{10}, \sigma_{20})$. Denote by $\widehat{\theta}_n$ the least squares estimator (LSE) of θ_0. Under some conditions (e.g., Koul and Qian (2002)), we have

$$n(\widehat{r}_n - r_0) \Longrightarrow M_- := s\text{-}\arg\min_{z\in\mathbb{R}} \mathcal{P}(z),$$

where

$$\mathcal{P}(z) = I(z \leq 0) \sum_{i=1}^{N_1(-z)} \zeta_i^{(1)} + I(z > 0) \sum_{i=1}^{N_2(z)} \zeta_i^{(2)}. \tag{3.2}$$

Here, $N_1(\cdot)$ and $N_2(\cdot)$ are independent Poisson processes with the same rate $\pi(r_0)$, which is the value of the density $\pi(\cdot)$ of z at r_0, and where $\{\zeta_k^{(1)} : k \geq 1\}$ is a sequence of i.i.d. random variables with the same distribution as the one induced by

$$\zeta^{(1)} = \{\mathbf{x}'(\beta_{10} - \beta_{20})\}^2 + 2\sigma_{10}\varepsilon\{\mathbf{x}'(\beta_{10} - \beta_{20})\} \quad \text{given } z = r_0^-,$$

and the sequence $\{\zeta_k^{(2)} : k \geq 1\}$ by

$$\zeta^{(2)} = \{\mathbf{x}'(\beta_{10} - \beta_{20})\}^2 - 2\sigma_{20}\varepsilon\{\mathbf{x}'(\beta_{10} - \beta_{20})\} \quad \text{given } z = r_0^+.$$

Here, $z = r_0^-$ and $z = r_0^+$ denote convergence to r_0 from below and from above respectively.

Clearly, $E\zeta^{(i)}$ is a function of $\beta_{10} - \beta_{20}$. To obtain an approximation of M_- by Theorem 1 when $\|\beta_{10} - \beta_{20}\|$ is small, we can introduce a new parameter $\gamma = E\zeta^{(1)}$ to re-parameterize the CPP (3.2). Note that unlike Hansen (1997), $\|\beta_{10} - \beta_{20}\|$ is fixed and not sample-size dependent.

By the definitions of m_γ and M_-, we have $m_\gamma = \gamma M_-$. Note that $E\zeta^{(1)} = E\zeta^{(2)} = \gamma$, $E\{\zeta^{(1)}\}^2 = 4\sigma_{10}^2\gamma + o(\gamma)$ and $E\{\zeta^{(2)}\}^2 = 4\sigma_{20}^2\gamma + o(\gamma)$. Then, by Corollary 1, it follows that

$$\gamma\pi(r_0)\, n(\widehat{r}_n - r_0) \rightsquigarrow T_1,$$

where $\rightsquigarrow$ means that T_1 is a usable approximation of $\gamma\pi(r_0)\, n(\widehat{r}_n - r_0)$ in the distribution sense, and the density of T_1 is

$$f_{T_1}(x;\sigma_{10},\sigma_{20}) = \begin{cases} g(-x;\ 1/(2\sigma_{10}),\ \sigma_{10}/(2\sigma_{20}^2)), & \text{for } x < 0, \\ \\ g(\,x;\ \ 1/(2\sigma_{20}),\ \sigma_{20}/(2\sigma_{10}^2)), & \text{for } x \geq 0. \end{cases}$$

In particular, if $\sigma_{10}^2 = \sigma_{20}^2 := \sigma^2$, then, by Corollary 2,

$$\frac{\gamma\pi(r_0)}{\sigma^2}\, n(\widehat{r}_n - r_0) \rightsquigarrow T_2.$$

In practice, $\pi(\cdot)$ can be estimated by the nonparametric kernel method. Then, on using the plug-in method, an estimate $\widehat{\pi}_n(\widehat{r}_n)$ of $\pi(r_0)$ can be obtained. An estimate of σ^2 can be got from the residuals. However, the estimate of γ is a little complicated since it is a conditional expectation, namely $\gamma = E(\{\mathbf{x}'(\beta_{10} - \beta_{20})\}^2 | z = r_0)$. Of course, if $\mathbf{x}$ and z are independent, then γ is an unconditional expectation. In this case, it is easy to estimate γ by $\widehat{\gamma}_n = n^{-1}\sum_{i=1}^n \{\mathbf{x}_i'(\widehat{\beta}_{1n} - \widehat{\beta}_{2n})\}^2$. If they are not independent, a good choice is to use the best linear predictor of $\{\mathbf{x}'(\beta_{10} - \beta_{20})\}^2$ based on z with $\widehat{\theta}_n$ in lieu of θ_0 to approximate γ; see (3.7) in the following Subsection 3.3. Once the estimates of γ, $\pi(r_0)$ and σ^2 are obtained, we can construct confidence intervals of r_0 by using the quantiles of T_2.

3.2. *Threshold AR model.* The TAR model is an important class of nonlinear time series models. The idea of threshold in the time series context was initially conceived around 1976, first appeared in Tong (1978) and was later formalized in Tong and Lim (1980). Fuller results can be found in the monograph of Tong (1990). For history and future outlook, see, e.g., Tong (2011, 2015). Chan (1993) is a significant contribution to the inference of TAR models. It is the first breakthrough in the asymptotic theory of the LSE of the threshold parameter in discontinuous two-regime TAR models. Other important contributions include Tsay (1989, 1998), Gonzalo and Pitarakis (2002), and others. Li and Ling (2012) first established the asymptotic theory of the LSE in multiple-regime TAR models.

A time series $\{y_t\}$ is said to follow a two-regime TAR model of order p if it satisfies

$$y_t = \begin{cases} \mathbf{y}_{t-1}'\beta_{10} + \sigma_{10}\varepsilon_t, & \text{if } y_{t-d} \le r_0, \\ \\ \mathbf{y}_{t-1}'\beta_{20} + \sigma_{20}\varepsilon_t, & \text{if } y_{t-d} > r_0, \end{cases} \tag{3.3}$$

where $\mathbf{y}_{t-1} = (1, y_{t-1}, ..., y_{t-p})'$, $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with zero mean and unit variance and ε_t independent of $\{y_{t-j} : j \ge 1\}$.

Suppose that $\{y_1, ..., y_n\}$ is a sample from the TAR model (3.3). Denote by $\widehat{r}_n$ the LSE of r_0. Under Conditions 1-4 in Chan (1993) or Assumptions 3.1-3.4 in Li and Ling (2012), and by Theorem 3.3

in Li and Ling (2012), we have

$$n(\widehat{r}_n - r_0) \Longrightarrow M_- := s\text{-}\arg\min_{z\in\mathbb{R}} \mathcal{P}(z),$$

where the left and the right jump distributions in the two-sided CPP $\mathcal{P}(\cdot)$ are induced by

$$\zeta_t^{(1)} = \{\mathbf{y}_{t-1}'(\beta_{10}-\beta_{20})\}^2 + 2\sigma_{10}\varepsilon_t\{\mathbf{y}_{t-1}'(\beta_{10}-\beta_{20})\} \quad \text{given } y_{t-d} = r_0^-$$

and

$$\zeta_t^{(2)} = \{\mathbf{y}_{t-1}'(\beta_{10}-\beta_{20})\}^2 - 2\sigma_{20}\varepsilon_t\{\mathbf{y}_{t-1}'(\beta_{10}-\beta_{20})\} \quad \text{given } y_{t-d} = r_0^+,$$

respectively. Both rates are the same, i.e., $\pi(r_0)$, which is the value of the density $\pi(\cdot)$ of y_t at r_0.

From the above expressions, we can set $\gamma = E(\{\mathbf{y}_{t-1}'(\beta_{10}-\beta_{20})\}^2|y_{t-d} = r_0)$, which is a function of $\beta_{10}-\beta_{20}$. Note that when $\|\beta_{10}-\beta_{20}\|$ is small, the range of M_- is large. In this case, we can use Theorem 1 to approximate M_-.

Note that $E(\zeta_t^{(1)}|y_{t-d} = r_0) = E(\zeta_t^{(2)}|y_{t-d} = r_0) = \gamma$, and $E(\{\zeta_t^{(1)}\}^2|y_{t-d} = r_0) = 4\sigma_{10}^2\gamma + o(\gamma)$ and $E(\{\zeta_t^{(2)}\}^2|y_{t-d} = r_0) = 4\sigma_{20}^2\gamma + o(\gamma)$. Thus, by Corollary 1, we have

$$\gamma\pi(r_0)\, n(\widehat{r}_n - r_0) \rightsquigarrow T_1 \tag{3.4}$$

with

$$f_{T_1}(x;\sigma_{10},\sigma_{20}) = \begin{cases} g(|x|;\ 1/(2\sigma_{10}),\ \sigma_{10}/(2\sigma_{20}^2)\,), & \text{for } x < 0, \\ \\ g(\,x;\ \ 1/(2\sigma_{20}),\ \sigma_{20}/(2\sigma_{10}^2)\,), & \text{for } x \geq 0. \end{cases}$$

In applications, in order to construct confidence intervals of r_0 by (3.4), we must estimate $\pi(r_0)$ and γ. Clearly, estimating $\pi(\cdot)$ is easy. For example, we can use the nonparametric kernel method and then use the plug-in method to get an estimate $\widehat{\pi}_n(\widehat{r}_n)$ of $\pi(r_0)$. However, it is rather difficult to estimate γ directly since it is a conditional expectation. An easy and good choice is to use the best linear predictor to replace it. Of course, the re-sampling method in Li and Ling (2012) is still helpful.

Now, by using Algorithms **B** and **C** in Li and Ling (2012), we can draw a new sample $\{\mathbf{y}_t^*\}$ with $y_{i-d}^* = \widehat{r}_n$ for each $\mathbf{y}_i^*$ and then use the new sample to estimate γ with $\widehat{\beta}_{1n} - \widehat{\beta}_{2n}$ in lieu of $\beta_{10} - \beta_{20}$.

In particular, we consider a simple discontinuous TAR(1) model:

$$y_t = \{\beta_{10} I(y_{t-1} \le r_0) + \beta_{20} I(y_{t-1} > r_0)\} y_{t-1} + \varepsilon_t,$$

where the notation is the same as in model (3.3), except for $\mathrm{var}(\varepsilon_t) = \sigma^2$. In this case, the jumps are unconditional and simple:

$$\zeta_t^{(1)} = \{r_0(\beta_{10} - \beta_{20})\}^2 + 2r_0(\beta_{10} - \beta_{20})\varepsilon_t$$

and

$$\zeta_t^{(2)} = \{r_0(\beta_{10} - \beta_{20})\}^2 - 2r_0(\beta_{10} - \beta_{20})\varepsilon_t.$$

Let $\gamma = \{r_0(\beta_{10} - \beta_{20})\}^2$. Then, $E\zeta_t^{(1)} = E\zeta_t^{(2)} = \gamma$, $E\{\zeta_t^{(1)}\}^2 = E\{\zeta_t^{(2)}\}^2 = 4\sigma^2\gamma + o(\gamma)$. By Corollary 2, it follows that

$$\frac{\gamma \pi(r_0)}{\sigma^2}\, n(\widehat{r}_n - r_0) \rightsquigarrow T_2.$$

In this simple case, we can estimate γ by $\widehat{\gamma}_n = \{\widehat{r}_n(\widehat{\beta}_{1n} - \widehat{\beta}_{2n})\}^2$ and $\pi(r_0)$ by $\widehat{\pi}_n(\widehat{r}_n)$, a nonparametric kernel estimate, and σ^2 by $\widehat{\sigma}_n^2 = n^{-1} \sum_{t=1}^n \widehat{\varepsilon}_t^2$, where $\{\widehat{\varepsilon}_t\}$ is the residual based on the LSE. Thus,

$$\mathrm{NS}_n := \frac{\widehat{\gamma}_n\, \widehat{\pi}_n(\widehat{r}_n)}{\widehat{\sigma}_n^2}\, n(\widehat{r}_n - r_0) \tag{3.5}$$

can be approximated by T_2 when $|r_0(\beta_{10} - \beta_{20})|$ is small.

3.3. *Threshold MA model.* The TMA model is an important class of threshold time series models. It is a natural generalization of linear MA models. The linear MA model was first introduced by Slutsky (1927) and since then it has been widely used in many areas such as business, economics, etc. It has played a prominent role in the development of time series analysis. However, nonlinear MA models have developed slowly and have been overshadowed by nonlinear AR models. The slow development was mostly due to difficulties in statistical inference for general nonlinear MA models;

see Robinson (1977). To-date, studies on nonlinear MA models mainly focus on TMA ones; see, e.g., Ling and Tong (2005), Ling, Tong and Li (2007), Li and Li (2008), Li, Ling and Tong (2012) and Li (2012). Recently, Li, Ling and Li (2013) studied the asymptotic theory of the LSE in TMA models and succeeded in obtaining the limiting distribution of the estimated threshold for the first time in the literature.

A time series $\{y_t\}$ is said to follow a TMA model of order 1 if it satisfies

$$y_t = \varepsilon_t + [\phi_0 I(y_{t-1} \le r_0) + \psi_0 I(y_{t-1} > r_0)]\varepsilon_{t-1},$$

where $\{\varepsilon_t\}$ is i.i.d. with mean zero and variance $\sigma_\varepsilon^2 \in (0, \infty)$, and ε_t is independent of $\{y_j : j < t\}$. Let $\theta = (\phi, \psi, r)'$ denote the parameter and θ_0 its true value.

Let $\widehat{\theta}_n$ be the LSE of θ_0. Li, Ling and Li (2013) showed that under their Assumptions 2.1-2.3

$$n(\widehat{r}_n - r_0) \Longrightarrow M_- := s\text{-}\arg\min_{z\in\mathbb{R}} \mathcal{P}(z),$$

where

$$\mathcal{P}(z) = I(z \le 0) \sum_{i=1}^{N_1(-z)} \zeta_i^{(1)} + I(z > 0) \sum_{i=1}^{N_2(z)} \zeta_i^{(2)}.$$

Here, $N_1(\cdot)$ and $N_2(\cdot)$ are independent Poisson processes with the same rate $\pi(r_0)$, which is the value of the density $\pi(\cdot)$ of y_t at r_0, and $\{\zeta_k^{(1)} : k \ge 1\}$ is an i.i.d. random variable with the same distribution as the one induced by

$$\begin{aligned}\zeta^{(1)} =& (\phi_0 - \psi_0)^2 \varepsilon_{t-1}^2 \sum_{j=0}^{\infty} \prod_{i=0}^{j-1} \{\phi_0^2 I(y_{t+i} \le r_0) + \psi_0^2 I(y_{t+i} > r_0)\} \\ &+ 2(\phi_0 - \psi_0)\varepsilon_{t-1} \sum_{j=0}^{\infty} \varepsilon_{t+j} \prod_{i=0}^{j-1} \{-\phi_0 I(y_{t+i} \le r_0) - \psi_0 I(y_{t+i} > r_0)\}\end{aligned}$$

given $y_{t-1} = r_0^-$. Similarly, for the sequence $\{\zeta_k^{(2)} : k \ge 1\}$, we have

$$\begin{aligned}\zeta^{(2)} =& (\phi_0 - \psi_0)^2 \varepsilon_{t-1}^2 \sum_{j=0}^{\infty} \prod_{i=0}^{j-1} \{\phi_0^2 I(y_{t+i} \le r_0) + \psi_0^2 I(y_{t+i} > r_0)\} \\ &- 2(\phi_0 - \psi_0)\varepsilon_{t-1} \sum_{j=0}^{\infty} \varepsilon_{t+j} \prod_{i=0}^{j-1} \{-\phi_0 I(y_{t+i} \le r_0) - \psi_0 I(y_{t+i} > r_0)\}\end{aligned}$$

given $y_{t-1} = r_0^+$. Here, the convention $\prod_{i=0}^{-1} \equiv 1$ is adopted.

When $|\phi_0 - \psi_0|$ is small, that is when the threshold effect is small, we can approximate M_- or $n(\widehat{r}_n - r_0)$ by Theorem 1. Note that, if $|\phi_0 - \psi_0| \doteq 0$, then $\phi_0 \doteq \psi_0$. Thus,

$$\begin{aligned}\gamma &= E(\zeta^{(1)}|y_{t-1} = r_0) = E(\zeta^{(2)}|y_{t-1} = r_0)\\ &= (\phi_0 - \psi_0)^2 \sum_{j=0}^{\infty} E\left\{\varepsilon_{t-1}^2 \prod_{i=0}^{j-1}[\phi_0^2 I(y_{t+i} \le r_0) + \psi_0^2 I(y_{t+i} > r_0)]\middle| y_{t-1} = r_0\right\}\\ &\doteq \frac{(\phi_0 - \psi_0)^2 E(\varepsilon_{t-1}^2|y_{t-1} = r_0)}{1 - \min(\phi_0^2, \psi_0^2)}\end{aligned}$$

and

$$E(\{\zeta^{(1)}\}^2|y_{t-1} = r_0) = E(\{\zeta^{(2)}\}^2|y_{t-1} = r_0) = 4\sigma_\varepsilon^2\gamma + o(\gamma).$$

Therefore, by Corollary 2, it follows that

$$\frac{\gamma\pi(r_0)}{\sigma_\varepsilon^2}\, n(\widehat{r}_n - r_0) \rightsquigarrow T_2. \tag{3.6}$$

In applications, $\pi(r_0)$ is readily estimated by the nonparametric kernel method, and σ_ε^2 by the residuals $\{\widehat{\varepsilon}_t\}$ based on the LSE. The real hard work is in estimating or approximating γ. The key point is how to approximate $E(\varepsilon_t^2|y_t = r_0)$. Here, we propose three ways to approximate this conditional expectation. One is the re-sampling method of Li, Ling and Li (2013). Similar to high-order TAR models in Subsection 3.2, we can draw a new sample satisfying the condition $y_{t-1} = \widehat{r}_n$ and then calculate the conditional expectation. This procedure is complicated and needs more computations. The second is to use nonparametric method to estimate it.

The third is to use the *best linear predictor* to replace $E(\varepsilon_t^2|y_t = r_0)$ as it is simple relatively. Note that the best linear predictor of Y based on X is

$$\mathcal{L}(Y|X) = EY + \frac{\text{cov}(X, Y)}{\text{var}(X)}(X - EX). \tag{3.7}$$

For small $|\phi_0-\psi_0|$, we can use $\varepsilon_t+((\phi_0+\psi_0)/2)\varepsilon_{t-1}$ to approximate y_t, i.e., $y_t \doteq \varepsilon_t+((\phi_0+\psi_0)/2)\varepsilon_{t-1}$. Hence, we have the following approximation

$$E(\varepsilon_t^2|y_t = r_0) \doteq \sigma_\varepsilon^2 + \frac{\kappa_3 r_0}{\sigma_\varepsilon^2(1 + (\phi_0 + \psi_0)^2/4)}.$$

where $\kappa_3 = E\varepsilon_t^3$. Therefore,

$$\left\{\widehat{\sigma}_\varepsilon^2 + \frac{\widehat{\kappa}_3 \widehat{r}_n}{\widehat{\sigma}_\varepsilon^2(1+(\widehat{\phi}_n+\widehat{\psi}_n)^2/4)}\right\} \frac{(\widehat{\phi}_n-\widehat{\psi}_n)^2\,\widehat{\pi}_n(\widehat{r}_n)}{\widehat{\sigma}_\varepsilon^2\{1-\min(\widehat{\phi}_n^2,\widehat{\psi}_n^2)\}}\, n(\widehat{r}_n - r_0) \tag{3.8}$$

can be approximated by T_2 by Corollary 2, where $\widehat{\sigma}_\varepsilon^2 = n^{-1}\sum_{t=1}^n \widehat{\varepsilon}_t^2$, $\widehat{\kappa}_3 = n^{-1}\sum_{t=1}^n \widehat{\varepsilon}_t^3$, and $\{\widehat{\varepsilon}_t\}$ is the residual.

In particular, if ε_t is symmetric, then $\kappa_3 = 0$ and in turn (3.8) reduces to

$$\mathrm{NS}_n := \frac{(\widehat{\phi}_n-\widehat{\psi}_n)^2\,\widehat{\pi}_n(\widehat{r}_n)}{1-\min(\widehat{\phi}_n^2,\widehat{\psi}_n^2)}\, n(\widehat{r}_n - r_0). \tag{3.9}$$

3.4. *Threshold ARMA model.* The TARMA model is a natural extension of TAR and TMA models. Like linear ARMA models, TARMA model can provide a parsimonious form for high-order TAR or high-order TMA models. Recently, Chan and Goracci (2018) studied the ergodicity of one-order TARMA models. However, in the literature to-date, there are few results on the statistical inference of TARMA models. Exceptions are Li and Li (2011) and Li, Li and Ling (2011), who considered the LSE and established its asymptotic theory.

A time series $\{y_t\}$ is said to follow a TARMA model of order (1,1) if it satisfies

$$y_t = \begin{cases} \mu_1 + \phi_1 y_{t-1} + \varepsilon_t + \psi_1\varepsilon_{t-1}, & \text{if } y_{t-1} \le r, \\ \\ \mu_2 + \phi_2 y_{t-1} + \varepsilon_t + \psi_2\varepsilon_{t-1}, & \text{if } y_{t-1} > r, \end{cases}$$

where $\{\varepsilon_t\}$ is i.i.d. with mean zero and variance $\sigma^2 \in (0,\infty)$, and ε_t is independent of $\{y_j : j < t\}$. Let $\theta = (\mu_1,\phi_1,\psi_1,\mu_2,\phi_2,\psi_2,r)'$ be the parameter and its true value be θ_0.

Li, Li and Ling (2011) showed that under their Assumptions 3.1-3.5

$$n(\widehat{r}_n - r_0) \Longrightarrow M_- := s\text{-}\arg\min_{z\in\mathbb{R}} \mathcal{P}(z),$$

where the left and right jump distributions are induced by

$$\begin{aligned}\zeta^{(1)} =& \delta_t^2 \sum_{j=0}^{\infty}\prod_{i=0}^{j-1}\{\psi_{10}^2 I(y_{t+i}\le r_0) + \psi_{20}^2 I(y_{t+i} > r_0)\} \\ &+ 2\delta_t \sum_{j=0}^{\infty}\varepsilon_{t+j}\prod_{i=0}^{j-1}\{-\psi_{10} I(y_{t+i}\le r_0) - \psi_{20} I(y_{t+i} > r_0)\}\end{aligned}$$

given $y_{t-1} = r_0^-$, and

$$\begin{aligned}\zeta^{(2)} =& \delta_t^2 \sum_{j=0}^{\infty} \prod_{i=0}^{j-1} \{\psi_{10}^2 I(y_{t+i} \leq r_0) + \psi_{20}^2 I(y_{t+i} > r_0)\} \\ & - 2\delta_t \sum_{j=0}^{\infty} \varepsilon_{t+j} \prod_{i=0}^{j-1} \{-\psi_{10} I(y_{t+i} \leq r_0) - \psi_{20} I(y_{t+i} > r_0)\}\end{aligned}$$

given $y_{t-1} = r_0^+$, where $\delta_t = (\mu_{10} - \mu_{20}) + (\phi_{10} - \phi_{20})r_0 + (\psi_{10} - \psi_{20})\varepsilon_{t-1}$.

When $|(\mu_{10} - \mu_{20}) + (\phi_{10} - \phi_{20})r_0| + |\psi_{10} - \psi_{20}|$ is small, we can approximate M_- or $n(\widehat{r}_n - r_0)$ by Theorem 1. Note that

$$\begin{aligned}\gamma &= E(\zeta^{(1)} | y_{t-1} = r_0^-) = E(\zeta^{(2)} | y_{t-1} = r_0^+) \\ &= \sum_{j=0}^{\infty} E\left\{ \delta_t^2 \prod_{i=0}^{j-1} \{\psi_{10}^2 I(y_{t+i} \leq r_0) + \psi_{20}^2 I(y_{t+i} > r_0) \Big| y_{t-1} = r_0 \right\} \\ &\doteq \frac{E(\delta_t^2 | y_{t-1} = r_0)}{1 - \min(\psi_{10}^2, \psi_{20}^2)}\end{aligned}$$

and

$$E(\{\zeta^{(1)}\}^2 | y_{t-1} = r_0) = E(\{\zeta^{(2)}\}^2 | y_{t-1} = r_0) = 4\sigma^2\gamma + o(\gamma).$$

Therefore, by Corollary 2, it follows that

$$\frac{\gamma \pi(r_0)}{\sigma^2} n(\widehat{r}_n - r_0) \rightsquigarrow T_2. \tag{3.10}$$

Similar to the procedure described in Subsection 3.3, we can estimate γ, $\pi(r_0)$ and σ^2. We omit the detail.

3.5. *T-CHARM.* To characterize the martingale difference structure implied in log-returns of assets in financial time series, Chan, et al. (2014) proposed a simple yet versatile model, called the conditional heteroscedastic AR model with thresholds (T-CHARM), which is a special case of Rabemananjara and Zakoïan (1993), Zakoïan (1994), Li and Ling (2012), Li, Ling and Zakoïan (2015) and Li, Ling and Zhang (2016).

A simple T-CHARM is defined as

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \sigma_{10}^2 I(y_{t-1} \le r_0) + \sigma_{20}^2 I(y_{t-1} > r_0),$$

where $\{\varepsilon_t\}$ is i.i.d. with zero mean and unit variance, $\sigma_{10}^2 \neq \sigma_{20}^2$.

Chan, et al. (2014) developed asymptotic theory on the quasi-maximum likelihood estimation (QMLE) of $(\sigma_{10}^2, \sigma_{20}^2, r_0)'$ under some assumptions, and proved that

$$n(\widehat{r}_n - r_0) \Longrightarrow M_- := s\text{-}\arg\min_{z\in\mathbb{R}} \mathcal{P}(z),$$

where the left and the right jumps in $\mathcal{P}(z)$ are respectively

$$\zeta_k^{(1)} = \log \frac{\sigma_{20}^2}{\sigma_{10}^2} + \Big(\frac{\sigma_{10}^2}{\sigma_{20}^2} - 1\Big)\varepsilon_k^2$$

and

$$\zeta_k^{(2)} = \log \frac{\sigma_{10}^2}{\sigma_{20}^2} + \Big(\frac{\sigma_{20}^2}{\sigma_{10}^2} - 1\Big)\varepsilon_k^2.$$

Let $\gamma = |\sigma_{10}^2 - \sigma_{20}^2|^2/(\sigma_{10}^4 + \sigma_{20}^4)$. If γ is small, then we have

$$E\zeta_1^{(1)} = E\zeta_1^{(2)} = \gamma + o(\gamma),$$
$$E\{\zeta_1^{(1)}\}^2 = E\{\zeta_1^{(2)}\}^2 = 2(\kappa_4 - 1)\gamma + o(\gamma),$$

where $\kappa_4 = E\varepsilon_t^4$. Thus, by Corollary 2,

$$\text{NS}_n := \frac{2\widehat{\gamma}_n \widehat{\pi}(\widehat{r}_n)}{\widehat{\kappa}_4 - 1}\, n(\widehat{r}_n - r_0) \rightsquigarrow T_2, \tag{3.11}$$

where $\widehat{\gamma}_n = |\widehat{\sigma}_{1n}^2 - \widehat{\sigma}_{2n}^2|^2/(\widehat{\sigma}_{1n}^4 + \widehat{\sigma}_{2n}^4)$, $\widehat{\kappa}_4 = n^{-1}\sum_{t=1}^n \widehat{\varepsilon}_t^4$, $\{\widehat{\varepsilon}_t\}$ is the residuals based on the QMLE, $\widehat{\pi}(\cdot)$ is the nonparametric kernel estimator of $\pi(\cdot)$.

For multiple-regime T-CHARM, approximations of the limiting distributions of the thresholds can be obtained similarly.

3.6. *Threshold DAR model.* The TDAR model is a significant extension of conditional heteroscedastic models, including the threshold ARCH model of Rabemananjara and Zakoïan (1993) and Zakoïan (1994). On TDAR models, recent work can be found in Li, Ling and Zakoïan (2015) and Li, Ling and Zhang (2016).

A time series $\{y_t\}$ is said to follow a TDAR model of order (1, 1) if

$$y_t = \begin{cases} \phi_0 + \phi_1 y_{t-1} + \varepsilon_t \sqrt{\alpha_0 + \alpha_1 y_{t-1}^2}, & \text{if } y_{t-1} \le r, \\ \\ \psi_0 + \psi_1 y_{t-1} + \varepsilon_t \sqrt{\beta_0 + \beta_1 y_{t-1}^2}, & \text{if } y_{t-1} > r, \end{cases}$$

where $\{\varepsilon_t\}$ is i.i.d. with zero mean and unit variance.

Li, Ling and Zakoïan (2015) and Li, Ling and Zhang (2016) studied the QMLE of TDAR model and discussed their asymptotics. Under Assumptions 3.1-3.5 in Li, Ling and Zhang (2016), we have

$$n(\widehat{r}_n - r_0) \Longrightarrow M_- := s\text{-}\arg\min_{z\in\mathbb{R}} \mathcal{P}(z),$$

where the left and the right jumps in $\mathcal{P}(z)$ are respectively

$$\zeta_k^{(1)} = \log \frac{\beta_0 + \beta_1 r^2}{\alpha_0 + \alpha_1 r^2} + \frac{\{(\phi_0 - \psi_0) + (\phi_1 - \psi_1)r + \varepsilon_k \sqrt{\alpha_0 + \alpha_1 r^2}\}^2}{\beta_0 + \beta_1 r^2} - \varepsilon_k^2$$

and

$$\zeta_k^{(2)} = \log \frac{\alpha_0 + \alpha_1 r^2}{\beta_0 + \beta_1 r^2} + \frac{\{(\phi_0 - \psi_0) + (\phi_1 - \psi_1)r - \varepsilon_k \sqrt{\beta_0 + \beta_1 r^2}\}^2}{\alpha_0 + \alpha_1 r^2} - \varepsilon_k^2$$

For simplicity, we assume that $\varepsilon_t \sim \mathcal{N}(0, 1)$ tentatively. Denote

$$\gamma = \frac{\{(\beta_0 - \alpha_0) + (\beta_1 - \alpha_1)r^2\}^2}{(\alpha_0 + \alpha_1 r^2)^2 + (\beta_0 + \beta_1 r^2)^2} + \frac{2\{(\phi_0 - \psi_0) + (\phi_1 - \psi_1)r\}^2}{(\alpha_0 + \alpha_1 r^2) + (\beta_0 + \beta_1 r^2)}.$$

By a simple calculation, we have

$$E\zeta_1^{(1)} = E\zeta_1^{(2)} = \gamma + o(\gamma),$$
$$E\{\zeta_1^{(1)}\}^2 = E\{\zeta_1^{(2)}\}^2 = 4\gamma + o(\gamma).$$

Thus, by Corollary 2,

$$\gamma\pi(r)\, n(\widehat{r}_n - r) \rightsquigarrow T_2.$$

In applications, γ and $\pi(r_0)$ can be estimated by their sample counterparts. For high-order cases, similar to high-order TAR models in Subsection 3.2, we can use the re-sampling method to estimate γ.

4. Simulation studies. In this section, we use simulations to assess the performance of the approximation in Section 2. The TAR(1), TMA(1) models and T-CHARM are used as typical cases. The error $\{\varepsilon_t\}$ is supposed to be i.i.d. $\mathcal{N}(0,1)$ for simplicity. For each model, the sample size is 500 and 2000 replications are used.

The TAR(1) model is defined as

$$y_t = \{0.5I(y_{t-1} \leq 1.5) + 0.9I(y_{t-1} > 1.5)\}y_{t-1} + \varepsilon_t. \tag{4.1}$$

Figure 2 shows the histogram and the empirical CDF of NS_n in (3.5) as well as those of T_2 in (2.2), from which we can see that the approximation performs well, even when the threshold effect is not small with $\gamma = |0.9 - 0.5| = 0.4$.

Hansen (1997, 2000) was probably the first to adopt a BM approximation approach to handle statistical inference in TAR models. His approach is based on a different setting from ours: he has effectively replaced the TAR model by a sequence of TAR models *indexed by the sample size* n, with n-dependent regression slopes, which coalesce (with a speed apparently not easily determined) as n goes to infinity. Let us call the difference between the regression slopes of the two regimes the *threshold effect*. On the other hand, for cases with *fixed threshold effects*, Li and Ling (2012) proposed a re-sampling method to simulate M_{-}. This method works well when the range of M_{-} is not very large, e.g., when the expectation of the jump is sufficiently large, implying a large threshold effect. However, the range becomes very large when the expectation of the jump is small associated with a small threshold effect. In this case the re-sampling method is not so accurate. We now take up the

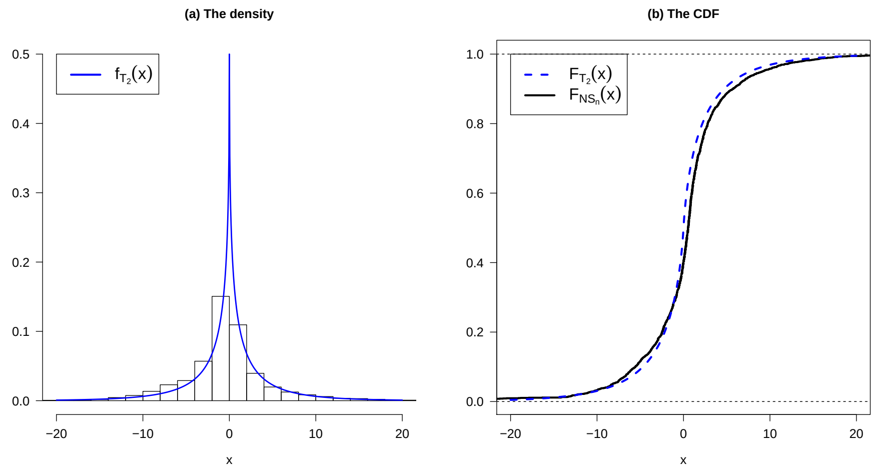

FIG 2. *The histogram* (*a*) *and the empirical CDF* (*b*) *of* NS_n *in* (3.5) *for TAR(1) model in* (4.1), *as well as the density and CDF of* T_2 *in* (2.2).

challenge of obtaining a BM approximation for the case with fixed (i.e. not n-dependent) but small threshold effects.

To compare the performance of likelihood ratio method in Hansen (1997) and ours, we compute the coverage probabilities of r_0 at 90% and 95% levels, respectively. The estimator of $\pi(\cdot)$ is obtained by two methods: one based on a nonparametric kernel method and the other the moving block bootstrapping (MBB) method. For this and other bootstrapping methods for dependent data, see Lahiri (2003). When the sample is small, the estimator $\widehat{\pi}_n(\widehat{r}_n)$ may have a larger bias and will affect the performance of the statistic NS_n. In this case, we recommend the MBB method. Table 2 reports the numerical results. Here, for each sample size, 1,000 replications are used. With each replication, 10 replicates are used for the MBB. From the table, we can see that Hansen's method over-estimates the coverage probability and becomes quite conservative when the sample size n is moderately large, like that in Hansen (1997, 2000). On the other hand, our method based on MBB performs well across all sample sizes; the method based on nonparametric kernels shows stable performance across all n, with good coverage probability for $n = 500$, but not as well as the MBB method for smaller n. Based on our experience, we recommend the nonparametric kernel method for large n, which will save computing costs, and the MBB for smaller n.

TABLE 2
The coverage probabilities of r_0 at 90% and 95% levels.

	90%			95%		
	Hansen's	Our-ker	Our-MBB	Hansen's	Our-ker	Our-MBB
$n = 50$	0.711	0.843	0.907	0.739	0.900	0.962
$n = 100$	0.912	0.818	0.899	0.934	0.871	0.952
$n = 200$	0.934	0.853	0.901	0.967	0.889	0.949
$n = 500$	0.935	0.895	-	0.964	0.948	-

For the TMA(1) model defined as

$$y_t = \varepsilon_t + [0.6I(y_{t-1} \le 0) + 0.9I(y_{t-1} > 0)]\varepsilon_{t-1}. \tag{4.2}$$

Figure 3 shows the histogram and the empirical CDF of NS_n in (3.9) as well as those of T_2 in (2.2),

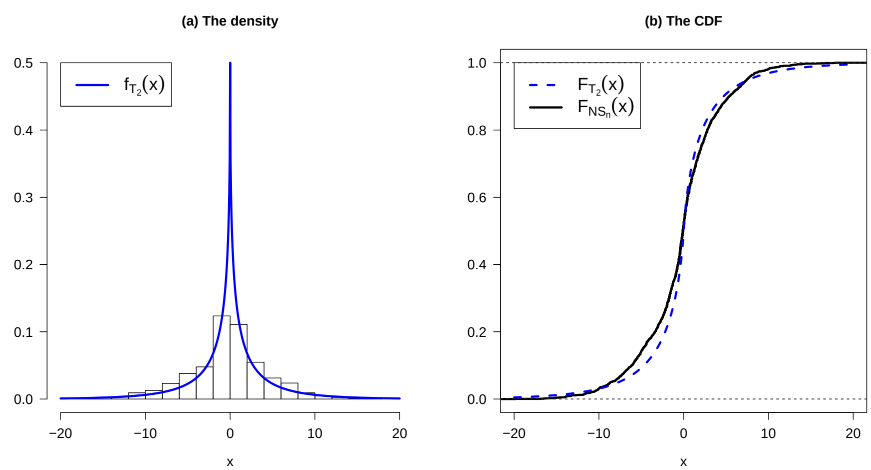

FIG 3. *The histogram (a) and the empirical CDF (b) of NS_n in (3.9) for TMA(1) model in (4.2), as well as the density and CDF of T_2 in (2.2).*

from which we can see that the approximation performs well.

For the T-CHARM defined as

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = 1I(y_{t-1} \le 0.5) + 2I(y_{t-1} > 0.5). \tag{4.3}$$

Figure 4 shows that the performance of approximation is good. Here, the ratio $\kappa := \sigma_2^2/\sigma_1^2 = 2$. When

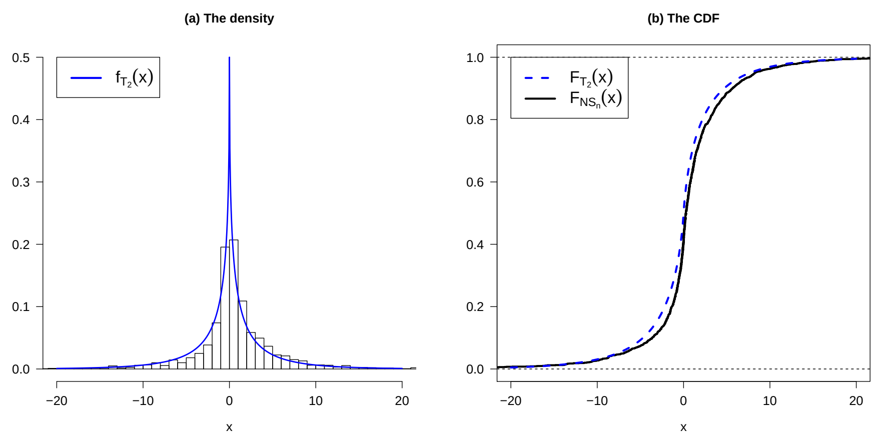

FIG 4. *The histogram (a) and the empirical CDF (b) of* NS_n *in* (3.11) *for T-CHARM model in* (4.3)*, as well as the density and CDF of* T_2 *in* (2.2).

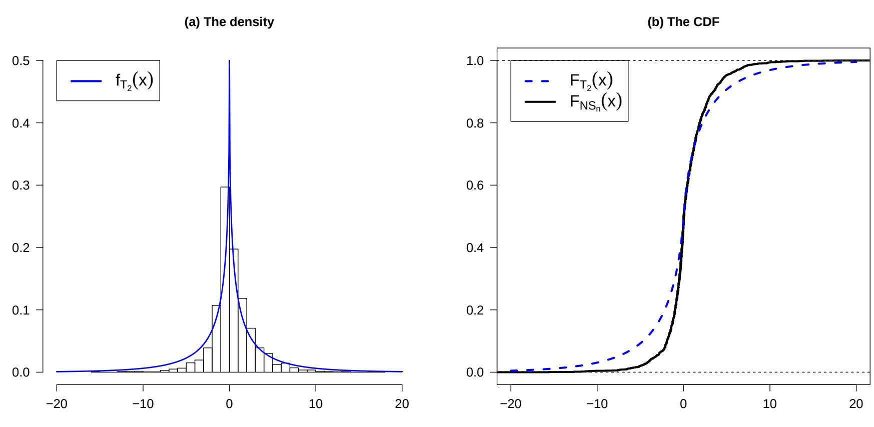

FIG 5. *The histogram (a) and the empirical CDF (b) of* NS_n *in* (3.11) *for T-CHARM model in* (4.4)*, as well as the density and CDF of* T_2 *in* (2.2).

κ increases, the performance of approximation may deteriorate. For example, consider a T-CHARM model defined as

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = 1I(y_{t-1} \leq 0.5) + 6I(y_{t-1} > 0.5). \tag{4.4}$$

Here, the ratio $\kappa = 6$.

Figure 5 shows the approximation for (4.4). Compared with Figure 4, the approximation is poorer as expected. Of course, when $\kappa > 5$, simulating a CPP will generally result in a better approximation for $n(\widehat{r}_n - r_0)$.

Unfortunately, there are no theoretical results to guide us on the choice between the resampling method in Li and Ling (2012) and our approximation method. However, our experience suggests the following procedure in practice. First, we use the resampling method in Li and Ling (2012) to simulate M_-. If the simulated numerical range of M_- is large, e.g., bigger than 50, then we use our approximation method instead.

5. An empirical example. The unemployment rate is an important index in measuring economic activity. Hansen (1997) explored the presence of nonlinearities in the business cycle through the use of a TAR model for U.S. unemployment rate among males age 20 and over. The sample is monthly from January 1959 through July 1996. There are 451 observations in total over the period, which is plotted in Figure 6.

Let $\{y_t\}$ be the rate. Hansen (1997) suggests the following fitted model

$$\triangle y_t = \begin{cases} \phi_0 + \sum_{i=1}^{12} \phi_i \triangle y_{t-i} + \sigma_1 \varepsilon_t, & \text{if } y_{t-1} - y_{t-12} \leq 0.302, \\ \\ \psi_0 + \sum_{i=1}^{12} \psi_i \triangle y_{t-i} + \sigma_2 \varepsilon_t, & \text{if } y_{t-1} - y_{t-12} > 0.302, \end{cases} \tag{5.1}$$

where $\triangle y_t = y_t - y_{t-1}$, $\sigma_1^2 = 0.154^2$, $\sigma_2^2 = 0.187^2$, and the estimates of the coefficients are summarized in Table 3. For more details, including the standard errors and 95% confidence intervals of the estimated coefficents, see Table 5 in Hansen (1997).

Fig 6. *Unemployment rate, Men 20+ years.*

Table 3
TAR estimates for unemployment rate.

			$y_{t-1} - y_{t-12} \leq 0.302$				
Variable	Intercept	$\triangle y_{t-1}$	$\triangle y_{t-2}$	$\triangle y_{t-3}$	$\triangle y_{t-4}$	$\triangle y_{t-5}$	$\triangle y_{t-6}$
ϕ	-.018	-.186	.084	.132	.165	.070	.027
Variable		$\triangle y_{t-7}$	$\triangle y_{t-8}$	$\triangle y_{t-9}$	$\triangle y_{t-10}$	$\triangle y_{t-11}$	$\triangle y_{t-12}$
ϕ		.062	.044	-.031	-.057	.091	-.136
			$y_{t-1} - y_{t-12} > 0.302$				
Variable	Intercept	$\triangle y_{t-1}$	$\triangle y_{t-2}$	$\triangle y_{t-3}$	$\triangle y_{t-4}$	$\triangle y_{t-5}$	$\triangle y_{t-6}$
ψ	.086	.241	.241	.124	-.026	-.020	-.084
Variable		$\triangle y_{t-7}$	$\triangle y_{t-8}$	$\triangle y_{t-9}$	$\triangle y_{t-10}$	$\triangle y_{t-11}$	$\triangle y_{t-12}$
ψ		-.151	-.035	.092	.103	-.114	-.412

From (3.4), using the estimated coefficients, we can obtain the density of T_1, which is displayed in Figure 7. The 2.5% and 97.5% quantiles of T_1 are -0.2477 and 0.3972, respectively.

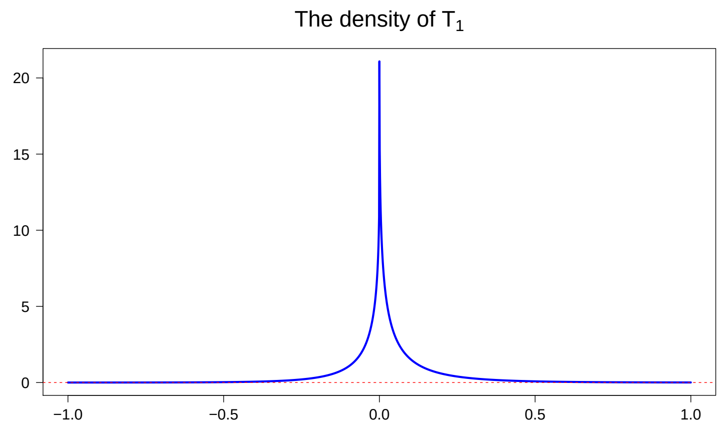

FIG 7. *The density of T_1 related to model (5.1).*

Now, using these quantiles, we can construct confidence intervals of the threshold parameter r_0 by our nonparametric kernel method with the MBB. This method gives the 95% confidence interval as $[0.255, 0.332]$. Here, the length of the moving block is 15 and the number of replicates is 50. The corresponding result using Hansen's method is $[0.213, 0.340]$, where the likelihood ratio is adjusted for residual heteroscedasticity by using a kernel estimator for the nuisance parameters. We note that Hansen's method has given a much wider confidence interval.

6. Conclusion and discussion. In this paper, we have developed an alternative approach to approximate two-sided CPPs by two-sided BMs. Significantly, we address the issue with small but *fixed* threshold effects. The new approach provides a simple yet efficacious tool to derive distributions of some functionals of the sample paths of CPPs, thus rendering statistical inference of the key threshold parameter in a threshold model, such as the construction of its confidence intervals, a practical proposition. Further, our approach continues to apply to threshold regressive/autoregressive models with multiple regimes since the distributions of all estimated threshold parameters are asymptotically independent; see, e.g., Li and Ling (2012), Chan, et al. (2014), Li, Ling and Zakoïan (2015). Thus, we can use our approach to construct confidence intervals for the thresholds one by one.

Our theory can be applied for other applied-oriented problems. For example, Hansen (1997, 2000) proposed a likelihood ratio-based statistic $LR_n(r_0)$ to test the null hypothesis $H_0 : r = r_0$ in threshold (auto)regression under his framework. However, under Tong's framework, i.e., the threshold effect is fixed, the limiting distribution of the related likelihood ratio-based statistic $LR_n(r_0)$ is a functional of two-sided compound Poisson process, which is hard to use for the same purpose. Our new theory can provide a usable approximation on $LR_n(r_0)$ and statistical inference for threshold can be realised.

Supplementary materials. The supplemental appendix is available from the authors.

References.

Chan, K.S., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. Ann. Statist. 21, 520–533.

Chan, K.S., Goracci, G., 2018. On the ergodicity of first-order threshold autoregressive moving-average processes. J. Time Series Anal. (In press).

Chan, K.S., Li, D., Ling, S., Tong, H., 2014. On conditionally heteroscedastic AR models with thresholds. Statist. Sinica 24, 625–652.

Chernoff, H., Rubin, H., 1956. The estimation of the location of a discontinuity in density. Proc. Third Berkeley Symp. Math. Statist. Probab. 1, 19–37. Univ. California Press, Berkeley.

Gonzalo, J., Pitarakis, J.Y., 2002. Estimation and model selection based inference in single and multiple threshold models. J. Econometrics 110, 319–352.

Hansen, B.E., 1997. Inference in TAR models. Stud. Nonlinear Dyn. Econom. 2, 1–14.

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

Hinkley, D.V., 1969. Inference about the intersection in two-phase regression. Biometrika 56, 495–504.

Hinkley, D.V., 1971. Inference in two-phase regression. J. Amer. Statist. Assoc. 66, 736–743.

Jacod, J., Shiryaev, A.N., 2003. Limit Theorems for Stochastic Processes, second ed. Springer, Berlin.

Koul, H., Qian, L., 2002. Asymptotics of maximum likelihood estimator in a two-phase linear regression model. J. Statist. Plann. Inference 108, 99–119.

Kushner, H.J., 1984. Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic Systems Theory. MIT Press, Cambridge.

Lahiri, S.N., 2003. Resampling Methods for Dependent Data. Springer, New York.

Li, D., 2012. A note on moving-average models with feedback. J. Time Series Anal. 33, 873–879.

Li, D., Li, W.K., Ling, S., 2011. On the least squares estimation of threshold ARMA models. Stat. Interface 4, 183–196.

Li, D., Ling, S., 2012. On the least squares estimation of multiple-regime threshold autoregressive models. J. Econometrics 167, 240–253.

Li, D., Ling, S., Li, W.K., 2013. Asymptotic theory on the least squares estimation of threshold moving-average models. Econometric Theory 29, 482–516.

Li, D., Ling, S., Tong, H., 2012. On moving-average models with feedback. Bernoulli 18, 735–745.

Li, D., Ling, S., Zakoïan, J.-M., 2015. Asymptotic inference in multiple-threshold double autoregressive models. J. Econometrics 189, 415–427.

Li, D., Ling, S., Zhang, R.M., 2016. On a threshold double autoregressive model. J. Bus. Econom. Statist. 34, 68–80.

Li, G.D., Li, W.K., 2008. Testing for threshold moving average with conditional heteroscedasticity. Statist. Sinica 18, 647–665.

Li, G.D., Li, W.K., 2011. Testing for linear time series models against its threshold extension. Biometrika 98, 243–250.

Ling, S., Tong, H., 2005. Testing for a linear MA model against threshold MA models. Ann. Statist. 33, 2529–2552.

Ling, S., Tong, H., Li, D., 2007. Ergodicity and invertibility of threshold moving-average models. Bernoulli 13, 161–168.

Quandt, R.E., 1958. The estimation of the parameters of a linear regression system obeying two separate regimes. J. Amer. Stat. Ass. 53, 873–880.

Rabemananjara, R., Zakoïan, J.-M., 1993. Threshold ARCH models and asymmetries in volatility. J. Appl. Econometrics 8, 31–49.

Robinson, P.M., 1977. The estimation of a nonlinear moving average model. Stochastic Process. Appl. 5, 81–90.

Seijo, E., Sen, B., 2011. Change-point in stochastic design regression and the bootstrap. Ann. Statist. 39, 1580–1607.

Skorokhod, A.V., Hoppensteadt, F.C., Salehi, H., 2002. Random Perturbation Methods with Applications in Science and Engineering. Springer, New York.

Slutsky, E., 1927. The summation of random causes as the source of cyclic processes. Voprosy Koniunktury 3, 34–64. {English translation in Econometrika 5, 105–146.}.

Stryhn, H., 1996. The location of the maximum of asymmetric two-sided Brownian motion with triangular drift, Statist. Probab. Lett. 29, 279–284.

Tong, H., 1978. On a threshold model. In *Pattern Recognition and Signal Processing*, ed. C.H. Chen, Sijthoff and noordhoff, Amsterdam, p.575–586.

Tong, H., 1990. Non-linear Time Series: A Dynamical System Approach. Oxford University Press, New York.

Tong, H., 2011. Threshold models in time series analysis – 30 years on. Stat. Interface 4, 107–118.

Tong, H., 2015. Threshold models in time series analysis – some reflections. J. Econometrics 189, 485–491.

Tong, H., Lim, K.S., 1980. Threshold autoregression, limit cycles and cyclical data (with discussion). J. Roy. Statist.

Soc. Ser. B 42, 245–292.

Tsay, R.S., 1989. Testing and modeling threshold autoregressive processes. J. Amer. Statist. Assoc. 84, 231–240.

Tsay, R.S., 1998. Testing and modeling multivariate threshold models. J. Amer. Statist. Assoc. 93, 1188–1202.

Yao, Y.C., 1987. Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. Ann. Statist. 15, 1321–1328.

Yu, P., 2012. Likelihood estimation and inference in threshold regression. J. Econometrics 167, 274–294.

Yu, P., 2015. Adaptive estimation of the threshold point in threshold regression. J. Econometrics 189, 83–100.

Zakoïan, J.-M., 1994. Threshold heteroscedastic models. J. Econom. Dynam. Control 18, 931–955.

Center for Statistical Science
and Department of Industrial Engineering
Tsinghua University
Beijing 100084, China
E-mail: malidong@tsinghua.edu.cn

Department of Mathematics
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
E-mail: maling@ust.hk

School of Mathematical Sciences
University of Electronic Science and Technology of China
Chengdu 611731, China;
and
Department of Statistics
London School of Economics and Political Science
London, UK.
E-mail: howell.tong@gmail.com

Department of Statistics
School of Economics
Jinan University
Guangzhou 510632, China
E-mail: tygr@jnu.edu.cn