Likelihood Ratio Test for Threshold MA Models *

BY SHIQING LING
Hong Kong University of Science and Technology

AND HOWELL TONG
London School of Economics and Political Science
and University of Hong Kong

Abstract

This paper investigates the (conditional) likelihood ratio test for the threshold in MA models. Under the hypothesis of no threshold, it is shown that the test statistic converges weakly to a function of the centred Gaussian process. Under local alternatives, it is shown that this test has nontrivial asymptotic power. The results are based on a new weak convergence of a linear marked empirical process, which is independently of interest. Other new results in this paper include an invertible expansion of the threshold MA (TMA) models, consistency and asymptotic normality of the estimated parameters in the TMA model with a known threshold, and the local asymptotic normality of the log-likelihood ratio of the TMA model.

Key words and phrases: Invertibility, likelihood ratio test, MA model, marked empirical process, threshold MA model, weak convergence.

AMS 1980 subject classifications: Primary 62F05, 62M10; secondary 60G10.

1 Introduction

Since Tong (1978), threshold autoregressive (AR) models have become a standard class of nonlinear time series models. Some fundamental results on the probabilistic structure of this class were given by Chan, Petruccelli, Tong and Woolford (1985), Chan and Tong (1985) and Tong (1990). The 1990s saw many more contributions

*We thank the Hong Kong Research Grants Commission (Grant #HKUST6113/02P and #HKUST4765/03H to SQL) and the University of Hong Kong (Distinguished Reseaerch Achievement Award to HT) for partial support.

including, for example, Chen and Tsay (1991), Brockwell, Liu and Tweedie (1992), Liu and Susko (1992), An and Huang (1996), An and Chen (1997), Liu, Li and Li (1997), Ling (1999) and others.

The likelihood ratio (LR) test for the threshold in AR models was investigated by Chan (1990, 1991) and Chan and Tong (1990). Tsay (1989, 1998) proposed some methods for testing the threshold in AR models and multivariate models, respectively. Lagrange multiplier tests were proposed by Wong and Li (1997, 2002) for TAR-ARCH and double TAR-ARCH models. The Wald test was studied by Hansen (2000) for TAR models. Testing threshold in nonstationary AR models was investigated by Caner and Hansen (2001). The asymptotic theory on the estimated threshold parameter in TAR models was established by Chan (1993) and Chan and Tsay (1998). Recently, Chan's result was extended to non-Gaussian errors TAR models by Qian (1998); see also Koul, Qian and Surgailis (2003) for threshold regression models. Hansen (2000) obtained a new limiting distribution for TAR models with changing threshold parameters; see also Koul (2000).

However, up to now, almost all the research in this area has been limited to the AR or AR-type models. Except for Brockwell, Liu and Tweedie (1992) and Ling (1999), it seems that threshold moving average (TMA) models have not attracted much attention in the literature. It is well known that in the linear case, moving average models are as important as the AR models. Now, the concept of threshold has been recognised as important for time series modeling. Therefore, the TMA models deserve fuller attention.

This paper investigates the LR test for threshold in MA models. Under the hypothesis of no threshold, it is shown that the test statistic converges weakly to a function of a centred Gaussian process. Under local alternatives, it is shown that this test has nontrivial asymptotic power. The results are based on a new weak convergence of a linear marked empirical process, which is independently of interest. Other new results in this paper include an invertible expansion of the TMA models, consistency and asymptotic normality of the estimated parameters in the

TMA model with a known threshold, and the local asymptotic normality (LAN) of the log-likelihood ratio of the TMA model.

This paper proceeds as follows. Section 2 gives the LR test and its null asymptotic distribution. Section 3 studies the asymptotic power under local alternatives. Sections 4-5 present the proofs of the results stated in Section 2.

2 LR Test and Its Null Limiting Distribution

The time series $\{y_t : t = 0, \pm 1, \cdots\}$ is said to follow a TMA model if it satisfied the following equation:

$$y_t = \sum_{i=1}^{p} \phi_i \varepsilon_{t-i} + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} I(y_{t-d} \le r) + \varepsilon_t, \tag{2.1}$$

where ε_t is a sequence of independent and identically distributed (i.i.d.) random variables, with mean zero and variance $0 < \sigma^2 < \infty$, and d is an integer. Assume that $p = q$. When $p \neq q$, we can add some terms with zero coefficients such that $p = q$. Denote $\phi = (\phi_1, \cdots, \phi_p)'$ and $\psi = (\psi_1, \cdots, \psi_p)'$. Here, $\lambda \equiv (\phi', \psi')'$ is the unknown parameter (vector) and its true value is $\lambda_0 = (\phi_0', \psi_0')'$. Let Θ and Θ_ψ be compact subsets of R^p and $\Theta_1 = \Theta \times \Theta_\psi$ be the parameter space. Assume that λ_0 is an interior point in Θ_1.

Given observations $y_1, \cdots, y_n$, we are interested in the following hypotheses:

$$H_0 : \psi_0 = 0 \;\; versus \;\; H_1 : \psi_0 \neq 0.$$

Under H_0, the true model (2.1) reduces to the usual linear MA model and $\{y_t\}$ is always strictly stationary and ergodic. In this case, its unknown parameter model can be written as follows.

$$y_t = \sum_{i=1}^{p} \phi_i \varepsilon_{t-i} + \varepsilon_t. \tag{2.2}$$

Note that the parameter r is absent under H_0, which renders the problem non-standard. Under H_1, Ling (1999) showed that there is always a strictly stationary solution $\{y_t\}$ to model (2.1) without any restriction on λ_0.

Under H_0 and H_1, the corresponding quasi-log-likelihood functions based on models (2.1) and (2.2) are respectively

$$L_{0n}(\phi) = \sum_{t=1}^{n} \varepsilon_t^2(\phi) \ \ and \ \ L_{1n}(\lambda, r) = \sum_{t=1}^{n} \varepsilon_t^2(\lambda),$$

where

$$\varepsilon_t(\phi) = y_t - \sum_{i=1}^{p} \phi_i \varepsilon_{t-i}(\phi) \ \ and \ \ \varepsilon_t(\lambda) = y_t - \sum_{i=1}^{p} \Big[\phi_i + \psi_i I(y_{t-d} \le r)\Big]\varepsilon_{t-i}(\lambda).$$

Here, $\varepsilon_t(\lambda)$ is the residual from the TMA model. To make it meaningful, we need to study the invertibility of this model, which is given in Section 4. Assumption 2.1 below is the condition for this.

Assumption 2.1. *$E\| \prod_{i=0}^{j}[\Phi + \Psi I(y_{t-i} \le r)]\|^2 = O(\rho^j)$ uniformly in t, where*

$$\Phi = \begin{pmatrix} \phi_1 & \cdots & \phi_p \\ I_{p-1} & O_{(p-1)\times 1} & \end{pmatrix}, \quad \Psi = \begin{pmatrix} \psi_1 & \cdots & \psi_p \\ & O_{(p-1)\times p} & \end{pmatrix},$$

I_k is the $k \times k$ identity matrix, $O_{k\times s}$ is the $k \times s$ zero matrix, and $\rho \in (0, 1)$.

This assumption is somewhat complicated. However, it is satisfied if $\sum_{i=1}^{p} \max \{|\phi_i|, |\phi_i + \psi_i|\} < 1$. In fact, let

$$\widetilde{\Phi} = \begin{pmatrix} \max\{|\phi_1|, |\phi_1 + \psi_1|\} & \cdots & \max\{|\phi_p|, |\phi_p + \psi_p|\} \\ I_{p-1} & O_{(p-1)\times 1} & \end{pmatrix}.$$

Then $E\| \prod_{i=0}^{j}[\Phi + \Psi I(y_{t-i} \le r)]\|^2 \le E\|\widetilde{\Phi}^j\|^2 = O(\rho^j)$ by Lemma 2.3 in Ling (1999). If all the roots of $z^p - \sum_{i=1}^{p} \phi_i z^{p-i} = 0$ lie inside the unit circle and $\{y_t\}$ is strictly stationary, then there exists a $\delta > 0$ such that Assumption 2.1 holds when $\sum_{i=1}^{p} |\psi_i| \le \delta$. This means that we can take Θ as large as that for the invertibility of the MA model under H_0 with a small Θ_ψ such that Assumption 2.1 holds.

Since there are only n observations, we need the initial values y_s when $s \le 0$ to calculate $\varepsilon_t(\phi)$ and $\varepsilon_t(\lambda)$. For simplicity, we assume $y_s = 0$ for $s \le 0$. Denote $\varepsilon_t(\phi)$ and $\varepsilon_t(\lambda)$ calculated with these initial values by $\widetilde{\varepsilon}_t(\phi)$ and $\widetilde{\varepsilon}_t(\lambda)$. Similarly, we modify the corresponding quasi-log-likelihood functions, respectively, to

$$\widetilde{L}_{0n}(\phi) = \sum_{t=1}^{n} \widetilde{\varepsilon}_t^2(\phi) \ \ and \ \ \widetilde{L}_{1n}(\lambda, r) = \sum_{t=1}^{n} \widetilde{\varepsilon}_t^2(\lambda).$$

Let $\hat{\phi}_n = \text{argmin}_{\Theta}\widetilde{L}_{0n}(\phi)$ and $\tilde{\lambda}_n = \text{argmin}_{\Theta_1}\widetilde{L}_{1n}(\lambda)$. We call $\hat{\phi}_n$ and $\tilde{\lambda}_n$ the conditional least squares estimators (LSE) of ϕ_0 and λ_0, respectively. The LR test for a given r is defined as

$$\widetilde{LR}_n(r) = -2[\widetilde{L}_{1n}(\tilde{\lambda}_n) - \widetilde{L}_{0n}(\hat{\phi}_n)].$$

Since the threshold parameter r is unknown, we consider the maximum of $\widetilde{LR}_n(r)$ on the finite interval $[a, b]$:

$$LR_n = \max_{r\in[a,b]} \widetilde{LR}_n(r)/\hat{\sigma}_n^2,$$

where $\hat{\sigma}_n^2 = \min_{r\in[a,b]} \widetilde{L}_{1n}(\tilde{\lambda}_n, r)/n$. Denote $U_{1t}(\lambda) = \partial\varepsilon_t(\lambda)/\partial\phi$, $U_{2t}(\lambda) = \partial\varepsilon_t(\lambda)/\partial\psi$, $U_t(\lambda) = [U_{1t}'(\lambda), U_{2t}'(\lambda)]'$, and $D_{1t}(\lambda) = U_{1t}(\lambda)\varepsilon_t(\lambda)$, $D_{2t}(\lambda) = U_{2t}(\lambda)\varepsilon_t(\lambda)$, and $D_t(\lambda) = [D_{1t}'(\lambda), D_{2t}'(\lambda)]'$. Let $P_t(\lambda) = U_t(\lambda)U_t'\lambda) + [\partial^2\varepsilon_t(\lambda)/\partial\lambda\partial\lambda']\varepsilon_t(\lambda)$. Furthermore, let $\Sigma_r = E[U_{2t}(\lambda_0)U_{2t}'(\lambda_0)]$, $\Sigma = \Sigma_\infty$, $\Sigma_{1r} = E[U_{1t}(\lambda_0)U_{2t}'(\lambda_0)]$ and $P_r = EP_t(\lambda_0)$. Here and in sequel, $o_p(1)$ denotes converging to zero in probability as $n \to \infty$. We first give the following theorem.

Theorem 2.1. *If Assumption 2.1 holds and ε_t has a positive and bounded density on R with $E|\varepsilon_t|^{2+\iota} < \infty$ for some $\iota > 0$, then under H_0, it follows that*

$$\begin{aligned}
&(a) \qquad \tilde{\lambda}_n = \lambda_0 + o_p(1),\\
&(b) \qquad \sqrt{n}(\tilde{\lambda}_n - \lambda_0) = -\frac{P_r^{-1}}{\sqrt{n}}\sum_{t=1}^{n} D_t(\lambda_0) + o_p(1),\\
&(c) \qquad \widetilde{LR}_n = T_n'(r)[\Sigma_r - \Sigma_{1r}'\Sigma^{-1}\Sigma_{1r}]^{-1}T_n(r) + o_p(1),
\end{aligned}$$

where $o_p(1)$ holds uniformly in $[a.b]$, and

$$T_n(r) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big[D_{2t}(\lambda_0) - \Sigma_{1r}'\Sigma^{-1}D_{1t}(\lambda_0)\Big].$$

Furthermore, (a) and (b) hold under H_1 if $\{y_t\}$ is strictly stationary and ergodic.

We note that (a) gives the uniform consistency of $\tilde{\lambda}_n$ when $r \in [a, b]$. Under H_1, $D_t(\lambda_0)$ is a martingale difference if $r = r_0$, the true value of the threshold parameter. Thus, (b) implies that $\tilde{\lambda}_n$ with $r = r_0$ is asymptotically normal. It is

not clear whether or not the method in Chan (1993) and Qian (1998) can be used to obtain the limiting distribution of the estimated r_0.

Under H_0, $D_{1t}(\lambda_0)=\varepsilon_t\partial\varepsilon_t(\phi_0)/\partial\phi$ and, by Lemma 4.1 in Section 4, $D_{2t}(\lambda_0)$ has the following expansion

$$D_{2t}(\lambda_0)=\Big[\sum_{i=0}^{\infty}(-\Phi)^i Z_{t-i-1}I(y_{t-d-i}\le r)\Big]\varepsilon_t \ a.s.\ and\ in\ L^2,$$

where $Z_t=(\varepsilon_t,\cdots,\varepsilon_{t-p+1})'$. $\{D_{2t}(\lambda_0): r\in R\}$ is a marked empirical process. We call $I(y_{t-d}\le r)$ a marker. This type of empirical process has been found to be very useful and were investigated by Chan (1993), Stute (1997), Koul and Stute (1999), Hansen (2000) and Ling (2003) for various purposes. However, all the processes in these papers have only one marker. The process $\{D_{2t}(\lambda_0): r\in R\}$ is a linear marked empirical process and includes infinitely many markers. To the best of our knowledge, it has never appeared in the statistical literature before.

Let $D^p[R_\gamma]=D[R_\gamma]\times\cdots\times D[R_\gamma]$ (p factors) which is equipped with the corresponding product Skorohod topology, where $R_\gamma=[-\gamma,\gamma]$. Weak convergence on $D^p[R]$, denoted by $\Longrightarrow$, is defined as weak convergence on $D^p[R_\gamma]$ for each $\gamma\in(0,\infty)$. We now state another assumption and give the weak convergence of $\{T_n(r): r\in R\}$.

Assumption 2.2. *ε_t has a continuous and positive density on R and $E\varepsilon_t^4<\infty$.*

Theorem 2.2. *If Assumption 2.2 holds and all the roots of $z^p-\sum_{i=1}^p\phi_i z^{p-i}=0$ lie inside the unit circle, then under H_0, it follows that*

$$T_n(r)\Longrightarrow\sigma G_p(r)\ \ in\ \ D^p[R],$$

where $\{G_p(r): r\in R\}$ is a $p\times 1$ vector Gaussian process with mean zero and covariance kernel $K_{rs}=\Sigma_{r\wedge s}-\Sigma_{1r}'\Sigma^{-1}\Sigma_{1s}$, and almost all the paths of $G_p(r)$ are continuous in r.

We should mention that the covariance kernel in Theorem 2.2 is essentially different from those of the empirical processes with one marker. Thus, it is a new weak convergence result. By (4.12) in Section 4, we can see that $\hat\sigma_n^2=\sigma^2+o_p(1)$. Furthermore, by Theorem 2.1(c), Theorem 2.2 and the continuous mapping theorem, we obtain the main results as follows.

Theorem 2.3. *If Assumptions 2.1-2.2 hold, then under* H_0,

$$LR_n \longrightarrow_{\mathcal{L}} \max_{r\in[a,b]} \left[G_p'(r)K_{rr}^{-1}G_p(r)\right],$$

where $\rightarrow_{\mathcal{L}}$ *stands for convergence in distribution as* $n \to \infty$.

When $d > p$, we have $\Sigma_r = \Sigma F_y(r)$ since Z_{t-1} and y_{t-d} are independent, where $F_y(r) = P(y_t \le r)$. Thus, the limiting distribution is the same as that of

$$\max_{\beta_1\le s\le\beta_2} \|B_p(s)\|^2/(s-s^2),$$

where $\beta_1 = F_y(a)$ and $\beta_2 = F_y(b)$, and $B_p(s)$ is a $p\times 1$ vector Gaussian process with mean zero and covariance kernel $r\wedge s - rs$. It is interesting that this distribution is the same as that of test statistics for change-points in Andrews (1993). Its critical values can be found in this paper. For other cases, the critical values of LR_n can be obtained via a simulation method which is simple in practice. Finally, we note that the test for TAR models in Chan and Tong (1990) and Chan (1991) cannot be used even when $p = 1$ since K_{rs} involves infinitely many markers.

3 Asymptotic Power Under Local Alternatives

This section investigates the asymptotically local power of LR_n by considering the local alternative hypothesis,

$$H_{1n} : \psi_0 = \frac{\gamma}{\sqrt{n}} \text{ for a constant vector } \gamma \in R.$$

For this, we need some basic concepts as follows. Let $P_{\lambda,f}$ be a probability measure on $(\mathcal{R}^Z, \mathcal{F}^Z)$, where $\mathcal{F}^Z$ is the Borel $\sigma-$field on $\mathcal{R}^Z$ with $Z = \{0, \pm1, \pm2, \cdots\}$, and let P_λ^n be the restriction of $P_{\lambda,f}$ on $\mathcal{F}_n$, the $\sigma-$field generated by $\{Y_0, y_1, \cdots, y_n\}$, where $Y_0 = \{y_0, y_{-1}, \cdots\}$. Suppose the errors $\{\varepsilon_1(\lambda), \varepsilon_2(\lambda), \cdots\}$ under P_λ^n are i.i.d. with density f and are independent of Y_0. From model (2.1), the distribution of initial value Y_0 is the same under both P_λ^n and $P_{\lambda_0}^n$. Thus, the log-likelihood ratio $\Lambda_n(\lambda_1,\lambda_2)$ of $P_{\lambda_2}^n$ to $P_{\lambda_1}^n$ is

$$\Lambda_n(\lambda_1,\lambda_2) = 2\sum_{t=1}^n \left[\log s_t(\lambda_2) - \log s_t(\lambda_1)\right],$$

where $s_t(\lambda) = \sqrt{f(\eta_t(\lambda))}$; see Koul and Schick (1997) and Ling and McAleer (2002) for details. We first introduce the following assumption.

Assumption 3.1. *The density f of ε_t is absolutely continuous with a.e.-derivative and the finite Fisher information*

$$0 < I(f) = \int_{-\infty}^{\infty} \Big(\frac{f'(x)}{f(x)}\Big)^2 f(x)dx < \infty.$$

The following theorem gives the LAN of $\Lambda_n(\lambda_1, \lambda_2)$ for the TMA model and the contiguity of $P^n_{\lambda_0}$ and $P^n_{\lambda_0+u_n/\sqrt{n}}$, where u_n is a bounded constant sequence in R^{2p}.

Theorem 3.1. *If Assumptions 2.1-2.2 and 3.1 hold, then under H_0,*

$$(a) \qquad \text{under } P^n_{\lambda_0},\ \Lambda_n(\lambda_0, \lambda_0 + \frac{u_n}{\sqrt{n}}) = \frac{u_n'}{\sqrt{n}} \sum_{t=1}^n U_t(\lambda_0)\xi_t - \frac{I(f)}{2} u_n' P_r u_n + o_p(1);$$

$$(b) \qquad P^n_{\lambda_0} \text{ and } P^n_{\lambda_0+u_n/\sqrt{n}} \text{ are contiguous,}$$

where $\xi_t = f'(\varepsilon_t(\lambda_0))/f(\varepsilon_t(\lambda_0))$ and P_r is defined as in Theorem 2.1. Furthermore, (a) and (b) hold under H_1 with $r = r_0$ if $\{y_t\}$ is strictly stationary and ergodic.

Proof. By Lemmas 4.2(b)-(c) and 4.3 in Section 4, P_r is finite and positive definite. Under H_0, $U_t(\lambda_0)\xi_t$ is a martingale difference in terms of $\mathcal{F}_t$. By the central limit theorem, we can show that $\sum_{t=1}^n U_t(\lambda_0)\xi_t/\sqrt{n} \to_{\mathcal{L}} N(0, I(f)P_r)$. Thus, by Theorem 2.1 and Remark 2.1 in Ling and McAleer (2003), for (a) and (b) it suffices to verify the following conditions under $P^n_{\lambda_0}$:

$$(3.1) \qquad \sum_{t=1}^n \Big[\varepsilon_t(\lambda_0 + \frac{u_n}{\sqrt{n}}) - \varepsilon_t(\lambda_0) - \frac{u_n'}{\sqrt{n}} U_t(\lambda_0)\Big]^2 = o_p(1),$$

$$(3.2) \qquad \frac{1}{n} \sup_{1\le t\le n} \|U_t(\lambda_0)\|^2 = o_p(1),$$

$$(3.3) \qquad \frac{1}{n} \sum_{t=1}^n \|U_t(\lambda_0)\|^2 = O_p(1),$$

$$(3.4) \qquad \frac{1}{n} \sum_{t=1}^n \Big\|[U_t(\lambda_0 + \frac{u_n}{\sqrt{n}}) - U_t(\lambda_0)]\Big\|^2 = o_p(1).$$

By Taylor's expansion, the left-hand-side of (3.1) is bounded by

$$A_t \equiv \frac{O(1)}{n^2} \sum_{t=1}^n \sup_{\lambda\in\Theta} \Big\|\frac{\partial^2 \varepsilon_t(\lambda)}{\partial\lambda\partial\lambda'}\Big\|^2.$$

By Lemma 4.2 (c) and the ergodic theorem, we readily see that $A_t = o_p(1)$. Similarly, we can show that (3.4) holds. Since $\{U_t(\lambda_0)\}$ is a strictly stationary sequence with $E\|U_t(\lambda_0)\|^2 < \infty$, it follows that $\sup_{1\le t\le n} \|U_t(\lambda_0)\|^2 /n \le [\sup_{1\le t\le n} \|U_t(\lambda_0)\| /\sqrt{n}]^2 = o_p(1)$, that is, (3.2) holds. Also (3.3) holds by the ergodic theorem. Under H_1 with $r = r_0$, $U_t(\lambda_0)\xi_t$ is also a martingale difference in terms of $\mathcal{F}_t$ and (3.1)-(3.4) holds. So, (a)-(b) holds under H_1. This completes the proof. □

The following theorem shows that LR_n has non-trivial local power under H_{1n}.

Theorem 3.2. *If Assumptions 2.1-2.2 and 3.1 hold, then, under H_{1n},*

$$\begin{aligned} &(a) \qquad T_n(r) \Longrightarrow \mu(r) + \sigma G_p(r) \text{ in } D^p[R], \\ &(b) \qquad LR_n \longrightarrow_{\mathcal{L}} \max_{r\in[a,b]} \left\{[\sigma^{-1}\mu(r) + G_p(r)]' K_{rr}^{-1} [\sigma^{-1}\mu(r) + G_p(r)]\right\}, \end{aligned}$$

where $\mu(r) = K_{rr}\gamma$ and $G_p(r)$ is a Gaussian process defined as in Theorem 2.1.

Proof. Under $P^n_{\lambda_0}$, by Theorem 3.1 (a) with $u_n = u = (0, \cdots, 0, \gamma')'$, we have

$$\Lambda_n(\lambda_0, \lambda_0 + \frac{u}{\sqrt{n}}) = \frac{1}{\sqrt{n}}\gamma' \sum_{t=1}^{n} U_{2t}(\lambda_0)\xi_t - \frac{I(f)}{2}\gamma'\Sigma_r\gamma + o_p(1).$$

Under H_0 and H_{1n}, the probability measures of $\{Y_0, y_1, \cdots, y_n\}$ are $P^n_{\lambda_0}$ and $P^n_{\lambda_0+u/\sqrt{n}}$, respectively. By Theorem 2.1, $\{T_n(r): r \in R_\gamma\}$ is tight under $P^n_{\lambda_0}$. By the continuity of $P^n_{\lambda_0}$ and $P^n_{\lambda_0+u/\sqrt{n}}$, we know that $\{T_n(r): r \in R_\gamma\}$ is tight under $P^n_{\lambda_0+u/\sqrt{n}}$. By the central limit theorem, we can show that the finite dimensional distributions of $[T_n'(r), \Lambda_n(\lambda_0, \lambda_0 + u/\sqrt{n})]'$ converge weakly to those of $G_J(r)$ in $D^p[R_\gamma] \times R$, where $G_J(r)$ is a Gaussian process, the mean and the covariance kernel of which are respectively

$$\begin{bmatrix} 0 \\ -\frac{1}{2}\gamma'\Sigma_r I(f)\gamma \end{bmatrix} \text{ and } \begin{bmatrix} \sigma^2 K_{rs} & -K_{rs}\gamma \\ -\gamma' K_{rs}' & \gamma'\Sigma_{r\wedge s} I(f)\gamma \end{bmatrix}.$$

By LeCam's third lemma in Van der Vaart and Wellner (1996), we can show that the finite dimensional distributions of $\{T_n(r): r \in R_\gamma\}$ converge weakly to those of $\{\mu(r) + \sigma G_p(r): r \in R_\gamma\}$ under $P^n_{\lambda_0+u/\sqrt{n}}$. Thus, (a) holds. Note that $\hat{\sigma}_n^2 = \sigma^2 + o_p(1)$ holds under H_{1n} since it holds under H_0. Thus, (b) holds by (a) and the continuous mapping theorem. This completes the proof. □

4 Proof of Theorem 2.1

We first give one lemma, which is on the invertibility of TMA models. As far as we know, the expansions in this lemma has never appeared in the literature before.

Lemma 4.1. *Let* $Z_t(\lambda) = [\varepsilon_t(\lambda), \cdots, \varepsilon_{t-p+1}(\lambda)]'$, $Z_{1t}(\lambda) = Z'_t(\lambda)I(y_{t-d+1} \le r)$ *and* $u = (1, 0, \cdots, 0)'$. *If Assumption 2.1 holds, then model (2.1) is invertible and the following expansions hold a.s. and in* L^2*:*

$$(a) \qquad \varepsilon_t(\lambda) = y_t + \sum_{j=0}^{\infty} u' \prod_{i=0}^{j} \Big[-\Phi - \Psi I(y_{t-d-i} \le r)\Big] u\, y_{t-j-1},$$

$$(b) \qquad \frac{\partial \varepsilon_t(\lambda)}{\partial \phi} = -Z_{t-1}(\lambda) + \sum_{j=0}^{\infty} \prod_{i=0}^{j} \Big[-\Phi - \Psi I(y_{t-d-i} \le r)\Big] Z_{t-j-2}(\lambda),$$

$$(c) \qquad \frac{\partial \varepsilon_t(\lambda)}{\partial \psi} = -Z_{1t-1}(\lambda) + \sum_{j=0}^{\infty} \prod_{i=0}^{j} \Big[-\Phi - \Psi I(y_{t-d-i} \le r)\Big] Z_{1t-j-2}(\lambda).$$

Proof. Let $A_t = \Phi + \Psi I(y_{t-d} \le r)$ and $Y_t = u y_t$. By the definition of $\varepsilon_t(\lambda)$, we can rewrite it in the vector form:

$$Z_t(\lambda) = Y_t - A_t Z_{t-1}(\lambda). \tag{4.1}$$

We iterate model (4.1) J steps:

$$Z_t(\lambda) = Y_t + \sum_{j=0}^{J-1} \prod_{i=0}^{j} (-A_{t-i}) Y_{t-j-1} + \prod_{i=0}^{J} (-A_{t-i}) Z_{t-J-1}(\lambda). \tag{4.2}$$

Let $S_J = Y_t + \sum_{j=0}^{J-1} \prod_{i=0}^{j} (-A_{t-i}) Y_{t-j-1}$. Since we always have that $\max_{-\infty<t<\infty} E y_t^2 < \infty$ for TMA models, by Assumption 2.1, it is not hard to see that, for any $J_1 < J_2$,

$$\begin{aligned} \|S_{J_1} - S_{J_2}\| &= \| \sum_{j=J_1+1}^{J_2} \prod_{i=0}^{j} (-A_{t-i}) Y_{t-j-1} \| \\ &\le O(1)[\sum_{j=J_1+1}^{J_2} (E\| \prod_{i=0}^{j} (-A_{t-i}) \|^2)^{1/2}]^2 = O(\rho^{(J_2-J_1)/2}). \end{aligned} \tag{4.3}$$

By (4.3), we know that $S_J \to S_\infty$ a.s. and in L^2. It is readily seen that S_∞ is a solution of (4.1). When $Z_t(\lambda) = S_\infty$, we can show that the second term in (4.2) converges to zero a.s. and in L^2. To see the uniqueness, suppose that there is another solution $Z^*_t(\lambda)$ a.s. and in L^2 for model (4.1). Let $V_t(\lambda) = Z_t(\lambda) - Z^*_t(\lambda)$.

$$V_t(\lambda) = A_t V_{t-1}(\lambda) = \cdots = \prod_{i=0}^{J} A_{t-i} V_{t-J-1}(\lambda).$$

Since $\max_{-\infty<t<\infty} E\|V_t(\lambda)\|^2 < \infty$, by the preceding equation and (4.2), we can see that $E\|V_t(\lambda)\|^2 = 0$ and hence $Z_t(\lambda) = Z_t^*(\lambda)$ a.s. and in L^2. Thus, (a) holds.

$$\frac{\partial \varepsilon_t(\lambda)}{\partial \phi} = -Z_{t-1}(\lambda) - A_t \frac{\partial \varepsilon_{t-1}(\lambda)}{\partial \phi},$$
$$\frac{\partial \varepsilon_t(\lambda)}{\partial \psi} = -Z_{t-1}(\lambda) I(y_{t-d} \le r) - A_t \frac{\partial \varepsilon_{t-1}(\lambda)}{\partial \psi}.$$

Similar to (a), we can show that (b) and (c) hold. This completes the proof. □

Lemma 4.2. *If Assumption 2.1 holds with $E|\varepsilon_t|^{2+\iota} < \infty$ for some $\iota > 0$, then under H_0 or under H_1 with $\{y_t\}$ being strictly stationary and ergodic, it follows that*

$$(a) \qquad E \sup_{\Theta_1} \sup_{r\in[a,b]} |\varepsilon_t(\lambda)|^{2+\iota} < \infty,$$
$$(b) \qquad E \sup_{\Theta_1} \sup_{r\in[a,b]} \Big\| \frac{\partial \varepsilon_t(\lambda)}{\partial \lambda} \Big\|^{2+\iota} < \infty,$$
$$(c) \qquad E \sup_{\Theta_1} \sup_{r\in[a,b]} \Big\| \frac{\partial^2 \varepsilon_t(\lambda)}{\partial \lambda \partial \lambda'} \varepsilon_t(\lambda) \Big\|^{1+\iota} < \infty.$$

Proof. From the proof of Lemma 4.1, we know that

$$\sup_{\Theta_1} |\varepsilon_t(\lambda)| \ \le \ O(1) \sum_{i=0}^{\infty} \rho^i |y_{t-i}|, \tag{4.4}$$

where $\rho \in (0,1)$. Since $E|\varepsilon_t|^{2+\iota} < \infty$, it is readily shown that $E|y_t|^{2+\iota} < \infty$. By Minkowskii's inequality, we can show that $E \sup_{\Theta_1} |\varepsilon_t(\lambda)|^{2+\iota} < \infty$. Thus, (a) holds. Similarly, we can show that (b) holds. As for (a), we can show that $E \sup_{\Theta_1} |\partial^2 \varepsilon_t(\lambda)/\partial\lambda\partial\lambda'|^{2+\iota} < \infty$. Furthermore, by (a) and Cauchy-Schwarz inequality, we can show that (c) holds. This completes the proof. □

Lemma 4.3. *If the assumption of Lemma 4.2 holds and ε_t has a positive density on R, then for each $\lambda \in \Theta_1$, $E\{[\partial\varepsilon_t(\lambda)/\partial\lambda][\partial\varepsilon_t(\lambda)/\partial\lambda']\} > 0$.*

Proof. Suppose that there are two $p \times 1$ constant vectors c_1 and c_2 such that

$$E\Big[c' \frac{\partial \varepsilon_t(\lambda)}{\partial \lambda} \frac{\partial \varepsilon_t(\lambda)}{\partial \lambda'} c\Big] = 0,$$

where $c = (c_1', c_2')'$. Then, $c_1'\partial\varepsilon_t(\lambda)/\partial\phi + c_2'\partial\varepsilon_t(\lambda)/\partial\psi = 0$ a.s.. This results in

$$\sum_{i=1}^{p} [c_{1i} + c_{2i} I(y_{t-d} \le r)] \varepsilon_{t-i}(\lambda)$$
$$= -\sum_{j=1}^{p} \Big\{[\phi_j + \psi_j I(y_{t-d} \le r)] \Big[c_1' \frac{\partial \varepsilon_{t-j}(\lambda)}{\partial \phi} + c_2' \frac{\partial \varepsilon_{t-j}(\lambda)}{\partial \psi}\Big]\Big\} = 0 \ \ a.s.,$$

where c_{ij} is the jth element of c_i for $i = 1, 2$. Thus,

$$\Big[\sum_{i=1}^{p} c_{1i}\varepsilon_{t-i}(\lambda)\Big] I(y_{t-d} > r) + \Big[\sum_{i=1}^{p}(c_{1i} + c_{2i})\varepsilon_{t-i}(\lambda)\Big] I(y_{t-d} \le r) = 0 \ \ a.s..$$

From this equation, we have that

$$\Big[\sum_{i=1}^{p} c_{1i}\varepsilon_{t-i}(\lambda)\Big] I(y_{t-d} > r) = 0 \ \ a.s., \tag{4.5}$$

$$\Big[\sum_{i=1}^{p}(c_{1i} + c_{2i})\varepsilon_{t-i}(\lambda)\Big] I(y_{t-d} \le r) = 0 \ \ a.s.. \tag{4.6}$$

If $c_{11} \neq 0$, for simplicity, let $c_{11} = 1$. Denote the event $A = \{\sum_{i=1}^{p} c_{1i}\varepsilon_{t-i}(\lambda) = 0\}$. Then $A = \{\varepsilon_{t-1}(\lambda) = \sum_{i=2}^{p} c_{1i}\varepsilon_{t-i}(\lambda)\}$. Let $g_{1t-1}(\lambda) = \sum_{i=1}^{p}[\phi_i + \psi_i I(y_{t-d-1} \le r)]\varepsilon_{t-i-1}(\lambda)$ and $g_{t-2} = g_{1t-1}(\lambda) - g_{1t-1}(\lambda_0) + \sum_{i=2}^{p} c_{1i}\varepsilon_{t-i}(\lambda)$. Then $A = \{\varepsilon_{t-1} = g_{t-2}\}$. Since ε_{t-1} and g_{t-2} are independent, we have that

$$\begin{aligned} P(A) &= E[I\{\varepsilon_{t-1} \in A\}] = E[I\{\varepsilon_{t-1} \in A\} I\{g_{t-2} \in A\}] \\ &= E[I\{\varepsilon_{t-1} \in A\}]E[I\{g_{t-2} \in A\}] = \{E[I\{\varepsilon_{t-1} \in A\}]\}^2 = P^2(A). \end{aligned}$$

Thus, $P(A) = 0$ or 1. If $P(A) = 1$, then $E\varepsilon_{t-1}^2 = E[\varepsilon_{t-1}^2 I\{\varepsilon_{t-1} \in A\}] = E[\varepsilon_{t-1} g_{t-2} I\{g_{t-2} \in A\}] = E\varepsilon_{t-1} E g_{t-2} = 0$. This is a contradiction. So, we can claim $P(A) = 0$. Using this, it follows that

$$\begin{aligned} &P(\{\omega : \Big[\sum_{i=1}^{p} c_{1i}\varepsilon_{t-i}(\lambda)\Big] I(y_{t-d} > r) = 0\Big\}) \\ &= P(\{\omega : \Big[\sum_{i=1}^{p} c_{1i}\varepsilon_{t-i}(\lambda)\Big] I(y_{t-d} > r) = 0\Big\} \cap A^c) \\ &= P(\{\omega : I(y_{t-d} > r) = 0\Big\} \cap A^c) \\ &= P(\{\omega : I(y_{t-d} > r) = 0\Big\}) \\ &= P(\Big\{\omega : \varepsilon_{t-d} \le r - g_{1t-d-1}(\lambda_0)\Big\}) = E\Big\{\int_{-\infty}^{r - g_{1t-d-1}(\lambda_0)} f(x)dx\Big\} > 0, \end{aligned}$$

since f is positive, where f is the density of ε_t. This contradicts (4.5). So, $c_{11} = 0$. Similarly, we can show that $c_{12} = \cdots = c_{1p} = 0$. Similarly, we can show that $c_{21} = \cdots = c_{2p}$ using (4.6). This completes the proof. □

Lemma 4.4. *If the assumption of Lemma 4.3 holds, then, for any $\eta > 0$,*

$$\inf_{\|\lambda - \lambda_0\| \ge \eta} \inf_{r \in [a,b]} E[\varepsilon_t^2(\lambda) - \varepsilon_t^2(\lambda_0)] > 0.$$

Proof. Let $V_{t-1}(\lambda) = \varepsilon_t(\lambda) - \varepsilon_t(\lambda_0)$. Then

$$\begin{aligned} V_{t-1}(\lambda) &= \sum_{i=1}^{p}[(\phi_i - \phi_{i0}) + (\psi_i - \psi_{i0})I(y_{t-d} \le r)]\varepsilon_{t-i}(\lambda) \\ &+ \sum_{i=1}^{p}\Big[(\phi_{i0} + \psi_{i0})I(y_{t-d} \le r)\Big]V_{t-i}(\lambda). \end{aligned} \tag{4.7}$$

Since $\varepsilon_t(\lambda_0) = \varepsilon_t$ is independent of $V_{t-1}(\lambda)$ and $\varepsilon_t(\lambda) = \varepsilon_t(\lambda_0) + V_{t-1}(\lambda)$, we have

$$E\varepsilon_t^2(\lambda) = E\varepsilon_t^2(\lambda_0) + EV_{t-1}^2(\lambda).$$

$EV_{t-1}^2(\lambda) = 0$ if and only if $V_{t-1}(\lambda) = 0$ a.s.. By (4.7), this occurs if and only if

$$\sum_{i=1}^{p}\Big[(\phi_i - \phi_{i0}) + (\psi_i - \psi_{i0})I(y_{t-d} \le r)\Big]\varepsilon_{t-i}(\lambda) = 0 \ a.s..$$

From the proof of Lemma 4.3, we know that the preceding equation holds if and only if $\lambda = \lambda_0$ for each $r \in [a, b]$. Let $c = \inf_{\{\|\lambda-\lambda_0\| \ge \eta\}\times[a,b]} EV_{t-1}^2(\lambda)$. Note that $EV_{t-1}^2(\lambda)$ is a continuous function of λ and r, and $\Theta_1 \times [a, b]$ is compact. If $c = 0$, then there exist a λ and an r in $\Theta_1 \times [a, b]$ such that $EV_{t-1}^2(\lambda) = 0$. But this is impossible since it has only one zero point at $\lambda = \lambda_0$. Thus, we can claim that $c > 0$. This completes the proof. □

Lemma 4.5. *If the assumption of Lemma 4.3 holds and the density of ε_t is bounded, then it follows that*

$$\lim_{n\to\infty} P(\frac{1}{n}\sup_{\Theta_1}\sup_{r\in[a,b]} |\sum_{t=1}^{n}[\varepsilon_t^2(\lambda) - E\varepsilon_t^2(\lambda)]| > \varepsilon) = 0.$$

Proof. To emphasize that $\varepsilon_t(\lambda)$ depends on r, we denote $\varepsilon_t(\lambda)$ by $\varepsilon_t(\lambda, r)$. Since Θ is bounded and compact, we can choose a collection of balls of radius $\delta > 0$ covering Θ and the number of such balls is a finite integer K_1. We take a point λ_i in the ith ball and denote this ball by V_{λ_i}. Similarly, we divide $[a, b]$ into K_2 parts such that $a = r_1 \le r_2 < \cdots < r_{K_2+1} = b$ with $|r_i - r_{i-1}| \le \delta$. For any $\varepsilon > 0$,

$$\begin{aligned} &P(\frac{1}{n}\sup_{\Theta_1}\sup_{r\in[a,b]} |\sum_{t=1}^{n}[\varepsilon_t^2(\lambda, r) - E\varepsilon_t^2(\lambda, r)]| > \varepsilon) \\ \le &\sum_{i=1}^{K_1}\sum_{j=1}^{K_2} P(\frac{1}{n}|\sum_{t=1}^{n}[\varepsilon_t^2(\lambda_i, r_j) - E\varepsilon_t^2(\lambda_i, r_j)]| > \frac{\varepsilon}{2K_1K_2}) \end{aligned}$$

$$+P(\sup_{1\le i\le K_1}\sup_{1\le j\le K_2}\sup_{\lambda\in V_{\lambda_i}}\sup_{r_j\le r\le r_{j+1}}|E[\varepsilon_t^2(\lambda_i,r_j)-\varepsilon_t^2(\lambda,r)]|>\frac{\varepsilon}{4})$$
$$+P(\frac{1}{n}\sup_{1\le i\le K_1}\sup_{1\le j\le K_2}\sup_{\lambda\in V_{\lambda_i}}\sup_{r_j\le r\le r_{j+1}}|\sum_{t=1}^{n}[\varepsilon_t^2(\lambda,r)-\varepsilon_t^2(\lambda_i,r_j)]|>\frac{\varepsilon}{4})$$
$$\equiv B_{1n}+B_{2n}+B_{3n}.$$

We first consider B_{3n}.

$$\begin{aligned}
B_{3n} &\le P(\frac{1}{n}\sup_{1\le i\le K_1}\sup_{\lambda\in V_{\lambda_i}}\sup_{r\in[a,b]}|\sum_{t=1}^{n}[\varepsilon_t^2(\lambda,r)-\varepsilon_t^2(\lambda_i,r)]|>\frac{\varepsilon}{8})\\
&\quad +P(\frac{1}{n}\sup_{\lambda\in\Theta_1}\sup_{1\le j\le K_2}\sup_{r_j\le r\le r_{j+1}}|\sum_{t=1}^{n}[\varepsilon_t^2(\lambda,r)-\varepsilon_t^2(\lambda,r_j)]|>\frac{\varepsilon}{8})\\
&\le O(1)E\sup_{1\le i\le K_1}\sup_{\lambda\in V_{\lambda_i}}\sup_{r\in[a,b]}|\varepsilon_t^2(\lambda,r)-\varepsilon_t^2(\lambda_i,r)|\\
&\qquad +E\sup_{\lambda\in\Theta_1}\sup_{1\le j\le K_2}\sup_{r_j\le r\le r_{j+1}}|\varepsilon_t^2(\lambda,r)-\varepsilon_t^2(\lambda,r_j)|.
\end{aligned}$$

By Taylor's expansion and Lemma 4.2(b), we have

$$E\sup_{1\le i\le K_1}\sup_{\lambda\in V_{\lambda_i}}\sup_{r\in[a,b]}|\varepsilon_t(\lambda,r)-\varepsilon_t(\lambda_i,r)|^2\le\delta^2E\sup_{\Theta_1}\sup_{r\in[a,b]}\|\frac{\partial\varepsilon_t(\lambda,r)}{\partial\lambda}\|^2=O(\delta^2).$$

Furthermore, by Lemma 4.2 (a) and Cauchy-Schwarz inequality, we can show that

$$\begin{aligned}
&E\sup_{1\le i\le K_1}\sup_{\lambda\in V_{\lambda_i}}\sup_{r\in[a,b]}|\varepsilon_t^2(\lambda,r)-\varepsilon_t^2(\lambda_i,r)|\\
&\le\{E\sup_{\lambda\in\Theta}\sup_{r\in[a,b]}\varepsilon_t^2(\lambda,r)E\sup_{1\le i\le K_1}\sup_{\lambda\in V_{\lambda_i}}\sup_{r\in[a,b]}|\varepsilon_t(\lambda,r)-\varepsilon_t(\lambda_i,r)|^2\}^{1/2}=O(\delta). \qquad (4.8)
\end{aligned}$$

For any $r'<r$, let $X_{t-1}=-\sum_{i=1}^{p}\psi_iI(r'\le y_{t-d}\le r)\varepsilon_{t-i}(\lambda,r')$. As for Lemma 4.1, we can show that

$$\begin{aligned}
\varepsilon_t(\lambda,r)-\varepsilon_t(\lambda,r') &= -\sum_{i=1}^{p}[\phi_i+\psi_iI(y_{t-d}\le r)][\varepsilon_{t-i}(\lambda,r)-\varepsilon_{t-i}(\lambda,r')]+X_{t-1}\\
&= X_{t-1}+u'\sum_{j=0}^{\infty}\prod_{i=0}^{j}(-A_{t-i})u\,X_{t-j-1}\ \textit{a.s. and in } L^2.
\end{aligned}$$

Since the density of ε_t is bounded, we can show that $EI(r'\le y_{t-d}\le r'+\delta)=O(\delta)$ uniformly in t. Furthermore, by Lemma 4.2(a) and Hölder's inequality, we have

$$\begin{aligned}
&E\sup_{\lambda\in\Theta_1}\sup_{r'\in[a,b]}\sup_{r'\le r\le r'+\delta}|X_{t-1}|^2\\
&\le O(1)E[I(r'\le y_{t-d}\le r'+\delta)(\sum_{i=1}^{p}\sup_{\lambda\in\Theta_1}\sup_{r\in[a,b]}|\varepsilon_{t-i}(\lambda,r)|)^2]\\
&\le O(1)\{E[I(r'\le y_{t-d}\le r'+\delta)]\}^{\frac{\iota_1}{2+\iota_1}}[E\sup_{\lambda\in\Theta_1}\sup_{r\in[a,b]}|\varepsilon_t(\lambda,r)|^{2+\iota_1}]^{\frac{2}{2+\iota_1}}=O(\delta^{\iota_1/4}),
\end{aligned}$$

for some $\iota_1 \in (0,2)$, which holds uniformly in t. By the preceding two equations and Minkowskii's inequality, it follows that

$$E \sup_{\lambda\in\Theta_1} \sup_{1\le j\le K_2} \sup_{r_j\le r\le r_{j+1}} |\varepsilon_t(\lambda,r)-\varepsilon_t(\lambda,r_j)|^2 \le O(1)(\delta^{\iota_1/8}+\sum_{i=0}^{\infty}\rho^i\delta^{\iota_1/8})^2 = O(\delta^{\iota_1/4}).$$

By this equation, Lemma 4.2(a) and Cauchy-Schwarz inequality,

$$E \sup_{\lambda\in\Theta_1} \sup_{1\le j\le K_2} \sup_{r_j\le r\le r_{j+1}} |\varepsilon_t^2(\lambda,r)-\varepsilon_t^2(\lambda,r_j)| = O(\delta^{\iota_1/8}). \tag{4.9}$$

By (4.8) and (4.9), we can take δ small enough such that $B_{3n} < \varepsilon/3$.

Again, by (4.8) and (4.9), we know that $B_{2n} < \varepsilon/3$. For this δ, K_1 and K_2 are fixed. By the ergodic theorem, $B_{1n} < \varepsilon/3$ for n large enough. Thus, we can claim that the conclusion holds. This completes the proof. $\Box$

Lemma 4.6. *If the assumption of Lemma 4.5 holds, then for any $\varepsilon > 0$, there is an $\eta > 0$ such that*

$$P(\frac{1}{n}\sup_{\|\lambda-\lambda_0\|\le\eta}\sup_{r\in[a,b]}\|\sum_{t=1}^{n}[P_t(\lambda)-P_r]\| > \varepsilon) < \varepsilon.$$

Proof. Denote $P_t(\lambda)$ by $P_t(\lambda,r)$. Since $P_t(\lambda,r)$ is continuous in λ, by Lemma 4.2(c) and the dominated convergence theorem, $E\sup_{r\in[a,b]}\sup_{\|\lambda-\lambda_0\|\le\eta}\|P_t(\lambda,r)-P_t(\lambda_0,r)\| \to 0$ as $\eta\to 0$. Thus, there exists an $\eta > 0$ such that

$$\begin{aligned} &P(\frac{1}{n}\sup_{\|\lambda-\lambda_0\|\le\eta}\sup_{r\in[a,b]}\|\sum_{t=1}^{n}[P_t(\lambda,r)-P_t(\lambda_0,r)]\| > \frac{\varepsilon}{2}) \\ &\le \frac{2}{\varepsilon}E\sup_{\|\lambda-\lambda_0\|\le\eta}\sup_{r\in[a,b]}\|P_t(\lambda,r)-P_t(\lambda_0,r)\| < \frac{\varepsilon}{2}. \end{aligned}$$

By the preceding equation, it is sufficient for the conclusion to show that

$$P(\frac{1}{n}\sup_{r\in[a,b]}\|\sum_{t=1}^{n}[P_t(\lambda_0,r)-P_r]\| > \frac{\varepsilon}{2}) < \frac{\varepsilon}{2}. \tag{4.10}$$

Using Lemma 4.2(b)-(c) and the notation as for Lemma 4.5, we can show that similar to (4.9)

$$\begin{aligned} &E\sup_{1\le j\le K_2}\sup_{r_j\le r\le r_{j+1}}\Big\|\frac{\partial\varepsilon_t(\lambda_0,r)}{\partial\lambda}\frac{\partial\varepsilon_t(\lambda_0,r)}{\partial\lambda'}-\frac{\partial\varepsilon_t(\lambda_0,r_j)}{\partial\lambda}\frac{\partial\varepsilon_t(\lambda_0,r_j)}{\partial\lambda'}\Big\| = O(\delta^{\iota_1/8}), \\ &E\sup_{1\le j\le K_2}\sup_{r_j\le r\le r_{j+1}}\Big\|\varepsilon_t(\lambda_0,r)\frac{\partial^2\varepsilon_t(\lambda_0,r)}{\partial\lambda\partial\lambda'}-\varepsilon_t(\lambda_0,r_j)\frac{\partial^2\varepsilon_t(\lambda_0,r_j)}{\partial\lambda\partial\lambda'}\Big\| = O(\delta^{\iota_1/8}). \end{aligned}$$

Thus, we can take δ small enough such that $E\sup_{1\le j\le K_2}\sup_{r_j\le r\le r_{j+1}}\|P_t(\lambda_0,r)-P_t(\lambda_0,r_j)\|\le\min\{\varepsilon^2/32,\varepsilon/8\}$. Thus,

$$\begin{aligned}
&P(\frac{1}{n}\sup_{1\le j\le K_2}\sup_{r_j\le r\le r_{j+1}}\|\sum_{t=1}^{n}[P_t(\lambda_0,r)-EP_t(\lambda_0,r)]\|>\frac{\varepsilon}{4})\\
&\le P(\frac{1}{n}\sup_{1\le j\le K_2}\sup_{r_j\le r\le r_{j+1}}\|\sum_{t=1}^{n}[P_t(\lambda_0,r)-P_t(\lambda_0,r_j)]\|\\
&\quad+\sup_{1\le j\le K_2}\sup_{r_j\le r\le r_{j+1}}E\|P_t(\lambda_0,r)-P_t(\lambda_0,r_j)\|>\frac{\varepsilon}{4})\\
&\le P(\frac{1}{n}\sup_{1\le j\le K_2}\sup_{r_j\le r\le r_{j+1}}\|\sum_{t=1}^{n}[P_t(\lambda_0,r)-P_t(\lambda_0,r_j)]\|>\frac{\varepsilon}{8})\\
&\le\frac{8}{\varepsilon}E\sup_{1\le j\le K_2}\sup_{r_j\le r\le r_{j+1}}\|P_t(\lambda_0,r)-P_t(\lambda_0,r_j)\|<\frac{\varepsilon}{4}.
\end{aligned}$$

For this δ, K_2 is fixed. Thus, by the preceding equation and using the ergodic theorem for each $P_t(\lambda_0,r_i)$,

$$\begin{aligned}
&P(\frac{1}{n}\sup_{r\in[a,b]}\|\sum_{t=1}^{n}[P_t(\lambda_0,r)-P_r]\|>\frac{\varepsilon}{2})\\
&\le\sum_{i=1}^{K_2}P(\frac{1}{n}\|\sum_{t=1}^{n}[P_t(\lambda_0,r_i)-P_{r_i}]\|>\frac{\varepsilon}{4K_2})\\
&\quad+P(\frac{1}{n}\sup_{1\le j\le K_2}\sup_{r_j\le r\le r_{j+1}}\|\sum_{t=1}^{n}[P_t(\lambda_0,r)-EP_t(\lambda_0,r)]\|>\frac{\varepsilon}{4})<\frac{\varepsilon}{2},
\end{aligned}$$

for large enough n, that is, (4.10) holds. This completes the proof. □

Lemma 4.7. *If the assumption of Lemma 4.2 holds, then it follows that*

$$\begin{aligned}
&(a)\qquad P(\frac{1}{n}\sup_{\Theta_1}\sup_{r\in[a,b]}|\sum_{t=1}^{n}[\varepsilon_t^2(\lambda)-\tilde{\varepsilon}_t^2(\lambda)]|>\varepsilon)=0,\\
&(b)\qquad P(\frac{1}{\sqrt{n}}\sup_{\Theta_1}\sup_{r\in[a,b]}\|\sum_{t=1}^{n}[D_t(\lambda)-\widetilde{D}_t(\lambda)]\|>\varepsilon)=0,\\
&(c)\qquad P(\frac{1}{n}\sup_{\Theta_1}\sup_{r\in[a,b]}\|\sum_{t=1}^{n}[P_t(\lambda)-\widetilde{P}_t(\lambda)]\|>\varepsilon)=0,
\end{aligned}$$

where typically $\widetilde{D}_t(\lambda)$ is $D_t(\lambda)$ with the initial values $y_s=0$ as $s\le 0$.

Proof . By Lemma 4.1(a), we have that

$$\sup_{\Theta_1}\sup_{r\in[a,b]}|\varepsilon_t(\lambda)-\tilde{\varepsilon}_t(\lambda)|\le O(1)\sum_{i=t}^{\infty}\rho^i|y_{t-i}|,$$

where ρ is a constant in (0,1). By Lemma 4.2 (a), $E\sup_{\Theta_1}\sup_{r\in[a,b]}\varepsilon_t^2(\lambda)<\infty$. It is readily shown that $E\sup_{\Theta_1}\sup_{r\in[a,b]}\tilde{\varepsilon}_t^2(\lambda)\le$ a constant independent of t. Thus,

by Cauchy-Schwarz inequality, it follows that

$$
\begin{aligned}
&P(\frac{1}{n}\sup_{\Theta_1}\sup_{r\in[a,b]}|\sum_{t=1}^{n}[\varepsilon_t^2(\lambda)-\tilde{\varepsilon}_t^2(\lambda)]|>\varepsilon)\\
&\le \frac{1}{n\varepsilon}\sum_{t=1}^{n}E\sup_{\Theta_1}\sup_{r\in[a,b]}|\varepsilon_t^2(\lambda)-\tilde{\varepsilon}_t^2(\lambda)| \le \frac{O(1)}{n\varepsilon}\sum_{t=1}^{n}\{E\sup_{\Theta_1}\sup_{r\in[a,b]}|\varepsilon_t(\lambda)-\tilde{\varepsilon}_t(\lambda)|^2\}^{1/2}\\
&\le \frac{O(1)}{n\varepsilon}\sum_{t=1}^{n}[E(\sum_{i=t}^{\infty}\rho^i|y_{t-i}|)^2]^{1/2} \le \frac{O(1)}{n\varepsilon}\sum_{t=1}^{n}\rho^t \to 0,
\end{aligned}
$$

as $n\to\infty$. Thus, (a) holds. Similarly, we can show that (b) and (c) hold. This completes the proof. □

Proof of Theorem 2.1. For any $\eta>0$, let $c=\inf_{\|\lambda-\lambda_0\|\ge\eta}\inf_{r\in[a,b]}E[\varepsilon_t^2(\lambda)-\varepsilon_t^2(\lambda_0)]$. By Lemma 4.4, $c>0$. Furthermore, by Lemma 4.5, we have that

$$
\begin{aligned}
&P\Big(\inf_{r\in[a,b]}\inf_{\|\lambda-\lambda_0\|\ge\eta}\{\sum_{t=1}^{n}[\varepsilon_t^2(\lambda)-\varepsilon_t^2(\lambda_0)]-\frac{cn}{2}\}<0\Big)\\
&=P\Big(\inf_{r\in[a,b]}\inf_{\|\lambda-\lambda_0\|\ge\eta}\{\sum_{t=1}^{n}[\varepsilon_t^2(\lambda)-E\varepsilon_t^2(\lambda)]\\
&\quad -\sum_{t=1}^{n}\Big[\varepsilon_t^2(\lambda_0)-E\varepsilon_t^2(\lambda_0)\Big]+n\Big[E\varepsilon_t^2(\lambda)-E\varepsilon_t^2(\lambda_0)\Big]-\frac{cn}{2}\}<0\Big)\\
&\le P\Big(\inf_{r\in[a,b]}\inf_{\Theta_1}\{-2\Big|\sum_{t=1}^{n}[\varepsilon_t^2(\lambda)-E\varepsilon_t^2(\lambda)]\Big|+cn-\frac{cn}{2}\}<0\Big)\\
&\le P\Big(\inf_{r\in[a,b]}\inf_{\Theta_1}\{-\Big|\frac{1}{n}\sum_{t=1}^{n}[\varepsilon_t^2(\lambda)-E\varepsilon_t^2(\lambda)]\Big|\}<-\frac{c}{4}\Big)\\
&=P\Big(\sup_{r\in[a,b]}\sup_{\Theta_1}\{\Big|\frac{1}{n}\sum_{t=1}^{n}[\varepsilon_t^2(\lambda)-E\varepsilon_t^2(\lambda)]\Big|\}>\frac{c}{4}\Big)\to 0,
\end{aligned}
$$

as $n\to\infty$. Furthermore, by Lemma 4.7(a), we have

$$
\begin{aligned}
&P\Big(\inf_{r\in[a,b]}\inf_{\|\lambda-\lambda_0\|\ge\eta}\{\sum_{t=1}^{n}[\tilde{\varepsilon}_t^2(\lambda)-\tilde{\varepsilon}_t^2(\lambda_0)]-\frac{cn}{4}\}<0\Big)\\
&\le P\Big(\inf_{r\in[a,b]}\inf_{\|\lambda-\lambda_0\|\ge\eta}\sum_{t=1}^{n}[\varepsilon_t^2(\lambda)-\varepsilon_t^2(\lambda_0)]\\
&\quad -2\sup_{r\in[a,b]}\sup_{\|\lambda-\lambda_0\|\ge\eta}\Big|\sum_{t=1}^{n}[\varepsilon_t^2(\lambda)-\tilde{\varepsilon}_t^2(\lambda_0)]\Big|-\frac{cn}{4}\}<0\Big)\\
&\le P\Big(\inf_{r\in[a,b]}\inf_{\|\lambda-\lambda_0\|\ge\eta}\sum_{t=1}^{n}[\varepsilon_t^2(\lambda)-\varepsilon_t^2(\lambda_0)]-\frac{cn}{2}<0\Big)\\
&\quad +P\Big(\sup_{r\in[a,b]}\sup_{\|\lambda-\lambda_0\|\ge\eta}\Big|\sum_{t=1}^{n}[\varepsilon_t^2(\lambda)-\tilde{\varepsilon}_t^2(\lambda_0)]\Big|>\frac{cn}{8}\Big)\to 0,
\end{aligned}
$$

as $n \to \infty$. Thus, for any $\epsilon > 0$, it follows that

$$
\begin{aligned}
&\lim_{n\to\infty} P(\sup_{r\in[a,b]} \|\tilde{\lambda}_n - \lambda_0\| > \epsilon) \\
&= \lim_{n\to\infty} P\Big\{\sup_{r\in[a,b]} \|\tilde{\lambda}_n - \lambda_0\| > \epsilon, \inf_{r\in[a,b]} \sum_{t=1}^{n} \Big[\tilde{\varepsilon}_t^2(\tilde{\lambda}_n) - \tilde{\varepsilon}_t^2(\lambda_0)\Big] \le 0\Big\} \\
&\le \lim_{n\to\infty} P\Big\{\inf_{r\in[a,b]} \inf_{\|\lambda-\lambda_0\|>\epsilon} \sum_{t=1}^{n} \Big[\tilde{\varepsilon}_t^2(\lambda) - \tilde{\varepsilon}_t^2(\lambda_0)\Big] \le 0\Big\} = 0.
\end{aligned}
$$

Thus, (a) holds under H_0 as well as under H_1 with $\{y_t\}$ being strictly stationary and ergodic. Applying Taylor's expansion to $\partial\tilde{\varepsilon}_t(\tilde{\lambda}_n)/\partial\lambda$, we have

$$
\sqrt{n}(\tilde{\lambda}_n - \lambda_0) = -\Big[\frac{1}{n}\sum_{t=1}^{n} \widetilde{P}_t(\tilde{\lambda}_n^*)\Big]^{-1} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \widetilde{D}_t(\lambda_0),
$$

where $\tilde{\lambda}_n^*$ lies between $\tilde{\lambda}_n$ and λ_0. By Lemma 4.6 and 4.7(b)-(c), we can show that (b) holds under H_0 as well as under H_1 with $\{y_t\}$ being strictly stationary and ergodic.

We now consider (c). First, note that $EP_t(\lambda_0) = E[U_t(\lambda_0)U_t'(\lambda_0)]$ under H_0. Denote $D_{1n} = \sum_{t=1}^{n} D_{1t}(\lambda_0)/\sqrt{n}$ and $D_{2n} = \sum_{t=1}^{n} D_{2t}(\lambda_0)/\sqrt{n}$. It is well known that $L_{0n}(\hat{\phi}_n)$ has the following expansion.

$$
\begin{aligned}
2[L_{0n}(\hat{\phi}_n) - L_{0n}(\phi_0)] &= -(\hat{\phi}_n - \phi_0)'\Sigma(\hat{\phi}_n - \phi_0) + o_p(1) \\
&= -D_{1n}'\Sigma^{-1}D_{1n} + o_p(1). \qquad (4.11)
\end{aligned}
$$

By (b) of this theorem and Lemma 4.6, using Taylor's expansion, it follows that

$$
\begin{aligned}
2[L_{1n}(\widetilde{\lambda}_n) - L_{1n}(\lambda_0)] &= -(\widetilde{\lambda}_n - \lambda_0)'P_r(\widetilde{\lambda}_n - \lambda_0) + o_p(1) \\
&= -D_n'P_r^{-1}D_n + o_p(1), \qquad (4.12)
\end{aligned}
$$

where $D_n = [D_{1n}'D_{2n}']'$. Denote $K_{0r} = \Sigma - \Sigma_{1r}\Sigma_r^{-1}\Sigma_{1r}'$. We can show that $P_r^{-1}D_t(\lambda_0)$ $= \{[D_{1t}(\lambda_0) - \Sigma_{1r}\Sigma_r^{-1}D_{2t}(\lambda_0)]'K_{0r}^{-1}, [D_{2t}(\lambda_0) - \Sigma_{1r}'\Sigma^{-1}D_{1t}(\lambda_0)]'K_{rr}^{-1}\}'$. Thus,

$$
\begin{aligned}
&2[L_{1n}(\widetilde{\lambda}_n) - L_{1n}(\lambda_0)] \\
&= -\Big[D_{1n}'K_{0r}^{-1}D_{1n} - D_{1n}'K_{0r}^{-1}\Sigma_{1r}\Sigma_r^{-1}D_{2n} \\
&\quad + D_{2n}'K_{rr}^{-1}D_{2n} - D_{2n}'K_{rr}^{-1}\Sigma_{1r}'\Sigma^{-1}D_{1n}\Big] + o_p(1) \\
&= -(D_{2n} - \frac{1}{2}CD_{1n})'K_{rr}^{-1}(D_{2n} - \frac{1}{2}CD_{1n}) \\
&\quad - D_{1n}'(K_{0r}^{-1} - \frac{1}{4}C'K_{rr}^{-1}C)D_{1n} + o_p(1), \qquad (4.13)
\end{aligned}
$$

where $C = \Sigma_{1r}'\Sigma^{-1} + K_{rr}\Sigma_r^{-1}\Sigma_{1r}'K_{0r}^{-1}$. After some algebra, we can show that $C = 2K_{rr}\Sigma_r^{-1}\Sigma_{1r}'K_{0r}^{-1} = 2\Sigma_{1r}'\Sigma^{-1}$ and $K_{0r}^{-1} - C'K_{rr}^{-1}C/4 = \Sigma$. Furthermore, by (4.11) and (4.13), the conclusion holds. This completes the proof. □

5 Proof of Theorem 2.2

To prove Theorem 2.2, we first introduce three lemmas.

Lemma 5.1. *If Assumption 2.2 holds, then when $r' \le r$, it follows that*

$$E[|\varepsilon_{t-j}|^k I(r' < y_{t-d} \le r)] \le C(r - r'),$$

where $k = 0, 1, 2, 3, 4$, $j \ge 1$ and C is some constant independent of t and j.

Proof. By the assumption given, $E|\varepsilon_{t-j}|^k < \infty$ for $k = 0, 1, 2, 3, 4$. Thus, there is a constant M such that $\sup_{|x|>M} |x|^k f(x) < 1$. Furthermore, since f is continuous, it follows that $\sup_{|x|\le M} |x|^k f(x) < \infty$. Thus, we can claim that

$$\sup_{x\in R} |x|^k f(x) < \infty \ for \ \ k = 0, 1, 2, 3, 4.$$

Let $g_t = \sum_{i=1}^p \phi_{0i}\varepsilon_{t-i}$. When $j \ne d$,

$$\begin{aligned} E[|\varepsilon_{t-j}|^k I(r' < y_{t-d} \le r)] &= E[|\varepsilon_{t-j}|^k I(r' - g_{t-d} < \varepsilon_{t-d} \le r - g_{t-d})] \\ &= E[|\varepsilon_{t-j}|^k \int_{r'-g_{t-d}}^{r-g_{t-d}} f(x)dx] \le C(r - r'). \end{aligned}$$

When $j = d$,

$$\begin{aligned} E[|\varepsilon_{t-d}|^k I(r' < y_{t-d} \le r)] &= E[|\varepsilon_{t-d}|^k I(r' - g_{t-d} < \varepsilon_{t-d} \le r - g_{t-d})] \\ &= E[\int_{r'-g_{t-d}}^{r-g_{t-d}} |x|^k f(x)dx] \le C(r - r'). \end{aligned}$$

By the preceding two equations, the conclusion holds. This completes the proof. □

Lemma 5.2. *Under the assumption of Theorem 2.2, when $r' \le r$, it follows that*

$$E\Big\|\frac{1}{\sqrt{n}}\sum_{t=1}^n \Big[\sum_{i=0}^\infty (-\Phi)^i Z_{t-i-1} I(r' < y_{t-d-i} \le r)\Big]\varepsilon_t\Big\|^4 \le C\Big[\frac{(r-r')^{1/2}}{\sqrt{n}} + (r - r')\Big]^2,$$

where C is some constant independent of n, and $r, r' \in [-\gamma, \gamma]$ for any given $\gamma > 0$.

Proof. Let $L_t = \sum_{i=0}^{\infty} \|\Phi^i\| \|Z_{t-i-1}\| I(r' < y_{t-d-i} \le r)$. By the assumption given, $\|\Phi^i\| = O(\rho^i)$ with $\rho \in (0,1)$. By Lemma 5.1,

$$E\varepsilon_t^2 L_t^2 \le C(r-r') \textit{ and } E(\varepsilon_t^2 L_t^2 - E\varepsilon_t^2 L_t^2)^2 \le C(r-r'),$$

where C is independent of t. Thus,

$$\begin{aligned}
& E\Big\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big[\sum_{i=0}^{\infty}(-\Phi)^i Z_{t-i-1} I(r' < y_{t-d-i} \le r)\Big]\varepsilon_t\Big\|^4 \\
& \le \frac{1}{n^2}\sum_{t=1}^{n}\sum_{t_1=1}^{n} E(\varepsilon_t^2 L_t^2 \varepsilon_{t_1}^2 L_{t_1}^2) \\
& = E\Big(\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2 L_t^2\Big)^2 \\
& = E\Big[\frac{1}{n}\sum_{t=1}^{n}(\varepsilon_t^2 L_t^2 - E\varepsilon_t^2 L_t^2) + E\varepsilon_t^2 L_t^2\Big]^2 \\
& \le \Big\{\frac{1}{n}\Big[E\Big(\sum_{t=1}^{n}(\varepsilon_t^2 L_t^2 - E\varepsilon_t^2 L_t^2)\Big)^2\Big]^{1/2} + C(r-r')\Big\}^2.
\end{aligned}$$

Since y_t is only p-dependent, $E[(\varepsilon_t^2 L_t^2 - E\varepsilon_t^2 L_t^2)(\varepsilon_{t_1}^2 L_{t_1}^2 - E\varepsilon_{t_1}^2 L_{t_1}^2)] = 0$ when $|t-t_1| > p$. Thus, it follows that

$$\begin{aligned}
E\Big(\sum_{t=1}^{n}(\varepsilon_t^2 L_t^2 - E\varepsilon_t^2 L_t^2)\Big)^2 &= \sum_{t=1}^{n} E(\varepsilon_t^2 L_t^2 - E\varepsilon_t^2 L_t^2)^2 \\
& \quad +2\sum_{t=1}^{n}\sum_{s=1}^{n-t} E[(\varepsilon_t^2 L_t^2 - E\varepsilon_t^2 L_t^2)(\varepsilon_{t+s}^2 L_{t+s}^2 - E\varepsilon_t^2 L_t^2)] \\
&= \sum_{t=1}^{n} E(\varepsilon_t^2 L_t^2 - E\varepsilon_t^2 L_t^2)^2 \\
& \quad +2\sum_{t=1}^{n}\sum_{s=1}^{\min\{n-t,p\}} E[(\varepsilon_t^2 L_t^2 - E\varepsilon_t^2 L_t^2)(\varepsilon_{t+s}^2 L_{t+s}^2 - E\varepsilon_t^2 L_t^2)] \\
&\le (2p+1)\sum_{t=1}^{n} E(\varepsilon_t^2 L_t^2 - E\varepsilon_t^2 L_t^2)^2 \le (p+1)nC(r-r').
\end{aligned}$$

By the preceding two equations, we can claim the conclusion holds. This completes the proof. □

Lemma 5.3. *Let $m_t = \|Z_{t-1}\| I(r' < y_{t-d} \le r)$. Under the assumption of Theorem 2.2, when $r' \le r$, it follows that*

$$(a) \qquad E\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{i=0}^{\infty}\|\Phi^i\|(m_{t-i} - Em_{t-i})\Big|^4 \le C\Big[\frac{r-r'}{n} + (r-r')^2\Big],$$

$$(b) \qquad E\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(|\varepsilon_t| - E|\varepsilon_t|)\sum_{i=0}^{\infty}\|\Phi^i\| m_{t-i}\Big|^4 \le C\Big[\frac{r-r'}{n} + (r-r')^2\Big],$$

where C is some constant independent of n, and $r', r \in [-\gamma, \gamma]$ for any given $\gamma > 0$.

Proof. Since m_t is strictly stationary, Em_t is independent of t. Furthermore, we have the following inequality

$$\begin{aligned}
&E\Big[\sum_{t=1}^{n}(m_{t-i} - Em_t)\Big]^4 \\
&\leq \sum_{t=1}^{n} E(m_{t-i} - Em_t)^4 + c_1\Big|\sum_{t=1}^{n}\sum_{s=1}^{n-t} E[(m_{t-i} - Em_t)^3(m_{t+s-i} - Em_{t+s})]\Big| \\
&\quad +c_2\Big|\sum_{t=1}^{n}\sum_{s=1}^{n-t} E[(m_{t-i} - Em_t)^2(m_{t+s-i} - Em_{t+s})^2]\Big| \\
&\quad +c_2\Big|\sum_{t=1}^{n}\sum_{t_1=1}^{n-t}\sum_{t_2=1}^{n-t-t_1}\sum_{t_3=1}^{n-t-t_1-t_2} E[(m_{t-i} - Em_t) \\
&\qquad \cdot(m_{t+t_1-i} - Em_t)(m_{t+t_1+t_2-i} - Em_t)(m_{t+t_1+t_2+t_3-i} - Em_t)]\Big| \\
&\equiv A_{1n} + c_1A_{2n} + c_2A_{3n} + c_3A_{4n},
\end{aligned}$$

where c_1, c_2 and c_3 are constants independent of n and i. By Lemma 5.1, we can see that $E(m_t - Em_t)^4 \leq C_1(r - r')$. Since $\{m_t - Em_t\}$ is a $p-$dependent sequence, $E[(m_t - Em_t)^3(m_{t_1} - Em_t)] = 0$ when $|t - t_1| > p$. Thus,

$$\begin{aligned}
A_{2n} &= \Big|\sum_{t=1}^{n}\sum_{s=1}^{\min\{n-t,p\}} E[(m_{t-i} - Em_t)^3(m_{t+s-i} - Em_{t+s})]\Big| \\
&\leq pnE(m_{t-i} - Em_t)^4 \leq npC_1(r - r').
\end{aligned}$$

Let $\widetilde{m}_t = (m_{t-i} - Em_t)^2 - E(m_{t-i} - Em_t)^2$. Then, by Lemma 5.1, we can show that $E\widetilde{m}_t^2 \leq C_2(r - r')$. Since $\{\widetilde{m}_t\}$ is a $p-$dependent sequence, we know that $E(\widetilde{m}_t\widetilde{m}_{t_1}) = 0$ when $|t - t_1| > p$. Furthermore, since $E(m_t - Em_t)^2 \leq C_3(r - r')$,

$$\begin{aligned}
A_{3n} &= \Big|\sum_{t=1}^{n}\sum_{s=1}^{n-t} E(\widetilde{m}_t\widetilde{m}_{t+s}) - \sum_{t=1}^{n}(n-t)[E(m_t - Em_t)^2]^2\Big| \\
&\leq \Big|\sum_{t=1}^{n}\sum_{s=1}^{\min\{n-t,p\}} E(\widetilde{m}_t\widetilde{m}_{t+s})\Big| + C_3n^2(r - r')^2 \leq C_2pn(r - r') + C_3n^2(r - r')^2.
\end{aligned}$$

Denote $\tilde{p} = \min\{n - t, p\}$. Similarly, we have that

$$\begin{aligned}
A_{4n} &= \Big|\sum_{t=1}^{n}\sum_{t_1=1}^{\tilde{p}}\sum_{t_2=1}^{\tilde{p}-t_1}\sum_{t_3=1}^{\tilde{p}-t_1-t_2} E[(m_{t-i} - Em_t) \\
&\quad \cdot(m_{t+t_1-i} - Em_t)(m_{t+t_1+t_2-i} - Em_t)(m_{t+t_1+t_2+t_3-i} - Em_t)]\Big| \\
&\leq p^3\sum_{t=1}^{n} E(m_{t-i} - Em_t)^4 \leq np^3C_4(r - r').
\end{aligned}$$

By the preceding four inequalities, we can claim that

$$E\Big[\sum_{t=1}^{n}(m_{t-i}-Em_t)\Big]^4 \le nC_5(r-r')+C_5n^2(r-r')^2.$$

In the preceding four equations, C_i, $i=1,\cdots,5$, are some constants independent of i and n. By the assumption given, $\|\Phi^i\|=O(\rho^i)$ with $\rho\in(0,1)$. Thus, by Minkowskii's inequality,

$$\begin{aligned}
&E\Big|\frac{1}{\sqrt{n}}[\sum_{t=1}^{n}\sum_{i=0}^{\infty}\|\Phi^i\|(m_{t-i}-Em_{t-i})\Big|^4\\
&\le \frac{1}{n^2}E\Big[\sum_{i=0}^{\infty}\|\Phi^i\|\Big|\sum_{t=1}^{n}(m_{t-i}-Em_{t-i})\Big|\Big]^4\\
&\le \frac{1}{n^2}\Big[\sum_{i=0}^{\infty}\|\Phi^i\|\Big\{E\Big|\sum_{t=1}^{n}(m_{t-i}-Em_{t-i})\Big|^4\Big\}^{1/4}\Big]^4\\
&\le \frac{1}{n^2}\Big\{[nC_5(r-r')+C_5n^2(r-r')^2]^{1/4}\sum_{i=0}^{\infty}\|\Phi^i\|\Big\}^4 \le \frac{C(r-r')}{n}+C(r-r')^2,
\end{aligned}$$

where C is some constant independent of n. Thus, (a) holds.

We now consider (b).

$$\begin{aligned}
&E\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(|\varepsilon_t|-E|\varepsilon_t|)\sum_{i=0}^{\infty}\|\Phi^i\|m_{t-i}\Big|^4\\
&\le \frac{O(1)}{n^2}E\Big[\sum_{t=1}^{n}\sum_{t_1=1}^{n}\Big(\sum_{i=0}^{\infty}\|\Phi^i\|m_{t-i}\Big)^2\Big(\sum_{i=0}^{\infty}\|\Phi^i\|m_{t_1-i}\Big)^2\Big]\\
&\le \frac{O(1)}{n^2}E\Big[\sum_{t=1}^{n}\Big(\sum_{i=0}^{\infty}\rho^i m_{t-i}\Big)^2\Big]^2\\
&\le \frac{O(1)}{n^2}E\Big\{\sum_{t=1}^{n}\Big[\sum_{i=0}^{\infty}\rho^i(m_{t-i}-Em_{t-i})\Big]^2+n\Big(\sum_{i=0}^{\infty}\rho^iEm_{t-i}\Big)^2\Big\}^2\\
&\le \frac{O(1)}{n^2}E\Big\{\sum_{t=1}^{n}\Big[\sum_{i=0}^{\infty}\rho^i(m_{t-i}-Em_{t-i})\Big]^2\Big\}^2+O((r-r')^4),
\end{aligned} \tag{5.1}$$

where $0<\rho<1$ is some constant. Denote $X_t=\sum_{i=0}^{\infty}\rho^i(m_{t-i}-Em_{t-i})$. For any $s\ge 1$, let $X_{1ts}=\sum_{i=0}^{s-1}\rho^i(m_{t+s-i}-Em_{t+s-i})$ and $X_{2ts}=\sum_{i=s}^{\infty}\rho^i(m_{t+s-i}-Em_{t+s-i})$. By Minkowskii's inequality and Lemma 5.1, it is not difficult to see that

$$EX_t^2=O(r-r'),\;\; EX_t^4=O(r-r') \;\; and \;\; EX_{1ts}^4=O(r-r').$$

When $s>p$, let $X_{1tsp}=\sum_{i=0}^{s-p-1}\rho^i(m_{t+s-i}-Em_{t+s-i})$ and $X_{2tsp}=\sum_{i=s-p}^{s-1}\rho^i$ $(m_{t+s-i}-Em_{t+s-i})$. Then, by Minkowskii's inequality and Lemma 5.1, we have

$$EX_{1tsp}^2=O(r-r'),\;\; EX_{1tsp}^4=O(r-r') \;\; and \;\; EX_{2tsp}^4\le O(\rho^s(r-r')).$$

$X_{2ts} = \rho^s \sum_{i=0}^{\infty} \rho^i (m_{t-i} - Em_{t-i}) = \rho^s X_t$ and X_t is independent of X_{1tsp} when $s > p$. Furthermore, since $X_{t+s} = X_{1ts} + X_{2ts}$ and $X_{1ts} = X_{1tsp} + X_{2tsp}$, we have that

$$
\begin{aligned}
E\Big(\sum_{t=1}^{n} X_t^2\Big)^2 &= \sum_{t=1}^{n} EX_t^4 + 2E\sum_{t=1}^{n}\sum_{s=1}^{n-t} X_t^2 X_{t+s}^2 \\
&\le \sum_{t=1}^{n} EX_t^4 + 4E\sum_{t=1}^{n}\sum_{s=1}^{n-t}(\rho^{2s}X_t^4 + X_{1ts}^2 X_t^2) \\
&\le O(1)nEX_t^4 + 4\sum_{t=1}^{n}\sum_{s=1}^{n-t} EX_{1ts}^2 X_t^2 \\
&\le O(1)n(r-r') + 4\sum_{t=n-p}^{n}\sum_{s=1}^{n-t} EX_{1ts}^2 X_t^2 + 4\sum_{t=1}^{n-p-1}\sum_{s=1}^{n-t} EX_{1ts}^2 X_t^2 \\
&\le O(1)n(r-r') + 8\sum_{t=1}^{n-p-1}\sum_{s=1}^{n-t}(EX_{1tsp}^2 EX_t^2 + EX_{2tsp}^2 X_t^2) \\
&\le O(1)n(r-r') + O(1)n^2(r-r')^2, \qquad (5.2)
\end{aligned}
$$

where we used the inequality $EX_{1ts}^2 X_t^2 \le (EX_{1ts}^4 EX_t^4)^{1/2} = O(r-r')$ and $E(X_{2tsp}^2 X_t^2) \le (EX_{2tsp}^4 EX_t^4)^{1/2} = O(\rho^{s/2}(r-r'))$. By (5.1)-(5.2), we can claim that (b) holds. This completes the proof. $\square$

Proof of Theorem 2.2. Let

$$
\widetilde{T}_n(r) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big[\sum_{i=0}^{\infty}(-\Phi)^i Z_{t-i-1} I(y_{t-d-i} \le r)\Big]\varepsilon_t.
$$

We first show that $\{\widetilde{T}_n(r) : r \in R_\gamma\}$ is tight for any $\gamma > 0$, where $R_\gamma = [-\gamma, \gamma]$.

For any given $\eta > 0$, we first choose (δ, n) such that $1 > \delta \ge n^{-1}$ and $\sqrt{n} \ge M/\eta$ and then choose an integer K such that $\delta n/2 \le K \le n\delta$, where M is determined later. Let $r_{k+1} = r_k + \delta/K$, where $r_1 = r'$ and $k = 1, \cdots, K$. Thus,

$$
\begin{aligned}
\sup_{r'<r\le r'+\delta} \|\widetilde{T}_n(r) - \widetilde{T}_n(r')\| &\le \sup_{1\le k\le K} \|\widetilde{T}_n(r_k) - \widetilde{T}_n(r')\| \\
&\quad + \sup_{1\le k\le K}\ \sup_{r_k<r\le r_k+\delta/K} \|\widetilde{T}_n(r) - \widetilde{T}_n(r_k)\|. \qquad (5.3)
\end{aligned}
$$

By Lemma 5.2, for any $1 \le i < j \le K$,

$$
E\|\widetilde{T}_n(r_i) - \widetilde{T}_n(r_j)\|^4 \le C[(r_j - r_i)^{1/2}/\sqrt{n} + (r_j - r_i)]^2 = C\Big(\sum_{k=i+1}^{j}\frac{\delta}{K}\Big)^2,
$$

where the last equation holds because $1/\sqrt{n} \le \sqrt{\delta/K}$ and $(r_j - r_i)^{1/2} = [(j - i)\delta/K]^{1/2} \le (j-i)\sqrt{\delta/K}$. Note that $\widetilde{T}_n(r_j) - \widetilde{T}_n(r_i) = \sum_{k=i+1}^{j}[\widetilde{T}_n(r_k) - \widetilde{T}_n(r_{k-1})]$.

By the preceding equation and Theorem 12.2 of Billingsley (1968, p.94), there exists a constant C_1 independent of K, n and r' such that

$$P\Big(\sup_{1\le k\le K}\|\widetilde{T}_n(r_k)-\widetilde{T}_n(r')\|>\frac{\eta}{2}\Big)\le\frac{CC_1}{\eta^4}\Big(\sum_{k=1}^{K}\frac{\delta}{K}\Big)^2=\frac{CC_1\delta^2}{\eta^4}. \tag{5.4}$$

Let $m_{kt}=\|Z_{t-1}\|I(r_k<y_{t-d}\le r_k+\delta/K)$. By Lemma 5.1 and the definition of K and η, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}E\Big[|\varepsilon_t|\Big[\sum_{i=0}^{\infty}\|\Phi^i\|m_{kt-i}\Big]\le\frac{C_2\sqrt{n}\delta}{K}\le\frac{2C_2\sqrt{n}\delta}{n\delta}\le\frac{2C_2}{\sqrt{n}}\le\frac{\eta}{4},$$

as $M\ge 8C_2$, where C_2 is a constant independent of k. By the preceding inequality, Lemma 5.3 and Markov's inequality,

$$\begin{aligned}
&P(\sup_{1\le k\le K}\sup_{r_k<r\le r_k+\delta/K}\|\widetilde{T}_n(r)-\widetilde{T}_n(r_k)\|>\frac{\eta}{2})\\
&\le\sum_{k=1}^{K}P(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big(|\varepsilon_t|\sum_{i=0}^{\infty}\|\Phi^i\|m_{kt}\Big)>\frac{\eta}{2})\\
&\le\sum_{k=1}^{K}P(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big[\Big(|\varepsilon_t|\sum_{i=0}^{\infty}\|\Phi^i\|m_{kt}\Big)-E\Big(|\varepsilon_t|\sum_{i=0}^{\infty}\|\Phi^i\|m_{kt}\Big)\Big]>\frac{\eta}{4})\\
&\le\frac{4^4}{\eta^4}\sum_{k=1}^{K}E\Big\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big[\Big(|\varepsilon_t|\sum_{i=0}^{\infty}\|\Phi^i\|m_{kt}\Big)-E\Big(|\varepsilon_t|\sum_{i=0}^{\infty}\|\Phi^i\|m_{kt}\Big)\Big]\Big\}^4\\
&\le\frac{4^4CK}{\eta^4}\Big(\frac{\delta}{nK}+\frac{\delta^2}{K^2}\Big)\le\frac{2^9C\delta^2}{\eta^4},
\end{aligned} \tag{5.5}$$

since $n>\delta/K$. Given $\varepsilon>0$ and $\eta>0$, let $\delta=\min\{\varepsilon\eta^4/(CC_1+2^9C),1\}$. We first select M such that $M\ge 8C_2$, and then select $N=\max\{\delta^{-1},M^2/\eta^2\}$. Then, for any r', as $n>N$, by (5.3)-(5.5), it follows that

$$P\Big(\sup_{r'<r\le r'+\delta}\|\widetilde{T}_n(r)-\widetilde{T}_n(r')\|>\eta\Big)\le\frac{2^9C\delta^2}{\eta^4}+\frac{CC_1\delta^2}{\eta^4}=\delta\varepsilon.$$

By Theorem 15.5 in Billingsely (1968) (also see the proof of his Theorem 16.1), we can claim that $\{\widetilde{T}_n(r):R_\gamma\}$ is tight.

It is well known that $\sum_{t=1}^{n}D_{1t}(\lambda_0)/\sqrt{n}\to_{\mathcal{L}}N(0,\sigma^2\Sigma)$ and hence it is tight. Furthermore, since Σ_{1r} is continuous in terms of r on R_γ, combining this with the tightness of $\{\widetilde{T}_n(r):R_\gamma\}$, we have that $\{T_n(r):R_\gamma\}$ is tight. By the central limit theorem [e.g. Corollary 2.1 in Hall and Heyde (1980)], we can show that

the finite dimensional distributions of $\{T_n(r) : x \in R_\gamma\}$ converge weakly to those of $\{\sigma G_p(x) : x \in R_\gamma\}$ as $n \to \infty$. By Prohorov's theorem in Billingsley (1968, p.37), $T_n(r) \Rightarrow \sigma G_p(x)$ on $D^p[R_\gamma]$ for each $\gamma \in (0, \infty)$. By Theorem 15.5 in Billingsely (1968), almost all the paths of $G_p(x)$ are continuous in terms of x. This completes the proof. $\square$

REFERENCES

An, H.Z. and Chen, S.G. (1997) A note on the ergodicity of non-linear autoregressive model. *Statist. Probab. Lett.* **34**, 365-372.

An, H.Z. and Huang, F C. (1996) The geometrical ergodicity of nonlinear autoregressive models. *Statist. Sinica* **6**, no. 4, 943-956.

Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* **61**, 821-855.

Billingsley, P. (1968) *Convergence of Probability Measures.* New York: Wiley.

Brockwell, P., Liu, J. and Tweedie, R.L. (1992) On the existence of stationary threshold autoregressive moving-average processes. *J. Time Ser. Anal.* **13**, 95-107.

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

Chan, K.S. (1990) Testing for threshold autoregression. *Ann. Statist.* **18** (1990), 1886-1894.

Chan, K.S. (1991) Percentage points of likelihood ratio tests for threshold autoregression.*J. Roy. Statist. Soc. Ser. B* **53**, 691-696.

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. 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.

Chan, K.S. and Tong, H. (1990) On likelihood ratio tests for threshold autoregression. *J. Roy. Statist. Soc. Ser. B* **52** (1990), 469-476.

Chan, K.S. and Tsay, R.S. (1998) Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika* **85** (1998), 413-426.

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.

Chen, R. and Tsay, R.S. (1991) On the ergodicity of TAR(1) processes. *Ann. Appl. Probab.* **1**, 613-634.

Hall, P. and Heyde, C. C. (1980) *Martingale Limit Theory and Its Applications.* Academic, San Diego.

Hansen, B.E. (2000) Sample splitting and threshold estimation. *Econometrica* **68** (2000), 575-603.

Koul, H.L. (2000) Fitting a two phase linear regression model. Special issue dedicated to Professor Vasant P. Bhapkar. *J. Indian Statist. Assoc.* **38**, 331-353.

Koul, H.L. and Schick, A. (1997) Efficient estimation in nonlinear autoregressive time-series models. *Bernoulli* **3**, 247-277.

Koul, H.L. and Stute, W. (1999) Nonparametric model checks for time series. *Ann. Statist.* **27**, 204-236.

Koul, H.L., Qian, L. and Surgailis, D. (2003) Asymptotics of M-estimators in two-phase linear regression models. *Stochastic Process. Appl.* **103**, 123-154.

Ling, S. (2003) Weak convergence for marked empirical processes of stationary sequences and goodness-of-fit tests for time series models. Working Paper in HKUST.

Ling, S. (1999) On probability properties of a double threshold ARMA conditional heteroskedasticity model. *Journal of Applied Probability* **36**, 688-705.

Ling, S. and McAleer, M. (2002b) On adaptive estimation in nonstationary ARMA models with GARCH errors. *Ann. Statist.* **31**, 642-674.

Liu, J. and Susko, E. (1992) On strict stationarity and ergodicity of a nonlinear ARMA model.*J. Appl. Probab.* **29** (1992), 363-373.

Liu, J., Li, W.K. and Li, C W. (1997) On a threshold autoregression with conditional heteroscedastic variances. *J. Statist. Plann. Inference* **62**, 279-300.

Qian, L. (1998) On maximum likelihood estimators for a threshold autoregression. *J. Statist. Plann. Inference* **75**, 21-46.

Stute, W. (1997) Nonparametric model checks for regression. *Ann. Statist.* **25**, 613-641.

Tong, H. (1978) On a threshold model. *Pattern recognition and signal processing.* (ed. C.H.Chen). Sijthoff and Noordhoff, Amsterdam.

Tong, H. (1990) *Nonlinear time series. A dynamical system approach.* The Clarendon Press, Oxford University Press, New York.

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.

Van der Vaart, A.W. and Wellner, J.A. (1996) Weak convergence and empirical processes. With applications to statistics. Springer Series in Statistics. Springer-Verlag, New York.

Wong, C.S. and Li, W.K. (1997) Testing for threshold autoregression with conditional heteroscedasticity. *Biometrika* **84**, 407-418.

Wong, C.S. and Li, W.K. (2000) Testing for double threshold autoregressive conditional heteroscedastic model. *Statist. Sinica* **10**, 173-189.