ON THE LEAST SQUARE ESTIMATION OF THRESHOLD AUTOREGRESSIVE AND MOVING AVERAGE MODELS

By Dong Li, Wai Keung Li and Shiqing Ling

Hong Kong University of Science and Technology
and University of Hong Kong

Dedicated to Professor Howell Tong on the occasion of his 65th birthday

This paper considers the least square estimation and establishes its asymptotic theory for threshold autoregressive and moving-average models. Under some mild conditions, it is shown that the estimator of the threshold is n-consistent and after normalization it converges weakly to the smallest minimizer of a compound Poisson process, while the estimators of other coefficients are strongly consistent and asymptotically multivariate normal. This paper also provides a numerical method to tabulate the limiting distribution of the estimated threshold in practice. Simulation studies are carried out to assess the performance of the least square estimation in finite samples.

1. Introduction. Since the seminal work of Tong (1978), the threshold model has received considerable attention and is extensively investigated by many researchers. Now it becomes a more or less standard model in nonlinear time series and has been widely used in diverse areas, including biological sciences, econometrics, environmental sciences, finance, hydrology, physics, population dynamics, and among others. In comparison with other nonlinear time series models existing in the literature, one of the leading reasons behind the success of threshold models is that piecewise linear functions can offer a relatively simple and easy-to-handle approximation to the complex nonlinear dynamics, perhaps more importantly, it can offer a reasonable model-interpretation. Threshold models can capture and be capable of producing many nonlinear phenomena, such as amplitude dependent frequencies, asymmetric limit cycle, chaos, jump resonance, harmonic distortion and so on. A fairly comprehensive survey is available in Tong (1990) and a selective review of the history of threshold models is given by Tong (2011).

Although threshold models have many important applications in practice, there does not have a complete theory and methodology as compared with

the linear ARMA models. Our knowledge on threshold models is still developing, in particular, for the threshold autoregressive and moving average (TARMA) model.

In the past 30 years or so, the study on the threshold model mainly focuses on threshold autoregressive (TAR) models and there are numerous existing literature related to this topic. For example, on the probabilistic structure of the TAR model, some basic results were given by Chan et al. (1985), Chan and Tong (1985) and Tong (1990). More related results can be found in An and Huang (1996), Brockwell et al. (1992), Chen and Tsay (1991), Ling (1999), Liu and Suskov (1992) and so on. On the other hand, the asymptotic theory of the least square estimation of the two-regime TAR model was established by Chan (1993) and Chan and Tsay (1998), see also Petruccelli (1986) and Qian (1998). The multi-regime TAR model was considered by Li and Ling (2010). It is worth pointing out that Chan (1993) established the limiting distribution of the estimated threshold, which is the smallest minimizer of a two-sided compound Poisson process, when the autoregressive function is not continuous.

In the development of threshold models, the threshold moving-average (TMA) model, which can be viewed as the superposition of the random impulses through the notion of thresholds, i.e. on-off feedback controllers, has been overshadowed by TAR ones. Earlier works on the TMA model mainly focus on the probabilistic structure, see Brockwell et al. (1992), Liu and Susko (1992) and Ling (1999). Recently, more works enrich the literature on TMA models. Ling et al. (2007) and Li, Ling and Tong (2011) considered the existence, strict stationarity, ergodicity and invertibility of TMA models. Gooijer (1998) proposed the maximum likelihood estimation when the driven impulse is normal. Li, Ling and Li (2010) developed the least square estimation and established the asymptotic theory for general TMA models. Ling and Tong (2005) studied a quasi-likelihood ratio test with linear MA model as the null hypothesis against the TMA model as the alternative. This test was extended to the heteroscedastic case by Li and Li (2008).

Up to date, however, it seems that TARMA models have not attracted too much attention. There are only a few results available in the literature, all of which mainly study probabilistic properties and give some sufficient conditions for the (geometric) ergodicity, strict stationarity and existence of the solution to TARMA models, although they are not easy to verify in practice, see Brockwell et al. (1992), Giovanni (2005), Liu and Susko (1992) and Ling (1999). Recently, Chan and Tong (2010) gave a sufficient condition for the invertibility of nonlinear ARMA models. Li and Li (2011) proposed a quasi-likelihood ratio test for ARMA models against their threshold ex-

tensions. To the best of our knowledge, the estimation of TARMA models has not been considered yet. This is one of our motivations in this paper.

In the application, Ghaddar and Tong (1981) applied a TAR(11) model to fit the annual sunspot data of the period 1700—1979, see also Tong (1990, page 420), and Gooijer (1998) used a TMA(17) to fit the quarterly US real GNP data covering the period 1947.I—1982.IV. More real examples can be found in Tong (1990), Tsay (1989) and so on. These examples have a common feature that the order of the model is too high. From a practical point of view, TARMA models should have more advantages over pure TAR or TMA models because they can provide a parsimonious form just like linear ARMA models. This motivates us to study further TARMA models. As a natural generalization of pure TAR and TMA models, it is necessary to establish a unified theory for estimation, which is our another motivation.

In this paper, we shall study the least square estimation (LSE) of TARMA models and establish its asymptotic theory. Under some mild conditions, it is shown that the estimator of the threshold is n-consistent and after normalization it converges weakly to the smallest minimizer of a two-sided compound Poisson process, while the estimators of other coefficients are strongly consistent and asymptotically multivariate normal. This paper also proposes a numerical method to tabulate the limiting distribution of the estimated threshold in practice. Simulation studies are carried out to assess the performance of the least square estimation in finite samples.

The remainder of the paper is organized as follows. Section 2 introduces the TARMA model and presents its LSE. Section 3 addresses some assumptions and states main results. Section 4 proposes a numerical approach on simulating the limiting distribution of the estimated threshold. Simulation studies are reported in Section 5 and some concluding remarks are given in Section 6. All proofs of Theorems are left in the Appendix.

Throughout the paper, some symbols are conventional. C is a positive constant, which may be different in different places. $a \vee b \equiv \max\{a, b\}$ for some a and b. The summation $\sum_{i=1}^{0} \equiv 0$ and the product $\prod_{i=1}^{0} \equiv 1$. $I(\cdot)$ is an indicator function. R^k is the Euclidean space of dimension k and $\|\cdot\|$ denotes the Euclidian norm. $o_p(1)$ $(O_p(1))$ denotes a sequence of random numbers converging to zero (bounded) in probability and $\xrightarrow{d}$ denotes convergence in distribution.

2. Model and least square estimation. A time series $\{y_t, t = 0, \pm 1, ...\}$ is said to be a TARMA model with order (p_1, p_2, q_1, q_2), abbreviated to

TARMA(p_1, p_2, q_1, q_2), if it satisfies

$$(2.1) \qquad y_t = \begin{cases} \phi_{10} + \sum_{i=1}^{p_1} \phi_{1i} y_{t-i} + \varepsilon_t + \sum_{i=1}^{q_1} \psi_{1i} \varepsilon_{t-i}, & \text{if } y_{t-d} \leq r, \\ \\ \phi_{20} + \sum_{i=1}^{p_2} \phi_{2i} y_{t-i} + \varepsilon_t + \sum_{i=1}^{q_2} \psi_{2i} \varepsilon_{t-i}, & \text{if } y_{t-d} > r, \end{cases}$$

where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (i.i.d) random variables with zero mean, $r \in R$, p_i and q_i are known nonnegative integers and d is a positive integer. Here, r is called the threshold parameter and d is the delay variable. When $q_1 = q_2 = 0$, model (2.1) reduces to a pure TAR model. When $p_1 = p_2 = 0$, model (2.1) is a pure TMA model. If all $\phi_{1i} = \phi_{2i}$ and $\psi_{1i} = \psi_{2i}$, then model (2.1) becomes a linear ARMA model.

Without loss of generality, we assume that there exist two finite constants $\underline{r}$ and $\bar{r}$ such that $r \in [\underline{r}, \bar{r}]$ since model (2.1) reduces to a linear ARMA model when $r = \pm\infty$, which is not of our interest here. In addition, d is an unknown parameter to be estimated and its true value is d_0 with $1 \leq d_0 \leq D_0$, where D_0 is a known positive integer. Let $\boldsymbol{\theta}_0 = (\boldsymbol{\lambda}_0', r_0, d_0)'$ be the true value of the parameter $\boldsymbol{\theta} = (\boldsymbol{\lambda}', r, d)' \equiv (\boldsymbol{\phi}_1', \boldsymbol{\psi}_1', \boldsymbol{\phi}_2', \boldsymbol{\psi}_2', r, d)'$ with $\boldsymbol{\phi}_i = (\phi_{i0}, ..., \phi_{ip_i})'$ and $\boldsymbol{\psi}_i = (\psi_{i1}, ..., \psi_{iq_i})'$ for $i = 1, 2$. The parameter space Θ is denoted by $\Theta = \Lambda \times [\underline{r}, \bar{r}] \times \{1, ..., D_0\}$, where Λ is a subset of $R^{p_1+p_2+q_1+q_2+2}$.

Assume that $\{y_1, ..., y_n\}$ is a sample from model (2.1) with sample size n. Given the initial value $Y_0^* \equiv \{y_s : s \leq 0\}$, the sum of square errors function $L_n(\boldsymbol{\theta})$ is defined as

$$L_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \varepsilon_t^2(\boldsymbol{\theta}),$$

where

$$\varepsilon_t(\boldsymbol{\theta}) = y_t - \Big(\phi_{10} + \sum_{i=1}^{p_1} \phi_{1i} y_{t-i}\Big) I(y_{t-d} \leq r) - \Big(\phi_{20} + \sum_{i=1}^{p_2} \phi_{2i} y_{t-i}\Big) I(y_{t-d} > r)$$

$$- \Big(\sum_{i=1}^{q_1} \psi_{1i} \varepsilon_{t-i}(\boldsymbol{\theta})\Big) I(y_{t-d} \leq r) - \Big(\sum_{i=1}^{q_2} \psi_{2i} \varepsilon_{t-i}(\boldsymbol{\theta})\Big) I(y_{t-d} > r).$$

The minimizer $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\lambda}}_n', \widehat{r}_n, \widehat{d}_n)'$ of $L_n(\boldsymbol{\theta})$ is called the LSE of $\boldsymbol{\theta}_0$, that is,

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta} \in \Theta} L_n(\boldsymbol{\theta}).$$

Note that $L_n(\boldsymbol{\theta})$ is discontinuous in r and d. The way to get $\widehat{\boldsymbol{\theta}}_n$ is as follows.

Algorithm A

Step A.1. For each fixed $r \in R$ and $d \in \{1, ..., D_0\}$, we minimize $L_n(\boldsymbol{\theta})$ and get its minimizer $\widehat{\boldsymbol{\lambda}}_n(r, d)$ and minimum $L_n^*(r, d) \equiv L_n(\boldsymbol{\theta})|_{\boldsymbol{\lambda} = \widehat{\boldsymbol{\lambda}}_n(r,d)}$.

Step A.2. Since $L_n^*(r, d)$ only takes on finite many possible values, we can get the minimizer $(\widehat{r}_n, \widehat{d}_n)'$ of $L_n^*(r, d)$ by grid search over the set of order statistics of $\{y_1, ..., y_n\}$ and $\{1, ..., D_0\}$.

Step A.3. Use a plug-in method, one can finally get $\widehat{\boldsymbol{\lambda}}_n(\widehat{r}_n, \widehat{d}_n)$ and $\widehat{\boldsymbol{\theta}}_n$.

Generally, there exist infinitely many r such that $L_n(\cdot)$ attains its global minimum. One can choose the smallest of these r's as an estimator of the threshold. According to the procedure for obtaining $\widehat{\boldsymbol{\theta}}_n$, it is not hard to show that $\widehat{\boldsymbol{\theta}}_n$ is the LSE of $\boldsymbol{\theta}_0$.

In practice, the initial value Y_0^* is not available and hence we have to replace it by some constants, e.g. $Y_0^* = x \equiv \{x_1, ...\}$. Since $\sup_{\boldsymbol{\theta} \in \Theta} \|\mathbf{H}_{t,j}(\boldsymbol{\theta})\| = O(\rho^j)$ a.s. by Theorem A.1 in Ling and Tong (2005) under some assumptions, where $\mathbf{H}_{t,j}(\boldsymbol{\theta})$ is defined in (3.1) below, we can show that

$$\sup_{\boldsymbol{\theta} \in \Theta} \left| \varepsilon_t^2(\boldsymbol{\theta}) - \varepsilon_t^2(\boldsymbol{\theta})|_{Y_0^* = x} \right| = O(\rho^t) \quad \text{a.s.}$$

for any given x and some $\rho \in (0, 1)$. Thus, the initial value will not affect the asymptotic properties of $\widehat{\boldsymbol{\theta}}_n$. For simplicity, in what follows, we assume that Y_0^* is from model (2.1). In this case, $\varepsilon_t(\boldsymbol{\theta}_0) = \varepsilon_t$.

3. Main results. Let $p = p_1 \vee p_2$ and $q = q_1 \vee q_2$. In what follows, Θ is assumed to be *compact*. We first give a sufficient condition for the invertibility of model (2.1). This condition is easy to verify in practice and similar to that of TMA models in Ling and Tong (2005).

ASSUMPTION 3.1. $\sum_{j=1}^{q} |\psi_{ij}| < 1$, where $\psi_{ij} = 0$ for $j > q_i$, $i = 1, 2$.

If $\psi_{1j} = \psi_{2j}$, $j = 1, ..., q$, then the invertible region of model (2.1) is the same as that of the ARMA model when $q = 1$, but is smaller than that of ARMA model when $q > 1$. This assumption can be relaxed when $q > 1$. A weaker condition available can be found in Chan and Tong (2010). Under Assumption 3.1 and the condition that $\{y_t\}$ is strictly stationary with $E|y_t| < \infty$, by Theorem A.2 in Ling and Tong (2005), the residual $\varepsilon_t(\boldsymbol{\theta})$ has the following representation:

$$\varepsilon_t(\boldsymbol{\theta}) = \sum_{j=0}^{\infty} [\mathbf{e}' \mathbf{H}_{t,j}(\boldsymbol{\theta}) \mathbf{e}] \Big\{ y_{t-j} - (\phi_{10} + \sum_{i=1}^{p_1} \phi_{1i} y_{t-i-j}) I(y_{t-d-j} \le r)$$
$$- (\phi_{20} + \sum_{i=1}^{p_2} \phi_{2i} y_{t-i-j}) I(y_{t-d-j} > r) \Big\},$$

where $\mathbf{e} = (1, 0, ..., 0)'_{q \times 1}$ and

$$(3.1) \quad \mathbf{H}_{t,j}(\boldsymbol{\theta}) = \prod_{i=1}^{j}[\boldsymbol{\Psi}_2 + (\boldsymbol{\Psi}_1 - \boldsymbol{\Psi}_2)I(y_{t-d-i+1} \leq r)], \quad j \geq 0,$$

with the convention $\prod_{i=1}^{0} = \mathbf{I}_q$, the identity matrix of size q, and

$$(3.2) \quad \boldsymbol{\Psi}_i = \begin{pmatrix} -\psi_{i1} & \cdots & -\psi_{iq} \\ \mathbf{I}_{q-1} & & \mathbf{0} \end{pmatrix}, \quad i = 1, 2.$$

The following result states the strong consistency of the estimator $\widehat{\boldsymbol{\theta}}_n$.

THEOREM 3.1. *Suppose that (i) Assumption 3.1 holds, (ii) $\{y_t\}$ satisfying (2.1) is strictly stationary and ergodic with $Ey_t^2 < \infty$, (iii) $\boldsymbol{\phi}_{10} \neq \boldsymbol{\phi}_{20}$ or $\boldsymbol{\psi}_{10} \neq \boldsymbol{\psi}_{20}$ and (iv) ε_t has a bounded, continuous and strictly positive density on R with $E\varepsilon_t = 0$ and $E\varepsilon_t^2 < \infty$. Then, $\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s. as $n \to \infty$.*

The condition (iii) in Theorem 3.1 is required to ensure the identification of the threshold. The strong consistency of $\widehat{\boldsymbol{\theta}}_n$ holds regardless if the autoregressive function is continuous or not. This result is similar to that of TAR models in Chan (1993) and Chan and Tsay (1998) and of TMA models in Li, Ling and Li (2010). From Theorem 3.1, we know that $\widehat{d}_n$ equals d_0 eventually. Thus, without loss of generality, we assume that d is known for the remainder and it is deleted from $\boldsymbol{\theta}$, i.e. $\boldsymbol{\theta} = (\boldsymbol{\lambda}', r)'$ and Θ is modified, accordingly. Denote $\widehat{\boldsymbol{\lambda}}_n(r) = \widehat{\boldsymbol{\lambda}}_n(r, d_0)$.

To obtain the convergence rate of $\widehat{r}_n$ and the uniformly asymptotic normality of $\widehat{\boldsymbol{\lambda}}_n \equiv \widehat{\boldsymbol{\lambda}}_n(\widehat{r}_n)$, we need another four assumptions as follows.

ASSUMPTION 3.2. *$\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with zero mean and $E\varepsilon_t^4 < \infty$. The random variable ε_1 has a bounded, continuous and positive density f_ε on R.*

ASSUMPTION 3.3. *$\{y_t\}$ is strictly stationary with $Ey_t^4 < \infty$.*

Let $\mathbf{Z}_t = (y_t, ..., y_{t-(p \vee d)+1}, \varepsilon_t, ..., \varepsilon_{t-q+1})'$ with $\{\varepsilon_t, ..., \varepsilon_k\} = \varnothing$ if $k > t$. Then we can rewrite model (2.1) as a Markovian vectorial representation, see equation (3) in Giovanni (2005) or (2.1)-(2.2) in Liu and Susko (1992), that is, $\{\mathbf{Z}_t\}$ is a Markov chain. Denote its l-step transition probability by $\mathcal{P}^l(\mathbf{z}, A)$, where $\mathbf{z} \in R^{(p \vee d)+q}$ and A is a Borel set.

ASSUMPTION 3.4. *$\{\mathbf{Z}_t\}$ admits a unique invariant measure $\Pi(\cdot)$ and is V-uniformly ergodic with $V(\mathbf{z}) = 1 + \|\mathbf{z}\|$.*

Actually, Assumption 3.4 is stronger than geometric ergodicity. For the notion of the V-uniform ergodicity, see Meyn and Tweedie (1993, Chapter 16). If the initial value $\mathbf{Z}_0$ is from the distribution $\Pi(\cdot)$, then Assumption 3.4 implies that $\{y_t\}$ is strictly stationary. For a pure TAR model, it is relatively easy to find a sufficient condition on the coefficients to ensure that Assumption 3.4 holds. See Chan (1989, 1993) and Chan and Tong (1985). While for a pure TMA model Assumption 3.4 is not necessary since the expression of the strictly stationary and ergodic solution to model (2.1) is available and we can find other weaker conditions to substitute it in studying the convergence rate of $\hat{r}_n$, see Li, Ling and Tong (2011) and Li, Ling and Li (2010). As is well known, it is very hard to verify the V-uniform ergodicity for general nonlinear time series models in practice. For TARMA model (2.1), if Assumption 3.2, the conditions $\sum_{j=1}^{p}|\phi_{ij}| < 1$ and $r \neq 0$ hold with $\phi_{10} = \phi_{20} = 0$, where $\phi_{ij} = 0$ for $j > p_i$, $i = 1, 2$, then there exists a function

$$V_1(\mathbf{Z}_t) = \max_{0 \leq i \leq (p \vee d)-1}\{\rho_0^i |y_{t-i}|\} + \sum_{j=0}^{q-1} L_j |\varepsilon_{t-j}| + 1$$

such that $E[V_1(\mathbf{Z}_t)|\mathbf{Z}_{t-1} = \mathbf{z}] \leq \rho_0 V_1(\mathbf{z}) + [1 + (L_0 + 1)E|\varepsilon_1|]$, which implies that $\{\mathbf{Z}_t\}$ is V_1-uniformly ergodic, where constants ρ_0 and $\{L_i\}$ satisfy

$$\begin{aligned}
\rho_0 &= \left(\max_{i=1,2}\sum_{j=1}^{p}|\phi_{ij}|\right)^{1/p} \in (0,1), \\
\rho_0 L_{q-1} &\geq b \equiv \max_{i=1,2}\max_{1 \leq j \leq q}|\psi_{ij}|, \\
\rho_0 L_{i-1} &\geq L_i + b, \qquad \text{for } i = 1, ..., q-1.
\end{aligned}$$

It is easy to show that there exist positive constants c_1 and c_2 such that $c_1 V(\mathbf{z}) \leq V_1(\mathbf{z}) \leq c_2 V(\mathbf{z})$. Thus, $\{\mathbf{Z}_t\}$ is also V-uniformly ergodic by the definition of V-uniform ergodicity. Unfortunately, for other cases, how to verify Assumption 3.4 for TARMA models is still an open problem since the irreducibility and T-continuity of $\{\mathbf{Z}_t\}$ has not been solved. For more details, see Giovanni (2005).

ASSUMPTION 3.5. *There exist nonrandom vectors* $\mathbf{w} = (1, w_1, ..., w_p)' \in R^p$ *with* $w_d = r_0$ *and* $\mathbf{a} \in R^q$ *such that* $(\boldsymbol{\phi}_{10} - \boldsymbol{\phi}_{20})'\mathbf{w} + (\boldsymbol{\psi}_{10} - \boldsymbol{\psi}_{20})'\mathbf{a} \neq 0$, *where* $\phi_{ij0} = 0$ *for* $j > p_i$ *and* $\psi_{ij0} = 0$ *for* $j > q_i$, $i = 1, 2$.

In Assumption 3.5, w_d may not be a component of $\mathbf{w}$ if $d > p$. In this case, Assumption 3.5 is equivalent to the condition $\|\boldsymbol{\phi}_{10} - \boldsymbol{\phi}_{20}\| + \|\boldsymbol{\psi}_{10} - \boldsymbol{\psi}_{20}\| > 0$. The latter is necessary and sufficient for the identification of the threshold.

When $p = 0$, that is, TARMA model (2.1) reduces to a pure TMA model, Assumption 3.5 is identical to Assumption 2.2 in Li, Ling and Li (2010) since $(\boldsymbol{\psi}_{10} - \boldsymbol{\psi}_{20})'\mathbf{a} \neq 0$ for some $\mathbf{a} \in R^q$ implies that $\|\boldsymbol{\psi}_{10} - \boldsymbol{\psi}_{20}\| > 0$ and in turn $\widehat{r}_n$ has an n-convergence rate. When $q = 0$, e.g., model (2.1) reduces to a pure TAR model, Assumption 3.5 is the same as Condition 4 in Chan (1993), which implies that the autoregressive function is not continuous over the hyperplane $y_d = r_0$ so that $\widehat{r}_n$ has an n-convergence rate. Assumption 3.5 plays a similar role in the convergence rate of $\widehat{r}_n$ for TARMA models.

For TAR models, from the results in Chan (1993) and Chan and Tsay (1998), we know that the convergence rate of $\widehat{r}_n$ heavily depends on the continuity of the autoregressive function over the hyperplane $y_d = r_0$. For TARMA models, however, the convergence rate of $\widehat{r}_n$ depends not only on the continuity of the autoregressive function, but also on the coefficients in the moving-average part. That is, the continuity of the autoregressive function is not a unique determinant for n-convergence rate of $\widehat{r}_n$. This point is extremely different from TAR models.

If Assumption 3.5 is violated and the threshold can be identified, then $d \leq p$ and model (2.1) becomes

$$(3.3)\quad y_t = \phi_0 + \sum_{\substack{i=1 \\ i \neq d}}^{p} \phi_i y_{t-i} + \phi_d^-(y_{t-d} - r)^- + \phi_d^+(y_{t-d} - r)^+ + \sum_{i=0}^{q} \psi_i \varepsilon_{t-i},$$

where $\phi_i, \psi_j,\ \phi_d^-$ and ϕ_d^+ are the parameters, $\psi_0 = 1$, $(y)^- = \min\{y, 0\}$ and $(y)^+ = \max\{y, 0\}$. Model (3.3) can be regarded as an extension of the TAR model of Chan and Tsay (1998) when the autoregressive function is continuous, under which $\widehat{r}_n$ has a $\sqrt{n}$-convergence rate like other coefficients and is asymptotically normal. We conjecture that similar results hold for model (3.3) under some suitable conditions. Here, we will not pursue this problem.

Theorem 3.2 below shows the convergence rates of $\widehat{r}_n$ and the uniformly asymptotic normality of $\widehat{\boldsymbol{\lambda}}_n(\widehat{r}_n)$.

THEOREM 3.2.   *If Assumptions 3.1-3.5 hold, then*

(i). $n(\widehat{r}_n - r_0) = O_p(1),$

(ii). $\sqrt{n} \sup_{|r-r_0| \leq B/n} \|\widehat{\boldsymbol{\lambda}}_n(r) - \widehat{\boldsymbol{\lambda}}_n(r_0)\| = o_p(1)$

for any fixed constant $B \in (0, \infty)$. Furthermore,

$$\sqrt{n}(\widehat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0) = \sqrt{n}(\widehat{\boldsymbol{\lambda}}_n(r_0) - \boldsymbol{\lambda}_0) + o_p(1) \xrightarrow{d} \mathcal{N}(0,\ \sigma^2 \Sigma^{-1}) \quad as\ n \to \infty,$$

where $\sigma^2 = E\varepsilon_1^2$ and $\Sigma = E[(\partial \varepsilon_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\lambda})(\partial \varepsilon_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\lambda}')].$

Generally, since σ^2 and Σ are unknown, we estimate them consistently by

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2(\widehat{\boldsymbol{\theta}}_n) \quad \text{and} \quad \widehat{\Sigma}_n = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial\varepsilon_t(\widehat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\lambda}}\frac{\partial\varepsilon_t(\widehat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\lambda}'}.$$

Using these estimators, one can easily construct approximate confidence regions for the coefficient parameters in TARMA models in terms of the limiting distribution in Theorem 3.2. For example, an approximate $(1-\alpha)$ confidence region for the coefficient vector $\boldsymbol{\lambda}$ is obtained as

$$\left\{\boldsymbol{\lambda} : (\widehat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda})'\widehat{\Sigma}_n(\widehat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}) \leq \widehat{\sigma}_n^2\chi_{p_1+p_2+q_1+q_2+2}^2(1-\alpha)/n\right\},$$

where $\chi_k^2(1-\alpha)$ is the 100αth percentile of the χ^2-distribution with k degrees of freedom.

To study the limiting distribution of $\widehat{r}_n$, we need to consider the following profile sum of square errors function:

$$\tilde{L}_n(z) = L_n\big(\widehat{\boldsymbol{\lambda}}_n(r_0 + \frac{z}{n}), r_0 + \frac{z}{n}\big) - L_n\big(\widehat{\boldsymbol{\lambda}}_n(r_0), r_0\big), \quad z \in R.$$

It is not hard to show that $\tilde{L}_n(z)$ can be approximated in $D(R)$, the space of all cadlag functions on R being equipped with the Skorokhod metric, by

$$
\begin{aligned}
\wp_n(z) =& I(z < 0)\sum_{t=1}^{n}\zeta_{1t}I(r_0 + \frac{z}{n} < y_{t-d} \leq r_0) \\
&+ I(z \geq 0)\sum_{t=1}^{n}\zeta_{2t}I(r_0 < y_{t-d} \leq r_0 + \frac{z}{n}),
\end{aligned}
$$

where

$$
\begin{aligned}
(3.4) \qquad \zeta_{1t} &= \Big\{\sum_{j=0}^{\infty}\big[\mathbf{e}'\mathbf{H}_{t+j,j}(\boldsymbol{\theta}_0)\mathbf{e}\big]^2\Big\}\delta_t^2 + 2\Big\{\sum_{j=0}^{\infty}\varepsilon_{t+j}\big[\mathbf{e}'\mathbf{H}_{t+j,j}(\boldsymbol{\theta}_0)\mathbf{e}\big]\Big\}\delta_t, \\
\zeta_{2t} &= \Big\{\sum_{j=0}^{\infty}\big[\mathbf{e}'\mathbf{H}_{t+j,j}(\boldsymbol{\theta}_0)\mathbf{e}\big]^2\Big\}\delta_t^2 - 2\Big\{\sum_{j=0}^{\infty}\varepsilon_{t+j}\big[\mathbf{e}'\mathbf{H}_{t+j,j}(\boldsymbol{\theta}_0)\mathbf{e}\big]\Big\}\delta_t
\end{aligned}
$$

and

$$\delta_t = (\phi_{100} - \phi_{200}) + \sum_{i=1}^{p}(\phi_{1i0} - \phi_{2i0})y_{t-i} + \sum_{i=1}^{q}(\psi_{1i0} - \psi_{2i0})\varepsilon_{t-i}.$$

Let $F_k(\cdot|r_0)$ be the conditional distribution of $\zeta_{k,d+1}$ given $y_1 = r_0$ for $k = 1, 2$. We define a two-sided compound Poisson process $\wp(z)$ as follows,

$$(3.5) \qquad \{\wp(z), z \in R\} = \{\wp^{(1)}(-z)I(z < 0) + \wp^{(2)}(z)I(z \geq 0), \ z \in R\},$$

where $\{\wp^{(1)}(z), z \geq 0\}$ and $\{\wp^{(2)}(z), z \geq 0\}$ are two independent Poisson processes with $\wp^{(1)}(0) = \wp^{(2)}(0) = 0$ a.s., with the same jump rate $\pi(r_0) > 0$, where $\pi(x)$ is the density of y_0, and with the jump distributions $F_1(\cdot|r_0)$ and $F_2(\cdot|r_0)$, respectively. Clearly, $\wp(z)$ goes to $+\infty$ a.s. when $|z|$ tends to infinity since $\int x F_k(dx|r_0) > 0$. Thus, there exists a unique random interval $[M_-, M_+)$ on which the process $\{\wp(z), z \in R\}$ attains its global minimum a.s.. Here, we work with the left continuous version for $\wp^{(1)}$ and the right continuous version for $\wp^{(2)}$. Now, we can state our another main result as follows.

THEOREM 3.3. *Suppose Assumptions 3.1-3.5 hold. Then $n(\widehat{r}_n - r_0) \overset{d}{\longrightarrow} M_-$ and $n(\widehat{r}_n - r_0)$ is asymptotically independent of $\sqrt{n}(\widehat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0)$ which is always asymptotically normally distributed.*

Now, we give some remarks on (3.4) to end this section. When $q = 0$, model (2.1) is a pure TAR model, that is, all $\psi_{ij} = 0$. Thus, (3.4) reduces to $\zeta_{1t} = \delta_t^2 + 2\varepsilon_t \delta_t$ and $\zeta_{2t} = \delta_t^2 - 2\varepsilon_t \delta_t$ with $\delta_t = (\phi_{100} - \phi_{200}) + \sum_{i=1}^{p}(\phi_{1i0} - \phi_{2i0})y_{t-i}$. If the condition $y_{t-d} = r_0$ is given, then this coincides with the jump sizes of the compound Poisson process used in Chan (1993). When $p = 0$, model (2.1) reduces to a pure TMA model and (3.4) does not change except for δ_t with $\delta_t = (\phi_{100} - \phi_{200}) + \sum_{i=1}^{q}(\psi_{1i0} - \psi_{2i0})\varepsilon_{t-i}$. This is identical to those in Li, Ling and Li (2010).

4. Numerical Implementation of M_-. In this section, we describe how to simulate the distribution of M_-. From (3.5), we know that two factors determine the density of M_-, that is, the jump rate $\pi(r_0)$ and the jump distributions $F_1(\cdot|r_0)$ and $F_2(\cdot|r_0)$. We can simulate M_- via simulating the compound Poisson process (3.5) on the interval $[-T, T]$ for any given $T > 0$ large enough since the expectations of the jumps are positive. Modifying Algorithm 6.2 in Cont and Tankov (2004, pp.174) for a one-sided compound Poisson process, we have an algorithm for a two-sided compound Poisson process as follows:

Algorithm B

Step B.1. Generate two i.i.d. Poisson random variables N_1 and N_2 with the parameter $\pi(r_0)T$ as the total number of jumps on the intervals $[-T, 0]$ and $[0, T]$, respectively.

Step B.2. Given N_1 and N_2, generate $\{U_1, \cdots, U_{N_1}\}$ and $\{V_1, \cdots, V_{N_2}\}$ as two independent jump time sequences , where $\{U_i\} \overset{i.i.d.}{\sim} U[-T, 0]$ and $\{V_i\} \overset{i.i.d.}{\sim} U[0, T]$. Here, $U[a, b]$ denotes the uniform distribution on the interval $[a, b]$.

Step B.3. Given N_1 and N_2, generate $\{Y_1, \cdots, Y_{N_1}\}$ and $\{Z_1, \cdots, Z_{N_2}\}$ as two independent jump-size sequences from $F_1(\cdot|r_0)$ and $F_2(\cdot|r_0)$, respectively.

For $z \in [-T, T]$, the trajectory of (3.5) is given by

$$(4.1) \quad \wp(z) = I(z < 0) \sum_{i=1}^{N_1} I(U_i > z)Y_i + I(z \geq 0) \sum_{j=1}^{N_2} I(V_j < z)Z_j.$$

Then, we take the smallest minimizer of $\wp(z)$ on $[-T, T]$ as M_-. By repeating the above algorithm B times and using the nonparametric kernel method, we can get the density of M_- numerically.

In the above algorithm, the key step is how to draw the jump-size sequences $\{Y_i\}$ and $\{Z_i\}$ from $F_1(\cdot|r_0)$ and $F_2(\cdot|r_0)$ in Step B.3, respectively. Note that $\zeta_{k,d+1}$'s in (3.4) are sums of infinite series and $\|\mathbf{H}_{d+1+j,j}(\boldsymbol{\theta}_0)\| = O(\rho^j)$ a.s., we truncate them and use the finite-term sums to approximate them as follows:

$$\zeta_{k,d+1}^{(m)} = \Big\{ \sum_{j=0}^{m} \Big(\mathbf{e}' \prod_{i=1}^{j} [\boldsymbol{\Psi}_{20} + (\boldsymbol{\Psi}_{10} - \boldsymbol{\Psi}_{20})I(y_{i+1} \leq r_0)]\mathbf{e}\Big)^2 \Big\} \delta_{d+1}^2$$

$$+ 2(-1)^{k+1} \Big\{ \sum_{j=0}^{m} \varepsilon_{d+1+j}\Big(\mathbf{e}' \prod_{i=1}^{j} [\boldsymbol{\Psi}_{20} + (\boldsymbol{\Psi}_{10} - \boldsymbol{\Psi}_{20})I(y_{i+1} \leq r_0)]\mathbf{e}\Big) \Big\} \delta_{d+1},$$

where $\boldsymbol{\Psi}_{i0}$ is defined in (3.2) with ψ_{ij}'s replaced by true values ψ_{ij0}'s, since the remaining terms are negligible when m is large enough, where $\rho \in (0, 1)$. Then,

$$F_k(x|r_0) \approx P(\zeta_{k,d+1}^{(m)} \leq x | y_1 = r_0)$$

$$= \int P(\zeta_{k,d+1}^{(m)} \leq x | y_1 = r_0, \mathbf{Z}_0 = \mathbf{z}) \frac{\pi(r_0|\mathbf{z})}{\pi(r_0)} G(d\mathbf{z})$$

$$\approx \sum_{i=1}^{K} P(\zeta_{k,d+1}^{(m)} \leq x | y_1 = r_0, \mathbf{Z}_0 = \mathbf{z}_i) \frac{\pi(r_0|\mathbf{z}_i)}{\sum_{l=1}^{K} \pi(r_0|\mathbf{z}_l)}$$

by the property of conditional expectation, strong law of large numbers (i.e., $K \to \infty$) and $E[\pi(r_0|\mathbf{Z}_0)] = \pi(r_0)$, where $\mathbf{Z}_0 = (y_0, ..., y_{1-(p \vee d)}, \varepsilon_0, ..., \varepsilon_{1-q})'$, $\mathbf{z}_i \in R^{(p \vee d)+q}$, $G(\cdot)$ is the distribution of $\mathbf{Z}_0$, $\pi(r_0|\mathbf{z})$ is the conditional density of y_1 given $\mathbf{Z}_0 = \mathbf{z}$. Clearly, there has a function $h(\cdot, \cdot)$ such that model (2.1) can be written as $y_t = \varepsilon_t + h(\mathbf{Z}_{t-1}, \boldsymbol{\theta}_0)$. Given $y_1 = r_0$ and $\mathbf{Z}_0 = \mathbf{z}$, we have $\varepsilon_1 = r_0 - h(\mathbf{z}, \boldsymbol{\theta}_0)$ and $\pi(r_0|\mathbf{z}) = f_\varepsilon(r_0 - h(\mathbf{z}, \boldsymbol{\theta}_0))$. Let $F_\varepsilon(\cdot)$ be the cumulative distribution function of ε_t. When $\boldsymbol{\theta}_0$, $\pi(r_0)$, $F_\varepsilon(\cdot)$ and $G(\cdot)$ are

known, the following algorithm describes how to sample Y_1 from $F_1(x|r_0)$ approximately.

Algorithm C:

Step C.1. Choose sufficiently large positive integers m and K and then draw a sample $\{\mathbf{z}_1, ..., \mathbf{z}_K\}$ from $G(\cdot)$.

Step C.2. For fixed i, generate an i.i.d. sequence $\{\varepsilon_t : 2 \leq t \leq d+1+m\}$ from $F_\varepsilon(\cdot)$ and iterate model (2.1) to generate $\{y_t : 2 \leq t \leq d+1+m\}$ with the initial value $y_1 = r_0$ and $\mathbf{Z}_0 = \mathbf{z}_i$. Here, the initial value of ε_1 is obtained by $\varepsilon_1 = r_0 - h(\mathbf{z}_i, \boldsymbol{\theta}_0)$. And use these values to calculate $\zeta_{1,d+1}^{(m)}$ and denote its value as $\xi_{1,d+1}^{(i)}$.

Step C.3. Calculate $\pi(r_0|\mathbf{z}_i)$'s and generate a random variable U such that $P(U = i) = \pi(r_0|\mathbf{z}_i)/[\sum_{l=1}^{K} \pi(r_0|\mathbf{z}_l)]$ for $i = 1, ..., K$, conditionally independent of $\{y_t, t \geq 2\}$ and $\{\varepsilon_t, t \geq 2\}$.

Step C.4. Obtain $Y_1 = \xi_{1,d+1}^{(U)}$.

It is not hard to prove that the distribution of Y_1 is $F_1(x|r_0)$ approximately. Similarly, we can sample Z_1 from $F_2(x|r_0)$ approximately. By **Algorithm B** and **C**, we can get an $\widetilde{M}_K^{(m)}$ as an approximation of M_- by Theorem 16 in Pollard (1984, pp. 134) and Lemma A.3 in Seijo and Sen (2010). Without loss of generality, in what follows, we regard $\widetilde{M}_K^{(m)}$ as M_-.

In practice, however, since only one sample $\mathfrak{X} \equiv \{y_1, ..., y_n\}$ is available, we can use it to estimate $\boldsymbol{\theta}_0$ and $\pi(r_0)$ a.s., denoting the estimators as $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\pi}(\widehat{r}_n)$, respectively, where $\widehat{\pi}(\cdot)$ is the kernel density estimator of $\pi(\cdot)$, and calculate the residuals $\{\widehat{\varepsilon}_t : k_0 + 1 \leq t \leq n\}$, where $k_0 = \max(p, d, q)$. Based on the residuals, we construct the estimators $\widehat{F}_\varepsilon(\cdot)$ and $\widehat{f}_\varepsilon(\cdot)$ of $F_\varepsilon(\cdot)$ and $f_\varepsilon(\cdot)$ such that $\|\widehat{F}_\varepsilon - F_\varepsilon\|_\infty \to 0$ and $\widehat{f}_\varepsilon \to f_\varepsilon$ a.s. as $n \to \infty$. Denote the conditional measure $P_\mathfrak{X}(\cdot|A) \equiv P(\cdot|A, \mathfrak{X})$.

Algorithm D:

Step D.1. Set $\hat{\mathbf{z}}_i = (y_i, ..., y_{i-(p\vee d)+1}, \widehat{\varepsilon}_i, ..., \widehat{\varepsilon}_{i-q+1})'$ for $i = k_0 + 1, ..., n$.

Step D.2. For fixed i, generate a conditionally i.i.d. sequence $\{\tilde{\varepsilon}_t : 2 \leq t \leq d+1+m\}$ from $\widehat{F}_\varepsilon$ and iterate model (2.1) to generate $\{\tilde{y}_t : 2 \leq t \leq d+1+m\}$ with the initial value $y_1 = \widehat{r}_n$ and $\mathbf{Z}_0 = \hat{\mathbf{z}}_i$. Here, the initial value of ε_1 is obtained by $\varepsilon_1 = \widehat{r}_n - h(\hat{\mathbf{z}}_i, \widehat{\boldsymbol{\theta}}_n)$. And use these values to

calculate $\tilde{\zeta}_{1,d+1}^{(m)}$ and denote its value as $\tilde{\xi}_{1,d+1}^{(i)}$, where

$$\tilde{\zeta}_{1,d+1}^{(m)} = \Big\{ \sum_{j=0}^{m} \Big(\mathbf{e}' \prod_{i=1}^{j} [\widehat{\mathbf{\Psi}}_{2n} + (\widehat{\mathbf{\Psi}}_{1n} - \widehat{\mathbf{\Psi}}_{2n})I(\tilde{y}_{i+1} \leq \widehat{r}_n)]\mathbf{e} \Big)^2 \Big\} \delta_{d+1}^{*\,2}$$

$$+ 2\Big\{ \sum_{j=0}^{m} \tilde{\varepsilon}_{d+1+j} \Big(\mathbf{e}' \prod_{i=1}^{j} [\widehat{\mathbf{\Psi}}_{2n} + (\widehat{\mathbf{\Psi}}_{1n} - \widehat{\mathbf{\Psi}}_{2n})I(\tilde{y}_{i+1} \leq \widehat{r}_n)]\mathbf{e} \Big) \Big\} \delta_{d+1}^{*}$$

and

$$\delta_{d+1}^{*} = (\widehat{\phi}_{10n} - \widehat{\phi}_{20n}) + \sum_{s=1}^{p}(\widehat{\phi}_{1sn} - \widehat{\phi}_{2sn})y_{d+1-s}^{*} + \sum_{s=1}^{q}(\widehat{\psi}_{1sn} - \widehat{\psi}_{2sn})\epsilon_{d+1-s}^{*}$$

with

$$y_j^* = \begin{cases} \tilde{y}_j, & \text{if } j \geq 2, \\ \widehat{r}_n, & \text{if } j = 1, \\ y_{i-j}, & \text{if } j \leq 0, \end{cases} \quad \text{and} \quad \epsilon_j^* = \begin{cases} \tilde{\varepsilon}_j, & \text{if } j \geq 2, \\ \widehat{r}_n - h(\mathbf{z}_i, \widehat{\boldsymbol{\theta}}_n), & \text{if } j = 1, \\ \widehat{\varepsilon}_{i-j}, & \text{if } j \leq 0, \end{cases}$$

Step D.3. Calculate $\widehat{\pi}(\widehat{r}_n|\mathbf{z}_i)$'s and generate a random variable U such that $P(U = i) = \widehat{\pi}(\widehat{r}_n|\mathbf{z}_i)/[\sum_{l=1}^{K} \widehat{\pi}(\widehat{r}_n|\mathbf{z}_l)]$ for $i = k_0 + 1, ..., n$, conditionally independent of $\{\tilde{y}_t, t \geq 2\}$ and $\{\tilde{\varepsilon}_t, t \geq 2\}$ given $\mathfrak{X}$.

Step D.4. Obtain $\tilde{Y}_1 = \tilde{\xi}_{1,d+1}^{(U)}$.

Denote $\widehat{M}_{n,K}^{(m)}$ obtained by **Algorithm B** and **D** as another approximation of M_-. Then, we have

THEOREM 4.1. *Suppose Assumptions 3.1-3.5 hold. Then*

$$\lim_{m\to\infty} \lim_{\substack{n\to\infty \\ K\to\infty}} |P_{\mathfrak{X}}(\widehat{M}_{n,K}^{(m)} \leq x) - P(M_- \leq x)| = 0$$

whenever $P(M_- = x) = 0$.

We now consider the following TARMA(1,1,1,1) model:

$$(4.2) \qquad y_t = \begin{cases} \phi_{10} + \phi_{11}y_{t-1} + \varepsilon_t + \psi_{11}\varepsilon_{t-1}, & \text{if } y_{t-1} \leq r_0, \\ \phi_{20} + \phi_{21}y_{t-1} + \varepsilon_t + \psi_{21}\varepsilon_{t-1}, & \text{if } y_{t-1} > r_0, \end{cases}$$

where $\boldsymbol{\theta}_0 = (\phi_{10}, \phi_{11}, \psi_{11}, \phi_{20}, \phi_{21}, \psi_{21}, r_0)' = (0.6, 0.6, -0.7, -1, 0.4, 0.5, 0.2)'$ and $\{\varepsilon_t\}$ is i.i.d standard normal. Since it is difficult to obtain the exact value of $\pi(r_0)$, we adopt the kernel method to estimate it, where the sample size is $N = 10,000$ and the Gaussian kernel is used. The bandwidth is chosen by

$$\widehat{h}_{\text{opt}}^{*} = \widehat{h}_{\text{opt}} \Big(1 + \frac{35}{48}\widehat{\gamma}_4 + \frac{35}{32}\widehat{\gamma}_3^2 + \frac{385}{1024}\widehat{\gamma}_4^2 \Big)^{-1/5},$$

where $\widehat{h}_{\mathrm{opt}} = 1.06sN^{-1/5}$ is the reference bandwidth selector, and $s, \widehat{\gamma}_3$ and $\widehat{\gamma}_4$ are the sample standard deviation, skewness and kurtosis, respectively. See Fan and Yao (2003, pp.201). To ensure the precision of the estimation, 100 replications are used. Finally, we get $\pi(r_0) \approx 0.3241$.
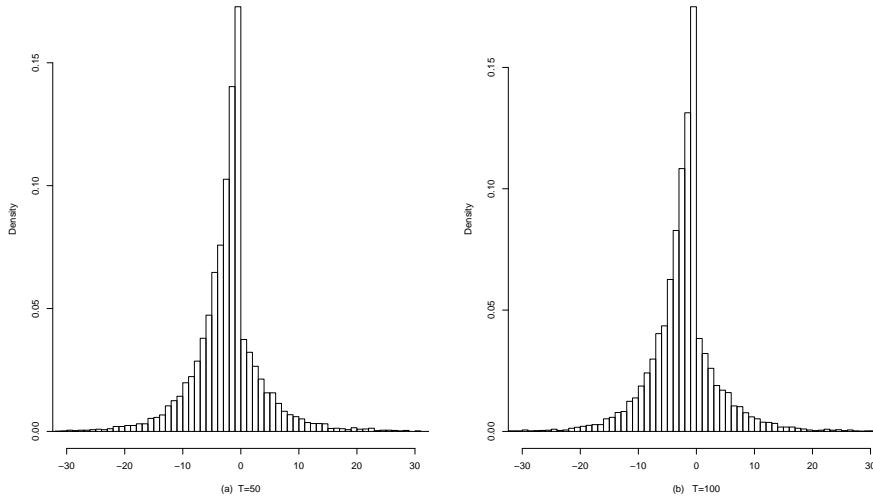

FIG 1. *The density of M_- for different values of T. (a). $T = 50$; (b). $T = 100$.*

Fig 1 gives the densities of M_- when $T = 50$ and 100 by using **Algorithm B** and **C**. 1001 observations (i.e., $K = 1000$) together with the driven noises generated by model (4.2) and $m = 200$ are used to calculate the jump sizes $\zeta_{k,d+1}^{(m)}$, $k = 1, 2$. Then, minimizing the process (4.1), we can obtain an observation of M_-. Here, 10,000 replications are used to get the density of M_-. From Fig 1, we can see that the densities of M_- are very close to each other for $T = 50$ and 100. In theory, the larger T, the more precise the density of M_-. How to select a suitable T is an interesting and open problem and there is no theory to support this in the literature.

Table 1 gives the quantiles of M_- under model (4.2) when the significance level $\alpha =$ 0.005, 0.01, 0.025, 0.05, 0.95, 0.975, 0.99 and 0.995.

TABLE 1
Quantiles for M_- under model (4.2).

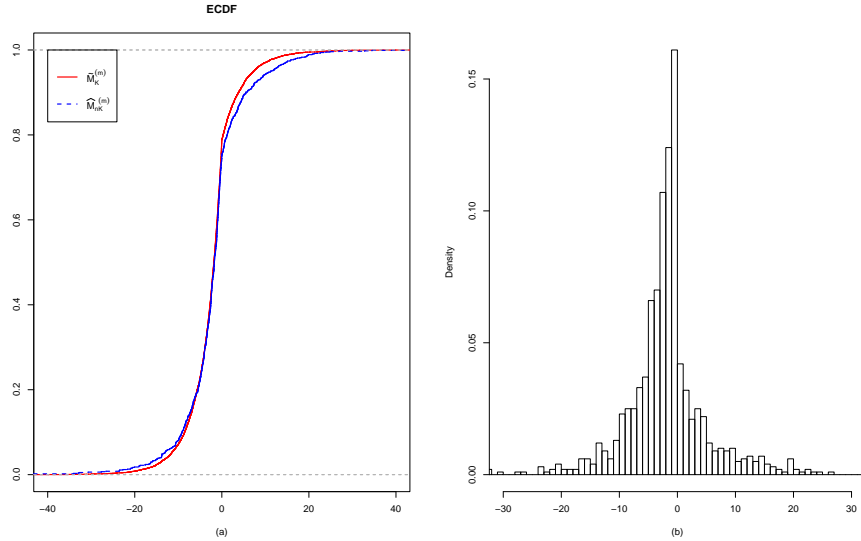| α	0.5%	1%	2.5%	5%	95%	97.5%	99%	99.5%
Q	-22.05	-19.18	-14.74	-11.61	7.27	10.88	15.86	20.06

FIG 2. *(a). The empirical cumulative distribution functions of $\widetilde{M}_K^{(m)}$ (M_-) and $\widehat{M}_{n,K}^{(m)}$. (b). The density of $\widehat{M}_{n,K}^{(m)}$, where the sample size is $n = 400$ and 1000 replications are used.*

For **Algorithm D**, simulation studies show that it works well in practice. Suppose that a sample $\{y_1, ..., y_{400}\}$ is from model (4.2) and fixed. Then we consistently estimate the quantities involved in Algorithm D. $m = 200$ and 1000 replications are used to obtain $\widehat{M}_{n,K}^{(m)}$. Fig 2 (a) shows the empirical cumulative distribution functions of $\widehat{M}_{n,K}^{(m)}$ and $\widetilde{M}_K^{(m)}$. Fig 2 (b) gives the density of $\widehat{M}_{n,K}^{(m)}$, which is close to those in Fig 1.

To further evaluate the performance of the simulation method above, we consider another example, e.g., a higher-order TARMA(3,3,0,0) model:

$$(4.3) \quad y_t = \begin{cases} 1 + 0.2y_{t-1} - 0.4y_{t-2} + 0.3y_{t-3} + \varepsilon_t, & y_{t-3} \leq 0, \\ -1 + 0.6y_{t-1} + 0.1y_{t-2} - 0.2y_{t-3} + \varepsilon_t, & y_{t-3} > 0, \end{cases}$$

where $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, 1)$. The simulation results are displayed in Fig 3. The density of $n(\widehat{r}_n - r_0)$ is given in Fig 3 (a), where the sample size is 600 and 10,000 replications are used. Fig 3 (b) shows the density of M_-, where 10,000 replications are used. Fig 3 (c)-(d) show the density of $\widehat{M}_-$ (obtained by Algorithm B and D) for a given and fixed sample $\{y_1, ..., y_{600}\}$ from model (4.3), where 1000 and 10,000 replications are used for (c) and (d), respectively. From Fig 3, we can see that (a), (b) and (d) are very matching.
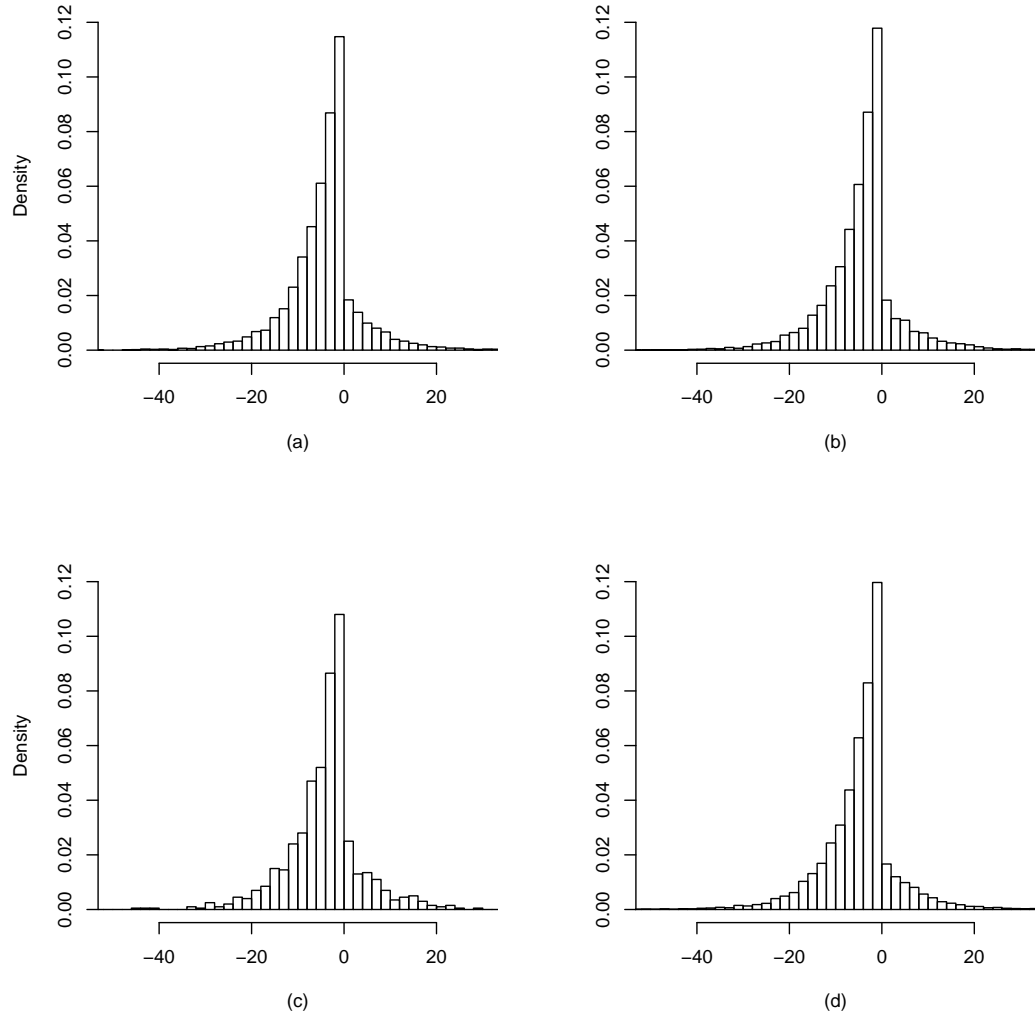
FIG 3. *(a). The density of $n(\widehat{r}_n - r_0)$. (b). The density of M_-. (c). The density of $\widehat{M}_-$ when 1000 replications are used. (d). The density of $\widehat{M}_-$ when 10,000 replications are used.*

5. Simulation studies. To assess the performance of the LSE of $\boldsymbol{\theta}_0$ in finite samples, we use sample sizes $n =$200, 400, 600 and 800, each with replications 1000 for model (4.2) with $\varepsilon_t \sim \mathcal{N}(0,1)$. In Table 2, we summarize the empirical mean, empirical standard deviation (ESD) and asymptotic standard deviation (ASD). Here, the ASD of $\widehat{\boldsymbol{\lambda}}$ are computed using Σ in Theorem 3.2 and the ASD of $\widehat{r}_n$ is obtained by simulating M_-.

TABLE 2
Simulation studies for model (4.2) with $\boldsymbol{\theta}_0 = (0.6, 0.6, -0.7, -1, 0.4, 0.5, 0.2)'$.

n		ϕ_{10}	ϕ_{11}	ψ_{11}	ϕ_{20}	ϕ_{21}	ψ_{21}	r
	EM	0.6149	0.6099	-0.7241	-1.0063	0.4046	0.5107	0.1877
200	ESD	0.1315	0.1920	0.1935	0.2103	0.2101	0.1936	0.0325
	ASD	0.1167	0.1636	0.1553	0.2003	0.1837	0.1535	0.0296
	EM	0.6083	0.6047	-0.7092	-1.0057	0.4076	0.4987	0.1953
400	ESD	0.0879	0.1215	0.1226	0.1529	0.1406	0.1196	0.0165
	ASD	0.0825	0.1157	0.1098	0.1416	0.1299	0.1085	0.0148
	EM	0.6086	0.6059	-0.7071	-1.0058	0.4073	0.4951	0.1965
600	ESD	0.0691	0.0952	0.0958	0.1128	0.1064	0.0944	0.0093
	ASD	0.0674	0.0945	0.0897	0.1156	0.1061	0.0886	0.0099
	EM	0.5999	0.6060	-0.7127	-1.0001	0.4026	0.4973	0.1967
800	ESD	0.0601	0.0836	0.0796	0.0987	0.0930	0.0804	0.0071
	ASD	0.0583	0.0818	0.0777	0.1001	0.0919	0.0767	0.0074

From Table 2, we see that the consistency of the estimators is shown by the empirical means and the closeness of the empirical standard deviations to the asymptotic standard deviations. We also see that the values of the ESDs for $\widehat{r}_n$ are about halfed each time when the value of n is doubled. This partially illustrates the n-consistency of the threshold estimator, under which the estimator of the threshold parameter would approach the true parameter much faster than the coefficient parameter estimators.

Table 3 reports the coverage probabilities of the estimator $\widehat{r}_n$ for $n =$200, 400, 600 and 800, respectively, based on the critical values in Table 1. From the table, we can see that the coverage probability is rather accurate even though the sample size n is 200.

TABLE 3
Coverage probabilities.

α	200	400	600	800
0.01	0.979	0.986	0.991	0.988
0.05	0.936	0.936	0.966	0.960
0.10	0.890	0.888	0.917	0.911

Fig 4 shows the density of $n(\widehat{r}_n - r_0)$ when the sample size $n = 200, 400,$ 600 and 800, respectively. From Fig 4, we see that the densities of $n(\widehat{r}_n - r_0)$ and M_- in Fig 1 are matching, which supports Theorem 3.3 empirically. We
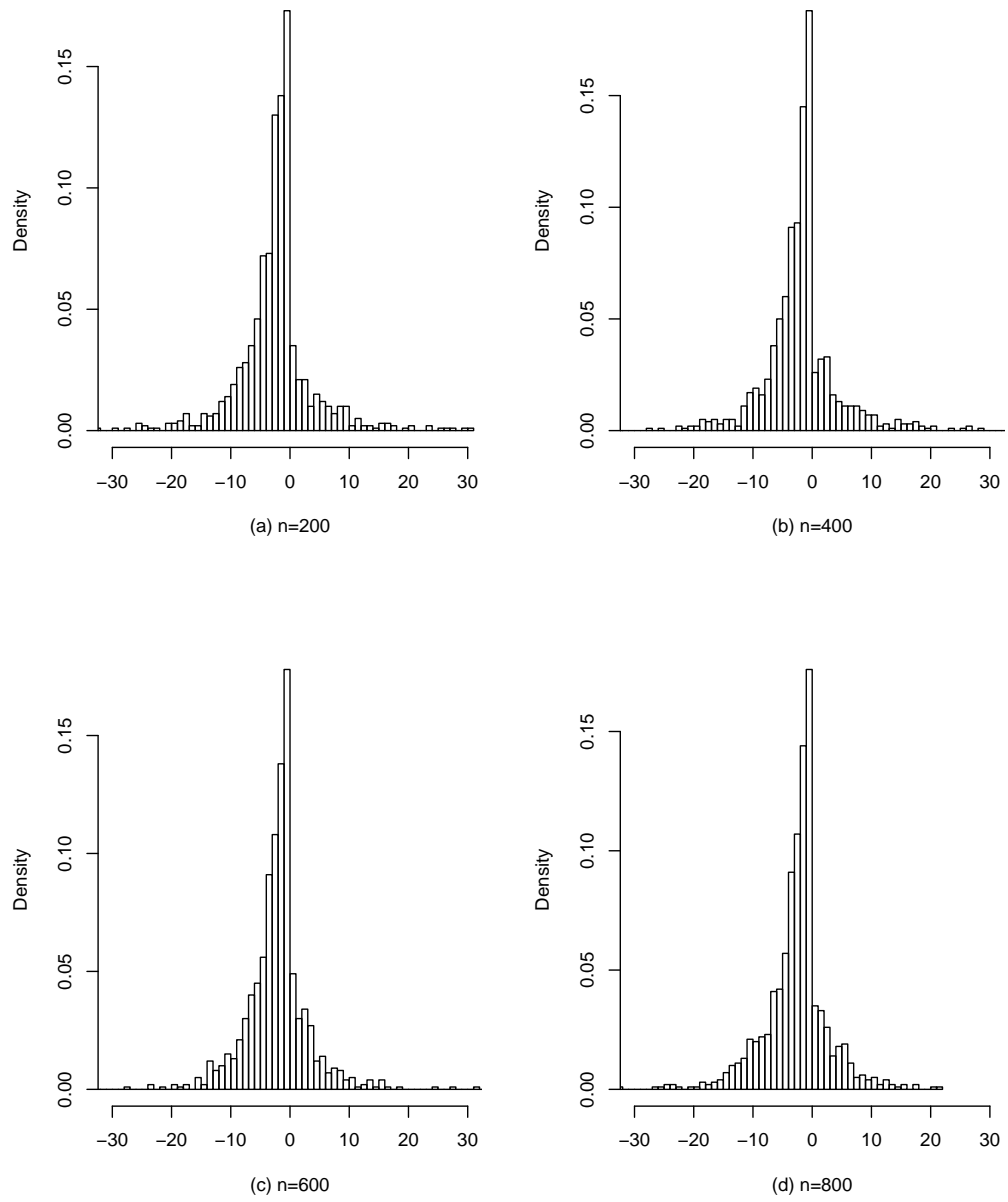
(a) n=200

(b) n=400

(c) n=600

(d) n=800

FIG 4. *The density of $n(\widehat{r}_n - r_0)$ when the sample size $n = 200, 400, 600$ and 800, respectively.*

also see that the density of $n(\widehat{r}_n - r_0)$ is leptokurtic and asymmetric. Here the skewness of M_- is 0.234 and the kurtosis is 8.96. Due to the skewness and kurtosis, we must be careful in constructing confidence intervals of the estimated threshold in practice.

6. Concluding remarks. This paper considers the general TARMA model which contains the pure TAR model and the pure TMA model as two special cases and has established the asymptotic theory of the least square estimation. The limiting distribution of the estimated threshold, which is the smallest minimizer of a two-sided compound Poisson process, has been derived. A numerical method is proposed to simulate this limiting distribution and simulation studies show that it does work well. These results are obtained under which the order (p_1, p_2, q_1, q_2) is known. In applications, since the order is generally unknown, one can use either the AIC or BIC to determine it and we do not pursue this problem anymore. As for TARMA model (3.3) where the autoregressive function is continuous and there is no threshold in the moving-average part, we conjecture that the estimated threshold has a $\sqrt{n}$-convergence rate and is asymptotically normal as well as other parameters. This conjecture corresponds to the results in Chan and Tsay (1998) on TAR models when the autoregressive function is continuous. For this case, further study will be needed in the future.

APPENDIX A: PROOF OF THEOREM 3.1

In this section, we give the proof for the case $p = q = 1$ without the constant terms in each regime. The general case can be proved similarly. Temporarily, $\boldsymbol{\phi}_i$ and $\boldsymbol{\psi}_i$ are abbreviated to ϕ_i and ψ_i. The proof of Theorem 3.1 is similar to that of Theorem 2.1 in Li et al. (2010) and can be complete by Huber's method and the following Lemma A.1. Hence, it is omitted.

LEMMA A.1. *If the conditions of Theorem 3.1 hold, then $E\varepsilon_t^2(\boldsymbol{\theta}) \geq \sigma^2$ and the equality holds if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.*

PROOF. By Theorem A.2 in Ling and Tong (2005), $E\varepsilon_t^2(\boldsymbol{\theta}) < \infty$. Clearly, $E\varepsilon_t^2(\boldsymbol{\theta}_0) = \sigma^2$ since $\varepsilon_t(\boldsymbol{\theta}_0) = \varepsilon_t$. A conditional argument yields

$$
\begin{aligned}
E\{\varepsilon_t^2(\boldsymbol{\theta}) - \varepsilon_t^2\} &= E\{[\varepsilon_t + \nabla_{t-1}(\boldsymbol{\theta})]^2 - \varepsilon_t^2\} \\
&= E\nabla_{t-1}^2(\boldsymbol{\theta}) + 2E\{\nabla_{t-1}(\boldsymbol{\theta})E(\varepsilon_t|\mathcal{F}_{-\infty}^{t-1})\} \\
&= E\nabla_{t-1}^2(\boldsymbol{\theta}) \geq 0,
\end{aligned}
$$

where $\nabla_{t-1}(\boldsymbol{\theta})$ is measurable with respect to $\mathcal{F}_{-\infty}^{t-1} = \sigma(\varepsilon_j : j \leq t-1)$.

If there exists a $\boldsymbol{\theta}_*$ such that $E\{\varepsilon_t^2(\boldsymbol{\theta}_*) - \varepsilon_t^2\} = 0$, then $\nabla_{t-1}(\boldsymbol{\theta}_*) = 0$ a.s. for each t since $\{\varepsilon_t(\boldsymbol{\theta}_*)\}$ is strictly stationary, and hence $\varepsilon_t(\boldsymbol{\theta}_*) \equiv \varepsilon_t + \nabla_{t-1}(\boldsymbol{\theta}_*) = \varepsilon_t$ a.s. for each t. Thus,

$$[(\phi_{20} - \phi_2^*) + (\phi_{10} - \phi_{20})I(y_{t-d_0} \leq r_0) - (\phi_1^* - \phi_2^*)I(y_{t-d^*} \leq r^*)]y_{t-1}$$
$$= [(\psi_2^* - \psi_{20}) + (\psi_1^* - \psi_2^*)I(y_{t-d^*} \leq r^*) - (\psi_{10} - \psi_{20})I(y_{t-d_0} \leq r_0)]\varepsilon_{t-1}.$$

which implies that

(A.1) $\quad (\phi_{20} - \phi_2^*) + (\phi_{10} - \phi_{20})I(y_{t-d_0} \leq r_0) - (\phi_1^* - \phi_2^*)I(y_{t-d^*} \leq r^*) = 0,$

(A.2) $\ (\psi_2^* - \psi_{20}) + (\psi_1^* - \psi_2^*)I(y_{t-d^*} \leq r^*) - (\psi_{10} - \psi_{20})I(y_{t-d_0} \leq r_0) = 0,$

since ε_{t-1} has a continuous and positive density on R and so does y_{t-1} by the convolution property, and y_{t-1} is not proportional to ε_{t-1}.

Now, we first prove that $d_* = d_0$. Otherwise, without loss of generality, assume that $d_0 > d_*$. Then, (A.1) reduces to

$$(\phi_{10} - \phi_1^*)I(y_{t-d_0} \leq r_0, y_{t-d^*} \leq r^*) + (\phi_{10} - \phi_2^*)I(y_{t-d_0} \leq r_0, y_{t-d^*} > r^*)+$$
$$(\phi_{20} - \phi_1^*)I(y_{t-d_0} > r_0, y_{t-d^*} \leq r^*) + (\phi_{20} - \phi_2^*)I(y_{t-d_0} > r_0, y_{t-d^*} > r^*) = 0.$$

Using the orthogonality among the indicators in the previous equation, it follows that

$$(\phi_{10} - \phi_1^*)^2 P(y_{t-d_0} \leq r_0, y_{t-d^*} \leq r^*) + (\phi_{10} - \phi_2^*)^2 P(y_{t-d_0} \leq r_0, y_{t-d^*} > r^*)+$$
$$(\phi_{20} - \phi_1^*)^2 P(y_{t-d_0} > r_0, y_{t-d^*} \leq r^*) + (\phi_{20} - \phi_2^*)^2 P(y_{t-d_0} > r_0, y_{t-d^*} > r^*) = 0.$$

We can write y_{t-d^*} into the form: $y_{t-d^*} = \varepsilon_{t-d^*} + \xi_{t-d^*-1}$, where ξ_{t-d^*-1} is a random variable and independent of ε_{t-d^*}. Since ε_{t-d^*} has a continuous and positive density on R and independent of y_{t-d_0},

$$P(y_{t-d^*} \leq r | y_{t-d_0} = x) = E\left[\int_{-\infty}^{r-\xi_{t-d^*-1}} h(z)dz \,\bigg|\, y_{t-d_0} = x\right] > 0$$

for $r \in [\underline{r}, \bar{r}]$ and any fixed x, where $h(x)$ is the density function of ε_t. Hence, all probabilities above are positive, which implies that $\phi_1^* = \phi_2^* = \phi_{10} = \phi_{20}$. Similarly, for (A.2), we have $\psi_1^* = \psi_2^* = \psi_{10} = \psi_{20}$. This is a contradiction with the condition (iii) in Theorem 3.1. Thus, $d_* = d_0$.

Secondly, without loss of generality, suppose that $r_* \leq r_0$. Then, using the preceding procedure, it follows that from (A.1)

$$(\phi_{10} - \phi_1^*)^2 P(y_{t-d_0} \leq r^*) + (\phi_{10} - \phi_2^*)^2 P(r^* < y_{t-d_0} \leq r_0)$$
$$+(\phi_{20} - \phi_2^*)^2 P(y_{t-d_0} > r_0) = 0.$$

Note that y_t has a continuous and positive density on R by Assumption 3.2, we have $\phi_1^* = \phi_{10}$, $\phi_2^* = \phi_{20}$ and $r_* = r_0$. Similarly, $\psi_1^* = \psi_{10}$, $\psi_2^* = \psi_{20}$. Thus, $\boldsymbol{\theta}_* = \boldsymbol{\theta}_0$. $\qquad\square$

APPENDIX B: PROOF OF THEOREM 3.2

In this section we give the proof for the case $q = 1$ and $1 \leq d \leq p$. For general case, the proof is similar. We first give a lemma.

LEMMA B.1. *Suppose that Assumptions 3.1-3.5 hold. Then, for any $\varepsilon > 0$ and $\eta > 0$, there exists a constant $B > 0$ such that for any fixed $\delta \in (0,1)$ small enough and any $n \geq 1$*

$$(B.1) \qquad P\left(\sup_{B/n < u < \delta} \left| \frac{\sum_{t=1}^n I^*(r_0 < y_{t-d} \leq r_0 + u)}{nG^*(u)} - 1 \right| < \eta \right) > 1 - \varepsilon,$$

$$(B.2) \qquad P\left(\sup_{B/n < u < \delta} \left| \frac{\sum_{t=1}^n A_t(0, u)}{nG(u)} \right| < \eta \right) > 1 - \varepsilon,$$

$$(B.3) \qquad P\left(\sup_{B/n < u < \delta} \left| \frac{\sum_{t=1}^n \varepsilon_t[\varepsilon_t(\boldsymbol{\lambda}_0, r_0 + u) - \varepsilon_t]}{nG(u)} \right| < \eta \right) > 1 - \varepsilon,$$

where $I^*(r_0 < y_{t-d} \leq r) = I(r_0 < y_{t-d} \leq r, \|\mathbf{Z}_{t-1} - \mathbf{Z}^*\| \leq \beta)$, $G^*(u) = P(r_0 < y_{t-d} \leq r_0 + u, \|\mathbf{Z}_{t-1} - \mathbf{Z}^*\| \leq \beta)$, $G(u) = P(r_0 < y_{t-d} \leq r_0 + u)$ and $A_t(u_1, u_2) = I(r_0 + u_1 < y_{t-d} \leq r_0 + u_2) \sum_{j=0}^\infty \rho^j \|\mathbf{Z}_{t-2-j}\| I(r_0 + u_1 < y_{t-1-d-j} \leq r_0 + u_2)$ for $u, u_1, u_2 \in [0, \delta]$. Here β is a fixed positive constant and $\mathbf{Z}^* = (\mathbf{w}', \mathbf{a}')'$ defined in Assumption 3.5.

PROOF. The proof of (B.1) is simple by Assumption 3.4, similar to that of (4.4a) in Chan (1993) and hence it is omitted. Since the proof of (B.3) is similar to that of (B.2), we only prove (B.2). It is sufficient to verify the following inequalities for $0 \leq u_1, u_2 \leq \delta$

$$(B.4) \qquad\qquad E[A_t(u_1, u_2)] \leq C[G(u_2) - G(u_1)],$$

$$(B.5) \qquad\qquad \mathrm{Var}(A_t(u_1, u_2)) \leq C[G(u_2) - G(u_1)],$$

$$(B.6) \qquad\qquad \mathrm{Var}\left(\sum_{t=1}^n A_t(u_1, u_2) \right) \leq Cn[G(u_2) - G(u_1)].$$

First, note that the following two facts. One is that for $k \geq 1$

$$P(r_0 < y_0 \leq r_0 + u, r_0 < y_k \leq r_0 + u) \leq Cu^2.$$

by a conditional argument. The other is that there exist two finite positive constants m_0 and M_0 such that

$$m_0(u_2 - u_1) \leq G(u_2) - G(u_1) \leq M_0(u_2 - u_1), \quad r_0 - 1 \leq u_1 \leq u_2 \leq r_0 + 1,$$

since the density of y_t is continuous and positive on R by Assumption 3.2 and the convolution property.

For (B.4), by Hölder's inequality and Assumptions 3.3-3.4, it follows that

$$(B.7) \qquad E[A_t(u_1, u_2)] \leq C(u_2 - u_1)^{3/2} \leq C[G(u_2) - G(u_1)].$$

The first inequality in (B.7) implies that it is not necessary to subtract $E[A_t(u_1, u_2)]$ from $A_t(u_1, u_2)$.

For (B.5), by Hölder's inequality and Assumptions 3.3 -3.4 again, we have

$$\text{Var}(A_t(u_1, u_2)) \leq E[A_t^2(u_1, u_2)] \leq C(E\|\mathbf{Z}_t\|^4)^{1/2}[G(u_2) - G(u_1)].$$

For (B.6), by a simple calculation, we have

$$|E(A_t|\mathcal{G}_0) - E(A_t)| \leq Ct\rho^t(1 + \|\mathbf{Z}_0\|) + C\rho^t + C\rho^t \sum_{j=0}^{\infty} \rho^j \|\mathbf{Z}_{-j}\|,$$

where $\mathcal{G}_0 = \sigma(\mathbf{Z}_t, t \leq 0)$, which implies that

$$\begin{aligned} |\text{Cov}(A_0(u_1, u_2), A_t(u_1, u_2))| &\leq E\{A_0|E(A_t|\mathcal{G}_0) - EA_t|\} \\ &\leq Ct\rho^t[G(u_2) - G(u_1)]. \end{aligned}$$

Thus,

$$\begin{aligned} &\text{Var}\Big(\sum_{t=1}^{n} A_t(u_1, u_2) \Big) \\ &= n\text{Var}(A_t(u_1, u_2)) + 2 \sum_{k=1}^{n-1} (n - k)\text{Cov}(A_0(u_1, u_2), A_k(u_1, u_2)) \\ &\leq Cn\Big\{ \sum_{k=1}^{\infty} k\rho^k \Big\}[G(u_2) - G(u_1)] \leq Cn[G(u_2) - G(u_1)]. \end{aligned}$$

The proof of is complete. $\qquad \square$

PROOF OF THEOREM 3.2 (i). Since $\widehat{\boldsymbol{\theta}}_n$ is consistent, we restrict Θ to an open neighborhood of $\boldsymbol{\theta}_0$, defined by $U_\delta = \{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\| < \delta, |r - r_0| < \delta\}$ for some $0 < \delta < 1$ to be determined later. Then it suffices to show that there exist constants $B > 0$, $\gamma > 0$, such that for n large enough

$$(B.8) \qquad P\Big(\inf_{\substack{B/n < |r - r_0| \leq \delta \\ \boldsymbol{\theta} \in U_\delta}} \frac{L_n(\boldsymbol{\lambda}, r) - L_n(\boldsymbol{\lambda}, r_0)}{nG(|r - r_0|)} > \gamma \Big) > 1 - 4\varepsilon.$$

Here, we only treat the case $r > r_0$. The proof for the case $r < r_0$ is similar.

Let $r = r_0 + u$ for some $0 < u < 1$. Decompose $L_n(\boldsymbol{\lambda}, r) - L_n(\boldsymbol{\lambda}, r_0)$ as

$$
\begin{aligned}
L_n(\boldsymbol{\lambda}, r) - L_n(\boldsymbol{\lambda}, r_0) &= \{[L_n(\boldsymbol{\lambda}, r) - L_n(\boldsymbol{\lambda}, r_0)] - [L_n(\boldsymbol{\lambda}_0, r) - L_n(\boldsymbol{\lambda}_0, r_0)]\} \\
&\quad + [L_n(\boldsymbol{\lambda}_0, r) - L_n(\boldsymbol{\lambda}_0, r_0)] \\
&\equiv L_n^{(1)}(\boldsymbol{\lambda}, r) + L_n^{(2)}(r).
\end{aligned}
$$

We first consider $L_n^{(2)}(r)$. Note that

$$
\begin{aligned}
\varepsilon_t(\boldsymbol{\lambda}_0, r) - \varepsilon_t &= f(\mathbf{Z}_{t-1}) + g(\mathbf{Z}_{t-1})[\varepsilon_{t-1}(\boldsymbol{\lambda}_0, r) - \varepsilon_{t-1}] \\
&= \sum_{j=0}^{\infty} \Big\{ \prod_{i=1}^{j} g(\mathbf{Z}_{t-i}) \Big\} f(\mathbf{Z}_{t-1-j}),
\end{aligned}
$$

where

$$
\begin{aligned}
f(\mathbf{Z}_{t-1}) &= \Big\{ \sum_{i=1}^{p} (\phi_{2i0} - \phi_{1i0}) y_{t-i} + (\psi_{20} - \psi_{10}) \varepsilon_{t-1} \Big\} I(r_0 < y_{t-d} \le r), \\
g(\mathbf{Z}_{t-1}) &= - [\psi_{20} + (\psi_{10} - \psi_{20}) I(y_{t-d} \le r)].
\end{aligned}
$$

By Assumption 3.5, there exists $\beta > 0$ such that $((\boldsymbol{\phi}_{10} - \boldsymbol{\phi}_{20})', (\boldsymbol{\psi}_{10} - \boldsymbol{\psi}_{20})')'\mathbf{Z}$ is bounded away from 0 for all $\mathbf{Z}$ satisfying $\|\mathbf{Z} - \mathbf{Z}^*\| \le \beta$ with $\mathbf{Z}^* = (\mathbf{w}', \mathbf{a}')'$. Thus,

$$
\begin{aligned}
\varepsilon_t^2(\boldsymbol{\lambda}_0, r) - \varepsilon_t^2 &\ge [f(\mathbf{Z}_{t-1})]^2 I(\|\mathbf{Z}_{t-1} - \mathbf{Z}^*\| \le \beta) + 2\varepsilon_t[\varepsilon_t(\boldsymbol{\lambda}_0, r) - \varepsilon_t] \\
&\quad + 2f(\mathbf{Z}_{t-1}) I(\|\mathbf{Z}_{t-1} - \mathbf{Z}^*\| \le \beta) g(\mathbf{Z}_{t-1})[\varepsilon_{t-1}(\boldsymbol{\lambda}_0, r) - \varepsilon_{t-1}].
\end{aligned}
$$

Clearly, $|f(\mathbf{Z}_{t-1})| \le C\|\mathbf{Z}_{t-1}\| I(r_0 < y_{t-d} \le r)$ and $|g(\mathbf{Z}_{t-1})| \le \rho < 1$ by the compactness of Θ. Then

$$
\begin{aligned}
&2|f(\mathbf{Z}_{t-1}) I(\|\mathbf{Z}_{t-1} - \mathbf{Z}^*\| \le \beta) g(\mathbf{Z}_{t-1})[\varepsilon_{t-1}(\boldsymbol{\lambda}_0, r) - \varepsilon_{t-1}]| \\
&\le C I(r_0 < y_{t-d} \le r) \sum_{j=0}^{\infty} \rho^j \|\mathbf{Z}_{t-2-j}\| I(r_0 < y_{t-1-d-j} \le r) = C A_t(0, u)
\end{aligned}
$$

and there exists a positive constant c_0 such that

$$
\begin{aligned}
[f(\mathbf{Z}_{t-1})]^2 I(\|\mathbf{Z}_{t-1} - \mathbf{Z}^*\| \le \beta) &\ge c_0 I(r_0 < y_{t-d} \le r, \|\mathbf{Z}_{t-1} - \mathbf{Z}^*\| \le \beta) \\
&= c_0 I^*(r_0 < y_{t-d} \le r),
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\frac{L_n^{(2)}(r)}{nG(u)} &\ge \frac{c_0 G^*(u)}{G(u)} \frac{\sum_{t=1}^{n} I^*(r_0 < y_{t-d} \le r_0 + u)}{nG^*(u)} - C\frac{\sum_{t=1}^{n} A_t(0, u)}{nG(u)} \\
&\quad + 2\frac{\sum_{t=1}^{n} \varepsilon_t[\varepsilon_t(\boldsymbol{\lambda}_0, r_0 + u) - \varepsilon_t]}{nG(u)}.
\end{aligned}
$$

Note that

$$\lim_{u\downarrow 0} \frac{G^*(u)}{G(u)} = P(\|\mathbf{Z}_{t-1} - \mathbf{Z}^*\| \leq \beta | y_{t-d} = r_0^+) > 0$$

by a conditional argument for the inequality above. Write the limit as a_0. Then there exists a $\delta > 0$ small enough such that $c_0 G^*(u)/G(u) \geq c_0 a_0/2$ on $[0, \delta]$. By Lemma B.1, it follows that

$$P\left(\inf_{B/n < u \leq \delta} \frac{L_n^{(2)}(r)}{nG(u)} > \frac{c_0 a_0}{4} \right) > 1 - 3\varepsilon.$$

Next, we consider $L_n^{(1)}(\boldsymbol{\lambda}, r)$. Clearly,

$$\frac{1}{n} L_n^{(1)}(\boldsymbol{\lambda}, r) = \frac{1}{n} \sum_{t=1}^{n} \int_0^1 \left[\frac{\partial \varepsilon_t^2(\boldsymbol{\lambda}_v, r)}{\partial \boldsymbol{\lambda}'} - \frac{\partial \varepsilon_t^2(\boldsymbol{\lambda}_v, r_0)}{\partial \boldsymbol{\lambda}'} \right] (\boldsymbol{\lambda} - \boldsymbol{\lambda}_0) dv,$$

where $\boldsymbol{\lambda}_v = \boldsymbol{\lambda}_0 + v(\boldsymbol{\lambda} - \boldsymbol{\lambda}_0)$. Similarly, we can prove that

$$\sup_{\substack{B/n < u \leq \delta \\ \boldsymbol{\theta} \in U_\delta}} \frac{|L_n^{(1)}(\boldsymbol{\lambda}, r)|}{nG(u)} = O_p\left(\sup_{\boldsymbol{\theta} \in U_\delta} \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\| \right) = O_p(\delta).$$

Clearly,

$$\inf_{\substack{B/n < u \leq \delta \\ \boldsymbol{\theta} \in U_\delta}} \frac{L_n(\boldsymbol{\lambda}, r) - L_n(\boldsymbol{\lambda}, r_0)}{nG(u)} \geq \inf_{B/n < u \leq \delta} \frac{L_n^{(2)}(r)}{nG(u)} - \sup_{\substack{B/n < u \leq \delta \\ \boldsymbol{\theta} \in U_\delta}} \frac{|L_n^{(1)}(\boldsymbol{\lambda}, r)|}{nG(u)}.$$

Let $\gamma = c_0 a_0/8 > 0$. Then, for sufficiently small $\delta > 0$, (B.8) holds. The proof of (i) is complete.

For the proof of Theorem 3.2(ii), it is similar to that of Theorem 2.2 (ii) in Li, Ling and Li (2010). Hence it is omitted. □

APPENDIX C: PROOFS OF THEOREMS 3.3 AND 4.1

PROOF OF THEOREM 3.3. Here, we give an outline of the proof, which is similar to that of Theorem 2.3 in Li, Ling and Li (2010). First of all, we prove that $\wp_n(z)$ converges weakly to a two-sided compound Poisson process $\wp(z)$ in $D(R)$. The tightness of $\{\wp_n(z)\}$ can be verified by Theorem 5 in Kushner (1984, pp.32). For $a > 0$ large enough we study the truncated process $\{\wp_n^{(a)}(z)\}$, defined similarly to $\{\wp_n(z)\}$ with the jumps being replaced by $\zeta_{1t}^{(a)} \equiv \zeta_{1t} I(|\zeta_{1t}| \leq a)$. Now we consider the weak convergence of $\{\wp_n^{(a)}(z)\}$

as $n \to \infty$ for each $a > 0$. This takes two steps. (i). Proving the *tightness* of $\{\wp_n^{(a)}(z)\}$. This can be done by Theorem 5 in Kushner (1984, pp.32) again; (ii). Characterizing the *convergence of finite dimensional distributions* of $\{\wp_n^{(a)}(z)\}$, which can be completed by verifying Assumptions A.1-A.4 in the Appendix in Li, Ling and Li (2010). Thus we have shown that $\{\wp_n^{(a)}(z)\}$ converges weakly to a two-sided compound Poisson process $\wp^{(a)}(z)$, having the jump rate $\kappa_1^{(a)}\pi(r_0)$ and the jump distribution $Q_1^{(a)}$ on the left side (i.e., $z < 0$) and the jump rate $\kappa_2^{(a)}\pi(r_0)$ and the jump distribution $Q_2^{(a)}$ on the right side (i.e., $z > 0$), where $\kappa_j^{(a)} = P(|\zeta_{jt}| \leq a|y_{t-d} = r_0)$ and $Q_j^{(a)}(A) = P(\zeta_{jt} \in A|y_{t-d} = r_0, |\zeta_{jt}| \leq a)$ for any Borel set A and $j = 1, 2$. Note that $\kappa_j^{(a)} \to 1$ and $Q_j^{(a)} \to F_j(\cdot|r_0)$ as $a \to \infty$, we have that $\wp^{(a)}(z)$ converges weakly to $\wp(z)$ in (3.5) by Theorem 16 in Pollard (1984, pp. 134). In addition, $\lim_{a\to\infty} \limsup_{n\to\infty} P(d(\wp_n^{(a)}(z), \wp_n^{(a)}(z)) > \epsilon) = 0$ for any $\epsilon > 0$, where $d(\cdot, \cdot)$ is the Skorokhod metric in $D(R)$. Hence, $\wp_n(z)$ converges weakly to $\wp(z)$ by Theorem 3.2 in Billingsley (1999, pp. 28). The remainder is the same as Theorem 2 in Chan (1993). $\qquad\square$

PROOF OF THEOREM 4.1. Here, we give an outline of the proof. By Theorem 16 in Pollard (1984, pp. 134) and Lemma A.3 (on the continuity of the smallest argmax functional) in Seijo and Sen (2010), it suffices to prove

$$(C.1) \quad P_{\mathfrak{X}}(\tilde{\zeta}_{k,d+1}^{(m)} \leq x|y_1 = \hat{r}_n, \mathbf{Z}_0 = \hat{\mathbf{z}}_i) \approx P(\zeta_{k,d+1}^{(m)} \leq x|y_1 = r_0, \mathbf{Z}_0 = \mathbf{z}_i)$$

in probability for each m. First of all, we have

$$
\begin{aligned}
P_{\mathfrak{X}}(\tilde{y}_2 \leq x|y_1 = \hat{r}_n, \mathbf{Z}_0 = \hat{\mathbf{z}}_i) &= P_{\mathfrak{X}}(\tilde{\varepsilon}_2 + h(\hat{\mathbf{v}}_i, \widehat{\boldsymbol{\theta}}_n) \leq x|y_1 = \hat{r}_n, \mathbf{Z}_0 = \hat{\mathbf{z}}_i) \\
&= \widehat{F}_\varepsilon(x - h(\hat{\mathbf{v}}_i, \widehat{\boldsymbol{\theta}}_n)) \\
&= \widehat{F}_\varepsilon(x - h(\hat{\mathbf{v}}_i, \widehat{\boldsymbol{\theta}}_n)) - F_\varepsilon(x - h(\hat{\mathbf{v}}_i, \widehat{\boldsymbol{\theta}}_n)) \\
&\quad + F_\varepsilon(x - h(\hat{\mathbf{v}}_i, \widehat{\boldsymbol{\theta}}_n)) - F_\varepsilon(x - h(\mathbf{v}_i, \boldsymbol{\theta}_0)) \\
&\quad + F_\varepsilon(x - h(\mathbf{v}_i, \boldsymbol{\theta}_0)) \\
&\approx P(\varepsilon_2 + h(\mathbf{v}_i, \boldsymbol{\theta}_0) \leq x) \\
&= P(\varepsilon_2 + h(\mathbf{v}_i, \boldsymbol{\theta}_0) \leq x|y_1 = r_0, \mathbf{Z}_0 = \mathbf{z}_i) \\
&= P(y_2 \leq x|y_1 = r_0, \mathbf{Z}_0 = \mathbf{z}_i),
\end{aligned}
$$

where $\hat{\mathbf{v}}_i = (\hat{r}_n, y_i, ..., y_{i-p\vee d+2}, \hat{r}_n - h(\hat{\mathbf{z}}_i, \widehat{\boldsymbol{\theta}}_n), \widehat{\varepsilon}_i, ..., \widehat{\varepsilon}_{i-q+2})'$ and $\mathbf{v}_i = (r_0, y_i, ..., y_{i-p\vee d+2}, r_0 - h(\mathbf{z}_i, \boldsymbol{\theta}_0), \varepsilon_i, ..., \varepsilon_{i-q+2})'$. Let $\tilde{H}_{[k]}(\cdot)$ be the conditional distribution of $(\tilde{y}_k, ..., \tilde{y}_2)'$ given $y_1 = \hat{r}_n$, $\mathbf{Z}_0 = \hat{\mathbf{z}}_i$ and $\mathfrak{X}$, and $H_{[k]}(\cdot)$ be the conditional distribution of $(y_k, ..., y_2)'$ given $y_1 = r_0$ and $\mathbf{Z}_0 = \mathbf{z}_i$. By the induction

over k ($\leq m + d + 1$), we have $\widetilde{H}_{[k]} \approx H_{[k]}$ as $n \to \infty$. On the other hand, There exists a measurable function $g(\cdot)$ such that $\tilde{\zeta}_{k,d+1}^{(m)} = g(\tilde{y}_2, ..., \tilde{y}_{m+d+1})$ and $\zeta_{k,d+1}^{(m)} = g(y_2, ..., y_{m+d+1})$. Note that $H_{[m+d+1]}(\cdot)$ is continuous and $H_{[m+d+1]}(D_g) = 0$, where $D_g = \{x : g \text{ is discontinous at } x\}$. Hence, (C.1) holds by Theorem 2.7 (continuous mapping theorem) in Billingsley (1999, pp.21). The proof is completed. $\square$

ACKNOWLEDGEMENTS

REFERENCES

[1] AN, H. Z. and HUANG, F.C. (1996). The geometrical ergodicity of nonlinear autoregressive models. *Statist. Sinica* **6** 943–956.

[2] BILLINGSLEY, P. (1999). *Convergence of Probability Measures* (2nd). John Wiley & Sons, New York.

[3] BROCKWELL, P.J., LIU, J. and TWEEDIE, R.L. (1992). On the existence of stationary threshold autoregressive moving-average processes. *J. Time Ser. Anal.* **13** 95–107.

[4] CHAN, K. S. (1989). A note on the geometric ergodicity of a Markov chain. *Adv. in Appl. Probab.* **21** 702–704.

[5] CHAN, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* **21** 520–533.

[6] CHAN, K.S., PETRUCCELLI, J.D., TONG, H. and WOOLFORD, S.W. (1985). A multiple-threshold AR(1) model. *J. Appl. Probab.* **22** 267–279.

[7] CHAN, K. S. and TONG, H. (1985). On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. in Appl. Probab.* **17** 666–678.

[8] CHAN, K. S. and TONG, H. (2010). A note on the invertibility of nonlinear ARMA models. *J. Statist. Plann. Inference* **140** 3709–3714.

[9] CHAN, K. S. and TSAY, R.S. (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika* **85** 413–426.

[10] CHEN, R. and TSAY, R.S. (1991). On the ergodicity of TAR(1) processes. *Ann. Appl. Probab.* **1** 613–634.

[11] CONT, R. and TANKOV, P. (2004). *Financial modelling with jump processes.* Chapman & Hall.

[12] FAN, J. and YAO, Q. (2003). *Nonlinear time series: Nonparametric and parametric methods.* Springer-Verlag, New York.

[13] GHADDAR, D.K. and TONG, H. (1981). Data transformation and self-exciting threshold autoregression. *J. R. Stat. Soc. Ser. C: Appl. Statist.* **30** 238–248.

[14] GIOVANNI, F. (2005). On the stability of nonlinear ARMA models. *Technical report.*

[15] GOOIJER, D. J. G. (1998). On threshold moving-average models. *J. Time Ser. Anal.* **19** 1–18.

[16] KUSHNER, H.J. (1984). *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory.* MIT.

[17] LI, D. and LING, S. (2010). Inference for multiple-regime threshold autoregressive models. *Working paper in HKUST*.

[18] LI, D., LING, S. and LI, W.K. (2010). Asymptotic theory on the least square estimation of threshold moving-average models. *Working paper in HKUST*.

[19] LI, D., LING, S. and TONG, H. (2011). On moving-average models with feedback. *To appear in Bernoulli*.

[20] LI, G.D. and LI, W. K. (2008). Testing for threshold moving average with conditional heteroscedasticity. *Statist. Sinica* **18** 647–665.

[21] LI, G.D. and LI, W. K. (2011). Testing for linear time series models against its threshold extension. *Biometrika* **98** 243–250.

[22] LING, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *J. Appl. Probab.* **36** 688–705.

[23] LING, S. and TONG, H. (2005). Testing for a linear MA model against threshold MA models. *Ann. Statist.* **33** 2529–2552.

[24] LING, S., TONG, H. and LI, D. (2007). Ergodicity and invertibility of threshold moving-average models. *Bernoulli* **13** 161–168.

[25] LIU, J. and SUSKO, E. (1992). On strict stationarity and ergodicity of a nonlinear ARMA model. *J. Appl. Probab.* **29** 363–373.

[26] MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov chains and stochastic stability*. Springer-Verlag.

[27] PETRUCCELLI, J.D. (1986). On the consistency of least squares estimators for a threshold AR(1) model. *J. Time Ser. Anal.* **7** 269–278.

[28] POLLARD, D. (1984). *Convergence of stochastic processes*. Springer-Verlag.

[29] SEIJO, E. and SEN, B. (2010). Change-Point in Regression and the Bootstrap. Available online at: http://www.stat.columbia.edu/˜ emilio (unpublished manuscript).

[30] QIAN, L. (1998). On maximum likelihood estimators for a threshold autoregression. *J. Statist. Plann. Inference* **75** 21–46.

[31] TONG, H. (1978). On a threshold model. In *Pattern recognition and signal processing*. (ed. C.H.Chen). Sijthoff and noordhoff, Amsterdam. 575–586.

[32] TONG, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford University. Press, New York.

[33] TONG, H. (2011). Threshold models in time series analysis — 30 years on. *To appear in Statistics and its Interface*.

[34] TSAY, R.S. (1989). Testing and modeling threshold autoregressive processes.*J. Amer. Statist. Assoc.* **84** 231–240.

DEPARTMENT OF MATHEMATICS
HONG KONG UNIVERSITY OF SCIENCE
  AND TECHNOLOGY
CLEAR WATER BAY
HONG KONG
E-MAIL: li20072010@gmail.com
       maling@ust.hk

DEPARTMENT OF STATISTICS
  AND ACTUARIAL SCIENCE
THE UNIVERSITY OF HONG KONG
POKFULAM ROAD, HONG KONG
E-MAIL: hrntlwk@hku.hk