Shiqing Ling

Time-Series Econometric Theory

December 5, 2011

Springer

Contents

Chapter 1
Introduction

A time-series data set is a sequence of observations ordered in time. The time unit may be second, day, month or year etc. For example, let Y_t be the daily closing price of Hang Seng Index (HSI). Figure 1 is the plot of Y_t from the year 1986-2010. Figure 2 is the plot of the corresponding log-return $x_t = \log Y_t - \log Y_{t-1}$.

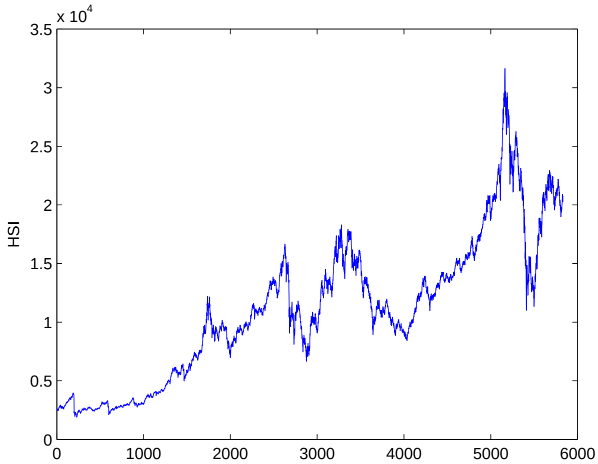


Fig. 1.1 Daily closing price of HSI from 31th Dec 1986 to 30th Jun 2010

If we ask what the HSI is tomorrow, the answer may 20,249, 20,500, $\cdots$, etc.. In fact, it can take any value in $(0, \infty)$, that is, HSI is a random variable (r.v.) tomorrow. Thus, we can understand that HSI is random each day and the closing price is just its observation. We define a time series as a sequence of r.v.s ordered in time, i.e., $\{Y_t : t \in T\}$, where T is the range of time t.

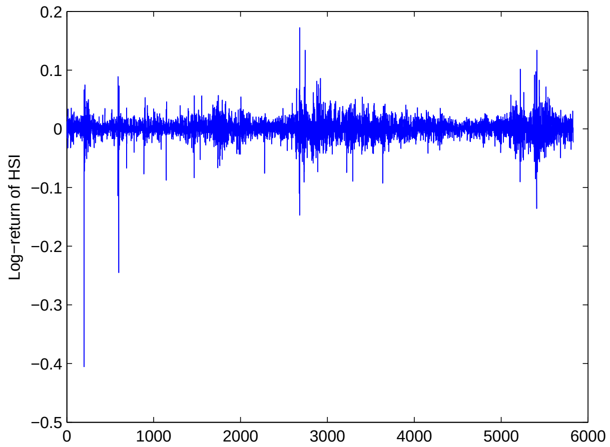
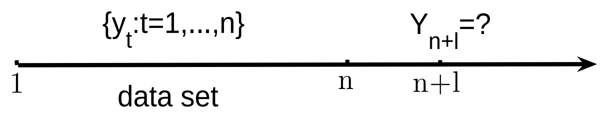


Fig. 1.2 Log-return of daily closing price of HSI from 31th Dec 1986 to 30th Jun 2010

In practice, we never can see $\{Y_t : t \in T\}$ rather than the data set $\{Y_t : t \in T\}$ which is the collection of observations of $\{Y_t : t \in T\}$. Suppose that we have a data set $\{Y_t : t = 1, \cdots, n\}$. What is Y_{n+l} ?



Nobody can answer this question. However, we can try to look for

$$\begin{aligned}&(1)\quad E\left(Y_{n+l}|Y_1, \cdots, Y_n\right),\\&(2)\quad P(a \le Y_{n+l} \le b) \text{ for some } a < b.\end{aligned}$$

These are the main targets in the field of time series.

We now define a general setup of time series. Let $(\Omega, \mathscr{F}, \mathscr{P})$ be a probability space and $\{\mathscr{F}_t : t \in T\}$ be a sequence of $\sigma-$fields with $\mathscr{F}_t \subseteq \mathscr{F}_{t+1} \subseteq \cdots \subseteq \mathscr{F}$. We simply understand $\mathscr{F}_t$ as the information available up to time t. Let $\{Y_t : t \in T\}$ be a time series defined on $(\Omega, \mathscr{F}, \mathscr{P})$ and $Y_t \in \mathscr{F}_t$ (i.e. Y_t is fully determined by the information up to time t). Given $\mathscr{F}_{t-1}$, the best predictor in mean square [i.e., the minimizer of $E(Y_t - c)^2$] is

$$\mu_t = E\left(Y_t|\mathscr{F}_{t-1}\right).$$

Let $\varepsilon_t = Y_t - \mu_t$. Then, $E(\varepsilon_t|\mathscr{F}_{t-1}) = 0$ and hence $\{\varepsilon_t\}$ is a sequence of martingale difference in terms of $\mathscr{F}_t$. Thus, we can decompose Y_t as

$$Y_t = \mu_t + \varepsilon_t. \tag{1.1}$$

Furthermore, we denote the conditional variance of Y_t by h_t, i.e.,

$$h_t = E\left[(Y_t - \mu_t)^2 |\mathscr{F}_{t-1}\right] = E\left(\varepsilon_t^2|\mathscr{F}_{t-1}\right),$$

and $\eta_t = \varepsilon_t/\sqrt{h_t}$. Then, we can further decompose Y_t as

$$Y_t = \mu_t + \eta_t\sqrt{h_t}, \tag{1.2}$$

where $\{\eta_t\}$ is a sequence of uncorrelated r.v.s with $E(\eta_t|\mathscr{F}_{t-1}) = 0$ and $E\left(\eta_t^2|\mathscr{F}_{t-1}\right) = 1$. To answer (1.1)-(1.2), we need to explore μ_t, h_t and η_t. We know that both μ_t and h_t are functions of $\mathscr{F}_{t-1}$. However, we don't know the forms of these functions in practice. We can try to use some simple functions to approximate μ_t and h_t. These different functions will create different so-called models. A lot of models have been proposed in the literature. We will discuss some models in details in other chapters.

In practice, all of models are some different approximations of the time series $\{Y_t\}$. We need to build up some reasonable criteria under which the model is close to $\{Y_t\}$ enough. There are many criteria in the literature, such as AIC, BIC and portmanteau tests, among many others. To make sure the selected model based on these criteria are reliable, we first need to study its structure. This is the fundamental part in the field of time series and will be discussed in the next chapter.

Chapter 2
Foundation and Model Structure

Time series in economics and finance is not as intuitive and convincing as statistics in other areas. The main difficulties are (i) the sample space is not visible and (ii) the sample is not reproducible. For example, one shoe company asks you to look for the average foot size of Chinese in China. It is clear that the sample space is all people in China. You can first get an i.i.d. sample and calculate its average, say 38cm. So you may claim that the population mean roughly equals to 38cm. If someone doesn't believe this, he/she can get another i.i.d. sample and calculate the average. This new average should be almost 38cm as long as the sample size is large enough and the sample design is almost the same. That is, 38cm is a convincing number as the average foot size of Chinese.

We consider the HSI again. Let Y_t be the daily log-return of HSI at day t with its own probability space $(\Omega_t, \mathscr{F}_t, \mathscr{P}_t)$, and

$$\mu_t \equiv EY_t = \int_{\Omega_t} Y_t d\mathscr{P}_t.$$

On day t, we have only one observation. How to estimate μ_t? No way ! We first need the following weak stationarity assumption:

Definition 2.1. A time series $\{Y_t\}$ is called weakly stationary if
(i). $\mu_t \equiv \mu$,
(ii). $\sigma^2 \equiv E(Y_t - \mu)^2$,
(ii). $\rho_k \equiv E(Y_t - \mu)(Y_{t+k} - \mu)/\sigma^2$.

That is, the mean, variance and the linear correlation of Y_t and Y_{t+k} are time-invariant. This assumption is reasonable if the world does not change too much. Usually, we use the historical data, say $y_1, \cdots, y_n$, to estimate μ, i.e.,

$$\mu \approx \frac{1}{n}\sum_{t=1}^{n} y_t. \tag{2.1}$$

Do you believe it? In mathematics, we can prove that the previous approximation is good enough as n is large if $\sum_{k=1}^{\infty} |\rho_k| < \infty$. However, we do not have an intuitive argument for this.

2.1 Strict Stationarity and Erogicity

We first assume that

$$(\Omega_t, \mathscr{F}_t, \mathscr{P}_t) \equiv (\Omega, \mathscr{F}, \mathscr{P})$$

—time invariant, i.e., $(\Omega_t, \mathscr{F}_t, \mathscr{P}_t) \xrightarrow{T} (\Omega, \mathscr{F}, \mathscr{P})$. Furthermore, we have the following definition:

Definition 2.2. $\{Y_t\}$ is said to be a (strictly) stationary sequence if, for every k and n, $(Y_0, \cdots, Y_n)$ and $(Y_k, \cdots, Y_{k+n})$ have the same distribution.

The strict stationarity means that the time series have the same distribution under the time transformation T. This and the time invariance of the space are not strictly correct but are relatively reasonable in many cases. Recall that any r.v. X on $(\Omega, \mathscr{F}, \mathscr{P})$ transforms its probability space into a real space, i.e.,

$$(\Omega, \mathscr{F}, \mathscr{P}) \xrightarrow{X} (R, \mathscr{B}, P).$$

By Kolmogorov's extension theorem, $\tilde{X} = (Y_0, Y_1, Y_2, \cdots)$ transforms the space $(\Omega, \mathscr{F}, \mathscr{P})$ into the sample space $(R^N, \mathscr{B}^N, P)$, i.e.,

$$(\Omega, \mathscr{F}, \mathscr{P}) \xrightarrow{\tilde{X}} (R^N, \mathscr{B}^N, P) \textit{ and } \tilde{X} : w \in \Omega \longrightarrow (w_0, w_1, \cdots) \in R^N.$$

For simplicity, we assume that

$$(\Omega, \mathscr{F}, \mathscr{P}) = (R^N, \mathscr{B}^N, P).$$

Then,

$$Y_0(w) = w_0 \text{ and } Y_n(w) = w_n \text{ for } n \geq 1.$$

We now define the shift operator T to be the mapping from $\Omega = R^N$ to R^N:

$$T : w = (w_0, w_1, \cdots) \to (w_1, w_2, \cdots).$$

Thus,

$$Y_n(w) = w_n = Y_0\left[(w_n, w_{n+1}, \cdots)\right] = Y_0\left(T^n w\right).$$

The shifts T^k define operators on r.v. Y on $(\Omega, \mathscr{F}, \mathscr{P})$ by

$$(T^k Y)(w) = Y\left[T^k(w)\right].$$

Thus, it follows that

$$Y_n = TY_{n-1} = \cdots = T^{n-1}Y_0$$

and

$$(Y_0, \cdots, Y_n) \xrightarrow{T} (Y_1, \cdots, Y_{n+1}) \xrightarrow{T} \cdots \xrightarrow{T} (Y_k, \cdots, Y_{k+n}),$$

i.e., the time transformation is the shift operator in mathematics. For $\forall A \in \mathscr{F}$, let $B = T^{-1}A$. Then,

$$\begin{aligned}
P(A) &= P(w \in A) \\
&= P[(w_0, w_1, \cdots) \in A] \\
&= P[w : (Y_0, Y_1, \cdots)(w) \in A] \\
&= P[w : (Y_1, Y_2, \cdots)(w) \in A] \text{ (stationarity)} \\
&= P[w : (w_1, w_2, \cdots) \in A] \\
&= P[w : T(w_0, w_1, \cdots) \in A] \\
&= P(B).
\end{aligned}$$

So we call the shift operator a measure preserving map. Thus, the stationarity time series $\{Y_t\}$ are determined by the shift operator and Y_0.

To see why and how (2.1) is reasonable, we draw a sample for the shoe example in the following way:

you randomly and continuously walk in China and get one sample point at 6:00 pm each day.

The sample foot size x_t depends on the region where you are on the t-th day, while the region on the t-th day depends on your place on the $(t-1)$-th day. Thus, x_t and x_{t-1} are not independent. The space and the distribution of $\{x_t\}$ are roughly time-invariant under the time transformation T. After n days, we can get a sample average. Can you claim that

$$\mu \approx \frac{1}{n}\sum_{t=1}^{n} x_t \; ?$$

A condition for this is that you can walk to everywhere or you cannot limit yourself to a region or you can access to anyone in Mainland China. This condition is so-called Ergodicity. If this ergodicity holds or almost holds, another samples got from this way should have almost the same average. We now formula this in the mathematical way.

Definition 2.3. The shift operator T on $(\Omega, \mathscr{F}, \mathscr{P})$ is said to be ergodic if there is not any $A \in \mathscr{F}$ such that $T^{-1}A = A$ [or $P(T^{-1}A = A) = 1$] except for $A = \Omega$ or $\emptyset$, and $\{Y_t\}$ is called ergodic if $Y_t = TY_{t-1}$.

Theorem 2.1. *(Ergodic Theorem) Assume that $\{Y_t\}$ is a sequence of strictly stationary and ergodic time series with $E|Y_t| < \infty$. Then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} Y_t = EY_t = \int_{\Omega} Y_t dP \;\; a.s..$$

The ergodic theorem plays an essential role in the field of time series.

2.2 Criteria for Strict Stationarity

We first give one simple criterion as follows.

Theorem 2.2. *(1) If $\{Y_t\}$ be a sequence of i.i.d. r.v.s, then $\{Y_t\}$ is stationary and ergodic.*
(2) If $\{X_t\}$ is stationary and ergodic, then $Y_t = g(X_t, X_{t-1}, \cdots)$ are stationary and ergodic for any measurable function g from $R^N \to R$.

Example 2.1. Let

$$Y_t = \varepsilon_t + a_1 \varepsilon_{t-1} + \cdots + a_i \varepsilon_{t-i} + \cdots ,$$

where $\sum_{i=1}^{\infty} a_i^2 < \infty$ and ε_t is iid r.v.'s or $\varepsilon_t = f(\eta_t, \eta_{t-1}, \cdots)$ with iid $\{\eta_t\}$. Then $\{Y_t\}$ is strictly stationary and ergodic.

Other criteria need to put the time series Y_t into a Markov chain setting. A time series $\{Y_t\}$ on $(\Omega, \mathscr{F}, \mathscr{P})$ is said to be a Markov Chain with respect to the σ-filed $\mathscr{F}_t$ if $Y_n \in \mathscr{F}_n$ and for all $B \in \mathscr{F}$,

$$P(Y_{n+1} \in B | \mathscr{F}_n) = P(Y_{n+1} \in B | Y_n) \equiv P(Y_n, B)$$

where $P(x, B)$ is called the transition probability. If there is a distribution π such that

$$\pi(B) = \int P(x, B) d\pi(x), \text{ for } \forall B \in \mathscr{F},$$

then π is called the invariant distribution (measure). If a Markov Chain $\{Y_t\}$ has an invariant distribution π and $Y_0 \sim \pi$, then $\{Y_t\}$ is stationary and ergodic, see Durrett (1995, page335).

Definition 2.4. A Markov Chain $\{Y_t\}$ is called φ-irreducible if there exists a measure φ on Ω such that, for $\forall A \in \mathscr{F}^+$, we have

$$\sum_{n=1}^{\infty} P^n(x, A) > 0$$

for all $x \in \Omega$, where $\mathscr{F}^+ = \{A : \varphi(A) > 0, A \in \mathscr{F}\}$ and $P^n(x, A)$ is the n-step transition of $\{Y_t\}$.

Definition 2.5. A set A is called recurrent if

$$E\left[\sum_{n=1}^{m} I(Y_n \in A) \Big| Y_0 = x\right] \to \infty$$

as $m \to \infty$, for all $x \in A$. An φ-irreducible Markov Chain $\{Y_t\}$ is called recurrent if

$$E\left[\sum_{n=1}^{m} I(Y_n \in A)\Big|Y_0 = x\right] \to \infty$$

as $m \to \infty$, for $\forall x \in \Omega$ and $A \in \mathscr{F}^+$. Otherwise, it is called transient.

If $\{Y_t\}$ is φ-irreducible and recurrent, then there exists a unique (up to constant multipler) invariant measure π, see Page 230 of Meyn and Tweedie (1993). To make $\pi(\Omega) < \infty$ under which we can get a unique invariant probability measure after a normalization, we need another concept.

Definition 2.6. A set A is called Harris recurrent if

$$P\left(\lim_{m\to\infty}\sum_{n=1}^{m} I(Y_n \in A) = \infty\Big|Y_0 = x\right) = 1,$$

for $\forall x \in A$. A φ-irreducible chain $\{Y_t\}$ is called Harris recurrent if each set in $\mathscr{F}^+$ is Harris recurrent.

A recurrent chain $\{Y_t\}$ differs only by φ-null set from a Harris recurrent chain, In the finite state space, recurrence is equivalent to Harris recurrence. If an irreducible chain $\{Y_t\}$ has an invariant probability measure π, then it is recurrent, see Page 231 of Meyn and Tweedie (1993). Thus, the recurrence is necessary for $\{Y_t\}$ to have an invariant probability measure. For an irreducible chain $\{Y_t\}$, there exists sets $D_1, \cdots, D_d \in \mathscr{F}$ such that

$$(i) \text{ for } x \in D_i, P(x, D_{i+1}) = 1, i = 0, \cdots, d-1,$$

$$(ii) \text{ the set } N = \left[\bigcup_{i=1}^{d} D_i\right]^c \text{ is } \varphi\text{-null and transient,}$$

where d is called the period of $\{Y_t\}$, see Page 188 of Meyn and Tweedie (1993). When $d = 1$, $\{Y_t\}$ is called aperiodic. An irreducible and periodic example is as follows:

	0	1	2	3
0	0	$\times$	$\times$	0
1	0	0	0	$\times$
2	0	0	0	$\times$
3	$\times$	0	0	0

$$\begin{tikzcd} 0 \arrow[r] \arrow[dr] & 1 \arrow[dl] \\ 3 \arrow[u] & 2 \arrow[l] \end{tikzcd} \qquad \begin{tikzcd} D_1 = \{0\} \arrow[r] & D_2 = \{1,2\} \arrow[d] \\ & D_3 = \{3\} \arrow[ul] \end{tikzcd}$$

Thus, we have three sub-Markov Chain $Y_t^k = Y_{kt}, k = 1,2,3$, and each Y_t^k is aperiodic and has its own invariant probability distribution. Thus, we only focus on the irreducible and aperiodic Markov Chain.

Lemma 2.1. *A necessary and sufficient condition for an irreducible $\{Y_t\}$ to be aperiodic is that there exists an $A \in \mathscr{F}^+$ and $B \subseteq A$ with $B \in \mathscr{F}^+$,*

$$P^n(x,B) > 0 \text{ and } P^{n+1}(x,B) > 0$$

for $\forall x \in B$ and some positive integer n.

In the finite space, an irreducible chain is recurrent as long as one state is recurrent. In the general spaces, a similar idea is there. If we can find out a recurrent $A \in \mathscr{F}^+$, then we can claim that the irreducible chain is recurrent, see Page 174 of Meyn and Tweedie (1993). What kind of sets are most likely recurrent? It is small set.

Definition 2.7. A set $C \in \mathscr{F}$ is called a small set is there exists an $m > 0$, and a non-trivial measure ν_m on $\mathscr{F}$ such that, for all $x \in C$, $B \in \mathscr{F}$,

$$P^m(x,B) \geq \nu_m(B).$$

The small set means that we start in C and can reach to any set in $\mathscr{F}$ uniformly in a finite step m. It just likes the center of a country. If $\{Y_t\}$ is a φ-irreducible Feller Chain [i.e., for each bounded continuous function g on R, given x, $E\left[g(Y_t)\middle|Y_{t-1} = x\right]$ is also continuous , see Feigin and Tweedie (1985)], then any relatively compact set A is small if $\varphi(A) > 0$. If $\{Y_t\}$ is μ_m-irreducible and aperiodic and takes the following form:

$$Y_t = T(Y_{t-1}) + \rho_t, \quad T : R^m \to R^m,$$

where $Y_t \in R^m$ and $(R^m, \mathscr{B}^m, \mu_m)$ is Lebesgue measure, then any non-null compact set is small if one of the following conditions holds.

(i) $\{\rho_t\}$ is i.i.d., the marginal distribution is absolutely continuous and has a positive pdf over R^m.

(ii) $\rho_t = (\eta_t, 0, \cdots, 0)$ with η_t's i.i.d. each having an absolutely continuous distribution and positive pdf everywhere in R.

Small set is the so-called petite set and a petite set is a small set if $\{Y_t\}$ is φ-irreducible and aperiodic. To check if a small set C is recurrent, we use a function $V(x)$ to "test" it. In many cases, $V(x)$ can look at as a distance from the starting point x to the "center C". When you hold the Markov Chain to walk around in Ω, the average distance is

$$\int_\Omega P(x,dy)V(y).$$

Starting outside of C, it should be less than $V(x)$ when C is recurrent, and larger than $V(x)$ when C is transient.

We now give a little strong drift criterion for the test function $V(x)$. It not only makes sure that $\{Y_t\}$ is Harris recurrent, but also $\pi(\Omega) < \infty$.

Theorem 2.3. *Suppose that $\{Y_t\}$ is φ-irreducible and aperiodic. If there exists some small set C, some $b < \infty$, and a non-negative function V finite at some one $x_0 \in \Omega$*

such that

$$(i) \quad \int P(x,dy)V(y) \le V(x) - 1 + bI_C(x), \quad x \in \Omega.$$
$$(ii) \quad C_V(M) = \{y : V(y) \le M\} \textit{ is a small set for } \forall M > 0,$$

then $\{Y_t\}$ *has a unique invariant probability measure* π *and*

$$\sup_{A \in \mathscr{F}} |P^n(x,A) - \pi(A)| \to 0, \tag{2.2}$$

as $n \to \infty$. *If (i) is strengthen as*

$$(iii) \quad \int P(x,dy)V(y) \le (1-\beta)V(x) + bI_C(x)$$

for some constants $\beta > 0$ *and* $b < \infty$, *then*

$$\rho^n \sup_{A \in \mathscr{F}} |P^n(x,A) - \pi(A)| \to 0, \tag{2.3}$$

for some $\rho \in (0,1)$ *as* $n \to \infty$.

Under the assumptions of Theorem 2.2, $\{Y_t\}$ is Harris recurrent with an invariant probability, by Theorem 17.1.7 of Meyn and Tweedie (1993), the invariant set $\mathscr{I}$ of $\{Y_t\}$ is trivial to the probability $P(x,\cdot)$ for all $x \in \Omega$. Thus, for $\forall A \in \mathscr{I}$, we have

$$\pi(A) = \int P(x,A)\pi(dx) = 0 \text{ or } 1.$$

Thus, $\{Y_t\}$ is stationary and ergodic. In the Markov Chain literature, $\{Y_t\}$ is called to be ergodic if (2.2) holds, which is equivalent to the ergodic in Theorem 2.1, and is called geometrically ergodic if(2.3) holds. If $\{Y_t\}$ is geometrically ergodic, then it is strongly mixing with the rate of convergence ρ^n.

2.3 ARMA Models

The process $\{Y_t\}$ is said to follow the general autoregressive [AR(p)] model if it satisfies the following equation:

$$\phi(B)Y_t = w_t, \tag{2.4}$$

where $\phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$ and $\{w_t\}$ is strictly stationary and ergodic.

Theorem 2.4. *Assume* $\{Y_t\}$ *is generated by model (2.4) with* $E|w_t|^\gamma < \infty$ *for some* $\gamma > 0$. *Then,* Y_t *has a stationary representation if and only if all the roots of* $\phi(z) = 0$ *lie outside the unit circle. Furthermore,* $\{Y_t\}$ *is unique and ergodic with* $E|Y_t|^\gamma < \infty$ *and its representation is:*

$$Y_t = \sum_{i=0}^{\infty} a_i w_{t-i},$$

where $a_i = O(\rho^i)$ *with* $\rho \in (0,1)$.

Proof. We first consider the case with $p = 1$. Note that

$$\begin{aligned} Y_t &= \phi^2 Y_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\ &= \cdots = \phi^m Y_{t-m} + \sum_{i=0}^{m-1} \phi^i \varepsilon_{t-i}. \end{aligned} \tag{2.5}$$

Let $S_{mt} = \sum_{i=0}^{m-1} \phi^i \varepsilon_{t-i}$. Then, for $\forall m, n > 0$,

$$E\left|S_{m+n,t} - S_{mt}\right|^\gamma = E\left|\sum_{i=m}^{m+n-1} \phi^i \varepsilon_{t-i}\right|^\gamma \le \left(\sum_{i=m}^{m+n-1} |\phi|^{i\gamma}\right) E|\varepsilon_t|^\gamma = O(\rho^n),$$

for some $\rho \in (0,1)$. By Cauchy criterion, we can show that

$$S_{mt} \to S_{\infty t}, \text{ a.s. and in } L^\gamma, \text{ as } m \to \infty.$$

It is easy to see that $S_{\infty t} = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}$ satisfies model (2.4), and it is stationary and ergodic by Theorem 2.1. Let $Y_{t-m} = S_{t-m,\infty}$ in (2.5). Then Y_t has the representation:

$$Y_t = \sum_{i=0}^{\infty} a_i w_{t-i},$$

where $a_i = O(\rho^i)$ with $\rho \in (0,1)$. To see the uniqueness, we suppose that there is another stationary $\{Y_t'\}$ to model (2.4) with $E|Y_t'|^\gamma < \infty$. Then,

$$\Delta Y_t \equiv Y_t - Y_t' = \phi(Y_{t-1} - Y_{t-1}') = \cdots = \phi^m (Y_{t-m} - Y_{t-m}').$$

Thus,

$$E|\Delta Y_t|^\gamma \le |\phi|^{\gamma m} \left(E|Y_t|^\gamma + E|Y_t'|^\gamma\right) \to 0,$$

as $m \to \infty$. Hence, $\Delta Y_t = 0$ a.s..

When $\phi = 1$, $\limsup_{t\to\infty} Y_t = Y_0 + \limsup_{t\to\infty} \sum_{i=1}^t w_i$ is invariant in terms of the invariant preserve measure T of $\{w_t\}$ and hence it has to be a constant in $[-\infty, \infty]$. Thus, we can show that $\limsup_{t\to\infty} Y_t$ is ∞ or $-\infty$ and hence $\{Y_t\}$ does not have a stationary representation. When $|\phi| > 1$, we have

$$Y_t = \sum_{i=0}^{t-1} \phi^i \varepsilon_{t-i} + \phi^t Y_0 \to \infty \; a.s.. \; when \; t \to \infty.$$

$\{Y_t\}$ does not have a stationary representation either.

When $p > 1$, we can rewrite model (2.4) in the vector form:

$$X_t = AX_{t-1} + \tilde{w}_t,$$

where $X_t = (Y_t, \cdots, Y_{t-p+1})'$, $\tilde{w}_t = (w_t, 0, \cdots, 0)'$ and

$$A = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

Using a similar method as for the case with $p = 1$, we can show that $Y_t\}$ has a unique stationary representation when $\rho(A) < 1$, which is equivalent to that all the root of $\phi(z) = 0$ lie outside the unit circle.

When there are some roots of $\phi(z) = 0$ not lying outside the unit circle, we can have the decomposition of $\phi(z)$:

$$\phi(z) = \phi^*(z)(1-z)^a(1+z)^b \prod_{k=1}^{l}(1 - 2\cos\theta_k z + z^2)^{d_k},$$

where a, b, l and d_k are nonnegative integers, $\theta_k \in (0,\ \pi)$, $\phi^*(z) = 1 - \sum_{i=1}^{p^*} \phi_i^* z^i$ with all roots outside the unit circle and $p^* = p - (a + b + 2d_1 + \cdots + 2d_l)$. Let

$$\phi^*(B)Y_t^* = w_t.$$

Then

$$(1-B)^a(1+B)^b \prod_{k=1}^{l}(1 - 2\cos\theta_k B + B^2)^{d_k} Y_t = Y_t^*.$$

By the sufficiency of this Theorem, Y_t^* has s stationary and ergodic representation. Using a similar method as for the case with $p = 1$, we can show that $\limsup_{t\to\infty} Y_t$ converges to $+\infty$ or $-\infty$. Thus, the necessity holds. This completes the proof.

The figure 2.1 below gives the sample path of AR(1) model with $\phi = 0.9$, 1.0 and 1.1 when $\{w_t\} \sim$ iid $N(0,1)$. If we are allowed to use the future information, we can have a stationary representation of Y_t in model (2.4) when $|\phi| > 1$ and $p = 1$. In fact, in this case, we ave

$$Y_t = \phi^{-1} Y_{t+1} - \phi^{-1}\varepsilon_{t+1} = \cdots = -\sum_{i=1}^{\infty} \phi^{-i}\varepsilon_{t+i}\ a.s..$$

When $w_t = \mu + \psi(B)\varepsilon_t$, that is,

$$\Phi(B)Y_t = \mu + \psi(B)\varepsilon_t, \tag{2.6}$$

the process $\{Y_t\}$ is said to follow the autoregressive moving-average [ARMA(p, q)] model, where $\psi(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ and $\{\varepsilon_t\}$ is strictly stationary time series. By Theorem 2.4, $\{Y_t\}$ is strictly stationary and ergodic if all the roots of $\phi(z) = 0$ lie outside the unit circle with $E|\varepsilon_t|^\gamma < \infty$. Furthermore, if all the roots of $\psi(z) = 0$

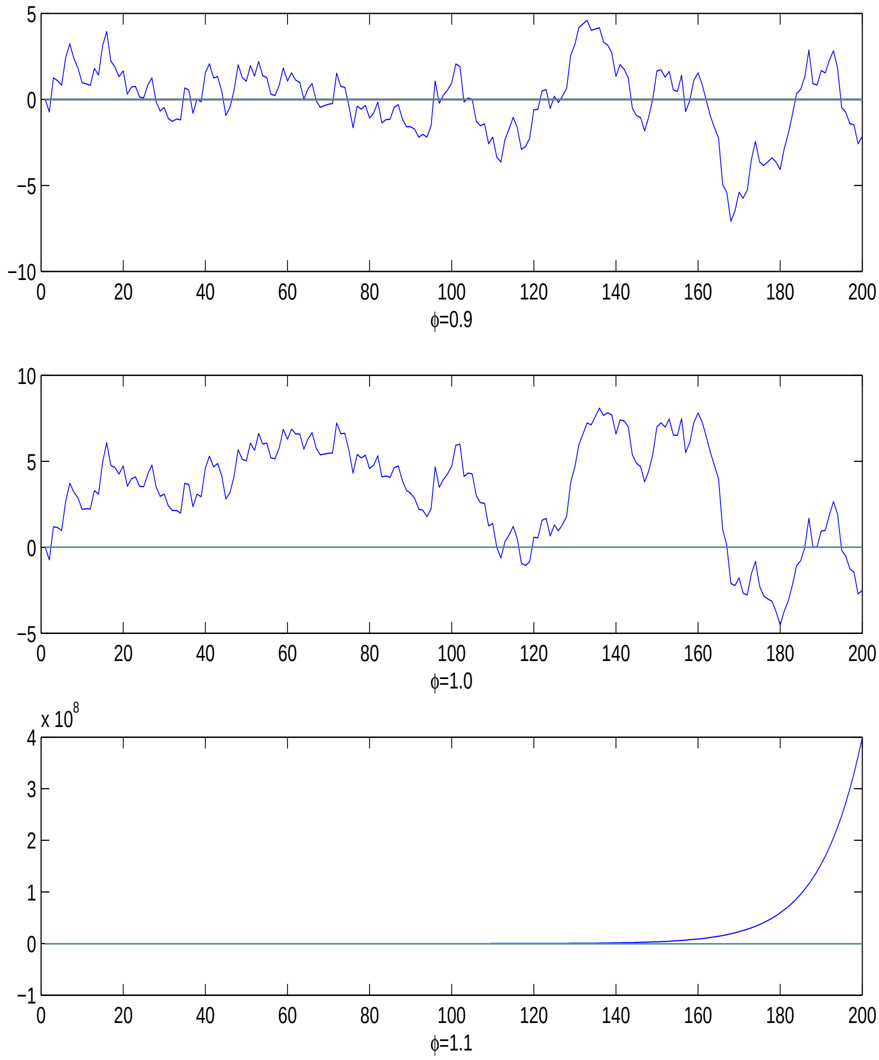


Fig. 2.1 Time series generated from a AR(1) model $Y_t = \phi Y_{t-1} + \varepsilon_t$, where $\phi = 0.9$ (top), 1.0 (middle) and 1.1 (bottom), respectively.

lie outside the unit circle, then we have the representation:

$$\varepsilon_t = \mu\psi^{-1}(1) + \sum_{i=0}^{\infty} b_i Y_{t-i},$$

where $b(i) = O(\rho^i)$ with $\rho \in (0,1)$. In this case, we say that ARMA model (2.6) is invertible. For model (2.6), we can show that $\gamma_k = O(\rho^k)$ with $\rho \in (0,1)$, that is, the dependence of $\{Y_t\}$ decays to zero, exponentially.

When $(1-B)^d w_t = \psi(B)\varepsilon_t$, that is,

$$\Phi(B)(1-B)^d Y_t = \psi(B)\varepsilon_t, \tag{2.7}$$

the process $\{Y_t\}$ is said to follow the fractionally ARMA [FARIMA(p, d, q)] model, where $\psi(z)$ and $\{\varepsilon_t\}$ is defined as model (2.6), $d \in (-1/2, 1/2)$ and

$$(1-z)^d = \sum_{i=0}^{\infty} a_{di} z^i \ \textit{and}\ a_{di} = \frac{(i+d-1)!}{i!(d-1)!}. \tag{2.8}$$

By Theorem 2.4, $\{Y_t\}$ is strictly stationary and ergodic if all the roots of $\phi(z) = 0$ lie outside the unit circle with $E\varepsilon_t^2 < \infty$ and model (2.7) is invertible if all the roots of $\psi(z) = 0$ lie outside the unit circle. Furthermore, the representations holds:

$$\begin{aligned} Y_t &= \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}, \\ \varepsilon_t &= \sum_{i=0}^{\infty} b_i Y_{t-i}, \end{aligned}$$

where $a_i = O(i^{d-1})$ and $b_i = O(i^{-d-1})$ since $a_{dk} = O(i^{d-1})$ by Stirling's formula. For model (2.7), we can show that $\gamma_k = O(k^{2d-1})$. Thus, the dependence of $\{Y_t\}$ decays at a much slower rate than that of $\{Y_t\}$ in model (2.6). In particular, when $d \in (0, 1/2)$, $\{Y_t\}$ from model (2.7) is called to be a long memory time series.

2.4 GARCH Models

We consider the first-order autoregressive heteroscedastic [GARCH(1,1)] model:

$$\varepsilon_t = \eta_t \sqrt{h_t} \ \textit{and}\ h_t = \alpha_0 + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{2.9}$$

where η_t is a sequence of iid r.v.s with $E\eta_t = 0$.

Theorem 2.5. *Assume $\{\varepsilon_t\}$ is generated by model (2.9). Then, ε_t has a stationary representation if and only if*

$$E \ln|\alpha\eta_t^2 + \beta| < 0. \tag{2.10}$$

Furthermore, $\{\varepsilon_t\}$ *is unique and ergodic with* $E|\varepsilon_t|^\gamma < \infty$ *for some* $\gamma > 0$ *and* h_t *has representation:*

$$h_t = \alpha_0 \left[1 + \sum_{j=1}^{\infty} \prod_{i=1}^{j} \left(\alpha \eta_{t-i}^2 + \beta \right) \right].$$

Proof. First, if (2.10) holds, then there exists a $\gamma > 0$ such that $\rho \equiv E|\alpha\eta_t^2 + \beta|^\gamma < 1$. We now write h_t in (2.9) as

$$\begin{aligned} h_t &= \alpha_0 + \left(\alpha \eta_{t-1}^2 + \beta \right) h_{t-1} \\ &= \alpha_0 + \alpha_0 \left(\alpha \eta_{t-1}^2 + \beta \right) + \left(\alpha \eta_{t-1}^2 + \beta \right) \left(\alpha \eta_{t-2}^2 + \beta \right) h_{t-2} \\ &= \cdots \\ &= \alpha_0 \left[1 + \sum_{j=1}^{m} \prod_{i=1}^{j} \left(\alpha \eta_{t-i}^2 + \beta \right) \right] + \prod_{i=1}^{m+1} \left(\alpha \eta_{t-i}^2 + \beta \right) h_{t-m-1}. \end{aligned}$$

Let

$$S_{mt} = \alpha_0 \left[1 + \sum_{j=1}^{m} \prod_{i=1}^{j} \left(\alpha \eta_{t-i}^2 + \beta \right) \right].$$

For any $m, n > 0$, we have

$$\begin{aligned} E\left| S_{m+n,t} - S_{mt} \right|^\gamma &= \alpha_0^\gamma E \left| \sum_{j=m+1}^{m+n} \prod_{i=1}^{j} \left(\alpha \eta_{t-i}^2 + \beta \right) \right|^\gamma \\ &\le \alpha_0^\gamma \sum_{j=m+1}^{m+n} \prod_{i=1}^{j} E \left| \alpha \eta_t^2 + \beta \right|^\gamma = O(\rho^n). \end{aligned}$$

Thus, by Cauchy criterion, we can show that

$$S_{mt} \to S_{\infty t} \equiv h_t, \text{ a.s. and in } L^\gamma, \text{ as } m \to \infty.$$

Let $h_{t-m-1} = S_{\infty, t-m-1}$. Then h_t has the stationary presentation in Theorem 2.5 and is ergodic with $Eh_t^\gamma < \infty$. To see the uniqueness, suppose that we have another stationary solution h_t^* to model (2.9). Then, $\Delta h_t \equiv h_t - h_t^* = (\alpha\eta_{t-1}^2 + \beta)\Delta h_{t-1} = \cdots = \prod_{i=1}^m (\alpha\eta_{t-i}^2 + \beta)\Delta h_{t-m}$ and

$$E\left| \Delta h_t \right|^\gamma \le \prod_{i=1}^{m} E \left| \alpha \eta_t^2 + \beta \right|^\gamma E \left| \Delta h_{t-m} \right|^\gamma \le C\rho^m \to 0, \text{ as } m \to \infty.$$

Thus, $\Delta h_t = 0$ a.s., i.e., $h_t = h_t^*$ a.s.. When $E \ln |\alpha\eta_t^2 + \beta| \ge 0$, we have

$$h_t \geq \alpha_0 \max_{1\leq j\leq m} \prod_{i=1}^{j} \left(\alpha\eta_{t-i}^2 + \beta\right)$$
$$= \alpha_0 e^{\max_{1\leq j\leq m} \sum_{i=1}^{j} \ln\left(\alpha\eta_{t-i}^2+\beta\right)} \to \infty, \text{ a.s., as } m \to \infty.$$

Thus, necessity holds. This completes the proof.

From Theorem 2.5, we can see that the moment of ε_t is highly liked to the parameter (α, β). The strict stationarity region and the regions for $E|\varepsilon_t|^{2\iota} < \infty$ are given in Figure 4 when $\eta_t \sim N(0,1)$.

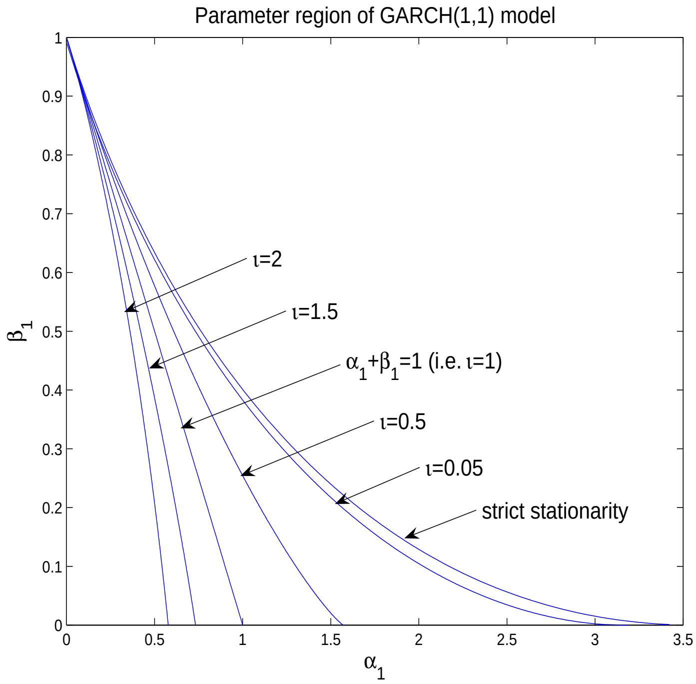


Fig. 2.2 The regions bounded by the indicated curves are for the strict stationarity and for $E|\varepsilon_t|^{2\iota} < \infty$ with $\iota = 0.05, 0.5, 1, 1.5$ and 2, respectively.

We now consider the GARCH (p, q) model:

$$\varepsilon_t = \eta_t \sqrt{h_t} \text{ and } h_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{s} \beta_i h_{t-i}, \tag{2.11}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$ $(i = 1, \cdots, r)$, $\beta_j \geq 0$ $(j = 1, \cdots, s)$, and η_t is deinfed as in (2.9). We can rewrite h_t in a vector form:

$$H_t = \alpha_0(\zeta_t + A_{t-1}\zeta_{t-1}),$$

where $\zeta_t = (\eta_t^2, 0, \cdots, 0, 1, \cdots, 0)'_{(r+s)\times 1}$ with the first component η_t^2 and the $(r+1)$th component 1, and

$$A_t = \left(\begin{array}{ccc|ccc} \alpha_1\eta_t^2 & \cdots & \alpha_r\eta_t^2 & \beta_1\eta_t^2 & \cdots & \beta_s\eta_t^2 \\ & I_{r-1} & O & & O & \\ \hline \alpha_1 & \cdots & \alpha_r & \beta_1 & \cdots & \beta_s \\ & O & & & I_{s-1} & O \end{array}\right),$$

Thus, $h_t = u_1' H_t$, where $u_i = (0, \cdots, 0, 1, \cdots, 0)'_{(r+s)\times 1}$ with the ith component 1.

Theorem 2.6. *Assume that $\{\varepsilon_t\}$ is generated by model (2.11). Then (i) If*

$$\text{there exists an integer } i_0 \text{ such that } E\|\prod_{k=0}^{i_0-1} A_k\|^{\iota} < 1, \tag{2.12}$$

for $\iota \in (0,1]$, then $\{\varepsilon_t\}$ is strictly stationary and ergodic with $E|\varepsilon_t|^{2\iota} < \infty$;

(ii) If $\tilde{\beta} \equiv \min\{\alpha_i, \beta_j : i = 0, 1, \cdots, r, j = 1, \cdots, s\} > 0$, and $\{\varepsilon_t\}$ is strictly stationary with $E|\varepsilon_t|^{2\iota} < \infty$, then (2.7) holds;

(iv) if $\tilde{\beta} > 0$,

$$\sum_{i=1}^{r} \alpha_i + \sum_{j=1}^{s} \beta_j = 1, \tag{2.13}$$

and η_t has a positive density on R such that $E|\eta_t|^{\tau} < \infty$ for all $\tau < \tau_0$ and $E|\eta_t|^{\tau_0} = \infty$ for some $\tau_0 \in (0, \infty]$, then $\lim_{x\to\infty} x^2 P(|\varepsilon_t| > x)$ exists and is positive,

(v). the necessary and sufficient condition for $\varepsilon_t^2 < \infty$ is

$$\sum_{i=1}^{r} \alpha_i + \sum_{i=1}^{s} \beta_i < 1. \tag{2.14}$$

Proof. The proof of stationarity and ergodicity is similar to that for Theorem 2.5 and henc eomitted. Further, let $B_t = \zeta_t + \sum_{j=1}^{i_0-1} \prod_{i=0}^{j-1} A_{t-i}\zeta_{t-j}$ and $\tilde{A}_t = \prod_{i=0}^{i_0-1} A_{t-i}$. We rewrite h_t as

$$h_t = \alpha_0 \Big[u'_{r+1} B_t + \sum_{k=1}^{\infty} u'_{r+1} \prod_{i_1=0}^{k-1} \tilde{A}_{t-i_0 i_1} B_{t-k i_0}\Big], \tag{2.15}$$

where $u'_{r+1}\zeta_t = 1$ is used. By (2.15), it follows that

$$Eh_t^{\iota} \le O(1) + O(1) \sum_{k=1}^{\infty} (E\|\tilde{A}_t\|^{\iota})^k < \infty.$$

Thus, $E|\varepsilon_t|^{2\iota} < \infty$ and hence (i) holds.

Denote $h_{Jt} = \alpha_{00}(1+\sum_{j=1}^{J} u'_{r+1}\prod_{i=0}^{j-1} A_{t-i}\xi_{t-j})$. Then $(h_{Jt}-h_{J-1,t})^{\iota} \to 0$ a.s. when $J \to \infty$. $\{h_{Jt}^{\iota}\}$ is an increasing sequence in terms of J and $E\sup_J h_{Jt}^{\iota} \le Eh_t^{\iota} < \infty$. Thus, by the dominated convergence theorem, we have $E(u'_{r+1}\prod_{i=0}^{J-1} A_{t-i}\zeta_{t-j})^{\iota} = E(h_{Jt}-h_{J-1,t})^{\iota}/\alpha_{00}^{\iota} \to 0$ as $J \to \infty$. Since $\tilde{\beta} > 0$, there exist J_1 and J_2 such that all the elements of $d_{1t} \equiv (u'_{r+1}\prod_{i=0}^{J_1-1} A_{t-i})'$ and $d_{2t} \equiv \prod_{i=0}^{J_2-1} A_{t-i}\zeta_{t-j}$ are a.s. positive, and $0 < Ed_{1it}^{\iota} < \infty$ and $0 < Ed_{2it}^{\iota} < \infty$, where d_{jit} is the i-th element of d_{jt} as $j = 1,2$. Note that

$$E\left(u'_{r+1}\prod_{i=0}^{J_1+i_0+J_2-1} A_{t-i}\zeta_{t-j}\right)^{\iota} = E\left[d'_{1t}(\prod_{i=J_1}^{J_1+i_0-1} A_{t-i})d_{2,J_1+i_0}\right]^{\iota} \to 0,$$

as $i_0 \to \infty$. Let a_{ijt} be the (i,j)-element of $\prod_{i=J_1}^{J_1+i_0-1} A_{t-i}$. From the preceding equation, we know that $E(d_{1it}a_{ijt}d_{2j,J_1+i_0})^{\iota} \to 0$ as $i_0 \to \infty$. Since d_{1it}, a_{ijt}, and d_{2j,J_1+i_0} are independent, we know that $Ea_{ijt}^{\iota} \to 0$ as $i_0 \to \infty$. Thus, there exists i_0 such that $E\|\prod_{i=J_1}^{J_1+i_0-1} A_{t-i}\|^{\iota} = E\|\prod_{i=0}^{i_0-1} A_{t-i}\|^{\iota} < 1$, i.e., (ii) holds.

(iii). Since (2.12) is the necessary and sufficient for $E\varepsilon_t^2 < \infty$, (2.14) implies $E\varepsilon_t^2 = \infty$. By (ii) of this theorem, $E\|\prod_{k=0}^{i_0-1} A_k\| \ge 1$ for any $i_0 \ge 1$. Thus,

$$\ln E\|\prod_{k=1}^{n} A_k\| \ge 0,$$

for all $n \ge 1$. Note that $u'_i\prod_{k=1}^{n} A_k u_j$ is the (i,j)th element of $\prod_{k=1}^{n} A_k$. We have

$$\begin{aligned}
E\|\prod_{k=1}^{n} A_k\| &= E[\sum_{i=1}^{r+s}\sum_{j=1}^{r+s}(u'_i\prod_{k=1}^{n} A_k u_j)^2]^{1/2} \\
&\le E[\sum_{i=1}^{r+s}\sum_{j=1}^{r+s}(u'_i\prod_{k=1}^{n} A_k u_j)] = \sum_{i=1}^{r+s}\sum_{j=1}^{r+s}(u'_i A^n u_j) \le (r+s)^2,
\end{aligned}$$

where $A = EA_t$. By the previous two inequalities, it follows that

$$\lim_{n\to\infty}\frac{1}{n}\ln E\|A_1\cdots A_n\| = 0.$$

By Theorems 2.4 and 3.1 (B) in Basrak, Davis and Mikosch (2002), the conclusion (iii) holds. (v) is in Exercise. This completes the proof. □

We can show that (2.12) and (2.13) are equivalent when $\iota = 1$. Under (2.14), model (2.11) is the IGARCH model with an infinite variance. (iii) implies that the tail index of the IGARCH(r, s) process is always 2.

2.5 Double AR Models

The process $\{Y_t\}$ is called to follow the double AR(1) model if it satisfies the equation:

$$Y_t = \phi Y_{t-1} + \eta_t \sqrt{w + \alpha Y_{t-1}^2}, \tag{2.16}$$

where $\phi \in R$, w and $\alpha > 0$ and $\{\eta_t\}$ are iid r.v.s with mean zero and variance 1.

Theorem 2.7. *Assume that η_t has a symmetric, continuous and positive density function $f(x)$ on R. Then, if and only if*

$$E \ln |\phi + \sqrt{\alpha}\eta_t| < 0, \tag{2.17}$$

$\{Y_t\}$ is geometric ergodic and hence has a unique stationary distribution and is strongly mixing with geometric rate of convergence.

Proof. It is easy to see that Y_t is a Markov Chain on $(R, \mathscr{B}, \nu)$ and it is a ν-irreducible Feller Chain since the transition probability function

$$P(x,A) = \int_A \frac{1}{\sqrt{w+\alpha x^2}} f\left(\frac{z-\phi x}{\sqrt{w+\alpha x^2}}\right) dz, \text{ for } A \in \mathscr{B},$$

is strictly positive and continuous, where ν is the Lebesgue measure on $(R, \mathscr{B})$. This implies that the set $[-M, M]$ is small for $\forall M > 0$. Thus, we only need to check the drift condition.

Let $h(u) = E\left|\phi + \sqrt{\alpha}\eta_t\right|^u$. Since $h'(0) < 0$, $\exists k > 0$ such that $h(u) < 1$ for every $u \in (0, k)$. Let $\delta \in (0, \min(k, 2))$ and $g(x) = 1 + |x|^\delta$. Since $f(x)$ is symmetric, we assume that $\phi > 0$ without loss of generality. Note that

$$\begin{aligned} E\left[g(Y_t)\middle|Y_{t-1} = x\right] &= 1 + E|\phi x + \eta_t \sqrt{w + \alpha x^2}|^\delta \\ &= 1 + \left(\frac{w}{\alpha} + x^2\right)^{\frac{\delta}{2}} E\left|\frac{\phi x}{\sqrt{\frac{w}{\alpha} + x^2}} + \sqrt{\alpha}\eta_t\right|^\delta. \end{aligned}$$

Since

$$\frac{\left(\frac{w}{\alpha} + x^2\right)^{\frac{\delta}{2}} E\left|\frac{\phi x}{\sqrt{w/\alpha + x^2}} + \sqrt{\alpha}\eta_t\right|^\delta}{|x|^\delta E\left|\phi + \sqrt{\alpha}\eta_t\right|^\delta} \to 1, \text{ as } x \to \infty,$$

we have

$$\begin{aligned} E\left[g(Y_t)\middle|Y_{t-1} = x\right] &= 1 + [1 + o(1)]|x|^\delta E\left|\phi + \sqrt{\alpha}\eta_t\right|^\delta \\ &= g(x) + |x|^\delta \left\{[1 + o(1)]E\left|\phi + \sqrt{\alpha}\eta_t\right|^\delta - 1\right\} \\ &\le g(x) + |x|^\delta (2C - 1), \end{aligned}$$

as $x \geq M$ and $0 < 2C < 1$. Thus,

$$\begin{aligned} E\left[g(Y_t)\middle|Y_{t-1} = x\right] &\leq g(x) - g(x)\frac{|x|^\delta(1-2C)}{1+|x|^\delta} \\ &\leq g(x) - \frac{g(x)(1-2C)}{2} \\ &\equiv (1-C_0)g(x), \end{aligned}$$

since $|x|^\delta/(1+|x|^\delta) \to 1$. When $|x| \leq M$, it is easy to see that

$$E\left[g(Y_t)\middle|Y_{t-1} = x\right] \leq C_1.$$

Thus, we have

$$E\left[g(Y_t)\middle|Y_{t-1} = x\right] \leq (1-C_0)g(x) + C_1 I_B(x),$$

where $B = [-M, M]$. Now the sufficiency follows from Theorem 2.3. We refer the proof of necessity to Borkovec and Kluppelberg (1998). This completes the proof.

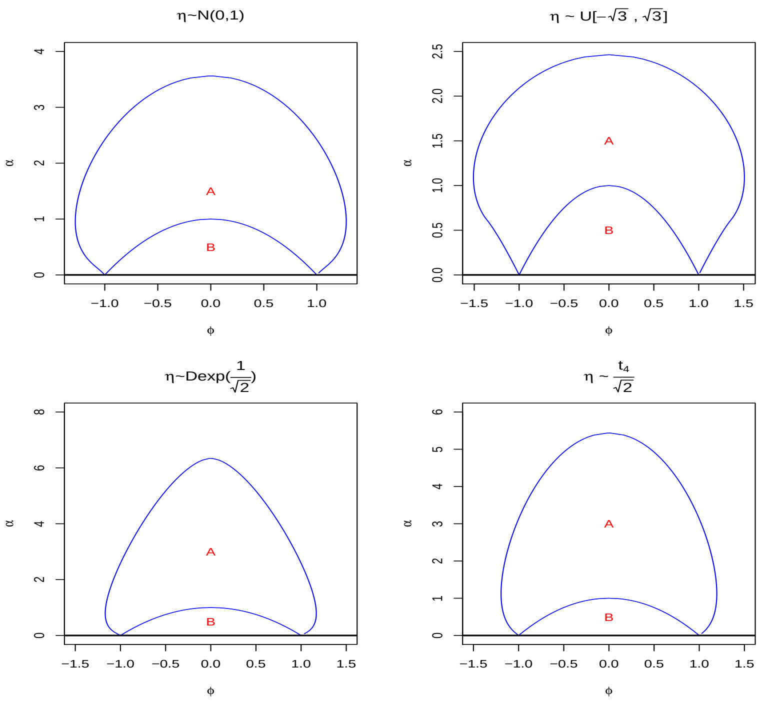


Fig. 2.3 The regions bounded by the indicated curves are $\{(\phi, \alpha)\}$ such that (2.17) holds.

We now consider the DAR(p) model:

$$Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + \eta_t \sqrt{\omega + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2}, \tag{2.18}$$

where $\omega, \alpha_i > 0$, and $\{\eta_t\}$ is defined as in model (2.16). Model (2.18) is a special case of the following general model:

$$Y_t = f(Y_{t-1}, \cdots, Y_{t-p}) + \eta_t \sqrt{h(Y_{t-1}, \cdots, Y_{t-p})},$$

where $h(Y_{t-1}, \cdots, Y_{t-p}) > 0$. Under this framework, the stationarity and the moment conditions have been extensively studied in the literature, e.g. Ango Nze (1992), Masry and Tjϕstheim (1995), Lu (1998), Cline and Pu (1999) and Lu and Jiang (2001). They obtained sufficient conditions for stationarity by directly checking the regular conditions in Theorem 2.3. For model (2.18), the weakest sufficient conditions for stationarity is

$$\sum_{i=1}^{p} |\phi_i| + E|\eta_t| \sum_{i=1}^{p} \alpha_i < 1,$$

which is also sufficient $E|y_t| < \infty$. Comparing this condition with that in Theorem 2.7, we can see that it is far away from the necessary condition. The necessary and sufficient condition for the stationarity of the higher DAR model is quite unclear up to date.

However, when $\eta_t \sim N(0,1)$, we can explore it in an undirect way. We first let $\xi_t = (\xi_{1t}, \cdots, \xi_{pt})$ be an independent $p \times 1$ standard normal vector sequence and independent of $\{\eta_t\}$ and let A_t be the $p \times p$ random matrix:

$$A_t = \begin{pmatrix} \phi_1 + \sqrt{\alpha_1}\xi_{1t} & \cdots & \phi_{p-1} + \sqrt{\alpha_{p-1}}\xi_{p-1,t} & \phi_p + \sqrt{\alpha_p}\xi_{pt} \\ & I_{p-1} & & O_{(p-1)\times 1} \end{pmatrix},$$

where I_r is the $r \times r$ identity matrix and $O_{r\times s}$ is the $r \times s$ zero matrix. We define the top Lyapounov exponen as

$$\gamma = \inf\{\frac{1}{n} E \ln \|A_1 \cdots A_n\|,\ n \geq 1\}. \tag{2.19}$$

Thus, we have the following theorem:

Theorem 2.8. *The necessary and sufficient condition for the existence of a strictly stationary solution to model (2.18) is $\gamma < 0$. Furthermore, the solution $\{Y_t : t \in \mathcal{N}\}$ is unique, geometrically ergodic and $E|Y_t|^u < \infty$ for some $u > 0$.*

Proof. Let $X_t = (Y_t, Y_{t-1}, \cdots, Y_{t-p+1})'$, $\mathscr{B}_p$ be the class of Borel sets of R^p and let ν_p be the Lebesgue measure on $(R^p, \mathscr{B}^p)$. Then $(R^p, \mathscr{B}^p, \nu_p)$ is the state space of the process $\{X_t\}$. Let $m : R^p \to R$ be the project map onto the first coordinate, i.e., $m(x) = x_1$ as $x \in R^p$. Then, $\{X_t\}$ is a homogeneous Markov chain with state space $(R^p, \mathscr{B}^p, \nu_p)$. It has the transition probability

$$P(x,A) = \int_{m(A)} \frac{1}{\sqrt{\lambda_2'\tilde{x}}} f\Big(\frac{z_1 - \lambda_1' x}{\sqrt{\lambda_2'\tilde{x}}}\Big) dz_1,\ x \in R^p \text{ and } A \in \mathscr{B}^p, \tag{2.20}$$

where $x = (x_p, \cdots, x_1)$, $\tilde{x} = (1, x_p^2, \cdots, x_1^2)$, and $f(x) = (2\pi)^{-0.5} e^{-x^2/2}$.

We first check that $\{X_t\}$ is ν_p-irreducibile, i.e., if $\sum_{n=1}^{\infty} P^n(x,A) > 0$ for every $x \in R^p$ whenever $\nu_p(A) > 0$, where

$$P^n(x,A) = \int_{R^p} P^{n-1}(y,A) P(x,dy),\ x \in R^p, A \in \mathscr{B}^p.$$

It is easy to see that the $p-$step transition probability of the Markov chain $\{Y_t\}$ is

$$P^p(x,A) = \int_A \prod_{i=1}^{p} \frac{1}{\sqrt{\lambda_2'\tilde{x}_{2i}}} f\Big(\frac{z_i - \lambda_1'\tilde{x}_i}{\sqrt{\lambda_2'\tilde{x}_{2i}}}\Big) dz_1 \cdots dz_p, \tag{2.21}$$

where $\tilde{x}_i = (z_i, \cdots, z_1, x_1, \cdots, x_{p-i})$ and $\tilde{X}_{2i} = (1, z_i^2, \cdots, z_1^2, x_1^2, \cdots, x_{p-i}^2)$. Since the transition density kernel in (2.22) is positive, we know that $\{X_t\}$ is ν_p-irreducible.

Let $\tilde{A}_t = A_t \cdots A_{t-s+1}$. By Exercise 2.2, there exists an integer, s, such that

$$E\|\tilde{A}_t\|^u < 1. \tag{2.22}$$

Using (2.22), we next prove that the $s-$step Markov chain $\{X_{ts}\}$ satisfies the drift condition of Theorem 4(ii) in Tweedie (1983), i.e., there exists a compact set K and a non-negative continuous function $g(x)$ such that $\nu_p(K) > 0$, $g(x) \geq 1$ on K, and

$$E(g(Y_{st})|Y_{(t-1)s} = x) \leq (1-\varepsilon) g(x),\ x \in K^c, \tag{2.23}$$

$$E(g(Y_{st})|Y_{(t-1)s} = x) \leq M,\ x \in K, \tag{2.24}$$

for some $\varepsilon > 0$. The key point is to find a function, g, such that (A.4)-(A.5) hold.

It is difficult to get g by a direct method. We first consider the RCAR(p) model:

$$\tilde{Y}_t = A_t \tilde{Y}_{t-1} + \tilde{\eta}_t, \tag{2.25}$$

where $\tilde{\eta}_t = (\sqrt{\omega}\eta_t, 0, \cdots, 0)'$ and $\tilde{Y}_t$ is independent of $\{\tilde{\eta}_{t'} : t' < t\}$. It is easy to see that $\{\tilde{Y}_t\}$ is a homogeneous Markov chain with state space $(R^p, \mathscr{B}^p, \nu_p)$ and its transition probability is

$$P(x,A) = \int_{m(A)} \frac{1}{\sqrt{\lambda_2'\tilde{x}}} f\Big(\frac{z_1 - \lambda_1' x}{\sqrt{\lambda_2'\tilde{x}}}\Big) dz_1,\ x \in R^p \textit{ and } A \in \mathscr{B}^p. \tag{2.26}$$

We choose $g(x) = 1 + \|x\|^u$, where $x \in R^p$ and u is defined as in (A.3). For a fixed s such that (A.3) holds, we iterate (A.6) to obtain the following expansion:

$$\tilde{Y}_{ts} = \Big(\tilde{\eta}_{ts} + \sum_{j=1}^{s-1} \prod_{r=0}^{j-1} A_{ts-r} \tilde{\eta}_{ts-j}\Big) + \tilde{A}_{ts} \tilde{Y}_{(t-1)s}. \tag{2.27}$$

By (A.8), we have

$$\begin{aligned} E(g(\tilde{Y}_{st})|\tilde{Y}_{(t-1)s}=x) &\leq 1+E\|\tilde{\eta}_{ts}+\sum_{j=1}^{s-1}\prod_{r=0}^{j-1}A_{ts-r}\tilde{\eta}_{ts-j}\|^u+E\|\tilde{A}_{ts}\|^u\|x\|^u \\ &= E\|\tilde{A}_{ts}\|^u\|x\|^u+C, \end{aligned} \tag{2.28}$$

where C is some constant. Let $K=\{x:\|x\|\leq L\}$ and L be a positive constant. It is easy to see that

$$E(g(\tilde{Y}_{st})|\tilde{Y}_{(t-1)s}=x)\leq M,\ x\in K, \tag{2.29}$$

for some constant M. Note that $E\|\tilde{A}_{ts}\|^u=E\|\tilde{A}_t\|^u$. As L is large enough and $x\in K^c$, by (A.3), there exists $\varepsilon>0$ such that

$$\begin{aligned} E(g(\tilde{Y}_{st})|\tilde{Y}_{(t-1)s}=x) &= g(x)+(E\|\tilde{A}_{ts}\|^u-1)\|x\|^u+C-1 \\ &\leq g(x)\{1-[(1-E\|\tilde{A}_{ts}\|^u)+\frac{C-1}{1+\|x\|^u}]\} \\ &\leq (1-\varepsilon)g(x). \end{aligned} \tag{2.30}$$

From (A.1) and (A.7), we know that $\{Y_t\}$ and $\{\tilde{Y}_t\}$ have the same transition kernel density. By (A.10)-(A.11), we know that (A.4)-(A.5) holds with the same $g(x)$ and K. For each bounded continuous function G on R^p, $E[G(Y_{st})|Y_{s(t-1)}=x]$ is continuous in x. Thus, $\{Y_{st}\}$ is a Feller chain. Furthermore, since Y_{st} is ν_p-irreducible, by Theorems 1-2 in Feigin and Tweedie (1985), we know that (i) Y_{st} is geometrically ergodic which ensures that there exists a unique stationary distribution π for $\{Y_{st}\}$, and (ii)

$$\int\|Y_{st}\|^u d\pi\leq\int_{R^p}g(x)\pi(dx)<\infty. \tag{2.31}$$

By Lemma 3.1 in Tjϕstheim (1990), Y_t is geometrically ergodic. Thus, Y_t has a unique stationary distribution, π. Let Y_0 be initialized from the stationary distribution π. Then $\{y_t:t\in\mathcal{N}\}$ is the unique stationary solution to model (1.1) and it is geometrically ergodic. Furthermore, by (A.12), $E|y_t|^u<\infty$. We refer the necessity to Ling (2007). This complete the proof.

.

2.6 Threshold Models

The process $\{Y_t\}$ is said to follow the first-order threshold AR [TAR(1)] model if the following equation is satisfies:

$$Y_t = \begin{cases} \phi_1 Y_{t-1} + \varepsilon_t, & \text{if } Y_{t-1} > 0, \\ \phi_2 Y_{t-1} + \varepsilon_t, & \text{if } Y_{t-1} \leq 0, \end{cases} \tag{2.32}$$

where ε_t is iid r.v.s with mean 0.

Theorem 2.9. *Assume that ε_t has a strictly positive density $f(\cdot)$ on R. Then, $\{Y_t\}$ from model (2.32) is stationary and ergodic if and only if*

$$\phi_1 < 1, \phi_2 < 1 \textit{ and } \phi_1 \phi_2 < 1. \tag{2.33}$$

The region described above is plotted in Figure 6.

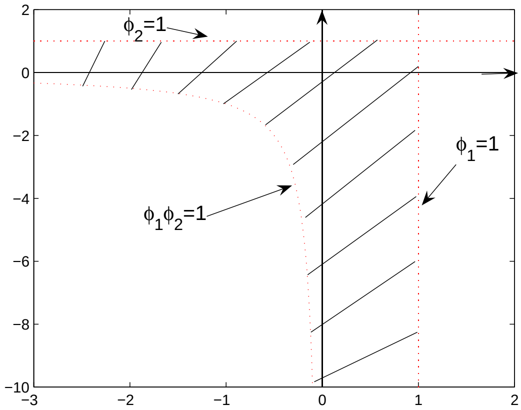


Fig. 2.4 The shaded region bounded by the dotted line is $\{(\phi_1, \phi_2) : \phi_1 < 1, \phi_2 < 1 \text{ and } \phi_1 \phi_2 < 1\}$

Proof. Note that $\{Y_t\}$ is a Markov Chain with state space $(R, \mathscr{B})$ and its transition density is given by

$$p(x,y) = f\left[y - \phi_1 x I(x > 0) - \phi_2 x I(x \leq 0)\right].$$

It is not hard to see that $\{Y_t\}$ is ν-irreducible and aperiodic, where ν is the Lebesgue measure. Since f is strictly positive, any non ν-null compact set $K \in \mathscr{B}$ is small. From (2.10), we can select positive constants a and b such that

$$-\frac{a}{b} < \phi_1 < 1 \text{ and } -\frac{b}{a} < \phi_2 < 1.$$

Choosing

$$g(x) = \begin{cases} ax, & \text{if } x > 0, \\ -bx, & \text{if } x \leq 0. \end{cases}$$

Then,

$$E\left[g(Y_t)|Y_{t-1}=x\right]=\begin{cases}Eg(\phi_1 x+\varepsilon_t), & \text{if } \ x>0,\\ Eg(\phi_2 x+\varepsilon_t), & \text{if } \ x\le 0.\end{cases}$$

When $x>0$ and $\phi_1<0$,

$$\begin{aligned}Eg(\phi_1 x+\varepsilon_t)&=ax+aE\left\{[(\phi_1-1)x+\varepsilon_t]\,I(\varepsilon_t>-\phi_1 x)\right\}\\&\qquad -bE\left\{\left[(\phi_1+ab^{-1})x+\varepsilon_t\right]I(\varepsilon_t\le-\phi_1 x)\right\}\\&\le ax-bE\left\{\left[(\phi_1+ab^{-1})x+\varepsilon_t\right]I(\varepsilon_t\le-\phi_1 x)\right\}+o(x)\\&\le ax-1,\ \text{as } x\to\infty.\end{aligned}$$

When $x>0$ and $\phi_1>0$,

$$\begin{aligned}Eg(\phi_1 x+\varepsilon_t)&\le ax+aE\left\{[(\phi_1-1)x+\varepsilon_t]\,I(\varepsilon_t>-\phi_1 x)\right\}+o(x)\\&\le ax-1,\ \text{as } x\to\infty.\end{aligned}$$

When $x<0$ and $\phi_2<0$,

$$\begin{aligned}Eg(\phi_2 x+\varepsilon_t)&=-bx-bE\left\{[(\phi_2-1)x+\varepsilon_t]\,I(\varepsilon_t\le-\phi_2 x)\right\}\\&\qquad +aE\left\{\left[(\phi_2+a^{-1}b)x+\varepsilon_t\right]I(\varepsilon_t>-\phi_2 x)\right\}\\&\le -bx-bE\left\{[(\phi_2-1)x+\varepsilon_t]\,I(\varepsilon_t\le-\phi_2 x)\right\}+o(x)\\&\le -bx-1,\ \text{as } x\to\infty.\end{aligned}$$

When $x>0$ and $\phi_2>0$,

$$\begin{aligned}Eg(\phi_2 x+\varepsilon_t)&\le -bx+aE\left\{\left[(\phi_2+a^{-1}b)x+\varepsilon_t\right]I(\varepsilon_t>-\phi_2 x)\right\}+o(x)\\&\le -bx-1,\ \text{as } x\to\infty.\end{aligned}$$

Combining the previous inequalities, we have

$$E\left[g(Y_t|Y_{t-1}=x)\right]\le g(x)-1,\ \text{when } x\notin[-M,M],$$

as M is large enough. It is not hard to see that there exists a constant c such that

$$E\left[g(Y_t|Y_{t-1}=x)\right]\le g(x)-1+cI_{[-M,M]}(x).$$

.Thus, there is a unique stationary and ergodic solution to (2.9). We refer the proof of the necessity to Petruccelli and Woolford (1984) and Chan et. al. (1985)

We next considers the m-regime TAR [MTAR] model($m\ge 2$) with order p :

$$Y_t=\sum_{j=1}^{m}(\mathbf{Y}'_{t-1}\boldsymbol{\beta}_j+\varepsilon_t)I(r_{j-1}<Y_{t-d}\le r_j), \tag{2.34}$$

where $\mathbf{Y}_{t-1} = (1, Y_{t-1}, ..., Y_{t-p})'$, $\boldsymbol{\beta}_j = (\phi_{j0}, \phi_{j1}, ..., \phi_{jp})' \in \mathbb{R}^{p+1}$, $\sigma_j > 0, j = 1, ..., m$; $-\infty = r_0 < r_1 < ... < r_m = \infty$ and $I(\cdot)$ is an indicator function. The number m of regimes and the order p of model (7.7) are positive integers. d is a positive integer called the delay parameter. $\{r_1, ..., r_{m-1}\}$ are threshold parameters. Assume that $\{\varepsilon_t\}$ satisfy the condition of Theorem 2.9. Vertorizing model (2.34), the Markov Chan is irreducible, see Section 2.2. Using the test function $g(x) = 1 + \max_i |x_i| p_i$ with $\max_i \sum_{j=1}^{m} |\phi_{ij}| p_1 / p_j < 1$ and $p_1 > p_2 \cdots > p_m > 0$, Chan and Tong (1985) show that, if

$$\max_i \sum_{j=1}^{m} |\phi_{ij}| < 1,$$

then $\{Y_t\}$ from model (2.34) is stationary and geometrically ergodic. This condition is much stronger than that in Theorem 2.9 and also stronger that of the AR(1) model. Except for the first-order MTAR model studied in Chan et. al. (1985), the necessary condition for the stationarity of model (2.34) is open.

It is natural to extend the threshold AR model to the threshold ARMA model. For the two regimes case, it can be defined as follows:

$$\begin{aligned} Y_t = & \left[\sum_{i=1}^{p} \phi_{1i} Y_{t-i} + \sum_{i=1}^{q} \psi_{1i} \varepsilon_{t-i} \right] I(Y_{t-d} \le r) \\ & + \left[\sum_{i=1}^{p} \phi_{2i} Y_{t-i} + \sum_{i=1}^{q} \psi_{2i} \varepsilon_{t-i} \right] I(Y_{t-d} > r) + \varepsilon_t . \end{aligned} \tag{2.35}$$

Ones can write model (2.35) into a vector Markov Chain. Howver, Except for the case with $\psi_{11} = \psi_{21} = 0$ in Liu and Susko (1992), no one shows that this Markov chain is irreducible. Thus, the stationarity and ergodicity of TARMA models is greatly unclear up to date.

However, when $p = 0$, model (2.35) reduces to the threshold MA (TMA) model. We can use Theorem 2.2 instead if Theorem 2.3. This is the following theorem.

Theorem 2.10. *Suppose that $\{\varepsilon_t\}$ is a sequence of iid r.v.s with $\mathbb{P}(a_n \le r, b_n \le r) + \mathbb{P}(a_n > r, b_n > r) \ne 0$, where $a_t = \varepsilon_t + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i}$ and $b_t = \varepsilon_n + \sum_{i=1}^{q} \phi_i \varepsilon_{t-i}$.. Then Y_t from model (2.35) with $p = 0$ has a unique strictly stationary and ergodic solution expressed by*

$$Y_t = \varepsilon_t + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} + \sum_{i=1}^{q} (\phi_i - \psi_i) \varepsilon_{t-i} \big] \alpha_{t-d}, \quad a.s.,$$

where

$$\alpha_{t-d} = \sum_{j=1}^{\infty} \Big[\big(\prod_{s=1}^{j-1} W_{t-sd} \big) U_{t-jd} \Big], \quad \text{in } L^1 \text{ and a.s.},$$

where $U_t = \mathbb{1}(a_t \le r)$ and $W_t = \mathbb{1}(b_t \le r) - \mathbb{1}(a_t \le r)$.

Proof. From model (2.35), $\mathbb{1}(Y_t \leq r) = U_n + W_n \mathbb{1}(Y_{t-d} \leq r)$. Iterating $k \geq 1$ steps, we have

$$\mathbb{1}(Y_t \leq r) = \sum_{j=0}^{k-1} \big[\big(\prod_{s=0}^{j-1} W_{t-sd}\big) U_{t-jd}\big] + \big(\prod_{i=0}^{k-1} W_{t-id}\big) \mathbb{1}(Y_{t-kd} \leq r)$$

with the convention $\prod_0^{-1} = 1$. Let

$$\alpha_{t,k} = \sum_{j=0}^{k-1} \big[\big(\prod_{s=0}^{j-1} W_{t-sd}\big) U_{t-jd}\big].$$

For given d and q, there exists a unique nonnegative integer m such that $md < \max(d, q+1) \leq (m+1)d$. Let $\delta = \mathbb{E}|W_1|$. Under the condition in Theorem 2.10, it is not difficult to prove that $0 \leq \delta < 1$. Observing that both $\{U_n\}$ and $\{W_n\}$ are q-dependent sequences, we can extract an independent subsequence $\{W_{t-j(m+1)d}, j = 0, 1, ..., [\frac{k-1}{m+1}]\}$ from the sequence $\{W_{t-id}, i = 0, 1, 2, ..., k-1\}$, where $[a]$ denotes the integral part of a. Since $|U_n| \leq 1$ and $|W_n| \leq 1$, it yields that

$$\mathbb{E}\Big|\big(\prod_{i=0}^{k-1} W_{t-id}\big) U_{t-kd}\Big| \leq (\mathbb{E}|W_1|)^{[\frac{k-1}{m+1}]},$$

implying

$$\sum_{j=1}^{\infty} \mathbb{E}\Big|\big(\prod_{i=0}^{j-1} W_{t-id}\big) U_{t-jd}\Big| \leq \sum_{j=1}^{\infty} \delta^{[\frac{j-1}{m+1}]} = (m+1) \sum_{k=0}^{\infty} \delta^k < \infty.$$

Using the above inequalities, we can prove that $\mathbb{E}|\alpha_{t,s} - \alpha_{t,t}| \to 0$ as $s, t \to \infty$ for each fixed n. By the Cauchy criterion, $\alpha_{t,k}$ converges in L^1 as $k \to \infty$. Write the limit as

$$\alpha_t = \sum_{j=0}^{\infty} \big[\big(\prod_{s=0}^{j-1} W_{t-sd}\big) U_{t-jd}\big].$$

Applying the inequalities above again, it is easy to get

$$\sum_{k=1}^{\infty} \mathbb{E}|\alpha_{t,k} - \alpha_t| \leq \sum_{k=1}^{\infty} \sum_{j=k}^{\infty} \delta^{[\frac{j-1}{m+1}]} < \infty,$$

yielding that

$$\lim_{k \to \infty} \alpha_{t,k} = \alpha_n, \quad \text{in } L^1 \text{ and a.s.}$$

Furthermore, recall that $U_n = \mathbb{1}(a_n \leq r)$ and $W_n = \mathbb{1}(b_n \leq r) - \mathbb{1}(a_n \leq r)$, where a_n and b_n are defined in (**??**), we have the iterative sequence: $\alpha_{t,1} = U_n$ and

$$\alpha_{t,k} = U_n + W_n \alpha_{t-d,k-1} = (1-\alpha_{t-d,k-1})\mathbb{1}(a_n \le r) + \alpha_{t-d,k-1}\mathbb{1}(b_n \le r)$$

for each n and $k \ge 1$. Note that $\alpha_{t,k}$ and $\alpha_{t-d,k}$ have the same distribution for fixed k since the error $\{e_i\}$ is i.i.d.. By induction over k, we have that $\alpha_{t,k}$ only takes two values 0 and 1 a.s. since $\alpha_{t,1}$ only takes 0 and 1. Thus, α_n at most takes two values 0 and 1 a.s., namely, $\alpha_n = \mathbb{1}(\alpha_n = 1)$ a.s.. Define a new sequence $\{S_n\}$

$$S_n = \mu_2 + e_n + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} + \left[(\mu_1 - \mu_2) + \sum_{i=1}^{q} (\phi_i - \psi_i)\varepsilon_{t-i}\right]\alpha_{t-d}.$$

By simple calculation, we have

$$\begin{aligned}
\mathbb{1}(S_n \le r) &= \mathbb{1}(a_n \le r)\mathbb{1}(\alpha_{t-d} = 0) + \mathbb{1}(b_n \le r)\mathbb{1}(\alpha_{t-d} = 1) \\
&= U_n + W_n \mathbb{1}(\alpha_{t-d} = 1) \\
&= U_n + W_n \alpha_{t-d} = \alpha_n, \quad \text{a.s..}
\end{aligned}$$

Hence,

$$S_n = \mu_2 + e_n + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} + \left[(\mu_1 - \mu_2) + \sum_{i=1}^{q} (\phi_i - \psi_i)\varepsilon_{t-i}\right]\mathbb{1}(S_{t-d} \le r), \quad \text{a.s..}$$

Thus, $\{S_t\}$ is the solution of model (2.1) which is strictly stationary and ergodic.

To uniqueness, suppose that $\tilde{S}_n$ is a solution to model (**??**), then

$$\mathbb{1}(\tilde{S}_t \le r) = U_n + W_n \mathbb{1}(\tilde{S}_{t-d} \le r).$$

Iterating the above equation, one can get for $k \ge 1$

$$\mathbb{1}(\tilde{S}_t \le r) = \alpha_{t,k} + \Big(\prod_{i=0}^{k-1} W_{t-id}\Big)\mathbb{1}(\tilde{S}_{t-kd} \le r).$$

We can show that the second term of the previous equation converges to zero a.s.. Thus, we have $\mathbb{1}(\tilde{S}_t \le r) = \alpha_n$ a.s.. Therefore,

$$\tilde{S}_n = \mu_2 + e_n + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} + \left[(\mu_1 - \mu_2) + \sum_{i=1}^{q} (\phi_i - \psi_i)\varepsilon_{t-i}\right]\alpha_{t-d}, \quad \text{a.s.,}$$

i.e., $\tilde{S}_n = S_t$ a.s. The proof is complete. □

2.7 Exercises

Exercise 2.1. Give the proof of Theorem 2.7(v).

Exercise 2.2. Let $\{A_t\}$ be iid random matrices with $E\|A_t\|^\delta < \infty$. Prove that the top Lyapounov exponent

$$\gamma = \inf\{\frac{1}{n}E\ln\|A_1\cdots A_n\|,\ n \geq 1\} < 0$$

if and only there exists an integer s such that $E\|A_1\cdots A_s\|^u < 1$ for some $u \in (0,1)$.

Exercise 2.3. $\{Y_t\}$ is said to follow a smooth TAR(1) model if it satisfies the equation:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-1} F\left(\frac{Y_{t-1} - r}{z}\right) + \varepsilon_t,$$

where $F(\cdot)$ is the distribution function of the standard normal r.v. and $\{\varepsilon_t\}$ is iid with mean zero and a finite variance σ^2. Show that if

$$\phi_1 < 1,\ \phi_1 + \phi_2 < 1 \textit{ and } \phi_1(\phi_1 + \phi_2) < 1,$$

then $\{Y_t\}$ from the smooth TAR(1) model is stationary and ergodic. Furthermore, $E|Y_t|^k < \infty$ if $E|\varepsilon_t|^k < \infty$.

Exercise 2.4. $\{Y_t\}$ is said to follow a bilinear model if it satisfies the equation:

$$Y_t = \phi Y_{t-1} + (\theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2})Y_{t-1} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is iid with mean zero and a finite variance σ^2. Show that if and only

$$E\ln|\phi + \theta_1\varepsilon_{t-1} + \theta_1\varepsilon_{t-2}| < 0,$$

then $\{Y_t\}$ from the bilinear model is stationary and ergodic.

Exercise 2.5. Prove that $\sum_{i=1}^s \beta_i < 1$ is equivalent to

$$0 \leq \rho(G) < 1, \text{ where } G = \begin{pmatrix} \beta_1 & \cdots & \beta_s \\ I_{s-1} & & O \end{pmatrix}, \tag{2.36}$$

I_k is the $k \times k$ identity matrix and $\rho(B)$ is the spectral radius of matrix B.

Chapter 3
Regular Estimation

3.1 A General Theory

Assume that the real $p \times 1$ vector time series $\{Y_t : t = 0, \pm 1, \cdots\}$ is $\mathscr{F}_t-$ measurable, strictly stationary and ergodic, and its conditional distribution is given by

$$Y_t | \mathscr{F}_{t-1} \sim G(\theta, Y_{t-1}), \tag{3.1}$$

where $\mathscr{F}_t$ is the σ-field generated by $\{Y_t, Y_{t-1}, \cdots\}$, $Y_t = (Y_t, \cdots, Y_{t-p+1})$ or $Y_t = (Y_t, Y_{t-1}, \cdots)$, and θ is an $m \times 1$ unknown parameter vector. The structure of the time series $\{Y_t\}$ is characterized by the distribution G and the parameter θ. We assume that the parameter space Θ is a compact subset of R^m, and the true value θ_0 of θ is an interior point in Θ. We use the following OF with the initial value Y_0 to estimate θ_0:

$$L_n(\theta) = \sum_{t=1}^{n} l_t(\theta),$$

where $l_t(\theta) = l(Y_t, \theta)$ is a measurable function with respect to Y_t and is continuous in terms of θ. The estimator of θ_0 the maximizer of $L_n(\theta)$ on Θ, i.e.

$$\hat{\theta}_n = \text{argmax}_{\Theta} L(\theta).$$

When the dimension of the initial value Y_0 is infinite, it need to be replaced by some constant $\widetilde{Y}_0$. To make it simple, we assume that Y_0 is available. We only need the identification condition for the consistency of $\hat{\theta}_n$ as follows.

Assumption 3.1 *$E \sup_{\theta \in \Theta} [l_t(\theta)] < \infty$, and $E[l_t(\theta)]$ has a unique maximizer at θ_0.*

Theorem 3.1. *If Assumptions 3.1 holds, then $\hat{\theta}_n \to \theta_0$ a.s..*

Proof. First, using the standard point-wise argument with the ergodic theorem, we can show that, for any $\eta > 0$,

$$\lim_{l\to\infty} P\Big(\max_{l\le n<\infty} \sup_{\|\theta-\theta_0\|\ge\eta} \sum_{t=1}^{n}[l_t(\theta)-l_t(\theta_0)]\ge 0\Big)=0. \tag{3.2}$$

Since $E[l_t(\theta)]$ has a unique maximum at θ_0, Θ is compact, and $El_t(\theta)$ is continuous, there exists a constant $c>0$, such that

$$\max_{\|\theta-\theta_0\|>\eta} E[l_t(\theta)-l_t(\theta_0)]\le -c, \tag{3.3}$$

for any $\eta>0$. By (3.1)-(3.2), it follows that

$$\begin{aligned}
&P\Big(\max_{l\le n<\infty} \sup_{\|\theta-\theta_0\|\ge\eta} \{\sum_{t=1}^{n}[l_t(\theta)-l_t(\theta_0)]+\frac{cn}{2}\}>0\Big)\\
&\le P\Big(\max_{l\le n<\infty} \sup_{\Theta}\{\Big|\frac{1}{n}\sum_{t=1}^{n}[l_t(\theta)-El_t(\theta)]\Big|\}>\frac{c}{4}\Big)\to 0,
\end{aligned}$$

as $l\to\infty$. By the previous inequality, for any $\varepsilon>0$, we have

$$\begin{aligned}
&\lim_{l\to\infty} P(\max_{l\le n<\infty}\|\hat\theta_n-\theta_0\|>\varepsilon)\\
&=\lim_{l\to\infty} P\Big\{\max_{l\le n<\infty}\|\hat\theta_n-\theta_0\|>\varepsilon,\ \max_{l\le n<\infty}\sum_{t=1}^{n}\Big[l_t(\hat\theta_n)-l_t(\theta_0)\Big]\ge 0\Big\}\\
&\le \lim_{l\to\infty} P\Big\{\max_{l\le n<\infty}\sup_{\|\theta-\theta_0\|>\varepsilon}\sum_{t=1}^{n}\Big[l_t(\theta)-l_t(\theta_0)\Big]\ge 0\Big\}=0.
\end{aligned}$$

Thus, the conclusion holds.□

For asymptotic normality, we assume that $l_t(\theta)$ has continuously twice differentiable almost surely (a.s.) in terms of θ. Denote

$$D_t(\theta)=\frac{\partial l_t(\theta)}{\partial\theta}\ \textit{and}\ P_t(\theta)=-\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'},$$

$\Sigma=E[P_t(\theta_0)]$ and $\Omega=E[D_t(\theta_0)D_t'(\theta_0)]$. We need one assumption as follows:

Assumption 3.2 *(i) $D_t(\theta_0)$ is a martingale difference in terms of $\mathscr{F}_t$,*
(ii) $0<\Sigma<\infty$ and $0<\Omega<\infty$,
(iii) $E\sup_{\|\theta-\theta_0\|<\eta}\|P_t(\theta)\|<\infty$ for some $\eta>0$.

We now state our second result as follows.

Theorem 3.2. *If $\hat\theta_n\to\theta_0$ a.s. and Assumptions 3.2 holds, then*

$$\begin{aligned}
&(i)\quad \hat\theta_n=\theta_0+O\Big[\Big(\frac{\log\log n}{n}\Big)^{1/2}\Big]\ a.s.,\\
&(ii)\quad \sqrt{n}(\hat\theta_n-\theta_0)\longrightarrow_{\mathscr{L}} N(0,\Sigma^{-1}\Omega\Sigma^{-1}).
\end{aligned}$$

Proof. Applying Taylor's expansion to $\partial l_t(\tilde{\theta}_n)/\partial\theta$ and using Lemma 2.2,

$$\begin{aligned}\hat{\theta}_n - \theta_0 &= -\Big[\frac{1}{n}\sum_{t=1}^{n} P_t(\hat{\theta}_n^*)\Big]^{-1}\frac{1}{n}\sum_{t=1}^{n} D_t(\theta_0)\\ &= -[\Sigma + o(1)]^{-1}\frac{1}{n}\sum_{t=1}^{n} D_t(\theta_0)\ a.s.,\end{aligned}$$

where $\hat{\theta}_n^*$ lies between $\hat{\theta}_n$ and θ_0 and $\hat{\theta}_n^* \to \theta_0$ a.s.. By the law of iterated logaritheorem, we can claim that (i) holds. By the central limit theorem, (b) holds. □

Usually, the initial value Y_0 does not effect the asymptotic results. If one wants rigor to prove this, the following conditions are enough in which $l_t(\theta)$ with the initial value $\widetilde{Y}_0$ is denoted by $\widetilde{l}_t(\theta)$ and similarly define $\widetilde{D}_t(\theta)$ and $\widetilde{P}_t(\theta)$:

Assumption 3.3

$$\begin{aligned}&(a)\ \ E\sup_{\Theta}|l_t(\theta) - \widetilde{l}_t(\theta)| = O(\frac{1}{t^{\nu}});\\ &(b)\ \ E\|D_t(\theta_0) - \widetilde{D}_t(\theta_0)\| = O(\frac{1}{t^{1/2+\nu}}),\\ &(c)\ \ E\sup_{\Theta}\|P_t(\theta) - \widetilde{P}_t(\theta)\| = O(\frac{1}{t^{\nu}}).\end{aligned}$$

for some $\nu > 0$.

The decay rates in these conditions are very low and are satisfied by most of time series models.

3.2 ARMA Models

Assume the random sample $\{Y_1, \cdots, Y_n\}$ is from model (2.6) with ε_t being a sequence of iid with mean zero and variance $\sigma^2 > 0$. Denote $\theta = (\mu, \phi_1, \cdots, \phi_p, \psi_1, \cdots, \psi_q)'$ and θ_0 as its true value. Assume that the parameter space Θ is a compact subset of R^m and θ_0 is an interior point in Θ, where $m = p+q+1$. The following assumption is for the stationarity, invertibility and identifiability of model (6.28).

Assumption 3.4 *$\phi(z) \neq 0$ and $\psi(z) \neq 0$ when $|z| \leq 1$, and $\phi(z)$ and $\psi(z)$ have no common root with $\phi_p \neq 0$ or $\psi_q \neq 0$.*

With initial values $\tilde{Y}_0 \equiv \{Y_0, Y_{-1}, \cdots\}$, we can write the parametric model as

$$\varepsilon_t(\theta) = Y_t - \mu - \sum_{i=1}^{p}\phi_i Y_{t-i} - \sum_{i=1}^{q}\psi_i\varepsilon_{t-i}(\theta). \tag{3.4}$$

Its first derivatives are as follows:

$$\frac{\partial \varepsilon_t(\theta)}{\partial u} = -1 - \sum_{i=1}^{q} \psi_i \frac{\partial \varepsilon_{t-i}(\theta)}{\partial u};$$
$$\frac{\partial \varepsilon_t(\theta)}{\partial \phi_j} = -y_{t-j} - \sum_{i=1}^{q} \psi_i \frac{\partial \varepsilon_{t-i}(\theta)}{\partial \phi_j}, \text{ for } 1 \le j \le p;$$
$$\frac{\partial \varepsilon_t(\theta)}{\partial \psi_j} = -\varepsilon_{t-j}(\theta) - \sum_{i=1}^{q} \psi_i \frac{\partial \varepsilon_{t-i}(\theta)}{\partial \psi_j}, \text{ for } 1 \le j \le q.$$

Similarly, we can write down $\partial^2 \varepsilon_t(\theta)/\partial\theta\partial\theta'$. Let $\mathscr{F}_{t-1} = \sigma\{\varepsilon_k : k \le t-1\}$. Using the conditional least squares estimation, we get the objective function is

$$L_n(\theta) = \sum_{t=1}^{n} \varepsilon_t^2(\theta). \tag{3.5}$$

The minimizer, $\hat{\theta}_n$, of $L_n(\theta)$ on Θ is called the conditional least squares estimator (LSE) of θ, i.e.,

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

We have the following theorem:

Theorem 3.3. *If Assumption 3.3 holds, then it follows that*

$$(i) \quad \hat{\theta}_n = \theta_0 + O\Big[\Big(\frac{\log\log n}{n}\Big)^{1/2}\Big] \ a.s.,$$
$$(ii) \quad \sqrt{n}(\hat{\theta}_n - \theta_0) \longrightarrow_{\mathscr{L}} N(0, \sigma^2 \Omega_0^{-1}),$$

where $\Omega_0 = E\,[\partial\varepsilon_t(\theta_0)/\partial\theta\partial\varepsilon_t(\theta_0)/\partial\theta']$.

Proof. To make it simple, we only present the proof for the case with $p = q = 1$ and without μ. In this case, the parameter space is

$$\Theta = \{(\phi_1, \psi_1)' : |\phi_1| \le c \quad \text{and} \quad |\psi_1| \le c\}$$

for some $0 < c < 1$, and $\varepsilon_t(\theta)$ has the following expansion:

$$\varepsilon_t(\theta) = \sum_{i=0}^{\infty} \psi^i (Y_{t-i} - \phi Y_{t-i-1}).$$

Thus, $\sup_{\theta\in\Theta} |\varepsilon_t(\theta)| \le c_0 \sum_{i=0}^{\infty} \rho^i |Y_{t-i}|$ and hence $E\sup_{\theta\in\Theta} |\varepsilon_t(\theta)|^2 < \infty$, where $\rho \in (0,1)$. Secondly, use the following expansion

$$\varepsilon_t(\theta) = \varepsilon_t - (\phi - \phi_0)Y_{t-1} + [\psi\varepsilon_{t-1}(\theta) - \psi_0\varepsilon_{t-1}],$$

it follows that

$$E\varepsilon_t^2(\theta) = \sigma^2 + E\{(\phi-\phi_0)Y_{t-1} - [\psi\varepsilon_{t-1}(\theta) - \psi_0\varepsilon_{t-1}]\}^2 \ge \sigma^2$$

and "=" holds if and only if

$$(\phi-\phi_0)Y_{t-1}-\psi\varepsilon_{t-1}(\theta)+\psi_0\varepsilon_{t-1}=0,$$

which holds if and only if $\psi=\psi_0$ and $\phi=\phi_0$. Thus, Assumption 3.1 holds and hence we can claim that $\hat{\theta}_n\to\theta_0$ by Theorem 3.1.

$$\begin{aligned}D_t(\theta)&=2\frac{\partial\varepsilon_t(\theta)}{\partial\theta}\varepsilon_t(\theta);\\ P_t(\theta)&=2\frac{\partial\varepsilon_t(\theta)}{\partial\theta}\frac{\partial\varepsilon_t(\theta)}{\partial\theta'}+2\frac{\partial^2\varepsilon_t(\theta)}{\partial\theta\partial\theta'}\varepsilon_t(\theta),\end{aligned}$$

where the first derivative of $\varepsilon_t(\theta)$ with respect to θ is

$$\begin{aligned}\frac{\partial\varepsilon_t(\theta)}{\partial\phi}&=-Y_{t-1}+\psi\frac{\partial\varepsilon_{t-1}(\theta)}{\partial\phi}=-\sum_{i=0}^{\infty}\psi^iY_{t-i-1};\\ \frac{\partial\varepsilon_t(\theta)}{\partial\psi}&=\varepsilon_{t-1}(\theta)+\psi\frac{\partial\varepsilon_{t-1}(\theta)}{\partial\psi}=\sum_{i=0}^{\infty}\psi^i\varepsilon_{t-i-1}(\theta).\end{aligned}$$

It is not hard to see that

$$\Omega=4\sigma^2\Omega_0 \textit{ and } \Sigma=\Omega_0.$$

Then, $\Sigma^{-1}\Omega\Sigma^{-1}=\sigma^2\Omega_0$. It is straightforward to show that $E\sup_{\Theta}\|P_t(\theta)\|<\infty$. To prove that Ω_0 is positive definite, see Exercise 3.1. Thus, Assumption 3.2 holds and the conclusion holds by Theorem 2.2. This complete the proof.

3.3 GARCH(1, 1) Models

Assume that the random sample $\{Y_1,\cdots,Y_n\}$ is from the GARCH(1, 1) model:

$$\varepsilon_t=\eta_t\sqrt{h_t}\text{ and }h_t=\alpha_0+\alpha\varepsilon_{t-1}^2+\beta h_{t-1},$$

where $\eta_t\sim$iid with mean zero and variance 1, $\alpha_0>0$, $\alpha>0$ and $\beta>0$. Let $\theta=(\alpha_0,\alpha,\beta)'$ and its true value is denoted by θ_0. We make the following assumption:

Assumption 3.5 *(a). $\Theta=\{\theta:E\ln(\beta+\alpha\eta_t^2)<0,\alpha_0\geq c,\alpha\geq c$ and $\beta\geq c\}$, where $c>0$ is a constant, and*
(b). η_t has a bounded density in some neighborhood of 0 with $E\eta_t^4<\infty$.

Using quasi-maximum likelihood estimation with the initial value $\tilde{\varepsilon}_0=\{\varepsilon_0,\varepsilon_{-1},\cdots\}$, we get the following objective function:

$$L_n = \sum_{t=1}^{n} l_t(\theta) = -\sum_{t=1}^{n} \left[\log h_t(\theta) + \frac{\varepsilon_t^2}{h_t(\theta)} \right],$$

where $h_t(\theta) = \alpha_0 + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}(\theta)$. The minimizer of L_n on Θ is called the quasi-maximum likelihood estimator (QMLE) of θ_0, denoted by $\hat{\theta}_n$.

Theorem 3.4. *(a) if Assumption 3.4(a) holds, then $\hat{\theta}_n \to \theta_0$ a.s.,*

(b) if Assumption 3.4 holds and θ_0 is an interior point in Θ, then

$$(i)\quad \hat{\theta}_n = \theta_0 + O\left[\left(\frac{\log\log n}{n}\right)^{1/2}\right] \text{ a.s.},$$
$$(ii)\quad \sqrt{n}(\hat{\theta}_n - \theta_0) \longrightarrow_{\mathscr{L}} N(0, \kappa\Omega_0^{-1}),$$

where $\Omega_0 = E\left[h_t^{-2}\partial h_t(\theta_0)/\partial\theta\partial h_t(\theta_0)/\partial\theta'\right]$ and $\kappa = (E\eta_4 - 1)/2$. Furthermore, when $\eta_t \sim N(0,1)$, $\kappa = 1$ and $\hat{\theta}_n$ is the MLE of θ_0

Proof. We first note that

$$\begin{aligned} El_t(\theta) &= -E\log h_t(\theta) - E\frac{h_t(\theta_0)}{h_t(\theta)} \\ &= -\left[-E\log\frac{h_t(\theta_0)}{h_t(\theta)} + E\frac{h_t(\theta_0)}{h_t(\theta)}\right] - \log h_t(\theta_0) \\ &= -\left[-E\log M_t + EM_t\right] - \log h_t(\theta_0), \end{aligned}$$

where $M_t = h_t(\theta_0)/h_t(\theta)$. Note that, for any $x > 0$, $f(x) \equiv -\log x + x \geq 1$ and hence

$$-E\log M_t + EM_t \geq 1.$$

When $M_t = 1$, we have $f(M_t) = f(1) = 1$. If $M_t \neq 1$, then

$$f(M_t) > f(1).$$

Thus, $Ef(M_t) \geq Ef(1)$ with equality only if $M_t = 1$ with probability 1. Thus, $El_t(\theta)$ reaches its maximum $-1 - \log h_t(\theta_0)$ and this occurs if and only if

$$h_t(\theta) = h_t(\theta_0),$$

which holds if and only if $\theta = \theta_0$, see exercise 3.2. Thus, the second part of Assumption 3.4(a) holds.

$$
\begin{aligned}
D_t(\theta) &= -\frac{1}{2h_t(\theta)}\frac{\partial h_t(\theta)}{\partial \theta}\left[1-\frac{\varepsilon_t^2}{h_t(\theta)}\right],\\
P_t(\theta) &= -\frac{1}{2h_t(\theta)}\frac{\partial h_t(\theta)}{\partial \theta}\frac{\partial h_t(\theta)}{\partial \theta'} - R_t(\theta),\\
R_t(\theta) &= \frac{1}{h_t(\theta)}\left[\frac{\partial^2 h_t(\theta)}{\partial\theta\partial\theta'} - \frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial \theta}\frac{\partial h_t(\theta)}{\partial \theta'}\right]\left[1-\frac{\varepsilon_t^2}{h_t(\theta)}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
h_t(\theta) &= \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}(\theta) = \omega + \alpha\sum_{i=0}^{\infty}\beta^i\varepsilon_{t-i-1}^2,\\
\frac{\partial h_t(\theta)}{\partial\omega} &= 1 + \beta\frac{\partial h_{t-1}(\theta)}{\partial\omega} = \frac{1}{1-\beta},\\
\frac{\partial h_t(\theta)}{\partial\alpha} &= \varepsilon_{t-1}^2 + \beta\frac{\partial h_{t-1}(\theta)}{\partial\alpha} = \sum_{i=0}^{\infty}\beta^i\varepsilon_{t-i-1}^2,\\
\frac{\partial h_t(\theta)}{\partial\beta} &= h_{t-1}(\theta) + \beta\frac{\partial h_{t-1}(\theta)}{\partial\beta} = \sum_{i=0}^{\infty}\beta^i h_{t-i-1}(\theta)\\
&= \frac{\omega}{1-\beta} + \alpha\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\beta^{i+j}\varepsilon_{t-i-j-2}^2.
\end{aligned}
$$

To check other assumptions, the difficulty is to show that, for any integer $m \geq 1$,

$$
E\sup_{\Theta}\left|\frac{h_t}{h_t(\theta)}\right|^m < \infty. \tag{3.6}
$$

We only give the proof when $m = 1$. Other case is similar. First,

$$
h_t = \omega_0 + \alpha_0\varepsilon_{t-1}^2 + \beta_0 h_{t-1} \leq c(1+\varepsilon_{t-1}^2+\varepsilon_{t-2}^2+\varepsilon_{t-3}^2+h_{t-3}).
$$

Thus, we have

$$
h_t(\theta) \geq c(1+\varepsilon_{t-1}^2+\varepsilon_{t-2}^2+\varepsilon_{t-3}^2).
$$

Furthermore, we have

$$
\varepsilon_{t-1}^2+\varepsilon_{t-2}^2+\varepsilon_{t-3}^2 \geq c(\eta_{t-1}^2+\eta_{t-2}^2+\eta_{t-3}^2)h_{t-3}.
$$

Thus,

$$
\frac{h_t}{h_t(\theta)} \leq c + \frac{ch_{t-3}}{\varepsilon_{t-1}^2+\varepsilon_{t-2}^2+\varepsilon_{t-3}^2} \leq c + \frac{c}{\eta_{t-1}^2+\eta_{t-2}^2+\eta_{t-3}^2}.
$$

$$\begin{aligned}
E\frac{1}{\eta_{t-1}^2+\eta_{t-2}^2+\eta_{t-3}^2} &= \int_0^\infty P(\eta_{t-1}^2+\eta_{t-2}^2+\eta_{t-3}^2<\frac{1}{y})dy \\
&\le N+\int_N^\infty P(\eta_{t-1}^2+\eta_{t-2}^2+\eta_{t-3}^2<\frac{1}{y})dy \\
&\le N+\int_N^\infty [P(\eta_{t-1}^2<\frac{1}{y})]^3 dy \\
&= N+\int_N^\infty \left[\int_{-\sqrt{y^{-1}}}^{\sqrt{y^{-1}}} f(x)dx\right]^3 dy \\
&\le N+c\int_N^\infty y^{-3/2}dy<\infty,
\end{aligned}$$

as N is large enough, where c is a constant. Thus, $E\sup_\Theta h_t/h_t(\theta)|<\infty$. Thus. the first part of Assumption 3.1 holds and Theorem 3.4(a) holds.

$$\begin{aligned}
\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\alpha} &= \sum_{i=0}^\infty \frac{\beta^i\varepsilon_{t-i-1}^2}{\omega+\alpha\beta^i\varepsilon_{t-i-1}^2} \\
&\le c\sum_{i=0}^\infty \frac{\beta^i\varepsilon_{t-i-1}^2}{1+\beta^i\varepsilon_{t-i-1}^2} \le c\sum_{i=0}^\infty \left[\frac{\beta^i\varepsilon_{t-i-1}^2}{1+\beta^i\varepsilon_{t-i-1}^2}\right]^\tau \\
&\le c\sum_{i=0}^\infty \beta^{i\tau}|\varepsilon_{t-i-1}|^{2\tau} \le c\sum_{i=0}^\infty \rho^i|\varepsilon_{t-i-1}|^{2\tau}
\end{aligned}$$

for any $\tau\in(0,1)$, where $\rho\in(0,1)$ and c are constants. Thus

$$E\sup_\Theta\left|\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\alpha}\right|^m \le c\left\{\sum_{i=0}^\infty \rho^{i\tau}\,[E|\varepsilon_{t-i}|^{m\tau}]^{\frac{1}{m}}\right\}^m<\infty \tag{3.7}$$

as τ is small enough.

$$\begin{aligned}
\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\beta} &\le c+c\sum_{i=0}^\infty\sum_{j=0}^\infty \frac{\beta^{i+j}\varepsilon_{t-i-j-2}^2}{\omega+\alpha\beta^{i+j+1}\varepsilon_{t-i-j-2}^2} \\
&\le c+c\sum_{i=0}^\infty\sum_{j=0}^\infty \frac{\beta^{i+j}\varepsilon_{t-i-j-2}^2}{1+\beta^{i+j}\varepsilon_{t-i-j-2}^2} \\
&\le c+c\sum_{i=0}^\infty\sum_{j=0}^\infty \left[\frac{\beta^{i+j}\varepsilon_{t-i-j-2}^2}{1+\beta^{i+j}\varepsilon_{t-i-j-2}^2}\right]^\tau \\
&\le c+c\sum_{i=0}^\infty\sum_{j=0}^\infty \rho^{(i+j)\tau}|\varepsilon_{t-i-j-2}|^\tau
\end{aligned}$$

for any $\tau \in (0,1)$, where $\rho \in (0,1)$ and c are constants. Thus, for any $m \geq 1$,

$$E \sup_{\Theta} \left| \frac{1}{h_t(\theta)} \frac{\partial h_t(\theta)}{\partial \beta} \right|^m \leq \left\{ c + c \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \rho^{(i+j)\tau} [E|\varepsilon_{t-i-j}|^{m\tau}]^{\frac{1}{m}} \right\}^m < \infty \quad (3.8)$$

as τ is small enough. Using (3.5) and (3.6), we can show that Assumption 3.2(ii)-(iii) holds. Thus, Theorem 3.4(b) holds and

$$\Omega = \kappa^2 \Omega_0 \text{ and } \Sigma = \frac{1}{2}\Omega_0.$$

Thus, $\Sigma^{-1}\Omega\Sigma^{-1} = \kappa\Omega_0^{-1}$. This completes the proof.

3.4 Double AR(1) Models

Assume that the random sample $\{Y_1, \cdots, Y_n\}$ is from the DAR(1) model:

$$Y_t = \phi Y_{t-1} + \varepsilon_t \text{ and } \varepsilon_t = \eta_t \sqrt{\omega + \alpha Y_{t-1}^2},$$

where $\omega, \alpha > 0$. We make the following assumption.

Assumption 3.6 *(a).* $\Theta = \{\theta \equiv (\phi, \omega, \alpha) : E \ln|\phi + \sqrt{\alpha}\eta_t| < 0$ *with* $|\phi| \leq \tilde{\phi}$, $\underline{\omega} \leq \omega \leq \tilde{\omega}$, *and* $\underline{\alpha} \leq \alpha \leq \tilde{\alpha}\ \}$, *where* $\tilde{\phi}$, $\underline{\omega}$, $\tilde{\omega}$, $\underline{\alpha}$ *and* $\tilde{\alpha}$ *are some finite positive constants.*

(b). η_t *has a symmetric, positive and continuous Lebesgue density in R with mean zero, variance 1 and with* $E\eta_t^4 < \infty$.

Using the quasi-maximum likelihood estimation, we get the objective function:

$$L_n(\theta) = \sum_{t=1}^{n} l_t(\theta) = -\frac{1}{2} \sum_{t=1}^{n} \left[\ln(\omega + \alpha Y_{t-1}^2) + \frac{(Y_t - \phi Y_{t-1})^2}{2(\omega + \alpha Y_{t-1}^2)} \right].$$

The QMLE of θ_0 is the maximizer of $L_n(\theta)$ on Θ, denoted by $\hat{\theta}_n$. We have the following theorem.

Theorem 3.5. *(a) if Assumption 3.5(a) holds, then* $\hat{\theta}_n \to \theta_0$ *a.s.,*

(b) if Assumption 3.5 holds and θ_0 *is an interior point in* Θ, *then*

$$(i) \quad \hat{\theta}_n = \theta_0 + O\Big[\Big(\frac{\log\log n}{n}\Big)^{1/2}\Big] \ a.s.,$$

$$(ii) \quad \sqrt{n}(\hat{\theta}_n - \theta_0) \longrightarrow_{\mathscr{L}} N(0, diag\{\Sigma_0^{-1}, \kappa\Omega_0^{-1}),$$

where $\Sigma_0 = E[Y_{t-1}^2/(\omega_0 + \alpha_0 Y_{t-1}^2)]$ *and* $\Omega = E\Big[(\omega_0 + \alpha_0 Y_{t-1}^2)^{-2} \begin{pmatrix} 1 & Y_{t-1}^2 \\ Y_{t-1}^2 & Y_{t-1}^4 \end{pmatrix}$ *and* $\kappa = (E\eta_4 - 1)/2$. *Furthermore, when* $\eta_t \sim N(0,1)$, $\kappa = 1$ *and* $\hat{\theta}_n$ *is the MLE of* θ_0

Proof. Let $\tilde{\omega}^* = \max\{1, \tilde{\omega}\}$. By Jensen's inequality, we have

$$\begin{aligned} E\ln(\tilde{\omega}^* + \tilde{\alpha} Y_{t-1}^2) &= \frac{2}{\delta} E\ln(\tilde{\omega}^* + \tilde{\alpha} Y_{t-1}^2)^{\delta/2} \\ &\leq \frac{2}{\delta} \ln(\tilde{\omega}^{*\delta/2} + \tilde{\alpha}^{\delta/2} E|Y_{t-1}|^{\delta}) < \infty, \end{aligned}$$

where the following elementary relation is used: $(a+b)^s \leq a^s + b^s$ for all $a, b > 0$ and $s \in [0,1]$. Thus,

$$\begin{aligned} E\sup_{\theta\in\Theta} |\ln(\omega + \alpha Y_{t-1}^2)| &\leq E\sup_{\theta\in\Theta} [I\{\omega + \alpha Y_{t-1}^2 \geq 1\} \ln(\omega + \alpha Y_{t-1}^2)] \\ &\quad + E\sup_{\theta\in\Theta} [-I\{\omega + \alpha Y_{t-1}^2 \leq 1\} \ln(\omega + \alpha Y_{t-1}^2)] \\ &\leq E\ln(\tilde{\omega}^* + \tilde{\alpha} Y_{t-1}^2) - I\{\underline{\omega} < 1\} \ln \underline{\omega} < \infty. \end{aligned}$$

Furthermore, since $Y_t - \phi Y_{t-1} = \varepsilon_t - (\phi - \phi_0) Y_{t-1}$, it can be shown that

$$\begin{aligned} &E\sup_{\theta\in\Theta} \Big[\frac{(Y_t - \phi Y_{t-1})^2}{\omega + \alpha Y_{t-1}^2}\Big] \\ &\leq 2E\sup_{\theta\in\Theta} \Big[\frac{(\phi - \phi_0)^2 Y_{t-1}^2}{\omega + \alpha Y_{t-1}^2}\Big] + 2E\sup_{\theta\in\Theta} \Big[\frac{\omega_0 + \alpha_0 Y_{t-1}^2}{\omega + \alpha Y_{t-1}^2}\Big] < \infty. \end{aligned}$$

Thus, $E\sup_{\theta\in\Theta} |l_t(\theta)| < \infty$

$$\begin{aligned} El_t(\theta) &= -\frac{1}{2} E\Big[\ln(\omega + \alpha Y_{t-1}^2) + \frac{(Y_t - \phi Y_{t-1})^2}{\omega + \alpha Y_{t-1}^2}\Big] \\ &= -\frac{1}{2}\Big\{E\ln(\omega + \alpha Y_{t-1}^2) + E\Big(\frac{\omega_0 + \alpha_0 Y_{t-1}^2}{\omega + \alpha Y_{t-1}^2}\Big)\Big\} - \frac{(\phi - \phi_0)^2}{2} E\Big(\frac{Y_{t-1}^2}{\omega + \alpha Y_{t-1}^2}\Big). \end{aligned}$$

The second term reaches its maximum at zero, and this occurs if and only if $\phi = \phi_0$. The first term is equal to

$$-\frac{1}{2}\{-E\ln M_t + EM_t\} - \frac{1}{2} E\ln(\omega_0 + \alpha_0 Y_{t-1}^2), \tag{3.9}$$

where $M_t = (\omega_0 + \alpha_0 Y_{t-1}^2)/(\omega + \alpha Y_{t-1}^2)$. For any $x > 0$, $f(x) \equiv -\ln x + x \geq 1$, and hence $-E\ln M_t + EM_t \geq 1$. $Ef(M_t) \geq Ef(1) = 1$ with equality only if $M_t = 1$ with probability one. Thus, (3.7) reaches its maximum $-1/2 - E\ln(\omega_0 + \alpha_0 Y_{t-1}^2)/2$, and this occurs if and only if

$$\omega_0 + \alpha_0 Y_{t-1}^2 = \omega + \alpha Y_{t-1}^2,$$

which holds if and only if $\omega = \omega_0$ and $\alpha = \alpha_0$. Thus, $El_t(\theta)$ is uniquely maximized at θ_0. Thus, Assumption 3.1 holds and hence (a) of Theorem 3.5 holds.

We next consider Assumption 3.2. By direct differentiation,

$$\begin{aligned}
\frac{\partial l_t(\theta)}{\partial \phi} &= \frac{Y_{t-1}(Y_t - \phi Y_{t-1})}{\omega + \alpha Y_{t-1}^2}, \\
\frac{\partial l_t(\theta)}{\partial \omega} &= -\frac{1}{2(\omega + \alpha Y_{t-1}^2)}\Big[1 - \frac{(Y_t - \phi Y_{t-1})^2}{\omega + \alpha Y_{t-1}^2}\Big], \\
\frac{\partial l_t(\theta)}{\partial \alpha} &= -\frac{Y_{t-1}^2}{2(\omega + \alpha Y_{t-1}^2)}\Big[1 - \frac{(Y_t - \phi Y_{t-1})^2}{\omega + \alpha Y_{t-1}^2}\Big], \\
\frac{\partial^2 l_t(\theta)}{\partial \phi^2} &= -\frac{Y_{t-1}^2}{\omega + \alpha Y_{t-1}^2}, \\
\frac{\partial^2 l_t(\theta)}{\partial \omega^2} &= \frac{1}{2(\omega + \alpha Y_{t-1}^2)^2}\Big[1 - \frac{2(Y_t - \phi Y_{t-1})^2}{\omega + \alpha Y_{t-1}^2}\Big], \\
\frac{\partial^2 l_t(\theta)}{\partial \alpha^2} &= \frac{Y_{t-1}^4}{2(\omega + \alpha Y_{t-1}^2)^2}\Big[1 - \frac{2(Y_t - \phi Y_{t-1})^2}{\omega + \alpha Y_{t-1}^2}\Big].
\end{aligned}$$

Similarly, we can have other second-order derivatives. It is easy to show that

$$\begin{aligned}
E\sup_{\theta\in\Theta}\Big[\frac{\partial l_t(\theta)}{\partial \phi}\Big]^2 &= E\sup_{\theta\in\Theta}\frac{Y_{t-1}^2[\varepsilon_t - (\phi - \phi_0)Y_{t-1}]^2}{(\omega + \alpha Y_{t-1}^2)^2} \\
&\le E\Big[\frac{8\tilde{\phi}^2 Y_{t-1}^4}{(\underline{\omega} + \underline{\alpha} Y_{t-1}^2)^2}\Big] + 2E\Big[\frac{Y_{t-1}^2\varepsilon_t^2}{(\underline{\omega} + \underline{\alpha} Y_{t-1}^2)^2}\Big] \\
&\le C_1 + 2E\Big[\frac{Y_{t-1}^2(\omega_0 + \alpha_0 Y_{t-1}^2)}{(\underline{\omega} + \underline{\alpha} Y_{t-1}^2)^2}\Big] \le C < \infty,
\end{aligned}$$

where C_1 and C are some constants. Similarly, we can show that $E\sup_{\theta\in\Theta}[\partial l_t(\theta)/\partial\omega]^2$, $E\sup_{\theta\in\Theta}[\partial l_t(\theta)/\partial\alpha]^2$, and other cross product terms are finite. Thus, we can show that Assumption 3.2 holds. Since the density of η_t is symmetric, we can show that

$$E\Big[\frac{\partial l_t(\theta_0)}{\partial \theta}\frac{\partial l_t(\theta_0)}{\partial \theta'}\Big] = \mathrm{diag}\Big\{\Sigma_0, \frac{\kappa\Omega}{4}\Big\}.$$

This completes the proof.

3.5 Long Memory FARIMA Models

Assume the random sample $\{Y_1, \cdots, Y_n\}$ is from model (2.7). The unknown parameter is $\theta = (d, \phi_1, \cdots, \phi_p, \psi_1, \cdots, \psi_q)'$ and its true value is θ_0. We assume that the parameter space Θ is a compact subset of R^{p+q+1} and θ_0 is an interior point in Θ,

With the initial value $\widetilde{Y}_0 = \{Y_0, Y_{-1}, \cdots\}$, the parametric model can be rewritten as follows:

$$\varepsilon_t(\theta) = \psi^{-1}(B)\phi(B)(1-B)^d y_t. \tag{3.10}$$

Its first and second derivatives are as follows:

$$\frac{\partial \varepsilon_t(\theta)}{\partial d} = \log(1-B)(1-B)^d Y_t = \sum_{i=1}^{\infty} a_{1i}(\theta) Y_{t-i}, \tag{3.11}$$

$$\frac{\partial^2 \varepsilon_t(\theta)}{\partial d^2} = \log^2(1-B)(1-B)^d Y_t = \sum_{i=1}^{\infty} a_{2i}(\theta) Y_{t-i}, \tag{3.12}$$

where $\sup_{\theta\in\Theta} |a_{ji}(\theta)| = O(i^{-1-\underline{d}})$ as $j = 1,2$. Using the conditional least squares estimation, we get the objective function

$$L_n(\theta) = \sum_{t=1}^{n} \tilde{\varepsilon}_t^2(\theta),$$

where $\tilde{\varepsilon}_t(\theta) = \varepsilon_t(\theta)|_{\tilde{Y}_0=0}$. The minimizer of $L_n(\theta)$ on Θ is denoted by $\hat{\theta}_n$. We have the following results:

Theorem 3.6. *If Assumption 3.4 holds and* $d \in (0,1/2)$, *then*

$$(a)\quad \hat{\theta}_n = \theta_0 + O\Big[\Big(\frac{\log\log n}{n}\Big)^{1/2}\Big] \ a.s.,$$

$$(b)\quad \sqrt{n}(\hat{\theta}_n - \theta_0) \longrightarrow_{\mathscr{L}} N\Big(0, \sigma^2 E^{-1}\Big[\frac{\partial \varepsilon_t(\theta_0)}{\partial \theta}\frac{\partial \varepsilon_t(\theta_0)}{\partial \theta'}\Big]\Big).$$

Proof. We verify Assumptions 3.1-3.2 with $l_t(\theta) = -\varepsilon_t^2(\theta)$. For simplicity, we only consider the case with $p = q = 0$, while the general case can be similarly verified.

First, Assumption 3.7 ensures that $\{Y_t\}$ is strictly stationary and ergodic with $EY_t^2 < \infty$, and the following expansions hold:

$$Y_t = \sum_{i=0}^{\infty} c_{0i}\varepsilon_{t-i} \ \textit{and} \ \varepsilon_t(\theta) = (1-B)^d Y_t = \sum_{i=0}^{\infty} a_i(\theta) Y_{t-i}, \tag{3.13}$$

where $c_{00} = a_0(\theta) = 1$, $c_{0i} = O(i^{-1+d_0})$ and $a_i(\theta) = O(i^{-1-d})$. Since Θ is compact, there are $\underline{d}$ and $\tilde{d}$ such that $0 < \underline{d} \le d \le \tilde{d} < 0.5$. Thus, we have $\sup_{\theta\in\Theta} |a_i(\theta)| = O(i^{-1-\underline{d}})$, and hence it follows that

$$\sup_{\Theta} |\varepsilon_t(\theta)| = \sup_{\Theta} |\sum_{i=0}^{\infty} a_i(\theta) Y_{t-i}| \le |Y_t| + O(1)\sum_{i=1}^{\infty} \frac{1}{i^{1+\underline{d}}}|Y_{t-i}|.$$

By the Cauchy-Schwarz inequality, we have $E\sup_{\Theta} |\varepsilon_t(\theta)|^2 < \infty$. It is not difficult to show that $-E[\varepsilon_t^2(\theta)]$ has a unique maximum on Θ. Thus, Assumption 2.1(i) holds.

$$D_t(\theta) = -2\varepsilon_t(\theta)\frac{\partial \varepsilon_t(\theta)}{\partial d} \text{ and } P_t(\theta) = 2\Big[\frac{\partial \varepsilon_t(\theta)}{\partial d}\Big]^2 + 2\varepsilon_t(\theta)\frac{\partial^2 \varepsilon_t(\theta)}{\partial d^2}.$$

Using these, it is straightforward to show that Assumption 3.1(ii)-(iii) holds.

We next consider Assumption 3.3. For simplicity, let $\widetilde{Y}_0 = (0, 0, \cdots)$. By (5.1),

$$\begin{aligned}
&E[\sup_{\Theta} |\varepsilon_t(\theta) - \tilde{\varepsilon}_t(\theta)|]^2 \\
&\quad = E\Big[\sup_{\Theta} |\sum_{i=t}^{\infty} a_i(\theta) Y_{t-i}|\Big]^2 \le CE\Big(\sum_{i=t}^{\infty} \frac{1}{i^{1+\underline{d}}} |Y_{t-i}|\Big)^2 = O(t^{-2\underline{d}}).
\end{aligned}$$

It is readily shown that $E \sup_{\Theta} \tilde{\varepsilon}_t^2(\theta)$ is bounded uniformly in t. Thus, by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&E \sup_{\Theta} |\varepsilon_t^2(\theta) - \tilde{\varepsilon}_t^2(\theta)| \\
&\quad \le \{E[\sup_{\Theta} |\varepsilon_t(\theta) + \tilde{\varepsilon}_t(\theta)|]^2 E[\sup_{\Theta} |\varepsilon_t(\theta) - \tilde{\varepsilon}_t(\theta)|]^2\}^{1/2} = O(t^{-\underline{d}}),
\end{aligned}$$

so that Assumption 2.2(i) holds. Similarly, we can show that Assumption 2.2(iii) holds.

We now verify Assumption 2.2(ii'). Denote

$$\begin{aligned}
A_t &= \varepsilon_t(\theta_0) - \tilde{\varepsilon}_t(\theta_0) = \sum_{i=t}^{\infty} a_i(\theta_0) Y_{t-i}, \\
A_{1t} &= \frac{\partial \varepsilon_t(\theta_0)}{\partial d} - \frac{\partial \tilde{\varepsilon}_t(\theta_0)}{\partial d}, \\
A_{2t} &= \frac{\partial \varepsilon_t(\theta_0)}{\partial d} - v_t = -\sum_{i=t}^{\infty} \frac{1}{i} \varepsilon_{t-i},
\end{aligned}$$

where $v_t = -\sum_{i=1}^{t-1} \varepsilon_{t-i}/i$. We first make the following decomposition:

$$\begin{aligned}
\widetilde{D}_t(\theta_0) - D_t(\theta_0) &= 2\varepsilon_t(\theta_0) \frac{\partial \varepsilon_t(\theta_0)}{\partial d} - 2\tilde{\varepsilon}_t(\theta_0) \frac{\partial \tilde{\varepsilon}_t(\theta_0)}{\partial d} \\
&= 2\varepsilon_t(\theta_0) A_{1t} + 2\frac{\partial \tilde{\varepsilon}_t(\theta_0)}{\partial d} A_t \\
&= 2\varepsilon_t(\theta_0) A_{1t} + 2A_t v_t + 2A_t A_{2t} - 2A_t A_{1t}. \qquad (3.14)
\end{aligned}$$

Since $E(Y_t Y_{t+r}) = O(|r|^{-1+2d_0})$, we have

$$\begin{aligned}
EA_t^2 &= \sum_{i=t}^{\infty} a_i^2(\theta_0) EY_{t-1-i}^2 + 2\sum_{i=t}^{\infty} \sum_{r=1}^{\infty} a_i(\theta_0) a_{i+r}(\theta_0) E(Y_{t-1-i} Y_{t-1-i-r}) \\
&= O(1)[\sum_{i=t}^{\infty} \frac{1}{i^{2(1+d_0)}} + 2\sum_{i=t}^{\infty} \sum_{r=1}^{\infty} \frac{1}{i^{1+d_0}(i+r)^{1+d_0} r^{1-2d_0}}] \\
&\le O(1)[\sum_{i=t}^{\infty} \frac{1}{i^{2(1+d_0)}} + 2\sum_{i=t}^{\infty} \frac{1}{i^{1+d_0}} \int_1^{\infty} \frac{1}{(i+x)^{1+d_0} x^{1-2d_0}} dx] \\
&\le O(1)[\sum_{i=t}^{\infty} \frac{1}{i^{2(1+d_0)}} + 2\sum_{i=t}^{\infty} \frac{1}{i^2} \int_0^{\infty} \frac{1}{(1+z)^{1+d_0} z^{1-2d_0}} dz] = O(t^{-1}). \qquad (3.15)
\end{aligned}$$

Using a similar method, we can show that

$$EA_{it}^2 = O(t^{-1}) \text{ and } E(A_t A_{it}) = O(t^{-1}) \text{ as } i = 1, 2. \tag{3.16}$$

As in the proof of (4.6), using (5.4), we can show that

$$\lim_{l\to\infty} P\Big(\max_{l\le n<\infty} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |A_t A_{it}| > \varepsilon\Big) = 0, \text{ as } i = 1, 2. \tag{3.17}$$

We next show that $\sum_{t=1}^{n} \varepsilon_t A_{1t}/\sqrt{n} = o(1)$ a.s. as $n \to \infty$, which is equivalent to

$$\lim_{l\to\infty} P\Big(\max_{l\le n<\infty} \frac{1}{\sqrt{n}} |\sum_{t=1}^{n} \varepsilon_t A_{1t}| > \varepsilon\Big) = 0. \tag{3.18}$$

By the Kronecker Lemma in Hall and Hedye (1981, p.31), it is sufficient to show that

$$S_k = \sum_{t=1}^{k} \frac{1}{\sqrt{t}} A_{1t}\varepsilon_t \text{ converges a.s..} \tag{3.19}$$

By (5.4), it follows that

$$E|\sum_{t=s}^{k} \frac{1}{\sqrt{t}} A_{1t}\varepsilon_t|^2 = O(1)\Big(\sum_{t=s}^{k} \frac{1}{t^2}\Big) = O(1)\Big(\sum_{t=s}^{k} \frac{1}{t^{1+\nu}}\Big)^{\alpha}, \tag{3.20}$$

for any integer $0 < s \le k$ and some $\alpha > 1$, where $O(1)$ holds uniformly in k and s. Consider the subsequence $\{S_{2^k} : k = 0, 1, \cdots\}$. By (5.8), we have

$$E|S_{2^{k+1}} - S_{2^k}| \le O(1)\Big(\sum_{t=2^k+1}^{2^{k+1}} \frac{1}{t^{1+\varepsilon}}\Big) \le O\Big(\frac{1}{2^{\varepsilon k}}\Big),$$

for some $\varepsilon > 0$. By this equation and the monotone convergence theorem, we have

$$E \lim_{n\to\infty} \sum_{k=0}^{n} |S_{2^{k+1}} - S_{2^k}| = \lim_{n\to\infty} E \sum_{k=0}^{n} |S_{2^{k+1}} - S_{2^k}| \le O\Big(\sum_{k=0}^{\infty} \frac{1}{2^{\varepsilon k}}\Big) < \infty.$$

Thus, $\sum_{k=0}^{n} |S_{2^{k+1}} - S_{2^k}|$ converges a.s. as $n \to \infty$, and hence

$$\lim_{n\to\infty} S_{2^{n+1}} = X_1 + \lim_{n\to\infty} \sum_{k=0}^{n} (S_{2^{k+1}} - S_{2^k}) \text{ converges a.s..} \tag{3.21}$$

By (5.8) and Theorem 12.2 in Billingsley (1968), it follows that, for any $\Delta > 0$,

$$P\Big(\max_{2^k<n\le 2^{k+1}} |S_n - S_{2^k}| \ge \Delta\Big) \le O(1)\Big(\sum_{t=2^k+1}^{2^{k+1}} \frac{1}{t^{1+\nu}}\Big)^{\alpha} = O\Big(\frac{1}{2^{\varepsilon_1 k}}\Big),$$

for some $\varepsilon_1 > 0$. By the Borel-Canteli Lemma, we can claim that

$$\max_{2^k < n \le 2^{k+1}} |S_n - S_{2^k}| \to 0 \; a.s. \; as \; k \to \infty. \tag{3.22}$$

By Lemma 2.3.1 in Stout (1974) and (5.9)-(5.10), we know that (5.7) holds.

Note that v_s is independent of A_{1t} as $s \ge 2$, so that

$$\begin{aligned} E|\sum_{t=s}^{k} \frac{1}{\sqrt{t}} A_{1t} \nu_t|^2 &= E|\sum_{t=s}^{k} \sum_{t_1=s}^{k} \frac{1}{\sqrt{t}} \frac{1}{\sqrt{t_1}} E(A_{1t} A_{1t_1}) E(\nu_t \nu_{t_1}) \\ &= O(1) \Big(\sum_{t=s}^{k} \frac{1}{t^{1+1/2}} \Big)^2 = O(1) \Big(\sum_{t=s}^{k} \frac{1}{t^{1+\nu}} \Big)^{\alpha}, \end{aligned}$$

for any integer $0 < s \le k$, where $O(1)$ holds uniformly in k and s. Using this equation and a similar method as for (5.6), we can show that

$$\lim_{l \to \infty} P(\max_{l \le n < \infty} \frac{1}{\sqrt{n}} | \sum_{t=1}^{n} v_t A_t | > \varepsilon) = 0. \tag{3.23}$$

By (5.2), (5.5)-(5.6) and (5.11), we can show that Assumption 2.2(ii') holds. □

3.6 Exercises

Exercise 3.1. Prove that Ω_0 in Theorem 2.3 is positive definite.

Exercise 3.2. Let $h_t(\theta)$ be defined as in Section 3.2. Show that $h_t(\theta) = h_t(\theta_0$ if and only $\theta = \theta_0$.

Exercise 3.3. For the smooth TAR(1) model in Exercise 2.3, show that the LSE of $\theta = (\phi_1, \phi_2, r, z)'$ is consistent and asymptotically normal.

Chapter 4
Model Diagnostic Checking

4.1 Ljung-Box Tests

Given the random sample $\{Y_1, Y_2, \cdots, Y_n\}$ and the initial value Y_0, we use the following model to fit the data:

$$Y_t = m(\theta, X_{t-1}) + \varepsilon_t, \tag{4.1}$$

where $X_t = (Y_t, Y_{t-1}, \cdots)$ and ε_t is iid with mean zero and variance σ^2. Assume that the conditions of Theorem 3.1 are satisfied. We can have the following expansion:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\frac{\Sigma^{-1}}{2\sqrt{n}} \sum_{t=1}^{n} D_t(\theta_0) + o_p(1).$$

After we estimate the model, we have the residuals:

$$\hat{\varepsilon}_t = \varepsilon_t(\hat{\theta}_n) = Y_t - m(\hat{\theta}_n, X_{t-1}).$$

The basic idea is that if model (4.1) is correct, then $\hat{\varepsilon}_t$ should be very close to ε_t. We know that ε_t is iid. Then, $\hat{\varepsilon}_t$ should be almost iiid or at least it should be uncorrelated. Thus, a simple way is to plot out the sample ACF of $\hat{\varepsilon}_t$ and see if they are close to zero. However, how close is close? We need to qualify this difference. First, we see the following expansion:

$$\begin{aligned}
\hat{\varepsilon}_t &= Y_t - m(\hat{\theta}_n, X_{t-1}) \\
&= \varepsilon_t + m(\theta_0, X_{t-1}) - m(\hat{\theta}_n, X_{t-1}) \\
&= \varepsilon_t - (\hat{\theta}_n - \theta_0)' \frac{\partial m(\theta_0, X_{t-1})}{\partial \theta} - (\hat{\theta}_n - \theta_0)' \frac{\partial^2 m(\xi^*, X_{t-1})}{\partial \theta \partial \theta'} (\hat{\theta}_n - \theta_0) \\
&= \varepsilon_t - (\hat{\theta}_n - \theta_0)' \frac{\partial m(\theta_0, X_{t-1})}{\partial \theta} + O_p\left(\frac{1}{n}\right).
\end{aligned}$$

Thus, the sample ACF of $\hat{\varepsilon}_t$ at the lag k is

$$\hat{\rho}_k = \frac{\sum_{t=k}^n \hat{\varepsilon}_{t-k}\hat{\varepsilon}_t}{\sum_{t=1}^n \hat{\varepsilon}_t^2}.$$

Note that

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^n \hat{\varepsilon}_t^2 &= \frac{1}{n}\sum_{t=1}^n \left[\varepsilon_t - (\hat{\theta}_n - \theta_0)'\frac{\partial m(\theta_0, X_{t-1})}{\partial \theta} + O_p\left(\frac{1}{n}\right)\right]^2 \\
&= \frac{1}{n}\sum_{t=1}^n \varepsilon_t^2 - \frac{2}{n}(\hat{\theta}_n - \theta_0)'\sum_{t=1}^n u_{t-1}\varepsilon_t \\
&\quad + \frac{1}{n}(\hat{\theta}_n - \theta_0)'\sum_{t=1}^n u_{t-1}u_{t-1}'(\hat{\theta}_n - \theta_0) + O_p\left(\frac{1}{n}\right) \\
&= \sigma^2 + o_p(1),
\end{aligned}$$

where $u_{t-1} = \partial m(\theta_0, X_{t-1})/\partial \theta$. Thus,

$$\sqrt{n}\hat{\rho}_k = \frac{1}{\sigma^2 + o_p(1)}\frac{1}{\sqrt{n}}\sum_{t=k}^n \hat{\varepsilon}_{t-k}\hat{\varepsilon}_t,$$

and

$$\begin{aligned}
\sum_{t=k}^n \hat{\varepsilon}_{t-k}\hat{\varepsilon}_t &= \sum_{t=k}^n \varepsilon_{t-k}\varepsilon_t - (\hat{\theta}_n - \theta_0)'\sum_{t=k}^n u_{t-k-1}\varepsilon_t \\
&\quad - (\hat{\theta}_n - \theta_0)'\sum_{t=k}^n u_{t-1}\varepsilon_{t-k} \\
&\quad + (\hat{\theta}_n - \theta_0)'\sum_{t=k}^n u_{t-k-1}u_{t-1}'(\hat{\theta}_n - \theta_0) + O_p(1) \\
&= \sum_{t=k}^n \varepsilon_{t-k}\varepsilon_t - n(\hat{\theta}_n - \theta_0)'A_k + O_p(1),
\end{aligned}$$

where $A_k = E\left(u_{t-1}\varepsilon_{t-k}\right)$. Thus,

$$\sigma^2\sqrt{n}\hat{\rho}_k = \frac{1}{\sqrt{n}}\sum_{t=k}^n \varepsilon_{t-k}\varepsilon_t - \sqrt{n}(\hat{\theta}_n - \theta_0)'A_k + o_p(1).$$

Usually, we need to check ACF up to M lags. Thus, we consider the vector as follows:

$$
\begin{aligned}
\sigma^2\sqrt{n}\begin{pmatrix}\hat{\rho}_1\\ \hat{\rho}_2\\ \cdot\\ \cdot\\ \hat{\rho}_M\end{pmatrix} &= \frac{1}{\sqrt{n}}\sum_{t=M}^{n}\begin{pmatrix}\varepsilon_{t-1}\\ \varepsilon_{t-2}\\ \cdot\\ \cdot\\ \varepsilon_{t-M}\end{pmatrix}\varepsilon_t - \begin{pmatrix}A_1'\\ A_2'\\ \cdot\\ \cdot\\ A_M'\end{pmatrix}\sqrt{n}(\hat{\theta}_n-\theta_0)+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{t=M}^{n}\left[\begin{pmatrix}\varepsilon_{t-1}\\ \varepsilon_{t-2}\\ \cdot\\ \cdot\\ \varepsilon_{t-M}\end{pmatrix} - \begin{pmatrix}A_1'\\ A_2'\\ \cdot\\ \cdot\\ A_M'\end{pmatrix}\Sigma^{-1}u_{t-1}\right]\varepsilon_t+o_p(1)\\
&\equiv \frac{1}{\sqrt{n}}\sum_{t=M}^{n}R_{t-1}\varepsilon_t+o_p(1)\\
&\to^d N\left[0,\sigma^2E\left(R_{t-1}R_{t-1}'\right)\right],
\end{aligned}
$$

as $n\to\infty$, by the central limiting theorem, where

$$E\left(R_{t-1}R_{t-1}'\right)=\sigma^2 I_M - A'\Sigma^{-1}A,$$

with $A=[A_1,A_2,\cdots,A_M]$. It follows that

$$\sqrt{n}\begin{pmatrix}\hat{\rho}_1\\ \hat{\rho}_2\\ \cdot\\ \cdot\\ \hat{\rho}_M\end{pmatrix}\to^d N\left[0,I_M-\frac{A'\Sigma^{-1}A}{\sigma^2}\right],$$

as $n\to\infty$. Thus,

$$\mathcal{Q}=n(\hat{\rho}_1,\cdots,\hat{\rho}_M)\left[I_M-A'\Sigma^{-1}A/\sigma^2\right]^{-1}(\hat{\rho}_1,\cdots,\hat{\rho}_M)'\to\chi^2(M),$$

as $n\to\infty$. $\mathcal{Q}$ is called the portmanteau test.

To look at the covariance of $\hat{\rho}_i$ closely, we consider the AR(1) model:

$$Y_t=\phi Y_{t-1}+\varepsilon_t.$$

In this case,

$$
\begin{aligned}
&u_{t-1}=Y_{t-1}=\sum_{i=0}^{\infty}\phi^i\varepsilon_{t-1-i},\\
&A_k=E\left(u_{t-1}\varepsilon_{t-k}\right)=\sigma^2\phi^{k-1},\\
&\Sigma=Eu_{t-1}^2=\sigma^2\sum_{i=0}^{\infty}\phi^{2i}=\frac{\sigma^2}{1-\phi^2}.
\end{aligned}
$$

Thus,

$$A'\Sigma^{-1}A = \sigma^2(1-\phi^2)\begin{pmatrix} 1 & \phi & \cdot\ \cdot & \phi^{M-1} \\ \phi & \phi^2 & \cdot\ \cdot & \phi^M \\ \cdot & & & \\ \cdot & & & \\ \phi^{M-1} & \phi^M & \cdot\ \cdot & \phi^{2M-2} \end{pmatrix}$$

$$= \sigma^2(1-\phi^2)Q'\begin{pmatrix} 1 & 0 & \cdot\ \cdot & 0 \\ 0 & \phi^2 & \cdot\ \cdot & 0 \\ \cdot & & & \\ \cdot & & & \\ 0 & 0 & \cdot\ \cdot & \phi^{2M-2} \end{pmatrix} Q,$$

where $Q'Q = I_M$. It follows that

$$I_M - A'\Sigma^{-1}A/\sigma^2 = Q'diag\{\phi^2, 1-\phi^2+\phi^4, \cdots, 1-\phi^{2M-2}+\phi^{2M}\}Q,$$

and its rank is only $M-1$. Thus,

$$\mathscr{Q} \to^d \chi^2_{M-1}, \text{ as } n \to \infty.$$

In general, we have

$$\mathscr{Q} \to^d \chi^2_{M-p}, \text{ as } n \to \infty,$$

where p is the number of estimated parameters. Note that ϕ^{2k} is very small, as k is large. Thus, to make it easy, Box and Pierce (1970) considered the following modified test

$$\mathscr{Q}_1(M) = n\sum_{k=1}^{M} \hat{\rho}_k^2 \sim \chi^2(M-p).$$

A further modification is Ljung and Box (1978) test as follows:

$$\mathscr{Q}_2(M) = n(n+2)\sum_{k=1}^{M} \frac{\hat{\rho}_k^2}{n-k} \sim \chi^2(M-p).$$

4.2 Li-Mak Tests

Given the random sample $\{Y_1, \cdots, Y_n\}$ and the initial value $\tilde{Y}_0 = \{Y_0, Y_{-1}, \cdots\}$, we fit the data by the model:

$$Y_t = m_t(\theta) + \eta_t\sqrt{h_t(\theta)} \tag{4.2}$$

where ηt is iid with zero mean and variance 1, $m_t(\theta) = m(\theta, X_{t-1})$ and $h_t(\theta) = h(\theta, X_{t-1}) > 0$ a.s.. Assume that the conditions in Theorem 3.1 are satisfied and $\hat{\theta}_n$ is the estimator in Theorem 3.1. We have the expansion:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{\Sigma^{-1}}{\sqrt{n}} \sum_{t=1}^{n} D_t(\theta_0) + o_p(1),$$

where Σ and D_t are defined as in Section 3.1. Now, the scaled residual, denoted by $\hat{\eta}_t$, is defined as

$$\hat{\eta}_t = \eta_t(\hat{\theta}_n) \text{ and } \eta_t(\theta) = \frac{Y_t - m(\theta, X_{t-1})}{\sqrt{h_t(\theta)}}.$$

Using Taylor's expansion, we have

$$\hat{\eta}_t = \eta_t - \frac{\partial \eta_t(\theta_0)}{\partial \theta'}(\hat{\theta}_n - \theta_0) + O_p\left(\frac{1}{n}\right) \frac{\partial^2 \eta_t(\theta^*)}{\partial \theta \partial \theta'},$$

where θ^* lies between θ_0 and $\hat{\theta}_n$. We can show that

$$\frac{1}{n}\sum_{t=1}^{n} \hat{\eta}_t^2 = \frac{1}{n}\sum_{t=1}^{n} \eta_t^2 + o_p(1) = 1 + o_p(1),$$
$$\frac{1}{\sqrt{n}}\sum_{t=k}^{n} \hat{\eta}_{t-k}\hat{\eta}_t = \frac{1}{\sqrt{n}}\sum_{t=k}^{n} \eta_{t-k}\eta_t - A_k'\sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1),$$

where

$$A_k = E\left(\frac{\partial \eta_t(\theta_0)}{\partial \theta}\eta_{t-k}\right).$$

Let

$$\hat{\rho}_k = \frac{\sum_{t=k}^{n} \hat{\eta}_{t-k}\hat{\eta}_t}{\sum_{t=1}^{n} \hat{\eta}_t^2}.$$

Assume that $E[\eta_t D_t(\theta_0] = 0$. Then,

$$\begin{aligned}
\sqrt{n}\begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \cdot \\ \cdot \\ \hat{\rho}_M \end{pmatrix} &= \frac{1}{\sqrt{n}}\sum_{t=M}^{n} \begin{pmatrix} \eta_{t-1} \\ \eta_{t-2} \\ \cdot \\ \cdot \\ \eta_{t-M} \end{pmatrix} \eta_t - \begin{pmatrix} A_1' \\ A_2' \\ \cdot \\ \cdot \\ A_M' \end{pmatrix} \sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{t=M}^{n} \left[\begin{pmatrix} \eta_{t-1} \\ \eta_{t-2} \\ \cdot \\ \cdot \\ \eta_{t-M} \end{pmatrix} \eta_t - \begin{pmatrix} A_1' \\ A_2' \\ \cdot \\ \cdot \\ A_M' \end{pmatrix} \Sigma^{-1} D_t(\theta_0) \right] \\
&\quad + o_p(1) \\
&\to^d N\left[0, I_M - A'\Sigma^{-1}\Omega\Sigma^{-1}A\right],
\end{aligned}$$

as $n \to \infty$, by the central limiting theorem, where $A = [A_1, A_2, \cdots, A_M]$. Thus,

$$\mathscr{Q}(M) = n(\hat{\rho}_1, \cdots, \hat{\rho}_M)\left[I_M - \frac{\kappa}{2}A'\Sigma^{-1}A\right]^{-1}(\hat{\rho}_1, \cdots, \hat{\rho}_M)' \to^d \chi^2(M),$$

as $n \to \infty$. The simple modification as the Ljung-Box test is

$$\mathscr{Q}(M) = n(n+2)\sum_{k=1}^{M}\frac{\hat{\rho}_k^2}{n-k}.$$

Li and Mak (1994) used the following statistic to construct test:

$$\hat{r}_k = \frac{\sum_{t=k}^{n}(\hat{\eta}_{t-k}^2 - 1)(\hat{\eta}_t^2 - 1)}{\sum_{t=1}^{n}(\hat{\eta}_t^2 - 1)^2}.$$

Under some regualr conditions, we can show that

$$\frac{1}{n}\sum_{t=1}^{n}(\hat{\eta}_t^2 - 1)^2 = \kappa + o_p(1),$$

where $\kappa = E\eta_t^4 - 1$. By Taylor's expansion, we have

$$\hat{\eta}_t^2 = \eta_t^2 - \frac{\partial \eta_t^2(\theta_0)}{\partial \theta}(\hat{\theta}_n - \theta_0)' + \frac{\partial^2 \eta_t^2(\theta^*)}{\partial \theta \partial \theta'}O_p(\frac{1}{n}),$$

where θ^* is between θ_0 and $\hat{\theta}^n$ and $O_p(1)$ holds uniformly in $t = 1, \cdots, n$. Thus,

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=k}^{n}(\hat{\eta}_{t-k}^2 - 1)(\hat{\eta}_t^2 - 1) &= \frac{1}{\sqrt{n}}\sum_{t=k}^{n}(\eta_{t-k}^2 - 1)(\eta_t^2 - 1) \\
&\quad - \frac{1}{\sqrt{n}}\sum_{t=k}^{n}(\eta_{t-k}^2 - 1)\frac{\partial \eta_t^2(\theta_0)}{\partial \theta}(\hat{\theta}_n - \theta_0)' + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{t=k}^{n}(\eta_{t-k}^2 - 1)(\eta_t^2 - 1) - B_k'\sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1),
\end{aligned}$$

where

$$B_k = E\left[(\eta_{t-k}^2 - 1)\frac{\partial \eta_t^2(\theta_0)}{\partial \theta}\right].$$

Thus,

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=k}^{n}(\hat{\eta}_{t-k}^2 - 1)(\hat{\eta}_t^2 - 1) &= \frac{1}{\sqrt{n}}\sum_{t=k}^{n}\left[(\eta_{t-k}^2 - 1)(\eta_t^2 - 1) - B_k'\Sigma^{-1}D_t(\theta_0)\right] \\
&\equiv \frac{1}{\sqrt{n}}\sum_{t=k}^{n}\tilde{R}_{kt}' + o_p(1).
\end{aligned}$$

Assume that $E[(\eta_t^2 - 1)D_t(\theta_0)] = 0$. By the central limiting theorem, we have

$$\sqrt{n}\begin{pmatrix}\hat{r}_1\\ \hat{r}_2\\ \cdot\\ \cdot\\ \hat{r}_M\end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{t=M}^{n}\begin{pmatrix}\tilde{R}'_{1t}\\ \tilde{R}'_{2t}\\ \cdot\\ \cdot\\ \tilde{R}'_{Mt}\end{pmatrix} + o_p(1)$$
$$\to^d N(0, \kappa EB_t), \text{ as } n \to \infty,$$

where

$$EB_t = E\left[(\tilde{R}_{1t}, \cdots, \tilde{R}_{Mt})'(\tilde{R}_{1t}, \cdots, \tilde{R}_{Mt})\right]$$
$$= \kappa I_M - B'\Sigma^{-1}\Omega\Sigma^{-1}B,$$

with $B = [B_1, B_2, \cdots, B_M]$. It follows that

$$\mathcal{Q}(M) = n(\hat{r}_1, \cdots, \hat{r}_M)\left[\kappa^2 I_M - \kappa B'\Sigma^{-1}\Omega\Sigma^{-1}B\right]^{-1}(\hat{r}_1, \cdots, \hat{r}_M)' \to \chi^2(M),$$

as $n \to \infty$. Especially, when $\eta_t \sim N(0,1)$, $\kappa = 3$.

4.3 Score Based Test

Given the random sample $\{Y_1, \cdots, Y_n\}$ from model (6.2) and the initial values $\{Y_s : s \le 0\}$, we first estimate θ_0 under the null H_0 and assume its estimator $\hat{\theta}_n$ satisfies:

Assumption 4.1

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = (\Sigma/\gamma)^{-1}\sum_{t=1}^{n} D_t(\theta_0)/\sqrt{n} + o_p(1), \tag{4.3}$$

where $\Sigma = E[D_t(\theta_0)D_t'(\theta_0)]$ and γ is a positive constant.

The estimator $\hat{\theta}_n$ can be quite flexible: it can be a QMLE, a LSE, an M-estimator or a LAD-estimator, among others. γ depends on the estimating method. When $\hat{\theta}_n$ is the MLE, $\gamma = 1$. Here, $D_t(\theta)$ is a relative score function. Our test statistic is based on a score-marked empirical process $\{T_n(x, \theta_0)\}$, where

$$T_n(x, \theta) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} D_t(\theta) I\{Y_{t-1} \le x\}. \tag{4.4}$$

Denote $\Sigma_x = E[D_t(\theta_0)D_t'(\theta_0)I\{Y_{t-1} \le x\}]$. Let $\hat{\Sigma}_{nx} = \sum_{t=1}^{n}[D_t(\hat{\theta}_n)D_t'(\hat{\theta}_n)I\{Y_{t-1} \le x\}]/n$ and $\hat{\Sigma}_n = \hat{\Sigma}_\infty$ be the estimators of Σ_x and Σ, respectively. Now, we define our general test statistic via the linear transformation of $T_n(x, \hat{\theta}_n)$ as follows:

$$S_n^a = \max_{a \le x \le \infty} \frac{[\beta'\hat{\Sigma}_{nx}^{-1}T_n(x, \hat{\theta}_n)]^2}{\beta'(\hat{\Sigma}_{na}^{-1} - \hat{\Sigma}_n^{-1})\beta}, \tag{4.5}$$

where β is a nonzero $p \times 1$ constant vector. When $p = 1$, S_n^a is equivalent to the weighted LR ratio test for $H_0 : Y_t = \mu_t(\theta) + \varepsilon_t$ against the alternative (1.4), with weight $(1 - \Sigma\Sigma_x^{-1})/(1 - \Sigma\Sigma_a^{-1})$. We can replaced the threshold variable Y_{t-1} by Y_{t-r} or $\hat{\theta}_n'(Y_{t-1}, \cdots, Y_{t-p})'$ used by Stute et al. (2006). In fact, it can be replaced by any function $\xi_{t-1} = g(Y_{t-1}, Y_{t-2}, \cdots)$ and our theory in the next section still holds as long as Assumption A.2 is satisfied. Our approach can be applied for the multivariate time series models with Y_{t-1} replaced by a suitable choice of ξ_{t-1}.

Typically, the quantity a is taken as corresponding an early quantile of the process values. It should ensure that $\hat{\Sigma}_{na}^{-1}$ exists. Note that $\max_\beta S_n^a$ does not have a limiting distribution as simple as that of S_n^a, see the discussion for Theorem 3.2 in Section 3. S_n^a will be invariant with respect to $\|\beta\|$. If we denote the normalized score $\hat{\Sigma}_{nx}^{-1} T_n(x, \hat{\theta}_n)$ by $U_n(x) = (u_1(x), \cdots, u_n(x))'$, then $\beta' U_n(x) = \sum_{i=1}^p \beta_i u_i(x)$ can be interpreted as the weighted score function. The optimal choice of β remains an open problem. A simple choice for β is $(1, \cdots, 1)'$, which means that we take equal weigh to each $u_i(x)$. The simulation in Section 4 shows that this choice with a being the 5% quantile of data can be recommended in terms of both size and power. Another possible choice is $\beta = \hat{\theta}_n$, but the simulation results in Section 4 show that this choice is not as good as the previous one.

We can construct the test S_n^a as long as we have $D_t(\theta)$. In fact, $D_t(\hat{\theta}_n)$ is always available when ones estimate model (1.1). S_n^a generally is just the maximum of n different values. It is easy to implement and is as simple as the Ljoung-Box test and the Meclod-Li test. (2.1) holds under usual regularity conditions in the time series. We next illustrate what the score $D_t(\theta)$ is via various time series models and how standard to allow the application of our test statistics.

To get the null distribution of S_n^a, we introduce two assumptions as follows.

Assumption 4.2 *$D_t(\theta_0)$ is a $\mathscr{F}_t$−measurable, strictly stationary and ergodic martingale difference with $E(\|D_t(\theta_0)\|^{2(1+\iota)}) < \infty$ for some $\iota > 0$.*

Assumption 4.3 *$D_t(\theta)$ has the expansion: $D_t(\theta) - D_t(\theta_0) = P_t(\theta^*)(\theta - \theta_0)'$, where θ^* lies between θ and θ_0, and, for any fixed $C > 0$,*

$$\sup_{\sqrt{n}\|\theta-\theta_0\| \le C} \left\| \frac{1}{n} \sum_{t=1}^{n} \left[\gamma P_t(\theta) - D_t(\theta_0) D_t'(\theta_0) \right] \right\| = o_p(1).$$

Here, $P_t(\theta)$ is the derivative of $D_t(\theta)$ if it is differentiable. Assumptions 2.1 and 3.2 hold for most of strictly stationary time series models. The moment condition in Assumption 3.1 is almost minimal. Using the assumptions 2.1 and 3.2, we can show that following lemma holds. The standard point-wise argument with Assumption A6.1, we can show that

$$\sup_{x \in R \cup \{\infty\}} \|\hat{\Sigma}_{nx} - \Sigma_x\| = o_p(1), \tag{4.6}$$

$$\sup_{x \in R} \left\| T_n(x, \hat{\theta}_n) - T_n(x, \theta_0) - \frac{\Sigma_x \Sigma^{-1}}{\sqrt{n}} \sum_{t=1}^{n} D_t(\theta_0) \right\| = o_p(1). \tag{4.7}$$

By (6.6)-(6.7) and Theorem A.1, we have the following weak convergence.

Theorem 4.1. *Suppose that Assumptions A6.1-A6.3 hold and η_t has a bounded density f. If Σ_x is positive definite for each $x \in R$, then we have*

$$T_n(x, \hat{\theta}_n) \Longrightarrow G_p(x) \ \text{in} \ D^p[R],$$

under H_0, where $\{G_p(x) : x \in R\}$ is a $p-$dimensional Gaussian process with mean zero and covariance kernel $K_{xy} = \Sigma_{x\wedge y} - \Sigma_x \Sigma^{-1} \Sigma_y$, almost all the paths of $G_p(x)$ are continuous in x, and $D^p[R]$ be defined as in Theorem A.1 in Appendix.

We first note that $\Sigma_x^{-1} T_n(x, \hat{\theta}_n) \Rightarrow G_{0p}(x)$ in $D^p[R]$ under H_0, where $\{G_{0p}(x)\}$ is a $p \times 1$ vector Gaussian process on $R \cup \{\pm\infty\}$ with mean zero and covariance kernel $K_{xy} = \Sigma_{x\vee y}^{-1} - \Sigma^{-1}$. An important observation is that $\{G_{0p}(x)\}$ has an independent increment with $E\{[G_{0p}(x) - G_{0p}(y)][G_{0p}(x) - G_{0p}(y)]'\} = \Sigma_y^{-1} - \Sigma_x^{-1}$ when $x > y$. For the marked empirical processes, the covariance kernel usually has the form $\sigma_{x\wedge y} - u_x' \Sigma^{-1} u_y$. In Theorem 3.1, we have $\sigma_{x\wedge x} = u_x = \Sigma_x$. This is the key for the process $\{G_{0p}(x)\}$ to have independent increments. For other marked processes such as the residual-marked process, since $\sigma_{x\wedge x} \neq u_x$, ones cannot obtain independent increment process for any normalization. We take an advantage from the estimator $\hat{\theta}_n$ and the score $D_t(\theta)$ plays a critical role in our approach.

Since the components of $G_{0p}(x)$ are dependent, its covariance kernel does not have a simple transformation and so is its quadratic form or the maxima of all its components. However, for any constant β, $\beta' G_{0p}(x)$ has a rather simple covariance kernel $\sigma_x \wedge \sigma_y$, where $\sigma_x = \beta'(\Sigma_x^{-1} - \Sigma^{-1})\beta$. For any finite constant a, σ_x/σ_a is a continuous and strictly decreasing function in terms of x and it runs through $[0, 1]$ when x runs from ∞ to a. Thus, $B(\tau) \equiv \beta' G_{0p}(x)/\sqrt{\sigma_a}$ is a standard Brownian motion on $\tau = \sigma_x/\sigma_a \in [0, 1]$. Let $b > a$ be a constant and

$$S_n^a(b) = \max_{a\le x\le b} \frac{[\beta' \hat{\Sigma}_{nx}^{-1} T_n(x, \hat{\theta}_n)]^2}{\beta'(\hat{\Sigma}_{na}^{-1} - \hat{\Sigma}_n^{-1})\beta}.$$

By Theorem 3.1 and the continuous mapping theorem, we have the main result:

Theorem 4.2. *If the assumptions of Theorem 6.1 hold and $\Sigma - \Sigma_x$ is positive definite for each $x \in R$, then, for any $p \times 1$ nonzero constant vector β, we have*

$$\lim_{b\to\infty} \lim_{n\to\infty} P[S_n^a(b) \le x] = P\left[\max_{\tau\in[0,1]} B^2(\tau) \le x\right]$$

for any $a > 0$ and any $x \in R$, where $B(\tau)$ is a standard Brownian motion on $C[0, 1]$.

From this theorem, the constant C_α such that $P[\max_{\tau\in[0,1]} B^2(\tau) \ge C_\alpha] = \alpha$ can be used as an approximating critical value of S_n^a for rejecting the null H_0 at the significance level α. From Shorack and Wellner (1986, p.34), we have

$$P\Big[\max_{\tau\in[0,1]} B^2(\tau) \ge x\Big] = 1 - \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\Big[-\frac{(2k+1)^2\pi^2}{8x}\Big],$$

for all $x > 0$, and $C_{0.1} = 3.83$, $C_{0.05} = 5.00$ and $C_{0.01} = 7.63$.

We now study the asymptotically local power of S_n^a. Let $r_{1t} = r_1(Y_{t-1}, Y_{t-2}, \cdots)$ and $r_{2t} = r_2(Y_{t-1}, Y_{t-2}, \cdots)$ are two $\mathscr{F}_{t-1}$−measurable random variables as $t = 0, \pm 1, \cdots$. Consider the local alternative hypothesis

$$H_{1n} : \left[\mu(\cdot) + \frac{r_1(\cdot)}{\sqrt{n}},\ h(\cdot) + \frac{r_2(\cdot)}{\sqrt{n}}\right] \in \mathscr{M}.$$

Assume that η_t is normal and independent of Y_s as $s \le 0$ under both H_0 and H_{1n}. Let P_{0n} and P_{1n} be the joint distributions of $(Y_1, \cdots, Y_n)$ under H_0 and H_{1n}, respectively. We need to show that P_{0n} and P_{1n} are contiguous in Le Cam's sense. Denote

$$m(x) = E[D_t(\theta_0)\zeta_t I\{Y_{t-1} \le x\}] - \Sigma_x \Sigma^{-1} E[D_t(\theta_0)\zeta_t],$$

where $\zeta_t = \eta_t r_{1t}/\sqrt{h_t(\theta_0)} + (1 - \eta_t^2) r_{2t}/h_t(\theta_0)$. We have the following result.

Theorem 4.3. *If the assumptions of Theorem 6.1 hold and $0 < Er_{1t}^2 + Er_{2t}^2 < \infty$ under H_0, then under H_{1n}, it follows that*

$$(a)\quad T_n(x, \hat{\theta}_n) \Longrightarrow m(x) + G_p(x) \text{ in } D^p[R],$$

$$(b)\quad \lim_{b\to\infty}\lim_{n\to\infty} P[S_n^a(b) \le x] = P\left[\max_{\tau\in[0,1]} [u(\tau) + B(\tau)]^2 \le x\right],$$

for any $x \in R$, where $u(\tau) = \beta'\Sigma_x^{-1} m(x)/[\beta'(\Sigma_a^{-1} - \Sigma^{-1})\beta]^{1/2}$ with x such that $\sigma_x = \tau$, and $G_p(x)$ and $B(\tau)$ are defined as in Theorems 3.1-3.2.

The proof of this theorem can be found in Ling and Tong (2006). It shows that S_n^a has non-trivial local power if $u(\tau) \ne 0$; otherwise it has no local power. It is unlikely to have $u(\tau) = 0$, except when $\zeta_t = \beta' D_t(\theta_0)$. When n and b are large, we have $P(S_n^a > C_\alpha) \approx P\Big\{\max_{\tau\in[0,1]}[u(\tau) + B(\tau)]^2 > C_\alpha\Big\} \to 1$ if $\max_{\tau\in[0,1]} |u(\tau)| \to \infty$.

4.4 Simulation Study

This section reports some simulation results which give some comparisons with Ljung-Box test and Li-Mak test. In all the experiments, we take a as the $5p\%$-quantile of data $\{Y_1, \cdots, Y_n\}$ and use 1000 independent replications. We first study the size of S_n^a when the null hypothesis is the ARMA(1,1) model, $Y_t = \phi Y_{t-1} + \psi\varepsilon_{t-1} + \varepsilon_t$, where ε_t is i.i.d. $N(0, 1)$. Table 1 summarizes the sizes of S_n^a.

To compare the power of Ljung-Box-test, Li-Mak-test and S_n^a, we consider the following two alternative models:

$$\text{TARMA Model}\quad Y_t = 0.5Y_{t-1} + 0.5\varepsilon_{t-1} - \theta(Y_{t-1} + \varepsilon_{t-1}) I\{Y_{t-1} \le 0\} + \varepsilon_t,$$

$$\text{BL Model}\quad Y_t = 0.5Y_{t-1} + 0.5\varepsilon_{t-1} - \theta Y_{t-2}\varepsilon_{t-1} + \varepsilon_t.$$

TABLE 1

Sizes of S_n^a for Null Hypothesis H_0:
ARMA(1,1) model
at Significance Level α(1000 replications)

	n=100			n=200			n=400		
α	.01	.05	.10	.01	.05	.10	.01	.05	.10
ϕ ψ	$\beta=(1,1)'$								
-.8 -.5	.019	.041	.066	.009	.029	.069	.008	.052	.097
-.5 -.5	.008	.036	.066	.010	.033	.079	.013	.037	.090
.0 -.5	.009	.037	.078	.006	.040	.085	.009	.030	.072
.8 -.5	.015	.051	.105	.009	.041	.080	.012	.036	.079
-.8 .5	.007	.035	.067	.009	.033	.072	.007	.043	.084
.0 .5	.006	.031	.068	.010	.043	.087	.007	.037	.084
.5 .5	.002	.030	.071	.008	.046	.085	.012	.047	.093
.8 .5	.013	.049	.100	.015	.044	.092	.013	.056	.097

We take $\theta = 0.1$, 0.2, 0.3, 0.4 and 0.5 and $n = 100$, 200 and 400. In Table 2, $Q_n(m)$ is Ljung-Box-test and $Q_n^2(m)$ is the Li-Mak-test. Table 2 reports the results when $m = 6$ and $\beta = (1,1)'$.

When the null hypothesis is the GARCH(1,1) model, $Y_t = \eta_t\sqrt{h_t}$ *and* $h_t = \alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1 h_{t-1}$, where η_t is i.i.d. $N(0,1)$. Table 3 summarizes the sizes of S_n^a when $\alpha_0 = 0.1$.

Two alternative models,

$$\text{TGARCH} \quad \sqrt{h_t} = 0.1 + 0.3|Y_{t-1}| + 0.4\sqrt{h_{t-1}} + \theta|Y_{t-1}|I\{Y_{t-1} \le 0\},$$
$$\text{NAGARCH} \quad h_t^{3/4} = 0.1 + 0.3|(\theta - \text{sgn}(\eta_t))Y_t|^{3/2} + 0.4h_{t-1}^{3/4}.$$

The first model is a threshold GARCH that is a special case of models proposed by Taylor (1986) and Schwert (1989). The second is a nonlinear asymmetric GARCH model proposed by Engle and Ng (1993). We take $\theta = 0.4$, 0.6, 0.8, 1.0 and 1.2. The sample sizes are $n = 100$, 200 and 400. Again, we compare the power of S_n^a with those of $Q_n(m)$ and $Q_n^2(m)$. The sizes of $Q_n(6)$ and $Q_n^2(6)$ are very close to their corresponding nominal values; see Li and Mak (1994) and Wong and Ling (2005) for simulation evidence. The results reported in Table 5 are for the significance level 0.05 when $\beta = (1,1)'$. In all cases, S_n^a is more powerful than $Q_n(6)$ and $Q_n^2(6)$. In particular, when the alternative is the NAGARCH model, S_n^a can reject GARCH with power reaching 50 percent, while both $Q_n(6)$ and $Q_n^2(6)$ have virtually no power. Again it seems that our test is more powerful against the TGARCH alternative than against the NAGARCH alternative. Similar conclusions hold when $\beta = (\hat{\alpha}_{0n}, \hat{\alpha}_{1n}, \hat{\beta}_{1n})'$ and $m = 12$. Details are available from the authors.

TABLE 2

Powers of S_n^a, $Q_n(m)$ and $Q_n^2(m)$ for Null Hypothesis H_0: ARMA(1,1) Model at Significance Level 0.05
[$\beta = (1,1)'$ and 1000 replications]

	θ	.0	.1	.2	.3	.4	.5
		H_1: TARMA Model					
	S_n^a	.026	.081	.146	.419	.681	.884
$n = 100$	$Q_n(6)$	.053	.051	.056	.070	.082	.102
	$Q_n^2(6)$	.027	.028	.027	.024	.029	.038
	S_n^a	.035	.118	.403	.780	.985	.997
$n = 200$	$Q_n(6)$	.066	.061	.077	.091	.118	.159
	$Q_n^2(6)$	.035	.044	.052	.061	.090	.128
	S_n^a	.035	.180	.704	.980	1.000	1.000
$n = 400$	$Q_n(6)$	.052	.057	.064	.095	.186	.324
	$Q_n^2(6)$	.034	.033	.036	.067	.120	.189
		H_1: BL Model					
	S_n^a		.052	.114	.203	.263	.289
$n = 100$	$Q_n(6)$		.053	.051	.045	.064	.081
	$Q_n^2(6)$		.030	.051	.104	.163	.229
	S_n^a		.073	.228	.406	.455	.440
$n = 200$	$Q_n(6)$		.072	.067	.069	.081	.098
	$Q_n^2(6)$		.041	.096	.206	.313	.408
	S_n^a		.165	.495	.738	.721	.634
$n = 400$	$Q_n(6)$		.044	.044	.045	.053	.095
	$Q_n^2(6)$		.044	.161	.393	.593	.626

TABLE 3

Sizes of S_n^a for Null Hypothesis H_0: GARCH(1,1) model at Significance Level α(1000 replications)

		n=100			n=200			n=400		
α		.01	.05	.10	.01	.05	.10	.01	.05	.10
α_1	β_1	$\beta = (1,1,1)'$								
.3	.4	.016	.031	.055	.005	.027	.056	.008	.035	.063
.3	.5	.007	.027	.051	.007	.025	.057	.007	.034	.064
.3	.6	.008	.029	.056	.009	.032	.062	.008	.033	.063
.2	.7	.009	.032	.058	.010	.036	.069	.008	.037	.069
.1	.8	.011	.041	.070	.011	.034	.071	.007	.037	.069
.3	.7	.010	.036	.068	.008	.038	.070	.011	.038	.073
.2	.8	.010	.042	.069	.008	.035	.068	.010	.037	.078
.1	.9	.015	.055	.090	.016	.043	.084	.012	.048	.092

TABLE 4

Powers of S_n^a, $Q_n(m)$ and $Q_n^2(m)$ for Null Hypothesis H_0: GARCH(1,1) Model at Significance Level 0.05

[$\beta = (1,1)'$ and 1000 replications]

θ		.4	.6	.8	1.0	1.2
		H_1: TGARCH Model				
	S_n^a	.310	.543	.614	.618	.670
$n = 100$	$Q_n(6)$	.153	.156	.176	.286	.457
	$Q_n^2(6)$	.072	.083	.069	.108	.177
	S_n^a	.478	.845	.766	.666	.751
$n = 200$	$Q_n(6)$	.138	.167	.196	.273	.502
	$Q_n^2(6)$	.085	.066	.070	.138	.286
	S_n^a	.680	.978	.896	.737	.834
$n = 400$	$Q_n(6)$	.125	.132	.146	.230	.512
	$Q_n^2(6)$	.145	.117	.080	.157	.479
		H_1: NAGARCH Model				
	S_n^a	.099	.143	.217	.322	.457
$n = 100$	$Q_n(6)$	.075	.082	.091	.103	.115
	$Q_n^2(6)$	.030	.040	.046	.050	.052
	S_n^a	.116	.173	.283	.454	.649
$n = 200$	$Q_n(6)$	.074	.079	.087	.100	.109
	$Q_n^2(6)$	.040	.040	.046	.060	.062
	S_n^a	.127	.214	.393	.630	.863
$n = 400$	$Q_n(6)$	.051	.055	.063	.068	.072
	$Q_n^2(6)$	.034	.044	.053	.065	.073

Chapter 5
Efficient and Adaptive Estimation

5.1 MLE and Efficient Estimators

In he general model (3.2), we assume that the error η_t is iid with the density f. Given $\tilde{Y}_o = \{Y_0, Y_{-1}, \cdots,\}$, the conditional MLE is the maximizer, denoted by $\hat{\theta}_n$, of the log-likelihood function:

$$L_n(\theta) = \sum_{t=1}^{n} \ln f(\eta_t(\theta)), \tag{5.1}$$

where

$$\eta_t(\theta) = [Y_t - m_t(\theta)]/\sqrt{h_t(\theta)},\ t = 1, 2, \cdots. \tag{5.2}$$

We first make the following assumption:

Assumption 5.1 *The density f is absolutely continuous with a.e.-derivative f' and*

$$I_1(f) = \int \xi_1^2(x) f(x) dx < \infty \text{ and } I_2(f) = \int \xi_2^2(x) f(x) dx < \infty,$$

where $\xi_1(x) = f'(x)/f(x)$ and $\xi_2(x) = 1 + x\xi_1(x)$.

Under the regular conditions of Theorem 3.2, we have the following expansion:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -S_n^{-1}(\theta_0) D_n(\theta_0) + o_p(1) \longrightarrow_L N(0, \Omega^{-1}), \tag{5.3}$$

where $\Omega = EX_t(\theta_0) I(f) X_t'(\theta_0)]$,

$$X_t(\theta) = \left[\frac{1}{\sqrt{h_t(\theta)}} \frac{\partial \varepsilon_t(\theta)}{\partial \theta}, \frac{1}{2h_t(\theta)} \frac{\partial h_t(\theta)}{\partial \theta}\right], \tag{5.4}$$

$$S_n(\theta) = \sum_{t=1}^{n} X_t(\theta)\zeta(\theta)\zeta'(\theta)X_t'(\theta) \text{ and } D_n(\theta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t(\theta)\zeta(\theta). \tag{5.5}$$

It can be shown that the asymptotic variance Ω^{-1} of $\hat{\theta}_n$ achieves the lowest CramrRao bound among the class of the unbiased estimators under Assumption 2.1. Note that all the estimators in the field of time series in the literature are asymptotically unbiased (i.i. consistency). It seems that no one has concluded that the Ω^{-1} is the lowest CramrRao bound among the class of the consistent estimators up to now. To see how good the MLE is, we consider the class of the $M-$estimators:

$$\left\{ \tilde{\theta}_n : \sum_{t=1}^{n} \psi_t(\tilde{\theta}_n) = 0, \right\}$$

where $\sum_{t=1}^{n} \psi_t(\theta) = 0$ is estimation equation and $\psi_t(\theta) = \psi(\theta, Y_t, Y_{t-1}, \cdots)$ is a martingale difference in terms of $\mathscr{F}_t$ when $\theta = \theta_0$. Assume that $\tilde{\theta}_n$ is asymptotically normal, i.e.,

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \Longrightarrow_L N(0, \Sigma_m^{-1}\Omega_m\Sigma_m^{-1}),$$

where $S_{mn}(\theta) = \Sigma_m = E\partial\psi_t(\theta_0)/\partial\theta$ and $\Omega_m = E[\psi_t(\theta)\psi_t'(\theta)]$. We first note that, when $(Y_t, \cdots, Y_1) = ((Y_t, \cdots, Y_1) \equiv \tilde{y}_t$, $f(\eta_t(\theta))$ and $\psi_t(\tilde{\theta}_n)$ are functions of $\tilde{y}_t$, θ_0 and $\tilde{Y}_0$, denoted by $f(\tilde{y}_t, \theta_0, \tilde{Y}_0)$ and $\psi(\tilde{y}_t, \theta_0, \tilde{Y}_0)$. Thus,

$$0 = E\sum_{t=1}^{n} \psi_t(\tilde{\theta}_n) = E\int \left(\sum_{t=1}^{n} \psi(\tilde{y}_t, \theta_0, \tilde{Y}_0) \right) \prod_{t=1}^{n} f(\tilde{y}_t, \theta_0, \tilde{Y}_0) d\tilde{y}_n.$$

Taking derivatives to both sides of this equation in terms of θ, we have

$$\begin{aligned}
0 &= E\int \left(\sum_{t=1}^{n} \frac{\partial\psi(\tilde{y}_t, \theta_0, \tilde{Y}_0)}{\partial\theta} \right) \prod_{t=1}^{n} f(\tilde{y}_t, \theta_0, \tilde{Y}_0) d\tilde{y}_n \\
&\quad + E\int \left(\sum_{t=1}^{n} \psi(\tilde{y}_t, \theta_0, \tilde{Y}_0) \right) \left(\sum_{t=1}^{n} \frac{\partial \ln f(\tilde{y}_t, \theta_0, \tilde{Y}_0))}{\partial\theta} \right) \prod_{t=1}^{n} f(\tilde{y}_t, \theta_0 \tilde{Y}_0) d\tilde{y}_n \\
&= n\Sigma_m + E\left\{ \left(\sum_{t=1}^{n} \psi_t(\theta_0) \right) \left(\sum_{t=1}^{n} X_t(\theta_0)\zeta_t \right) \right\} \\
&= n\Sigma_m + E[\psi_t(\theta_0) X_t(\theta_0)\zeta_t].
\end{aligned}$$

From this, we can show that $\Omega^{-1} \le \Sigma_m^{-1}\Omega_m\Sigma_m^{-1})$. Thus, the MLE achieves its lowest bound among the $M-$estimators. An estimator is called $M-$**efficient** or simply **efficient** if its asymptotic variance is Ω^{-1}.

To obtain an efficient estimator, ones do not have to look for the global MLE as in Section 3.2 and this may need more restrictive conditions. If one $\sqrt{n}-$consistent estimator, say, $\bar{\theta}_n$ is available, the estimator via the following one-step iteration is efficient:

$$\tilde{\theta}_n = \bar{\theta}_n - S_n^{-1}(\bar{\theta}_n) D_n(\bar{\theta}_n)/\sqrt{n}, \tag{5.6}$$

if $S_n(\bar{\theta}_n)/n = \Omega + o_p(1)$ *and* $D_n(\bar{\theta}_n) = D_n(\theta_0) + o_p(\sqrt{n})$. How to check these conditions is not easy in some case. We introduce the discretized initial estimator.

Definition 5.1. A sequence of estimators $\{\bar{\theta}_n\}$ measurable in terms of $\mathscr{F}_n$ is said to be discretized and $\sqrt{n}-$consistent if, for any small $\varepsilon > 0$, there exists a constant $\Delta > 0$ and an integer $K > 0$ such that

$$P_{\theta_0,f}(\|\sqrt{n}(\bar{\theta}_n - \theta_0)\| < \Delta) > 1 - \varepsilon$$

uniformly in n and, for each n, $\bar{\theta}_n$ takes on at most K different values in $\Theta_n = \{\theta \in R^l : \|\sqrt{n}(\theta - \theta_0)\| \leq \Delta\}$.

This discretization method was proposed first by LeCam (1960) and has become an important tool in the construction of efficient estimators. Using the estimator $\bar{\theta}_n$ and the one-step Newton-Raphson iteration,

$$\tilde{\theta}_n = \bar{\theta}_n + n^{-1/2} S_n^{-1}(\bar{\theta}_n) W_n(\bar{\theta}_n), \tag{5.7}$$

by Lemma 4.4 in Kreiss (1987) and Theorem 2.1(ii), we know that, under $P^n_{\theta_0}$, $\sqrt{n}(\tilde{\theta}_n - \theta_0) = S_n^{-1}(\theta_0) W_n(\theta_0) + o_p(1) \rightarrow_{\mathscr{L}} N(0, \Omega^{-1})$ as $n \rightarrow \infty$, and hence $\tilde{\theta}_n$ is an efficient estimator.

When f is unknown, we need to estimate the density f in (4.4). Since f is an infinite-dimensional nuisance parameter, its estimator may be effect the efficiency of $\hat{\theta}_n$. Thus, ones need to explore if the efficient estimator is available in this case. If not, what is the lowest bound and how achieve it.

5.2 Adaptivity and Efficiency

To explain efficiency and adaptivity for models (1.1)-(1.2), we need to parameterize the density f. This technique is discussed carefully in Bickel et al. (1993). Denote the unknown density by $\chi \in \mathscr{D}$, where $\mathscr{D}$ is a class of Lebesgue densities. Let $\mathscr{B} = \{P_{\theta,\chi} : (\theta, \chi) \in \Theta \times \mathscr{D}\}$ be a family of probability measures on $(\mathscr{R}^Z, \mathscr{F}^Z)$. Assume that the rescaled errors $\{\eta_1(\theta), \cdots, \eta_n(\theta)\}$ in (2.1)-(2.2) under $P_{\theta,\chi}$ are i.i.d. with density χ and independent of Y_0. As in Koul and Schick (1996, 1997), we give the following definition:

Definition 5.2. Let $q : c \rightarrow f_c$ be a map from a neighborhood Δ of the origin in R^2 into $\mathscr{D}$ such that $f_0 = f$. We say that $q : c \rightarrow f_c$ is a *regular* path if there exists a measurable function ζ from R to R^2 such that $\int \|\zeta(x)\|^2 f(x)dx < \infty$, $V_{\zeta\zeta} \equiv \int \zeta(x)\zeta'(x)f(x)dx$ is nonsingular, and

$$\int \left[\sqrt{f_c(x)} - \sqrt{f(x)} - \frac{1}{2} c' \zeta(x) \sqrt{f(x)} \right]^2 dx = o(\|c\|^2). \tag{5.8}$$

Let $P^n_{\theta,\chi}$ be the restriction of $P_{\theta,\chi}$ on $\mathscr{F}_n$. The distribution of the initial value Y_0 in (2.1)-(2.2) is the same under both $P^n_{\theta,f}$ and P^n_{θ,f_c}. The log-likelihood ratio $\Lambda_n(\theta_1,\theta_2, c)$ of $P^n_{\theta_2,f_c}$ to $P^n_{\theta_1,f}$ is that:

$$\Lambda_n(\theta_1,\theta_2,c) = 2\sum_{t=1}^{n}\left[\log s_{c,t}(\theta_2) - \log s_t(\theta_1)\right], \tag{5.9}$$

where $s_{c,t}(\theta) = \sqrt{f_c(\eta_t(\theta))}/\sqrt[4]{h_t(\theta)}$ and $s_t(\theta) = s_{0,t}(\theta)$. Define the $(l+2)-$dimensional random vector and the $(l+2)\times(l+2)$ matrix, respectively, by

$$\tilde{W}_n(\theta) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix} X_t(\theta)\xi(\eta_t(\theta)) \\ \zeta(\eta_t(\theta))\end{pmatrix} \; and \; \tilde{\Omega}(\zeta) = \begin{pmatrix} \Omega & \nu V_{\xi\zeta} \\ V'_{\xi\zeta}\nu' & V_{\zeta\zeta}\end{pmatrix},$$

where $\nu = EX_t(\theta_0)$ and $V_{\xi\zeta} = E[\xi(\eta_t)\zeta'(\eta_t)]$.

Let $\theta_n = \theta_0 + u_n/\sqrt{n}$ and and $\tilde{u}_n = (u'_n, v'_n)' \in R^{l+2}$ are bounded sequences.

Assumption 5.2 *(a). The path $q: c \to f_c$ is regular such that $\tilde{\Omega}(\zeta)$ is positive definite and (b). under $P_{\theta_n,f}$,*

$$\Lambda_n(\theta_0,\theta_n,\frac{v_n}{\sqrt{n}}) = \tilde{u}'_n\tilde{W}_n(\theta_0) - \frac{1}{2}\tilde{u}'_n\tilde{\Omega}(\zeta)\tilde{u}_n + o_p(1), \tag{5.10}$$

and, as $n\to\infty$,

$$\tilde{W}_n(\theta_0) \to_{\mathscr{L}} N(0,\tilde{\Omega}(\zeta)).$$

The maximizer of (4.7) is

$$\tilde{u}_n = \tilde{\Omega}(\zeta)^{-1}\tilde{W}_n(\theta_0)/\sqrt{n} + 0_p(1).$$

Solve this equation, we have

$$u_n = I_q^{-1}\sum_{t=1}^{n}[X_t\xi(\eta_t(\theta)) - \nu V_{\zeta\zeta}^{-1}\zeta(\eta_t)]/\sqrt{n} + o_p(1),$$

where

$$I(q) = \Omega - \nu V_{\xi\zeta}V_{\zeta\zeta}^{-1}V'_{\xi\zeta}\nu'.$$

The asymptotic covariance matrix of u_n is

$$I(q)^{-1} + I(q)^{-1}\nu(V_{\xi\zeta} - V_{\zeta\zeta})V_{\zeta\zeta}^{-1}(V_{\xi\zeta} - V_{\zeta\zeta})'\nu' I(q)^{-1} \geq I(q)^{-1}.$$

When $\nu = 0$, u_n is the efficient estimator for any path q. When $\nu \neq 0$, we need to look for a ζ such that the equation holds which if and only if

$$E\{[\xi(\eta_t) - \zeta(\eta_t)]\zeta'(\eta_t)\} = 0, \tag{5.11}$$

i.e., $\zeta(x)$ is the projection of $\xi(x)$ on the q-direction. Furthermore, we need to identify a class of path q denoted by $\mathscr{Q}$, and a path $q^* \in \mathscr{Q}$ such that

$$I(q^*) = \sup_{q^* \in \mathscr{Q}} I(q). \tag{5.12}$$

The path q^* is called least favorable since it contains the least amount of information about θ. u_n achieves the smallest asymptotic variance in the class $\mathscr{Q}$. An estimator is called $\mathscr{Q}-$efficient or efficient if its asymptotic variance $I(q^*)^{1}$.

We let $\mathscr{Q}$ be the set of regular paths such that each path, $q : c \to f_c$, in $\mathscr{Q}$ satisfies:

$$\int (1+x^2) f_c(x)dx \longrightarrow \int (1+x^2) f(x)dx, \ as \ c \to 0, \tag{5.13}$$

and $T_{\mathscr{Q}}$ be the two-dimensional functional space:

$$T_{\mathscr{Q}} = \Big\{ (a,b) \in L_2^2(F) : \int (a(x), b(x))'(1, x, x^2) f(x)dx = 0 \Big\}.$$

By Lemma 3 of Schick (2002), each $\zeta(x) = (a(x), b(x))'$ that appears in (5.7) must also be in $T_{\mathscr{Q}}$ if (5.8) holds.

For each $(a,b) \in T_{\mathscr{Q}}$, we can construct a regular path, $q : c \to f_c$, such that (5.8) holds. In fact, let

$$f_c(x) = f(x)\Psi(c_1 a(x) + c_2 b(x)) \Big/ \int f(x)\Psi(c_1 a(x) + c_2 b(x))dx,$$

where $\Psi(y) = 2/(1+e^{-2y})$. It is easy to show that $f_c(x)$ satisfies (5.7) and (5.8) when $\|c\| < \varepsilon$ with ε small enough. $(a,b)'$ is called the characteristics of q.

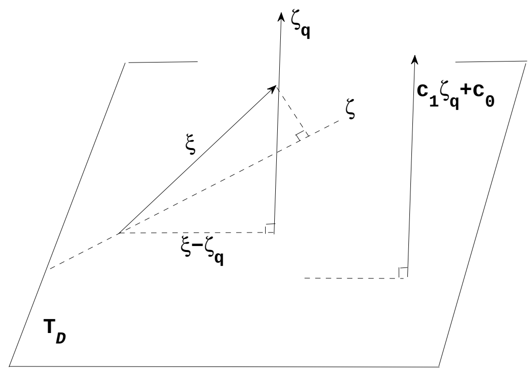


Fig. 5.1 Geometric Figure.

As argued by Schick (2002), it can be shown that

$$\int \xi(x)f(x)dx = 0 \textit{ and } \int (x,x^2-1)'\xi(x)f(x)dx = I. \tag{5.14}$$

Thus, $\xi \notin T_{\mathscr{Q}}$ and hence we can show that $\tilde{\Omega}(\zeta)$ is positive definite for each $\zeta \in T_{\mathscr{Q}}$ if Assumptions 1-3 hold. To have (5.8)-(5.9), the ζ corresponding to the path q^* must be the projection of ξ into $T_{\mathscr{Q}}$, denoted by ζ_*. From the definition of $T_{\mathscr{Q}}$, there exists a constant vector C_0 and a constant 2×2 matrix C_1 such that

$$\xi - \zeta_* = C_1(x,x^2)' + C_0.$$

Since $E\xi(\eta_t) = E\zeta_*(\eta_t) = 0$, we have $C_0 = -C_1(0,1)'$. Thus, we have

$$\xi - \zeta_* = C_1(x,x^2-1)' = C_1\zeta_q,$$

where $\zeta_q = (x,x^2-1)'$. Since $\xi - \zeta_*$ is the projection of ξ on the ζ_q, we have

$$\begin{aligned}
\zeta_* &= \xi - \Big[\int \xi(x)\zeta_q'(x)f(x)dx\Big]\Big[\int \zeta_q(x)\zeta_q'(x)f(x)dx\Big]^{-1}\zeta_q, \\
&= \xi - \Big[\int \zeta_q(x)\zeta_q'(x)f(x)dx\Big]^{-1}\zeta_q,
\end{aligned}$$

where we use

$$\int \xi(x)\zeta_q'(x)f(x)dx = \mathrm{diag}(1,1).$$

Of course, ones can verify (5.8)-(5.9) in a direction way for this ζ_*. On this path, its information for estimating θ_0 is

$$I_* = \Omega - \nu\{V_{\xi\xi} - E^{-1}[\zeta_q(\eta_t)\zeta_q'(\eta_t)]\}\nu'. \tag{5.15}$$

When an estimator has the asymptotic covariance matrix I_*^{-1}, it is called $\mathscr{Q}-$efficient or optimal. Note that Ω is the information matrix when the density f is known.

$$I_d \equiv \Omega - I_* = \nu(V_{\xi\xi} - CE^{-1}[\zeta_q(\eta_t)\zeta_q'(\eta_t)]C)\nu'$$

is the difference of information in the cases with known and unknown f. If $I_d = 0$, $\hat{\theta}_n$ is said to be $\mathscr{Q}-$adaptive, or simply adaptive, since there is no information loss with unknown density f.

Let the loss function L be a Borel measurable function from R^{l+2} to $[0,\infty)$ such that $L(0) = 0$, $L(x) = L(-x)$, $x \in R$, and L is nondecreasing on $[0,\infty)$. The following theorem state how good is I_* from the view of the asymptotic minimax bound.

Theorem 5.1. *If Assumptions 1-3 hold and $\hat{\theta}_n$ is an estimator of θ_0, then*

$$\sup_{q\in\mathscr{Q}} \lim_{C\to\infty} \liminf_{n\to\infty} \sup_{\|\theta-\theta_0\|+\|c\|\le C/\sqrt{n}} \int L[\sqrt{n}(\hat{\theta}_n - \theta_0)]dP_{\theta,f_c} \ge \int L dN(0, I_*^{-1})$$

for every bound function L. Furthermore, if under $P_{\theta_0,f}$, $\hat{\theta}_n$ satisfies

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} I_*^{-1}[X_t(\theta_0)\xi_t(\eta_t) - \nu\zeta_*(\eta_t)] + o_p(1), \tag{5.16}$$

then, under $P_{\theta_n, f_{c_n/\sqrt{n}}}$, $\sqrt{n}(\hat{\theta}_n - \theta_n) \longrightarrow_{\mathscr{L}} N(0, I_*^{-1})$ *as* $n \to \infty$ *for every local sequence* θ_n, *every* $q \in \mathscr{Q}$ *and every bounded sequence* u_n *in* R^l, *and this implies*

$$\sup_{q\in\mathscr{Q}} \lim_{C\to\infty} \limsup_{n\to\infty} \sup_{\|\theta-\theta_0\|+\|c\|\le C/\sqrt{n}} \int L[\sqrt{n}(\hat{\theta}_n - \theta_0)]dP_{\theta,f_c} = \int LdN(0, I_*^{-1})$$

for every bounded loss function L.

This theorem is an extension of Theorem 6 in Fabian and Hannan (1982). Its proof is similar to that of Theorem 3.2 in Koul and Schick (1997) and hence is omitted. This theorem shows that I_*^{-1} is the asymptotic minimax bound that may be achieved when the density f is unknown.

5.3 LAN of FARIMA-GARCH Models

Assume that the random sample $\{Y_1, \cdots, Y_n\}$ is generated by ARMA model (2.7) with ε_t being from GARCH model (2.10). We call model (2.7) and (2.10) the FARIMA-GARCH model. Based on the initial value $\tilde{Y}_0 = (Y_0, Y_{-1}, \cdots)$, the conditional residuals $\{\eta_t(\theta)\}$ satisfy the following equations:

$$\phi(B)(1-B)^d Y_t = \psi(B)\varepsilon_t(\theta), \tag{5.17}$$

$$\varepsilon_t(\theta) = \eta_t(\theta)\sqrt{h_t(\theta)} \text{ and } h_t(\theta) = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2(\theta) + \sum_{i=1}^{s} \beta_i h_{t-i}(\theta), \tag{5.18}$$

where $d \in (0, 1/2)$. Denote $l = p+q+r+s+1$. The true parameter $\theta = \theta_0$, where $\theta = (\gamma', \delta')'$, $\gamma = (d, \phi_1, \cdots, \phi_p, \psi_1, \cdots, \psi_q)'$ and $\delta = (\alpha_0, \alpha_1, \cdots, \alpha_r, \beta_1, \cdots, \beta_s)'$. Assume that the parameter space Θ is an open subset of R^l and $\theta_0 \in \Theta$.

Assumption 5.3 $\alpha_i > 0$, $i = 0, \cdots, r$, $\beta_j > 0$, $j = 1, \cdots, s$, $\sum_{i=1}^{r} \alpha_i + \sum_{i=1}^{s} \beta_i < 1$, *and* $\sum_{i=1}^{r} \alpha_i z^i$ *and* $1 - \sum_{i=1}^{s} \beta_i z^i$ *have no common root.*

From Theorem 2.1 below, we can see that the effect of Y_0 on the likelihood ratio is ignorable, and hence we can simply let $Y_i = 0$ as $i \le 0$ in practice. Our method is similar to that of Hallin et al. (1999).

Theorem 5.2. *Suppose that* $d \in (0, 1/2)$, *Assumptions 3.4, 5.1 and 5.3 hold. If Assumption 5.2(a) is satisfied, then Assumption 5.2 (b) holds.*

Let $P_{\theta,f}$ be a probability measure on $(\mathscr{R}^Z, \mathscr{F}^Z)$, where $\mathscr{F}^Z$ is the Borel $\sigma-$field on $\mathscr{R}^Z$ with $Z = \{0, \pm1, \pm2, \cdots\}$, and let P_θ^n be the restriction of $P_{\theta,f}$ on $\mathscr{F}_n$, the $\sigma-$field generated by $\{\tilde{Y}_0, Y_1, \cdots, Y_n\}$. Note that the rescaled errors $\{\eta_t(\theta) : t = 1, \cdots, n\}$ under P_θ^n are i.i.d. with the same density as η_t and are independent of Y_0.

It is clear that (2.1)-(2.2) under $P^n_{\theta_0}$ are exactly the same as models (1.1)-(1.2) with $\theta = \theta_0$. From (2.1)-(2.2), the distribution of the initial value $\tilde{Y}_0$ is the same under both P^n_θ and $P^n_{\theta_0}$. Thus, the log-likelihood ratio $\Theta_n(\theta_1,\theta_2)$ of $P^n_{\theta_2}$ to $P^n_{\theta_1}$ is

$$\Lambda_n(\theta_1,\theta_2) = 2\sum_{t=1}^{n}[\log s_t(\theta_2) - \log s_t(\theta_1)],$$

where $s_t(\theta) = \sqrt{f(\eta_t(\theta))}/\sqrt[4]{h_t(\theta)}$. Let u_n in R^l be any bounded vector and $\theta_n = \theta_0 + n^{-1/2}u_n$ be a local sequence. The following corollary gives the LAN of $\Theta_n(\theta_0,\theta_n)$:

Corollary 5.1. *Suppose that $d \in (0,1/2)$, Assumptions 3.4, 5.1 and 5.3 hold. Then, it follows that,*

(i) under $P^n_{\theta_0}$, $\Theta_n(\theta_0,\theta_n) = u_n' W_n(\theta_0) - u_n'\Omega u_n/2 + o_p(1)$;

(ii) still under $P^n_{\theta_0}$, $W_n(\theta_0) \to_{\mathscr{L}} N(0,\Omega)$ as $n\to\infty$, and

$$\sum_{t=1}^{n}[W_t(\theta_n) - W_t(\theta_0)] + \Omega[\sqrt{n}(\theta_n - \theta_0)] = o_p(1);$$

(iii) $P^n_{\theta_0}$ and $P^n_{\theta_n}$ are contiguous.

When h_t is a constant, the LAN of model (1.1) was proved by Hallin et al. (1999). When f is symmetric, $\Omega = \mathrm{diag}\{\Omega_m, \Omega_\delta\}$, where

$$\Omega_\gamma = E\Big[\frac{1}{h_t(\theta_0)}\frac{\partial\varepsilon_t(\theta_0)}{\partial\gamma}\frac{\partial\varepsilon_t(\theta_0)}{\partial\gamma'}\Big]I_1(f) + E\Big[\frac{1}{4h_t^2(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\gamma}\frac{\partial h_t(\theta_0)}{\partial\gamma'}\Big]I_2(f)\tag{5.19}$$

$$\Omega_\delta = E\Big[\frac{1}{4h_t^2(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\delta}\frac{\partial h_t(\theta_0)}{\partial\delta'}\Big]I_2(f).\tag{5.20}$$

In this case, $\nu = 0$ and

$$I_* = \mathrm{diag}\Big\{\Omega_\gamma,\ \Omega_\delta I_2(f) - I_{d\delta}\Big\},\tag{5.21}$$

where $I_{d\delta} = [I_2(f) - (E\eta_t^4 - 1)^{-1}]E(\partial h_t/\partial\delta)E(\partial h_t/\partial\delta')$. Thus, $\hat{\gamma}_n$ is adaptive, but $\hat{\delta}_n$ is not unless f is normal. When f is asymmetric, both $\hat{\gamma}_n$ and $\hat{\delta}_n$ are not adaptive, since $\nu \neq 0$, generally. However, when h_t is a constant, $\nu = 0$ even if f is asymmetric and, hence, $\hat{m}_n$ is adaptive in this case.

If f is symmetric and $\mathscr{D}$ contains only densities symmetric about zero, then each (a,b) satisfying (A.1) is even. Let $T_{\mathscr{D}} = \Big\{(a,b) \in L_2^2(F) : (a(x),b(x)) = (a(-x),b(-x)) \int(a(x),b(x))'(1,x^2)f(x)dx = 0\Big\}$. The projection of ξ into $T_{\mathscr{D}}$ is

$$\zeta_* = [0, \xi_2 - 2(x^2-1)(E\eta_t^4 - 1)^{-1}]',\tag{5.22}$$

which is least favorable. The corresponding information for estimating θ_0 is the same as I_* in (A.6). If reparameterized δ_0 as $(\alpha_{00},\tilde{\delta}_0) = (\alpha_{00},\tilde{\alpha}_{01},\cdots,\tilde{\alpha}_{0r},\tilde{\beta}_{01},\cdots,$

$\tilde{\beta}_{0s})'$ with $\tilde{\alpha}_{0i} = \alpha_{0i}/\alpha_{00}$ and $\tilde{\beta}_{0i} = \beta_{0i}/\alpha_{00}$, then $\tilde{\delta}_0$ is adaptively estimable as discussed by Linton (1993), Drost et al. (1997), Drost and Klaassen (1997), and Ling and McAleer (2003). Since α_{00} is unknown, it has to be estimated. Its optimal estimator was given by Drost and Klaassen (1997). By Theorem 5.1, the estimator of δ_0 obtained through this procedure is not superior to the estimator using the path in (5.9).

5.4 Adaptive Estimation

Let $K(x) = e^{-x}/(1+e^{-x})^2$ be the logistic kernel and $\bar{K}(x, \eta_t(\theta)) = K([x+\eta_i(\theta)]/a) + K([x-\eta_i(\theta)]/a)$. $f(x)$ and $\xi_1(x)$ are estimated by the kernel method:

$$\hat{f}_{a,j}(x,\theta) = \frac{1}{2a(n-1)} \sum_{i=1, i\neq j}^{n} \bar{K}(x, \eta_i(\theta)) \;\; and \;\; \hat{\xi}_{1n,j}(x,\theta) = \frac{\hat{f}'_{a_n,j}(x,\theta)}{b_n + \hat{f}_{a_n,j}(x,\theta)} \tag{5.23}$$

where $j = 1, \cdots, n$, and a_n and b_n satisfy the condition that $na_n^3 b_n \to \infty$. Define $\hat{\xi}_{2n,j}(x,\theta) = 1 + x\hat{\xi}_{1n,j}(x,\theta)$. $I_1(f)$ and $I_2(f)$ are respectively estimated by

$$\hat{I}_{1n}(\theta) = \frac{1}{n}\sum_{t=1}^{n} \hat{\xi}^2_{1n,t}(\eta_t(\theta), \theta) \;\; and \;\; \hat{I}_{2n}(\theta) = \frac{1}{n}\sum_{t=1}^{n} \hat{\xi}^2_{2n,t}(\eta_t(\theta), \theta).$$

We estimate the score function and the information matrix by:

$$\hat{W}_{\gamma n}(\theta) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big[\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial \varepsilon_t(\theta)}{\partial \gamma}\hat{\xi}_{1n,t}(\eta_t(\theta)) - \frac{1}{2h_t(\theta)}\frac{\partial h_t(\theta)}{\partial \gamma}\hat{\xi}_{2n,t}(\eta_t(\theta))\Big],$$

$$\hat{S}_{\gamma n}(\theta) = \frac{1}{n}\sum_{t=1}^{n}\Big[\frac{1}{h_t(\theta)}\frac{\partial \varepsilon_t(\theta)}{\partial \gamma}\frac{\partial \varepsilon_t(\theta)}{\partial \gamma'}\hat{I}_{1n}(\theta) + \frac{1}{4h_t^2(\theta)}\frac{\partial h_t(\theta)}{\partial \gamma}\frac{\partial h_t(\theta)}{\partial \gamma'}\hat{I}_{2n}(\theta)\Big].$$

Using the estimated score function and information matrix and letting

$$\hat{\gamma}_n = \bar{\gamma}_n + n^{-1/2}\hat{S}^{-1}_{\gamma n}(\bar{\theta}_n)\hat{W}_{\gamma n}(\bar{\theta}_n), \tag{5.24}$$

we have the following theorem.

Theorem 5.3. *Suppose that $\bar{\theta}_n$ is a discretized and $\sqrt{n}-$consistent estimator, f is symmetric and Assumptions 1-3 hold. Then, under $P^n_{\theta_0}$,*

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) \longrightarrow_{\mathscr{L}} N(0, \Omega_\gamma^{-1}),$$

as $n \to \infty$, and hence $\hat{\gamma}_n$ is an adaptive estimator, where Ω_γ is defined by (2.3).

When f is asymmetric, γ_0 is not adaptively estimable. Furthermore, δ_0 is not adaptively estimable regardless of the symmetry of f. We now use the least favorable path in (5.9) and the split sample technique in Schick (1986) and Drost et al. (1997) to construct an efficient estimator for θ_0.

Let k_n be an integer such that $k_n/n \to \tau \in (0,1)$. Split the residuals $\{\eta_1(\theta), \cdots, \eta_n(\theta)\}$ into two parts: $\{\eta_1(\theta), \cdots, \eta_{k_n}(\theta)\}$ and $\{\eta_{k_n+1}(\theta), \cdots, \eta_n(\theta)\}$. We estimate the density $f(x)$ by the kernel method:

$$\hat{f}_a^{(1)}(x,\theta) = \frac{1}{ak_n}\sum_{i=1}^{k_n} K\Big(\frac{x+\eta_i(\theta)}{a}\Big) \ and\ \hat{f}_a^{(2)}(x,\theta) = \frac{1}{a(n-k_n)}\sum_{i=k_n+1}^{n} K\Big(\frac{x+\eta_i(\theta)}{a}\Big).$$

Denote $\hat{\xi}_{1nt}^{(i)} = \hat{f}_{a_n}^{(i)\prime}(\eta_t(\bar{\theta}_n),\bar{\theta}_n)/[b_n+\hat{f}_{a_n}^{(i)}(\eta_t(\bar{\theta}_n),\bar{\theta}_n)]$, $\hat{\xi}_{2nt}^{(i)} = 1+\eta_t(\bar{\theta}_n)\hat{\xi}_{1nt}^{(i)}$, $\hat{\xi}_{nt}^{(i)} = (\hat{\xi}_{1nt}^{(i)}, -\hat{\xi}_{2nt}^{(i)})'$ and $\hat{\zeta}_{*nt}^{(i)} = \hat{\xi}_{nt}^{(i)} - diag\{1,1/2\}\hat{A}_n^{(i)-1}\hat{\zeta}_{nt}$, where $i=1,2$, $\hat{\zeta}_{nt} = (\eta_t(\bar{\theta}_n), [\eta_t^2(\bar{\theta}_n)-1]/2)'$, $\hat{A}_n^{(1)} = k_n^{-1}\sum_{t=1}^{k_n}(\hat{\zeta}_{nt}\hat{\zeta}_{nt}')$, $\hat{A}_n^{(2)} = (n-k_n)^{-1}\sum_{t=k_n+1}^{n}(\hat{\zeta}_{nt}\hat{\zeta}_{nt}')$ and $n^{-1}a_n^{-3}b_n^{-1}\to 0$. Furthermore, denote

$$\hat{\psi}_{1nt} = X_t(\bar{\theta}_n)\hat{\xi}_{nt}^{(2)} - \hat{v}_{1n}\hat{\zeta}_{*nt}^{(2)} \ and\ \hat{\psi}_{2nt} = X_t(\bar{\theta}_n)\hat{\xi}_{nt}^{(1)} - \hat{v}_{2n}\hat{\zeta}_{*nt}^{(1)},$$

where $\hat{v}_{1n} = k_n^{-1}\sum_{t=1}^{k_n}X_t(\bar{\theta}_n)$ and $\hat{v}_{2n} = (n-k_n)^{-1}\sum_{t=k_n+1}^{n}X_t(\bar{\theta}_n)$. Let

$$\hat{\theta}_n = \bar{\theta}_n + \Big[\sum_{t=1}^{k_n}(\hat{\psi}_{1nt}\hat{\psi}_{1nt}') + \sum_{t=k_n+1}^{n}(\hat{\psi}_{2nt}\hat{\psi}_{2nt}')\Big]^{-1}\Big(\sum_{t=1}^{k_n}\hat{\psi}_{1nt} + \sum_{t=k_n+1}^{n}\hat{\psi}_{2nt}\Big). \tag{5.25}$$

We have the following theorem.

Theorem 5.4. *Suppose that $\bar{\theta}_n$ is a discretized and $\sqrt{n}-$consistent estimator and Assumptions 1-3 hold. Then, under $P_{\theta_0}^n$,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \longrightarrow_{\mathscr{L}} N(0, I_*^{-1}),$$

as $n\to\infty$, and hence $\hat{\theta}_n$ is an efficient estimator.

The adaptive estimator in Theorem 5.4 uses the full sample without splitting and truncation. When f is symmetric, the estimators of m_0 in Theorems 5.3 and 5.4 are asymptotically equivalent. However, the one in Theorem 5.3 can provide smaller standard errors in the finite sample than that in Theorem 5.4. When f is asymmetric, the estimator $\hat{\theta}_n$ in Theorem 5.4 is more efficient than the CLSE and QMLE. But it is not as efficient as the estimator given by (5.1) when f is known. González-Rivera and Drost (1999) compared the efficiency of the MLE, the QMLE and the semi-parametric estimator with unknown f. The conclusion here is consistent with their results.

We next discusses Wald-tests for hypotheses:

$$H_0:\ d = d_0\ vs \tag{5.26}$$

$$H_1:\ d = d_n = u/\sqrt{n}. \tag{5.27}$$

in the case when f is unknown and symmetric. When f is unknown and asymmetric, Wald-tests can be similarly constructed using the efficient estimator in Theorem 5.4. When $H_0 : d = 0$ is rejected, y_t is a long-memory time series. We first consider the simple ARFIMA$(0,d,0)$-GARCH model. The Wald-test statistic is defined by

$$W_{An} = n(\hat{d}_n - d_0)^2 \hat{\sigma}_{dn}^2 \ \textit{and}$$
$$\hat{\sigma}_{dn}^2 = \frac{1}{n}\sum_{t=1}^{n}\Big[\frac{1}{\sqrt{h_t(\hat{\theta}_n)}}\frac{\partial \varepsilon_t(\hat{\theta}_n)}{\partial d}\Big]^2 \hat{I}_{1n}(\hat{\theta}_n) + \frac{1}{4n}\sum_{t=1}^{n}\Big[\frac{1}{\sqrt{h_t(\hat{\theta}_n)}}\frac{\partial h_t(\hat{\theta}_n)}{\partial d}\Big]^2 \hat{I}_{2n}(\hat{\theta}_n),$$

where $\hat{\theta}_n = (\hat{d}_n, \hat{\delta}_n)$, $\hat{d}_n$ is an adaptive estimator of d_0, and $\hat{\delta}_n$ is a $\sqrt{n}-$consistent estimator of δ_0. The following theorem gives the asymptotic distribution under the null hypothesis. It comes directly from Theorem 4.3.

Theorem 5.5. *If the assumptions of Theorem 3.1 hold, then*
(i) under H_0, $W_{An} \to_{\mathscr{L}} \chi_1^2$, as $n \to \infty$,
(ii) under H_1, $W_{An} \to_{\mathscr{L}} \chi_1^2(u^2\sigma_d^2)$ as $n \to \infty$, and hence it is asymptotically LMP, where σ_d^2 is defined as Ω_m in (2.3) with $m = d$

When h_t is a constant, we have $\sigma_d^2 = 6/\pi^2 I_1(f)$. When ε_t has GARCH, d_0 can be estimated by the CLSE, and its asymptotic variance is $\sigma_{dL}^2 = 36E(\varepsilon_t \partial\varepsilon_t/\partial d)^2/\pi^4$. The QMLE can be used to estimate d_0, and its asymptotic variance is $\sigma_{dQ}^2 = s^{-4}\sigma^2$, where $s^2 = E(h_t^{-1/2}\partial\varepsilon_t/\partial d)^2 + E(h_t^{-1/2}\partial h_t/\partial d)^2/2$ and $\sigma^2 = E(h_t^{-1/2}\partial\varepsilon_t/\partial d)^2 + E(h_t^{-1/2}\partial h_t/\partial d)^2(E\eta_t^4 - 1)/4$. Denote Wald-tests based on the CLSE and QMLE as W_{Ln} and W_{Qn}, respectively. It can be shown that their local powers have the asymptotic relationship, $W_{Ln} \le W_{Qn} \le W_{An}$, where the first equation holds if and only if h_t is a constant and the second equation holds if and only if f is normal.

For the general ARFIMA(p,d,q)-GARCH models, the Wald-test statistic is defined by

$$W_{An} = n(\hat{d}_n - d_0)^2(\hat{\sigma}_{dn}^2 - D_n'\Omega_{\phi n}D_n) \ \textit{and}$$
$$D_n = \frac{1}{n}\sum_{t=1}^{n}\Big[\frac{1}{h_t(\hat{\theta}_n)}\frac{\partial \varepsilon_t(\hat{\theta}_n)}{\partial d}\frac{\partial \varepsilon_t(\hat{\theta}_n)}{\partial \phi}\hat{I}_{1n}(\hat{\theta}_n) + \frac{1}{4h_t(\hat{\theta}_n)}\frac{\partial h_t(\hat{\theta}_n)}{\partial d}\frac{\partial h_t(\hat{\theta}_n)}{\partial \phi}\hat{I}_{2n}(\hat{\theta}_n)\Big],$$
$$\Omega_{\phi n} = \frac{1}{n}\sum_{t=1}^{n}\Big[\frac{1}{h_t(\hat{\theta}_n)}\frac{\partial \varepsilon_t(\hat{\theta}_n)}{\partial \phi}\frac{\partial \varepsilon_t(\hat{\theta}_n)}{\partial \phi'}\hat{I}_{1n}(\hat{\theta}_n) + \frac{1}{4h_t(\hat{\theta}_n)}\frac{\partial h_t(\hat{\theta}_n)}{\partial \phi}\frac{\partial h_t(\hat{\theta}_n)}{\partial \phi'}\hat{I}_{2n}(\hat{\theta}_n)\Big],$$

where $\phi = (\phi_1, \cdots, \phi_p, \psi_1, \cdots, \psi_q)'$, $\hat{\sigma}_{dn}^2$ is defined as in Section 4.1, $\hat{d}_n$ is an adaptive estimator of d_0, and $(\hat{\phi}_n, \hat{\delta}_n)$ is a $\sqrt{n}-$consistent estimator of (ϕ_0, δ_0). The following theorem comes from Theorem 3.1 and gives the asymptotic distribution under the null hypothesis.

Theorem 5.6. *If the assumptions of Theorem 3.1 hold, then,*
(i) under H_0, $W_{An} \to_{\mathscr{L}} \chi_1^2$, as $n \to \infty$,
(ii) under H_1, $W_{An} \to_{\mathscr{L}} \chi_1^2[u^2(\sigma_d^2 - D'\Omega_\phi^{-1}D)]$ as $n \to \infty$, and hence it is asymptotically LMP, where

$$D = E[\frac{1}{h_t(\theta_0)} \frac{\partial \varepsilon_t(\theta_0)}{\partial d} \frac{\partial \varepsilon_t(\theta_0)}{\partial \phi}] I_1(f) + E[\frac{1}{4h_t(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial d} \frac{\partial h_t(\theta_0)}{\partial \phi}] I_2(f),$$

$$\Omega_\phi = E[\frac{1}{h_t(\theta_0)} \frac{\partial \varepsilon_t(\theta_0)}{\partial \phi} \frac{\partial \varepsilon_t(\theta_0)}{\partial \phi'}] I_1(f) + E[\frac{1}{4h_t(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial \phi} \frac{\partial h_t(\theta_0)}{\partial \phi'}] I_2(f).$$

From Theorem 4.4, we see that the Wald-test in the case with $p \geq 1$ has lower power than in the case with $p = 0$, since $\hat{d}_n$ is asymptotically correlated with $\hat{\phi}_n$.

5.5 Simulation Study

This section examines the performance of the adaptive estimator in Theorem 3.1 in finite samples through Monte Carlo experiments. The following ARFIMA(0,d,0)-GARCH(1,1) and ARFIMA(1,d,0)-GARCH(1,1) models are used:

$$(1-B)^d y_t = \varepsilon_t, \tag{5.28}$$

$$(1-\phi B)(1-B)^d y_t = \varepsilon_t, \tag{5.29}$$

$$\varepsilon_t = \eta_t \sqrt{h_t} \textit{ and } h_t = \alpha_0 + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{5.30}$$

where $\eta_t = \xi_t / \sqrt{10}$ and ξ_t follows a mixed normal density $f(x) = [0.5e^{-(x-3)^2/2} + 0.5e^{-(x+3)^2/2}]/\sqrt{2\pi}$. This density function has been frequently used for investigating the finite-sample behavior of adaptive estimators, as in Kreiss (1987) and Shin and So (1999). 500 independent replications were used in each experiment. The experiments are carried out using Fortran 77 and the optimization algorithm from the Fortran subroutine DBCOAH in the IMSL library was used.

In the experiment, the bandwidth was taken to be

$$a_n = c_2^{-2/5} \Big\{ \int_{-\infty}^{\infty} K^2(x) dx \Big\}^{1/5} \Big\{ \int_{-\infty}^{\infty} f''^2(x) dx \Big\}^{-1/5} n^{-1/5}, \tag{5.31}$$

where $c_2 = \int_{-\infty}^{\infty} x^2 K(x) dx$ and $K(x)$ is the kernel as defined in Section 3. The bandwidth (5.4) was chosen because it asymptotically minimizes the mean integrated square error, $E \int_{-\infty}^{\infty} \{\hat{f}_n(x) - f(x)\}^2 dx$ [see Silverman (1986, pp.40)]. According to Silverman (1986, pp.50), the ingredient $\int_{-\infty}^{\infty} f''^2(x) dx$ in (5.4) can be consistently approximated by

$$\frac{1}{n^2 a_0} \sum_{i=2}^{n} \sum_{j=2}^{n} exp\Big\{ - \frac{[\eta_i(\bar{\theta}) - \eta_j(\bar{\theta})]^2}{4a_0^2} \Big\}, \tag{5.32}$$

where a_0 is an initial bandwidth. As suggested by Silverman (1986), (5.4)-(5.5) were used for iteration, starting with an initial value a_0. The optimal bandwidth a_n obtained through this data-driven method is stable in terms of the initial value a_0. As in Shin and So (1999), we take $a_0 = 0.4$ and $b_n = 0.001 a_n^{1/3}$. In our experiments, the adaptive estimator was found to be not sensitive to the values of a_0 and b_n.

We compare first the biases and standard deviations (SD) of the CLSE, the QMLE, the adaptive estimator (AE) and the MLE. The MLE was obtained by using the true score for iteration as in (3.1). The true observations were generated through models (5.1)-(5.3) and (5.2)-(5.3) with parameters: $(\alpha_0, \alpha, \beta) = (0.4, 0.2, 0.7)$, and $d = -0.3, 0.3, 0.7, 1.0, 1.3$ and 2.2 for models (5.1)-(5.3), and $(d, \phi) = (-0.3, 0.5), (0.3, 0.5)$ and $(1.3, 0.5)$ for models (5.2)-(5.3). Tables 1 and 2 summarize the empirical means and SDs of the estimators of d and (d, ϕ). [hcp]

TABLE 1

Empirical means and standard deviations of the estimated d in ARFIMA(0,d,0)-GARCH(1,1) models

$\alpha_0 = 0.4,\ \alpha = 0.2,\ \beta = 0.7$ and 500 replications

	$d = -0.3$	d=0.3	d=0.7	d=1.0	d=1.3	d=2.2
n=250						
CLSE Mean	-0.303	0.297	0.695	0.991	1.296	2.190
SD	0.055	0.054	0.055	0.055	0.054	0.055
QMLE Mean	-0.302	0.297	0.695	0.991	1.295	2.189
SD	0.048	0.047	0.048	0.047	0.047	0.044
AE Mean	-0.296	0.305	0.704	1.000	1.304	2.201
SD	0.023	0.021	0.023	0.020	0.021	0.021
MLE Mean	-0.296	0.305	0.704	1.000	1.304	2.201
SD	0.021	0.020	0.021	0.018	0.020	0.019
n=400						
CLSE Mean	-0.303	0.297	0.697	0.993	1.296	2.192
SD	0.041	0.041	0.041	0.041	0.041	0.041
QMLE Mean	-0.302	0.298	0.697	0.993	1.296	2.193
SD	0.036	0.036	0.036	0.036	0.036	0.034
AE Mean	-0.297	0.303	0.703	0.999	1.303	2.201
SD	0.018	0.017	0.018	0.016	0.017	0.016
MLE Mean	-0.297	0.304	0.703	1.000	1.303	2.201
SD	0.016	0.016	0.017	0.015	0.016	0.015

Tables 1 and 2 show that all the biases of these estimators are almost the same. The SDs of the AEs are much smaller than those of the QMLEs and the CLSEs and are very close to those of the MLEs. This is consistent with the theoretical results. As the sample size n is increased from 250 to 400, all the SDs become smaller. In each case, the AE can improve the QMLE by about 52% for models (5.1)-(5.3). The corresponding improvement for models (5.2)-(5.3) is about 42%. When ε_t is i.i.d. with a constant variance σ^2, the asymptotic SD of the CLSE of d in model (5.1) is $\sqrt{6/n\pi^2}$, which is 0.049 and 0.039 when $n = 250$ and 400, respectively. From Table 1, we see that all the SDs of CLSEs are larger than 0.049 and 0.039 for the sample sizes n =250 and 400 because of the GARCH effect. However, all the corresponding [hcp] SDs of the QMLEs, AEs and MLEs are smaller than 0.049 and 0.039. It should be noted that the QMLE improves just a little bit, but the improvements of the AE and MLE are quite significant. Some simulation results not reported here for cases with different true values of $(\alpha_0, \alpha, \beta)$ and for cases with the density f being the t^5-distribution and the double exponential distribution also support the same conclusion as above, except that the QMLE becomes much more efficient than the CLSE when α is larger. These results are available upon request from the author.

We now compare the size and power of the Wald-test based on the AE (W_{An}) as in Section 4 with those of the Wald-tests based on the CLSE (W_{Ln}) and based on the QMLE (W_{Qn}). The true observations are generated through models (5.1)-(5.3) and (5.2)-(5.3) with true parameters $(\alpha_0, \alpha, \beta) = (0.4, 0.2, 0.7)$, $d - d_0 = -0.2, -0.1, 0.0, 0.1, 0.2$ and $\phi = 0.5$ (only for model (5.2)). The rejection frequencies of these test statistics are their sizes and powers, respectively, when $d - d_0 = 0$ and $d - d_0 \neq 0$. Tables 3 and 4 summarize the empirical sizes and powers of these test statistics at the upper-tailed significance levels 5% and 10%. [hcp]

Table 3 shows that all of these tests yield similar sizes and these sizes tend to the corresponding nominal 5% and 10% when the sample size n is increased from 250 to 400. The powers of W_{An} are uniformally better than those of W_{Qn}, while the powers of W_{Qn} are uniformally better than those of W_{Ln}. The powers of all these test statistics are increased when the sample size n is increased from 250 to 400. Table 4 shows that the sizes of all these tests seem very nice when the sample size $n = 250$, but the W_{Ln} has no power for cases with $d - d_0 = -0.2$ and -0.1. In these two cases, W_{Qn} also loses some power, but the loss is not as serious as that of W_{Ln}. With the sample size $n = 400$, the power of W_{Ln} is increased, but the size is distorted at both the significance levels 5% and 10%. The overall performance of W_{Qn} is obviously better than that of W_{Ln}. W_{An} is significantly more powerful than W_{Qn}, even though its size is smaller than that of W_{Qn} when $n = 400$. [hcp]

These simulation results indicate that the performance of the adaptive estimator and the Wald-test based on this estimator is quite satisfactory in finite samples and that our theory and method should be useful in some practical applications.

5.6 Appendix: Proof of Theorems 5.2-5.5

Lemma 5.1. *If $d \in (0, 1/2)$ and Assumptions 3.4 and 5.3 hold, then there exists a bound subspace $\Theta_0 \subset \Theta$ such that, under $P^n_{\theta_0}$,*

TABLE 2

Empirical means and standard deviations of the estimated d and ϕ in ARFIMA(1,d,0)-GARCH(1,1) models

$\alpha_0 = 0.4,\ \alpha = 0.2,\ \beta = 0.7$ and 500 replications

	$d=-0.3$	$\phi=0.5$	$d=0.3$	$\phi=0.5$	$d=1.3$	$\phi=0.5$
n=250						
CLSE Mean	-0.301	0.483	0.289	0.493	1.285	0.498
SD	0.131	0.145	0.136	0.150	0.132	0.147
QMLE Mean	-0.289	0.473	0.292	0.491	1.285	0.497
SD	0.113	0.125	0.119	0.131	0.117	0.128
AE Mean	-0.294	0.488	0.307	0.483	1.312	0.486
SD	0.086	0.083	0.103	0.089	0.075	0.079
MLE Mean	-0.295	0.489	0.306	0.485	1.309	0.489
SD	0.083	0.077	0.099	0.082	0.073	0.079
n=400						
CLSE Mean	-0.303	0.490	0.292	0.495	1.289	0.498
SD	0.109	0.121	0.118	0.127	0.115	0.125
QMLE Mean	-0.291	0.479	0.293	0.495	1.289	0.499
SD	0.093	0.103	0.104	0.111	0.100	0.107
AE Mean	-0.294	0.494	0.311	0.489	1.310	0.490
SD	0.057	0.060	0.052	0.056	0.053	0.057
MLE Mean	-0.293	0.493	0.310	0.490	1.310	0.491
SD	0.052	0.055	0.048	0.052	0.048	0.052

TABLE 3

Sizes and powers of Wald-tests for the null hypothesis, H_0: $d = d_0$, in FARIMA(0,d,0)-GARCH(1,1) models

$d_0 = 1,\ \alpha_0 = 0.4,\ \alpha = 0.2,\ \beta = 0.7$ and 500 replications

	upper-tailed 5%					upper-tailed 10%				
$d-d_0=$	-0.2	-0.1	0.0	0.1	0.2	-0.2	-0.1	0.0	0.1	0.2
n=250										
W_{Ln}	.974	.486	.070	.452	.950	.986	.642	.116	.574	.972
W_{Qn}	.986	.556	.066	.462	952	.994	.672	.132	.572	.976
W_{An}	.994	.658	.070	.600	.988	1.00	.800	.120	.690	.992
n=400										
W_{Ln}	1.00	.712	.066	.638	.994	1.00	.792	.104	.738	1.00
W_{Qn}	.998	.732	.058	.642	.996	.998	.828	.116	.748	1.00
W_{An}	1.00	.864	.046	.794	1.00	1.00	.920	.094	.882	1.00

$$(i)\, E\sup_{\Theta_0}\varepsilon_t^2(\theta)<\infty,$$

$$(ii)\, E\sup_{\Theta_0}\Big\|\frac{\partial\varepsilon_t(\theta)}{\partial\gamma}\Big\|^2<\infty,$$

$$(iii)\, E\sup_{\Theta_0}\Big\|\frac{\partial^2\varepsilon_t(\theta)}{\partial\gamma\partial\gamma'}\Big\|^2<\infty,$$

$$(iv)\, E\sup_{\Theta_0}\Big\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial h_t(\theta)}{\partial\theta}\Big\|^2<\infty,$$

$$(v) E\sup_{\Theta_0}\Big\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial^2 h_t(\theta)}{\partial\theta\partial\theta'}\Big\|^2<\infty.$$

Proof. Let $d\in(0,c]\subset(0,d_0+0.5(0.5-d_0))$. Then, it follows that

$$|\varepsilon_t(\theta)|\le C\sum_{k=0}^{\infty}\frac{1}{k^{1+d}}|y_{t-k}|\le C\sum_{k=0}^{\infty}\frac{1}{k^{1+c}}|y_{t-k}|, \tag{5.33}$$

where C is some constant independent of θ. Thus, $E\sup_{\Theta_0}\varepsilon_t^2(\theta)<\infty$. Thus, (i) holds. It is straightforward to show that (ii) and (ii) hold.

$$\begin{aligned}\frac{\partial h_t(\theta)}{\partial\gamma}&=2\sum_{i=1}^{r}\alpha_i\varepsilon_{t-i}(\theta)\frac{\partial\varepsilon_{t-i}(\theta)}{\partial\gamma}+\sum_{i=1}^{s}\beta_i\frac{\partial h_{t-i}(\theta)}{\partial\gamma},\\ \frac{\partial h_t}{\partial\delta}&=\tilde{\varepsilon}_t+\sum_{i=1}^{s}\beta_i\frac{\partial h_{t-i}(\theta)}{\partial\delta},\end{aligned}$$

where $\tilde{\varepsilon}_t=(1,\varepsilon_{t-1}^2(\gamma),\cdots,\varepsilon_{t-r}^2(\gamma),h_{t-1}(\theta),\cdots,h_{t-s}(\theta))'$. Furthermore, we can show that

$$\begin{aligned}\Big\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial h_t(\theta)}{\partial\gamma}\Big\|&\le C\sum_{k=1}^{\infty}\rho^k\Big\|\frac{\partial\varepsilon_{t-k}(\theta)}{\partial\gamma}\Big\|\\ \Big\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial h_t}{\partial\delta}\Big\|&\le C\sum_{k=1}^{\infty}\rho^k|\varepsilon_{t-k}(\theta)|,\end{aligned}$$

where C and ρ are constants independent of θ and $\rho\in(0,1)$. By (i)-(ii) of this lemma, we know that (iv) holds. Similarly, we can show that (v) holds. □

Proof of Theorem 5.2. By Theorem 5.7 and Remark 5.1 in Appendeix 5.7, it is sufficient for Theorem 5.2 to prove that the following hold:

$$
\begin{aligned}
&(i) && \left[\inf_{1\le t\le n}\sqrt{h_t(\theta_0)}\right]^{-1}=O_p(1),\\
&(ii) && \sum_{t=1}^{n}\left[g_t(\tilde\theta_n)-g_t(\theta_0)-(\tilde\theta_n-\theta_0)'U_t(\theta_0)\right]^2=o_p(1),\\
&(iii) && \sup_{1\le t\le n}\left\|\frac{1}{\sqrt{n}}U_t(\theta_0)\right\|^2=o_p(1),\\
&(iv) && \frac{1}{n}\sum_{t=1}^{n}\|U_t(\theta_0)\|^2=O_p(1),\\
&(v) && \sum_{t=1}^{n}\left\|\frac{1}{\sqrt{n}}[U_t(\theta_n)-U_t(\theta_0)]\right\|^2=o_p(1),
\end{aligned}
$$

under $P^n_{\theta_0}$, where $g_t(\theta)=[\varepsilon_t(\theta),\sqrt{h_t(\theta)}]$ and $U_t(\theta)=\partial g_t(\theta)/\partial\theta$.

It is obvious that (i) holds and that (ii) comes from (v). $U_t(\theta)$ is also strictly stationary and ergodic since it is a fixed function of $\bar{y}_t$. Furthermore, $E\|U_t(\theta_0)\|^2<\infty$ by Lemma 5.1 and hence (iii) holds. By the ergodic theorem, (iv) holds. It remains to prove (v). As n is large enough, it follows that

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\Big\|\frac{1}{\sqrt{h_t(\theta_n)}}\frac{\partial h_t(\theta_n)}{\partial\theta}-\frac{1}{\sqrt{h_t(\theta_0)}}\frac{\partial h_t(\theta_0)}{\partial\theta}\Big\|^2\\
&\le\frac{O(1)}{n^2}\sum_{t=1}^{n}\Big(\Big\|\frac{1}{\sqrt{h_t(\theta_n^*)}}\frac{\partial h_t(\theta_n^*)}{\partial\theta}\Big\|^4+\Big\|\frac{1}{\sqrt{h_t(\theta_n^*)}}\frac{\partial^2 h_t(\theta_n^*)}{\partial\theta\partial\theta'}\Big\|^2\Big)\\
&\le O(1)\Big[\frac{1}{n^2}\sum_{t=1}^{n}\sup_{\theta\in\Theta_0}\Big\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial h_t(\theta)}{\partial\theta}\Big\|^4+\frac{1}{n^2}\sum_{t=1}^{n}\sup_{\theta\in\Theta_0}\Big\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial^2 h_t(\theta)}{\partial\theta\partial\theta'}\Big\|^2\Big]\\
&\equiv O(1)(A+B),
\end{aligned}
\tag{5.34}
$$

where θ_n^* is on the segment between θ_0 and θ_n, and Θ_0 is defined as in Lemma 5.1 such that (i)-(v) hold in that lemma. Note that

$$
g_t=\sup_{\theta\in V_{d_0}}\|h_t^{-1/2}(\theta)\partial h_t(\theta)/\partial\theta\|
$$

is a function in terms of $\bar{y}_t$ and hence it is strictly stationary and ergodic with $Eg_t^2<\infty$ (by Lemma 5.1). Thus, $[\max_{1\le t\le n}(n^{-1/2}g_t)]^2=o_p(1)$ under $P^n_{\theta_0}$. Furthermore, $\sum_{t=1}^n g_t^2/n=Eg_t^2+o_p(1)$ by the ergodic theorem and hence

$$
A\le[\max_{1\le t\le n}(n^{-1/2}g_t)]^2(n^{-1}\sum_{1}^{n}g_t^2)=o_p(1)
$$

under $P^n_{\theta_0}$. Similarly, we can show that $B=o_p(1)$ and

$$\frac{1}{n}\sum_{t=1}^{n}\Big\|\frac{\partial\varepsilon_t(\theta_n)}{\partial\theta}-\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\Big\|^2=o_p(1), \tag{5.35}$$

under $P^n_{\theta_0}$. Using (5.34) and (5.35), we can complete the proof. □

Proof of Theorem 5.3. By Corollary 5.1(iii), $P^n_{\theta_0}$ and $P^n_{\theta_n}$ are contiguous. Since $\hat{W}_n(\theta_n)$ and $W_n(\theta_n)$ are measurable in terms of $\mathscr{F}_n$, it is sufficient to prove the results under $P^n_{\theta_n}$. Ling and McAleer (2003) showed that $\hat{I}_{in}(\theta_n)=I_i(f)+o_p(1)$ under $P^n_{\theta_n}$. Furthermore, by Lemma 4.4 of Kreiss (1987), we only need to prove that

$$\hat{W}_{mn}(\theta_n)=W_{mn}(\theta_n)+o_p(1), \tag{5.36}$$

under $P^n_{\theta_n}$, where θ_n is a local sequence.

By (6.7) of Ling and McAleer (2003), $\Xi_{1,n}\equiv E_{\theta_n}\int(1+x^2)[\hat{\xi}_{1n}(x,\theta_n)-\xi_1(x)]^2 f(x)dx=o(1)$, where E_{θ_n} denotes the expectation under $P^n_{\theta_n}$. Let $c_n=\max\{5\tilde{c}n^{-2}a_n^{-4}b_n^{-2},\Xi_{1,n}^{1/4}\}$ for some constant $\tilde{c}$, $A_{n,t}=\{\int(1+x^2)[\hat{\xi}_{1n,t}(x,\theta)-\xi_1(x)]^2f(x)dx\le c_n\}$ and $A_n=\cap_{t=1}^n A_{n,t}$. Denote $\hat{\xi}_{in,t}(\eta_t(\theta_n),\theta_n)$ by $\hat{\xi}_{it}$, $i=1,2$. From the proof of Theorem 4.2 of Ling and McAleer (2003), it is sufficient to prove that

$$B_{1n}\equiv E_{\theta_n}\Big\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big[\frac{1}{\sqrt{h_t(\theta_n)}}\frac{\partial\varepsilon_t(\theta_n)}{\partial m}(\hat{\xi}_{1t}-\xi_1(\eta_t(\theta_n)))\Big]I_{A_n}\Big\}^2=o(1), \tag{5.37}$$

$$B_{2n}\equiv E_{\theta_n}\Big\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big[\frac{1}{h_t(\theta_n)}\frac{\partial h_t(\theta_n)}{\partial m}(\hat{\xi}_{2t}-\xi_2(\eta_t(\theta_n)))\Big]I_{A_n}\Big\}^2=o(1). \tag{5.38}$$

The proof of (B.10) is similar to that of (6.14) in Ling and McAleer (2003). It remains only to prove that (B.11) holds.

Note that $\hat{\xi}_{2t}$ and $\xi_2(\eta_t(\theta_n))$ are symmetric functions of $\eta_t(\theta_n)$. We use the symmetry of f and consider the cross-product terms in the expansion of B_{2n}. Denote $\xi_t^*=[\eta_t(\theta_n)\hat{\xi}_{1t}-\eta_t(\theta_n)\xi_1(\eta_t(\theta_n))]I_{A_n}$. Since f is symmetric, it can be shown that

$$\begin{aligned}
H_{nti}\ &\equiv E_{\theta_n}\Big[\frac{1}{h_{t+i}(\theta_n)h_t(\theta_n)}\frac{\partial h_{t+i}(\theta_n)}{\partial m}\frac{\partial h_t(\theta_n)}{\partial m'}\xi^*_{t+i}\xi^*_t\Big]\\
&=E_{\theta_n}\Big\{\frac{2}{h_{t+i}(\theta_n)h_t(\theta_n)}[\sum_{j=1}^{i}\upsilon_{\alpha_n\beta_n}(j)\varepsilon_{t+i-j}(\theta_n)\frac{\partial\varepsilon_{t+i-j-1}(\theta_n)}{\partial m}]\frac{\partial h_t(\theta_n)}{\partial m'}\xi^*_{t+i}\xi^*_t\\
&\quad+\frac{4}{h_{t+i}(\theta_n)h_t(\theta_n)}[\sum_{j=i+1}^{\infty}\upsilon_{\alpha_n\beta_n}(j)\varepsilon_{t+i-j}(\theta_n)\frac{\partial\varepsilon_{t+i-j-1}(\theta_n)}{\partial m}]\\
&\quad\cdot[\sum_{j=1}^{\infty}\upsilon_{\alpha_n\beta_n}(j)\varepsilon_{t-j}(\theta_n)\frac{\partial\varepsilon_{t-j-1}(\theta_n)}{\partial m'}]\xi^*_{t+i}\xi^*_t\Big\}\\
&=E_{\theta_n}\Big\{\frac{4}{h_{t+i}(\theta_n)h_t(\theta_n)}[\sum_{j=1}^{\infty}\upsilon_{\alpha_n\beta_n}(j+i)\upsilon_{\alpha_n\beta_n}(j)\varepsilon^2_{t-j}(\theta_n)\\
&\quad\frac{\partial\varepsilon_{t-j-1}(\theta_n)}{\partial m}\frac{\partial\varepsilon_{t-j-1}(\theta_n)}{\partial m'}]\xi^*_{t+i}\xi^*_t\Big\},
\end{aligned} \tag{5.39}$$

where the last equation holds because each term in the above equation is an odd function of $\eta_{t+i-j}(\theta_n)$ when $i+1 \le j \le \infty$, and $\upsilon_{\alpha\beta}(k)$ is the coefficient in $(\sum_{i=1}^r \alpha_i z^i)(1-\sum_{i=1}^s \beta_i z^i)^{-1} = \sum_{k=1}^\infty \upsilon_{\alpha\beta}(k) z^k$. Similar to Lemma 6.3 of Ling (2002), it can be shown that

$$\|H_{nti}\| \le O(1)\rho^i \sum_{j=1}^{\infty} \rho^j E_{\theta_n}\left[\left\|\frac{\partial \varepsilon_{t+i-j-1}(\theta_n)}{\partial m}\right\|^2 |\xi_{t+i}^* \xi_t^*|\right] \le O(c_n)\rho^i, \quad (5.40)$$

where $O(\cdot)$ holds uniformly for all t and $\rho \in (0,1)$ and the last inequality holds by Lemma B.1(ii) and because $E_{\theta_n}(\xi_t^{*2}|\mathscr{F}_{t-1}) \le c_n^2$. By (B.13), we have that

$$B_{2n} = \frac{1}{n}\sum_{t=1}^{n} H_{nt0} + \frac{2}{n}\sum_{t=1}^{n-1}\sum_{i=1}^{n-t} H_{nti} = o(1).$$

This completes the proof. □

5.7 Appendix: A General LAQ Criterion

This section presents a general LAQ criterion for (5.9), which is a generalization of the criteria in Jeganathan (1995), DKW (1997) and Koul and Schick (1997). Our discussion follows the fashion of Koul and Schick (1997).

Assumption 5.4 *There is a sequence G_n of invertible diagonal $k \times k$ matrices with $G_n^{-1} \to 0$ and $k \times 2$ random matrices $U_t(\theta)$ such that, for all bounded sequence θ_n and ϑ_n, the following statements are true with $\theta_n = \theta_0 + G_n^{-1}\theta_n$ and $\tilde{\theta}_n = \theta_n + G_n^{-1}\vartheta_n$:*

$$(i) \qquad \left[\inf_{1\le t\le n} \sqrt{h_t(\theta_n)}\right]^{-1} = O_{\theta_n}(1),$$

$$(ii) \qquad \sum_{t=1}^{n}\left[g_t(\tilde{\theta}_n) - g_t(\theta_n) - (\tilde{\theta}_n - \theta_n)' U_t(\theta_n)\right]^2 = o_{\theta_n}(1),$$

$$(iii) \qquad \sup_{1\le t\le n} \left\|G_n^{-1} U_t(\theta_n)\right\|^2 = o_{\theta_n}(1),$$

$$(iv) \qquad \sum_{t=1}^{n} \left\|G_n^{-1} U_t(\theta_n)\right\|^2 = O_{\theta_n}(1).$$

Assumption 5.5 *$\int |q_{\theta,f_c}(x) - q_{\theta_0,f}(x)|dx = o(1)$ as $||\theta - \theta_0|| = o(1)$ and $||c|| = o(1)$, where $f_c(x)$ is defined as in Definition 2.1.*

We further introduce the following notation:

$$W_n(\theta) = G_n^{-1}\sum_{t=1}^{n} X_t(\theta)\xi(\eta_t(\theta)),\ W_{\zeta n}(\theta) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\zeta(\eta_t(\theta)),$$

$$S_n(\theta) = G_n^{-1}\sum_{t=1}^{n} X_t(\theta)V_{\xi,\xi}X_t'(\theta)G_n^{-1},\ S_{\zeta n}(\theta) = \frac{G_n^{-1}}{\sqrt{n}}\sum_{t=1}^{n} X_t(\theta),$$

$$\tilde{W}_n(\theta) = \begin{pmatrix} W_n(\theta) \\ W_{\zeta n}(\theta) \end{pmatrix},\ \tilde{S}_n(\theta) = \begin{pmatrix} S_n(\theta) & S_{\zeta n}(\theta)V_{\zeta,\xi}' \\ V_{\zeta,\xi}S_{\zeta n}'(\theta) & V_{\zeta,\zeta} \end{pmatrix},$$

where $X_t(\theta) = U_t(\theta)/\sqrt{h_t(\theta)}$, $\xi = (\xi_1, -\xi_2)'$, $V_{\xi,\xi} = E[\xi(\eta_t)\xi'(\eta_t)]$, $V_{\zeta,\xi} = E[\xi(\eta_t)\zeta'(\eta_t)]'$ and $V_{\zeta,\zeta} = E[\zeta(\eta_t)\,\zeta'(\eta_t)]$. Now, we give the general LAQ criterion as follows.

Theorem 5.7. *Suppose that the path $c \to f_c$ is regular and that Assumptions 2.1-2.3 hold. Let $u_n = (\vartheta_n', v_n')'$ and v_n be a bounded sequence in R^l. Then:*
(a) $\Theta_n(\theta_n, \tilde{\theta}_n, v_n/\sqrt{n}) = u_n'\tilde{W}_n(\theta_n) - u_n'\tilde{S}_n(\theta_n)u_n/2 + o_{\theta_n}(1)$,
(b) $P_{\theta_0,n}$ and $P_{\theta_n,n}$ are contiguous,
(c) $\tilde{S}_n(\theta_n) = \tilde{S}_n(\theta_0) + o_{\theta_0}(1)$ and $\tilde{W}_n(\theta_n) = \tilde{W}_n(\theta_0) - \tilde{S}_n(\theta_0)\begin{pmatrix}\theta_n \\ 0\end{pmatrix} + o_{\theta_0}(1)$.

Remark 5.1. If the LAQ of $\Theta_n(\theta_n, \tilde{\theta}_n, v_n/\sqrt{n})$ is LAN, LAMN or LABF, then (b) automatically holds (see Kallianpur 1980, Ch. 7, and Jeganathan 1995, p. 850). In this case, it is sufficient to verify Assumption 2.2 with $\theta_n = \theta_0$ and that

$$\sum_{t=1}^{n} \left\| G_n^{-1}[U_t(\theta_n) - U_t(\theta_0)] \right\|^2 = o_{\theta_0}(1). \tag{5.41}$$

Assumption 2.3 means that the starting conditions have a negligible effect. If Y_0 is assumed to be independent of (θ, χ), as in the next section, then this assumption holds.

Before giving the proof of Theorem 2.1, we introduce the following notation and lemma. Let $\tilde{\xi} = (\xi', \zeta')'$, $+\tilde{G}_n =$diag $(G_n, \sqrt{n}I_l)$, $\tilde{U}_t(\theta) =$diag$(X_t(\theta), I_l)$, and $Y(\theta) = [y - Z_t(\theta)]/\sqrt{h_t(\theta)}$. For simplicity, we denote $Y(\theta_n)$ by Y and $Y(\tilde{\theta}_n)$ by Y_n. Similarly, denote h_t, h_{nt}, Z_t, Z_{nt}, g_t, g_{nt}, η_t, η_{nt}, $\tilde{U}_t$ and $\tilde{U}_{nt}$.

Lemma 5.2. *Under the assumptions of Theorem 2.1, it follows that:*

$$(a)\quad \sum_{t=1}^{n} E_{\theta_n}\left[(u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t))^2|\mathscr{F}_{t-1}\right] = O_{\theta_n}(1),$$

$$(b)\quad \sum_{t=1}^{n} E_{\theta_n}\left[(u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t))^2 I_{\{|u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t)|>\varepsilon\}}|\mathscr{F}_{t-1}\right] = o_{\theta_n}(1),$$

$$(c)\quad \left|\sum_{t=1}^{n}\{(u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t))^2 - E_{\theta_n}[(u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t))^2|\mathscr{F}_{t-1}]\}\right| = o_{\theta_n}(1),$$

$$(d)\quad \sum_{t=1}^{n}\int\left\{\frac{\sqrt{f_{v_n/\sqrt{n}}(Y_n)}}{\sqrt[4]{h_{nt}}} - \frac{\sqrt{f(Y)}}{\sqrt[4]{h_t}} - \frac{1}{2}u_n'\tilde{G}_n^{-1}\tilde{U}_t\frac{(\tilde{\xi}\sqrt{f})(Y)}{\sqrt[4]{h_t}}\right\}^2 dy = o_{\theta_n}(1).$$

Proof. The basic idea of the proof comes from LeCam (1970), Fabian and Hannan (1982), BKRW (1993), and DKW (1997).

Let $T_{nt} = 2[s_{v_n/\sqrt{n},t}(\tilde{\theta}_n)/s_t(\theta_n) - 1]$ and $B_n = \{\max_{1\le t\le n}|T_{nt}| < \varepsilon\}$ for some enough small $\varepsilon > 0$. Then, on the event B_n, the log-LR has the Taylor expansion:

$$\begin{aligned}\Theta_n(\theta_n, \tilde{\theta}_n, \frac{v_n}{\sqrt{n}}) &= 2\sum_{t=1}^{n}\log(1+\frac{1}{2}T_{nt}) + \Theta_{n0}\\ &= \sum_{t=1}^{n}T_{nt} - \frac{1}{4}\sum_{t=1}^{n}T_{nt}^2 + \frac{1}{6}\sum_{t=1}^{n}\alpha_{nt}T_{nt}^3 + \Theta_{n0},\end{aligned}$$

where $|\alpha_{nt}| < 1$ and $\Theta_{n0} = \log[q_{\tilde{\theta}_n, f_{v_n/\sqrt{n}}}(Y_0)/q_{\theta_n,f}(Y_0)] = o_{\theta_n}(1)$ by Assumption 2.3.

To prove (a), it is sufficient to show that

$$\sum_{t=1}^{n}\{T_{nt} - u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t) + \frac{1}{4}E_{\theta_n}[(u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t))^2|\mathscr{F}_{t-1}]\} = o_{\theta_n}(1), \tag{5.42}$$

$$\sum_{t=1}^{n}\{T_{nt}^2 - E_{\theta_n}[(u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t))^2|\mathscr{F}_{t-1}]\} = o_{\theta_n}(1), \tag{5.43}$$

$$\max_{1\le t\le n}|T_{nt}| = o_{\theta_n}(1) \text{ and } \sum_{t=1}^{n}T_{nt}^3 = o_{\theta_n}(1). \tag{5.44}$$

Note that $\int[s^y_{v_n/\sqrt{n},t}(\tilde{\theta}_n) - s^y_t(\theta_n)]^2dy = -E_{\theta_n}(T_{nt}|\mathscr{F}_n)$, where $s^y_t(\theta)$ is defined as $s_t(\theta)$ with $\eta_t(\theta)$ replaced by Y, and similarly define $s^y_{v_n/\sqrt{n},t}(\theta)$. By Lemma A.1 (a) and (d), and the inequality $|a^2 - b^2| \le (1+\alpha)(a-b)^2 + b^2/\alpha$ with $\alpha > 0$ and $a, b \in R$,

$$\begin{aligned}&\Big|\sum_{t=1}^{n}\{E_{\theta_n}(T_{nt}|\mathscr{F}_{t-1}) + \frac{1}{4}E_{\theta_n}[(u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t))^2|\mathscr{F}_{t-1}]\}\Big|\\ &\le (1+\alpha)\sum_{t=1}^{n}\int\Big[s^y_{v_n/\sqrt{n},t}(\tilde{\theta}_n) - s^y_t(\theta_n) - \frac{1}{2}u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(Y_t)s^y_t(\theta_n)\Big]^2dy\\ &\quad + \frac{1}{4\alpha}\sum_{t=1}^{n}E_{\theta_n}[(u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t))^2|\mathscr{F}_{t-1}] = o_{\theta_n}(1+\alpha) + O_{\theta_n}(\frac{1}{\alpha}) = o_{\theta_n}(1),\end{aligned} \tag{5.45}$$

where the last equation holds by first letting $n \to \infty$ and then letting $\alpha \to \infty$.

Let $D_{nt} = T_{nt} - u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t)$. $\sum_{t=1}^{n} E_{\theta_n}\{[D_{nt} - E_{\theta_n}(D_{nt}|\mathscr{F}_n)]^2|\mathscr{F}_{t-1}\} \le \sum_{t=1}^{n} E_{\theta_n}(D_{nt}^2|\mathscr{F}_{t-1}) = o_{\theta_n}(1)$ by Lemma A.1(d), and hence $\sum_{t=1}^{n}[D_{nt} - E_{\theta_n}(D_{nt}|\mathscr{F}_n)] = o_{\theta_n}(1)$ by Remark 3.7 (iii) in Fabian and Hannan (1982). Note that $E_{\theta_n}\tilde{\xi}(\eta_t) = 0$. We have $\sum_{t=1}^{n}[D_{nt} - E_{\theta_n}(T_{nt}|\mathscr{F}_n)] = o_{\theta_n}(1)$. Furthermore, by (A.5), we know that (A.2) holds. To (A.3), by Lemma A.1 (c), it is sufficient to show that

$$\Big|\sum_{t=1}^{n}\{T_{nt}^2 - [u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t)]^2\}\Big| = o_{\theta_n}(1). \tag{5.46}$$

Note that $\sum_{t=1}^{n} E_{\theta_n}[D_{nt}^2 I_{\{|D_{nt}|>\varepsilon\}}|\mathscr{F}_{t-1}] \le \sum_{t=1}^{n} E_{\theta_n}(D_{nt}^2|\mathscr{F}_{t-1}) = o_{\theta_n}(1)$ by Lemma A.1(d). By (3.15) of McLeish (1974), $\sum_{t=1}^{n} D_{nt}^2 = o_{\theta_n}(1)$. Now, by Lemma A.1 and using a similar argument as for (A.5), we can show that (A.6) holds. By Lemma A.1 (b) and (d), and following the steps in DKW (1997, p.794), we can show that $\max_{1\le t\le n}|T_{nt}| = o_{\theta_n}(1)$. By (A.3) and Lemma A.1(a), we have $\sum_{t=1}^{n} T_{nt}^2 = O_{\theta_n}(1)$, and hence $\sum_{t=1}^{n} T_{nt}^3 = o_{\theta_n}(1)$. Thus, (A.4) holds.

By (a) of this theorem, $\Theta_n(\theta_0,\theta_n,0) = u_n'\tilde{W}_n(\theta_0) - u_n'\tilde{S}_n(\theta_0)u_n/2 + o_{\theta_0}(1)$, and $\Theta_n(\theta_0,\theta_n,0) = -\Theta_n(\theta_n,\theta_n+G_n^{'-1}(-\theta_n),0) = -[\tilde{u}_n'\tilde{W}_n(\theta_n) - \tilde{u}_n'\tilde{S}_n(\theta_n)\tilde{u}_n/2] + o_{\theta_n}(1)$ with $\tilde{u}_n = (-\theta_n',0)'$. By Assumptions 2.1 and 2.2, we can show that $\tilde{W}_n(\theta_n) = O_{\theta_n}(1)$ and $\tilde{S}_n(\theta_n) = O_{\theta_n}(1)$. Note that $\tilde{W}_n(\theta_n)$ and $\tilde{S}_n(\theta_n)$ are measurable in terms of $\mathscr{F}_n$ and hence they are bounded under $P_{\theta_n,n}$. Thus, $\Theta_n(\theta_0,\theta_n,0)$ is bounded under both $P_{\theta_0,n}$ and $P_{\theta_n,n}$, which implies (b). The first part of (c) holds by Assumption 2.2 and the second part holds by exploring the equation: $\Theta_n(\theta_0,\theta_n,0) + \Theta_n(\theta_n,\theta_n + G_n^{'-1}\vartheta_n, v_n/\sqrt{n}) - \Theta_n(\theta_0,\theta_n + G_n^{'-1}\vartheta_n, v_n/\sqrt{n}) = 0$. This completes the proof. □

Proof. By Assumptions 4 and A.1 (i), (iv) and (vi), and the finiteness of $\int ||\zeta(x)||f(x)dx$, (a) holds. By Assumptions 4 and A.1, and using similar argument as in Koul and Schick (1997, p253), we can show that (b) holds. By (3.15) of McLeish (1974) and (a)-(b) of this lemma, (c) holds.

The proof of (d) is similar to that for (2.15) in Koul and Schick (1997). The right-hand-side of (d) is bounded by $3(T_{1n}+T_{2n}+T_{3n})$, where

$$T_{1n} = \sum_{t=1}^{n}\int h_{nt}^{-\frac{1}{2}}(f_{\tilde{v}_n/\sqrt{n}}^{\frac{1}{2}} - f^{\frac{1}{2}} - \frac{\tilde{v}_n'}{2\sqrt{n}}\zeta f^{\frac{1}{2}})^2(Y_n)dy$$
$$= n\int (f_{\tilde{v}_n/\sqrt{n}}^{\frac{1}{2}} - f^{\frac{1}{2}} - \frac{\tilde{v}_n'}{2\sqrt{n}}\zeta f^{\frac{1}{2}})^2(x)dx = o(1)$$

by Definition 3.1;

$$\begin{aligned} T_{2n} &= \frac{||\tilde{v}_n||}{n}\sum_{t=1}^{n}\int \left\| h_{nt}^{-\frac{1}{4}}(\zeta f^{\frac{1}{2}})(Y_n) - h_t^{-\frac{1}{4}}(\zeta f^{\frac{1}{2}})(Y)\right\|^2 dy \\ &\le O(1)R_{2n}^2\int ||(\zeta f^{\frac{1}{2}})(x)||^2 dx \\ &\quad + O(1)\sup_{|s_1|\le R_{1n},|s_2|\le R_{2n}}\int ||(\zeta f^{\frac{1}{2}})(x(1+s_2)+s_1) - (\zeta f^{\frac{1}{2}})(x)||^2 dx \\ &= o_{\lambda_n}(1), \end{aligned}$$

where $R_{1n} = [\max_{1\le t\le n}(|Z_{nt} - Z_t|h_{nt}^{-\frac{1}{2}})^2]^{1/2} = o_{\lambda_n}(1)$ and $R_{2n} = [\max_{1\le t\le n}(|\sqrt{h_{nt}} - \sqrt{h_t}|h_{nt}^{-\frac{1}{2}})^2]^{1/2} = o_{\lambda_n}(1)$ by Assumption A.1 (i)-(iii), and the above equation holds by $\int ||\zeta(x)||f(x)dx < \infty$ and Lemma 19 in Jeganathan (1995); and

$$T_{3n} = \sum_{t=1}^{n}\int \left[h_{nt}^{-\frac{1}{4}}f^{\frac{1}{2}}(Y_n) - h_t^{-\frac{1}{4}}f^{\frac{1}{2}}(Y) - \frac{1}{2}\tilde{\hbar}_n' G_n^{'-1}U_t h_t^{-\frac{3}{4}}(\varpi f^{\frac{1}{2}})(Y)\right]^2 dy.$$

In order to show that $T_{3n} = o_{\lambda_n}(1)$, denote $U_{nt}^* = g_{nt} - g_t$, $Y_n^* = [y - Z_t + u(Z_{nt} - Z_t)]h_{nt}^{*-\frac{1}{2}}$ and $h_{nt}^* = [h_t^{\frac{1}{2}} + u(h_{nt}^{\frac{1}{2}} - h_t^{\frac{1}{2}})]^2$. By Assumption 4 and using Cauchy's form of Taylor's theorem, T_{3n} is bounded by

$$\begin{aligned}
&\sum_{t=1}^{n} \int_0^1 \int \left[U_{nt}^* h_{nt}^{*-\frac{3}{4}} (\varpi f^{\frac{1}{2}})(Y_n^*) - \tilde{\hbar}_n' G_n'^{-1} U_t h_t^{-\frac{3}{4}} (\varpi f^{\frac{1}{2}})(Y) \right]^2 dydu \\
&\quad \le \sum_{t=1}^{n} \int_0^1 \int \left\{ \tilde{\hbar}_n' G_n'^{-1} U_t \left[h_{nt}^{*-\frac{3}{4}} (\varpi f^{\frac{1}{2}})(Y_n^*) - h_t^{-\frac{3}{4}} (\varpi f^{\frac{1}{2}})(Y) \right] \right\}^2 dydu \\
&\quad + \sum_{t=1}^{n} \left\| U_{nt}^* - \hbar_n' G_n'^{-1} U_t \right\|^2 h_t^{-1} \int \left\| (\varpi f^{\frac{1}{2}})(x) \right\|^2 dx, \qquad (5.47)
\end{aligned}$$

where the second term is $o_{\lambda_n}(1)$ by Assumption 4 and A.1(i)-(ii), and the first term is bounded by

$$\begin{aligned}
&\Big[O(1) \sup_{|s_1| \le R_{1n}, |s_2| \le R_{2n}} \int \left\| (\varpi f^{\frac{1}{2}})(x(1+s_2)+s_1)) - (\varpi f^{\frac{1}{2}})(x) \right\|^2 dx \\
&\quad + O(1) R_{2n}^2 \int \left\| (\varpi f^{\frac{1}{2}})(x) \right\|^2 dx \Big] \sum_{t=1}^{n} \left\| G_n'^{-1} U_{nt} \right\|^2 = o_{\lambda_n}(1),
\end{aligned}$$

by Assumptions 4 and A.1 (vi), and Lemma 19 in Jeganathan (1995). Thus, $T_{3n} = o_{\lambda_n}(1)$ and hence (d) holds. This complete the proof. $\square$

TABLE 4

Sizes and powers of Wald-tests for the null hypothesis, H_0: $d = d_0$, in ARFIMA(1,d,0)-GARCH(1,1) models

$d_0 = 1,\ \alpha_0 = 0.4,\ \alpha = 0.2,\ \beta = 0.7$ and 500 replications

	upper-tailed 5%					upper-tailed 10%				
$d - d_0$=	−0.2	−0.1	0.0	0.1	0.2	−0.2	−0.1	0.0	0.1	0.2
n=250										
W_{Ln}	.038	.022	.052	.168	.288	.170	.094	.116	.226	.382
W_{Qn}	.144	.144	.056	.148	.316	.338	.212	.104	.212	.394
W_{An}	.818	.226	.058	.260	.776	.890	.372	.092	.386	.850
n=400										
W_{Ln}	.266	.074	.092	.190	.404	.524	.188	.126	.254	.522
W_{Qn}	.356	.152	.058	.184	.450	.538	.258	.090	.266	.544
W_{An}	.978	.484	.024	.450	.934	.990	.674	.040	.572	.964

Chapter 6
LAD-type Estimation

In the regression setup, one of advantages of the least absolute deviation estimator (LAD) is that it does not require any moment condition on the errors to obtain the asymptotic normality, see Bassett and Koenker (1982), Koenker and Bassett (1982) and Portnoy and Koenker (1989). However, when we used this method to the time series models, such as in Koenker and Zhao (1996), Davis and Dunsmuir (1997), Mukherjee (1999), and Ling and McAleer (2003), this advantage is disappeared. This chapter considers the self-weighted LAD estimator proposed by Ling (2005).

6.1 Self-weighted LAD for AR Models

Assume that the random sample $\{Y_1, \cdots, Y_n\}$ is from the AR(p) model:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t, \tag{6.1}$$

where $\{\varepsilon_t\}$ is i.i.d. errors with a common distribution F. Denote $\theta = (\phi_0, \phi_1, \cdots, \phi_p)'$. When $E\varepsilon_t^2 < \infty$, it is well known that all kinds of the estimators of θ are asymptotically normal and various methods are available to do inferences for the model.

When $E\varepsilon_t^2 = \infty$, model (6.1) is called the infinite variance AR (IVAR) model. This kind of models displaying the features of heavy tails are encountered in several fields, such as teletraffic engineering in Duffy, et al. (1994), hydrology in Castilo (1988), and economics and finance in Koedijk, et al. (1990) and Janson and de Vries (1991). Kanter and Steiger (1974) showed the weak consistency of the least squares estimator (LSE) of θ. Furthermore, Hannan and Kanter (1977) proved its strong consistency with a convergent rate $n^{1/\delta}$, where n is the sample size, $\delta > \alpha$ and $\alpha \in (0, 2)$ is the stable index of ε_t. The limiting distribution of the LSE had not been available until Davis and Resnick (1986). Based on the point processes, they showed that the LSE converges weakly to the ratio of two stable random variables with the rate $n^{1/\alpha} L_\nu(n)$, where $L_\nu(n)$ is a slowly varying function. The LAD was considered by Gross and Steiger (1979) and its strong consistency was proved. An and Chen (1982) showed that a convergent rate of the LAD is $n^{1/\delta}$. The asymptotic theory of the LAD and M-estimator of θ was completely established by Davis, et al.(1992). They showed that these estimators converge weakly to the minimum of a stochastic process with the rate $a_n = \inf\{x{:}P(|\varepsilon_t| > x) \le n^{-1}\}$. It is hard to do statistical inference in this case.

We use the following objective function:

$$L_n(\theta) = \sum_{t=1}^{n} w_t |Y_t - X'_{t-1}\theta|,$$

where $X_t = (1, Y_t, \cdots, Y_{t-p+1})'$ and $w_t = g(Y_{t-1}, \cdots, Y_{t-p})$ with a given real function g. The minimizer of $L_n(\theta)$ is called the self-weighted LAD estimator $\hat{\theta}_n$ of θ_0. We state the regularity conditions as follows.

Assumption 6.1 *The characteristic polynomial* $1 - \phi_1 z - \cdots - \phi_p z^p$ *has all roots outside the unit circle and* $E|\varepsilon_t|^\alpha < \infty$ *for some* $\alpha > 0$.

Assumption 6.2 $E(w_t\|X_t\|^2) < \infty$ *and* $E(w_t^2\|X_t\|^2) < \infty$.

Assumption 6.3 $E(w_t\|X_{t-1}\|^3) < \infty$ *and* $F(x)$ *has a differential density* $f(x)$ *everywhere in* R *with* $\sup_{x\in R}|f'(x)| < \infty$.

The condition 6.2 is to downweight the big *leverage* points in the covariance matrix. Technically, we need to linearize $L_n(\theta)$ around θ_0. But as shown in Davis, Knight and Liu (1992), the usual Taylor's series expansion does not work for IVAR models since the remainder term from the quadratic form does not go to zero. The condition 6.3 is to overcome this difficulty, see (3.? in in Section 3.

Theorem 6.1. *If Assumptions 6.1-6.3 are satisfied and* ε_t *has zero median with* $f(0) > 0$*, then it follows that*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \longrightarrow_{\mathscr{L}} N\Big(0, \frac{1}{4f^2(0)}\Sigma^{-1}\Omega\Sigma^{-1}\Big),$$

where $\Sigma = E(w_t X_{t-1}X'_{t-1})$ *and* $\Omega = E(w_t^2 X_{t-1}X'_{t-1})$.

Proof. Denote $\hat{\gamma}_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$ and

$$\widetilde{L}_n(u) = \sum_{t=1}^{n} w_t\Big[|\varepsilon_t - \frac{1}{\sqrt{n}}u'X_t| - |\varepsilon_t|\Big],$$

where $u \in R^{p+1}$. Then $\hat{\gamma}_n$ is the minimizer of $\widetilde{L}_n(u)$ on R^{p+1}. Using the identity:

$$|x-y| - |x| = -y[I(x>0) - I(x<0)] + 2\int_0^y [I(x \le s) - I(x \le 0)]ds,$$

which holds when $x \neq 0$ [see Knight(1998)], it follows that

$$\widetilde{L}_n(u) = -u'T_n + T_{1n}(u), \tag{6.2}$$

where $T_n = \sum_{t=1}^{n} w_t X_t [I(\varepsilon_t > 0) - I(\varepsilon_t < 0)]/\sqrt{n}$,

$$T_{1n}(u) = 2\sum_{t=1}^{n} \xi_t(u) \text{ and } \xi_t(u) = w_t \int_0^{u'X_t/\sqrt{n}} [I(\varepsilon_t \le s) - I(\varepsilon_t \le 0)]ds.$$

We can write $T_{1n}(u)$ as

$$T_{1n}(u) = 2\sum_{t=1}^{n} E[\xi_t(u)|\mathscr{F}_{t-1}] + 2\sum_{t=1}^{n}\{\xi_t(u) - E[\xi_t(u)|\mathscr{F}_{t-1}]\}.$$

Denote the distribution of ε_t by $F(x)$. By Taylor's expansion, we have

$$\begin{aligned}\sum_{t=1}^{n} E[\xi_t(u)|\mathscr{F}_{t-1}] &= \sum_{t=1}^{n} w_t \int_0^{u'X_t/\sqrt{n}} [F(s) - F(0)]ds \\ &= \sum_{t=1}^{n} w_t \int_0^{u'X_t/\sqrt{n}} [sf(0) + \frac{1}{2}s^2 f'(s^*)]ds \\ &= u'\Big[\frac{f(0)}{2n}\sum_{t=1}^{n} w_t X_t X_t'\Big]u + \sum_{t=1}^{n} w_t \int_0^{u'X_t/\sqrt{n}} [s^2 f'(s^*)]ds, \quad (6.3)\end{aligned}$$

where $s^* \in (0, s)$. Thus, we have

$$\widetilde{L}_n(u) = -u'T_n + f(0)u'P_n u + R_n(u), \tag{6.4}$$

where $P_n = \sum_{t=1}^{n} w_t X_t X_t'/n$ and

$$R_n(u) = 2\sum_{t=1}^{n}\{\xi_t(u) - E[\xi_t(u)|\mathscr{F}_{t-1}]\} + 2\sum_{t=1}^{n} w_t \int_0^{u'X_t/\sqrt{n}} [s^2 f'(s^*)]ds.$$

Let $a_t = w_t X_t [I(\varepsilon_t > 0) - I(\varepsilon_t < 0)]$. By Assumption 6.2, we can claim that $E(a_t a_t') = \Omega < \infty$. By Assumptions 6.1-6.2 and the ergodic theorem, we have

$$P_n \to \Sigma \text{ and } \frac{1}{n}\sum_{t=1}^{n}(w_t^2 X_t X_t') \to \Omega, \tag{6.5}$$

almost surely, as $n \to \infty$. Since $\{a_t\}$ is strictly stationary and ergodic with $E(a_t|\mathscr{F}_{t-1}) = 0$, by the central limit theorem, it follows that

$$T_n \longrightarrow_{\mathscr{L}} \Phi, \tag{6.6}$$

where $\Phi \sim N(0, \Omega)$. By Assumptions 6.2-6.3, for each u, it follows that

$$\Big|\sum_{t=1}^{n} w_t \int_0^{u'X_t/\sqrt{n}} [s^2 f'(s^*)]ds\Big| \le \sup_x |f'(x)| \frac{O(1)}{n\sqrt{n}} \sum_{t=1}^{n} w_t \|X_t\|^3 \to 0, \tag{6.7}$$

almost surely, as $n \to \infty$. As for (6.2), we can show that

$$
\begin{aligned}
E\xi_t^2(u) &= E\Big[w_t\int_0^{u'X_t/\sqrt{n}}|F(s)-F(0)|ds\Big]^2\\
&\le E\Big[w_t\int_0^{u'X_t/\sqrt{n}}|F(s)-F(0)|^{1/2}ds\Big]^2\\
&\le \max_x f(x)E\Big[w_t\int_0^{u'X_t/\sqrt{n}}|s|^{1/2}ds\Big]^2 \le n^{-1.5}\max_x f(x)\|u\|^3E(w_t^2\|X_t\|^3),
\end{aligned}
$$

where $\max_x f(x)<\infty$ and $E(w_t^2\|X_t\|^3)<\infty$ by Assumptions 6.2-6.3. For each u,

$$
\begin{aligned}
E\Big(\sum_{t=1}^n\{\xi_t(u)-E[\xi_t(u)|\mathscr{F}_{t-1}]\}\Big)^2 &= \sum_{t=1}^n E\{\xi_t(u)-E[\xi_t(u)|\mathscr{F}_{t-1}]\}^2\\
&\le 2\sum_{t=1}^n E\xi_t^2(u)\to 0, \qquad (6.8)
\end{aligned}
$$

as $n\to\infty$. By the preceding equation and (6.6), we know that $R_n(u)=o_p(1)$ for each u. Furthermore, by (6.3)-(6.5), for each u, it follows that

$$\widetilde{L}_n(u)\longrightarrow_{\mathscr{L}}\widetilde{L}(u)\equiv -u'\Phi+f(0)u'\Sigma u.$$

$\widetilde{L}(u)$ has a unique minimum at $u=[2f(0)]^{-1}\Sigma^{-1}\Phi$ almost surely. As for Corollary 2 in Knight (1998), by the convexity of $\widetilde{L}_n(u)$ for each n, we can show that

$$\hat{\gamma}_n=\sqrt{n}(\hat{\theta}_n-\theta_0)\longrightarrow_{\mathscr{L}} N(0,\frac{1}{4f^2(0)}\Sigma^{-1}\Omega\Sigma^{-1}).$$

This completes the proof. $\square$

When $E\varepsilon_t^2<\infty$, we can take $\omega_t=1$. The SLAD reduces to the standard LAD estimator. When $E\varepsilon_t^2=\infty$, the results in Theorem 6.1 is surprising and entirely different from those in the literature. Using the logistic kernel $K(x)=e^{-x}/(1+e^{-x})^2$ and the bandwidth $b_n=c/n^{\nu}$ with $\nu\in(0,1/2)$ and constant $c>0$, we can estimate $f(0)$ by

$$\hat{f}_n(0)=\frac{1}{\hat{\sigma}_w nb_n}\sum_{t=1}^n w_tK\Big(\frac{y_t-\hat{\theta}_n'X_t}{b_n}\Big), \qquad (6.9)$$

where $\hat{\sigma}_w=n^{-1}\sum_{t=1}^n w_t$. Σ and Ω can be estimated, respectively, by

$$\hat{\Sigma}_n=\frac{1}{n}\sum_{t=1}^n(w_tX_tX_t') \text{ and } \hat{\Omega}_n=\frac{1}{n}\sum_{t=1}^n(w_t^2X_tX_t'). \qquad (6.10)$$

Theorem 2.1 and (2.1)-(2.2) now can be used to do statistical inferences for the IVAR model. Here, we consider the p_1 linear hypothesis of the form:

$$H_0:\Gamma\theta_0=\gamma,$$

in the usual notation. The Wald test statistic for the hypothesis H_0 is defined as

$$W_n(p_1) = 4n\hat{f}_n^2(0)(\Gamma\hat{\theta}_n - \gamma)'\Big(\Gamma\hat{\Sigma}_n^{-1}\hat{\Omega}_n\hat{\Sigma}_n^{-1}\Gamma'\Big)^{-1}(\Gamma\hat{\theta}_n - \gamma).$$

The following theorem gives the limiting distribution of $W_n(p_1)$:

Theorem 6.2. *If Assumptions 6.1-6.3 and $b_n = O(1/n^{\nu})$ with $\nu \in (0,1/2)$, then under H_0, it follows that $W_n(p_1) \longrightarrow_{\mathscr{L}} \chi^2_{p_1}$.*

Proof. Denote $A_t(u) = \int K(y)f(b_n y + u'X_t/\sqrt{n})dy$, where $u \in R^{p+1}$. Then, we have

$$E\Big\{\frac{1}{b_n}K\Big(\frac{\varepsilon_t - u'X_t/\sqrt{n}}{b_n}\Big)\Big]\Big|\mathscr{F}_{t-1}\Big\} = \frac{1}{b_n}\int K\Big(\frac{x - u'X_t/\sqrt{n}}{b_n}\Big)f(x)dx = A_t(u).$$

Using this with $\int K(x)dx = 1$ and $\int K^2(x)dx < \infty$, for each u, it follows that

$$\begin{aligned}
& E\Big\{\frac{1}{n}\sum_{t=1}^n w_t\Big[\frac{1}{b_n}K\Big(\frac{\varepsilon_t - u'X_t/\sqrt{n}}{b_n}\Big) - A_t(u)\Big]\Big\}^2 \\
&= \frac{1}{n^2}\sum_{t=1}^n E\Big\{w_t^2\Big[\frac{1}{b_n}K\Big(\frac{\varepsilon_t - u'X_t/\sqrt{n}}{b_n}\Big) - A_t(u)\Big]\Big\}^2 \\
&\le \frac{2}{n^2 b_n}\sum_{t=1}^n E\Big[\frac{w_t^2}{b_n}K^2\Big(\frac{\varepsilon_t - u'X_t/\sqrt{n}}{b_n}\Big)\Big] + \frac{2}{n^2}\sum_{t=1}^n E[w_t^2 A_t^2(u)] \\
&\le \frac{2}{n^2 b_n}\sum_{t=1}^n E\Big[w_t^2\int K^2(y)f(b_n y + u'X_t/\sqrt{n})dy\Big] + \frac{2}{n^2}\sum_{t=1}^n E[w_t^2 A_t^2(u)] \\
&\le \frac{2\max_x f(x)Ew_t^2}{nb_n}\int K^2(y)dy + \frac{2\max_x f(x)Ew_t^2}{n} = o(1), \qquad (6.11)
\end{aligned}$$

by Assumptions 6.2-6.3. By (6.8) and the continuity of f, we can further show that

$$\sup_{\|u\|\le M}\Big|\frac{1}{n}\sum_{t=1}^n w_t\Big[\frac{1}{b_n}K\Big(\frac{\varepsilon_t - u'X_t/\sqrt{n}}{b_n}\Big) - A_t(u)\Big| = o_p(1).$$

Let $\hat{\theta}_n = \sqrt{n}(\hat{\phi}_n - \phi_0)$. Since $\hat{\theta}_n = O_p(1)$, by the preceding equation, it follows that

$$\frac{1}{n}\sum_{t=1}^n w_t\Big[\frac{1}{b_n}K\Big(\frac{\varepsilon_t - \hat{\theta}_n' X_t/\sqrt{n}}{b_n}\Big) - A_t(\hat{\theta}_n)\Big] = o_p(1).$$

Since $\int |y|K(y)dy < \infty$, by Taylor's expansion and Assumptions 6.2-6.3, we have

$$
\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n} w_t\Big|A_t(\hat{\theta}_n)-f(0)\Big| &= \frac{1}{n}\sum_{t=1}^{n} w_t\Big|\int K(y)[f(b_n y+\hat{\theta}_n' X_t/\sqrt{n})-f(0)]dy\Big| \\
&= \frac{1}{n}\sum_{t=1}^{n} w_t\Big|\int K(y)(b_n y+\hat{\theta}_n' X_t/\sqrt{n})]f'(\xi_t^*)dy\Big| \\
&\leq \max_x |f'(x)|\Big[\frac{b_n}{n}\sum_{t=1}^{n} w_t\int |y|K(y)dy+\frac{\|\hat{\theta}_n\|}{n^{3/2}}\sum_{t=1}^{n} w_t\|X_t\|\Big] \\
&= O_p(b_n)+O_p(\frac{1}{\sqrt{n}})=o_p(1),
\end{aligned}
$$

where ξ_t^* lies between 0 and $b_n y+\hat{\theta}_n' X_t$. By the preceding two inequalities, we can readily show that $\hat{f}_n(0)=f(0)+o_p(1)$. Finally, by Theorem 2.1 and (6.4), the conclusion holds. This completes the proof. $\square$

To use the results in this section, we need to select the weight w_t in practice. It seems reasonable to use the following weight analog to Huber's influence function:

$$
w_t=\begin{cases} 1 & if\ a_t=0, \\ C^3/a_t^3 & if\ a_t\neq 0, \end{cases} \tag{6.12}
$$

where $a_t=\sum_{i=1}^{p}|y_{t-i}|(|y_{t-i}|\geq C)$ and $C>0$ is a constant. This weight satisfies Assumption 2.2. It downweights the covariance matrix with leverage points but take full advantage of all those without leverage points. As Huber's estimator for regression models, we need to choose C. When $P(\varepsilon_t>x)=P(\varepsilon_t<-x)=x^{-\alpha}$ with $\alpha<2$ for $x\geq 1$, by (2.7) in Davis and Resnick (1985), we can show that

$$
\Sigma^{-1}\Omega\Sigma^{-1}=O(1)(E[y_{t-1}^2 I(|y_{t-1}|\leq C)]+C^{2-\alpha})^{-1},
$$

as $p=1$ and $C\to\infty$. Thus, the larger is C, the smaller is the asymptotic variance. However, for a large C, the distribution of $\hat{\phi}_n$ may not be well approximated by its limiting distribution as n is small. We still have not a theory to support the choice of C. However, our simulation results in Section 3 show that it works well when C is the 95%-quantile of data $\{y_1,\cdots,y_n\}$. Also we note that a small tailed index α results in a small asymptotic variance. Obviously, there are a lot of other weights, such as $w_t=(1+C\|X_t\|^2)^{-3/2}$ and $I(\max_{1\leq i\leq p}|y_{t-1}|\leq C)$, that satisfy Assumption 2.2. However, our simulation results not reported in this paper show that the SLAD based on w_t in (2.3) is much more efficient than that based on these weights.

Lemma (Davis, et al, 1992) Let $\{V_n(u)\}$ and $\{V(u)\}$ be stocchastic processes on R^p and suppose that

$$
V_n(u)\Longrightarrow_L V(u)\ on\ C[a,b]^p,
$$

for every $a<b\in R^p$. Let ξ_n minimize $V_n(u)$, and ξ minimize $V(u)$. If $V_n(u)$ is convex for each n and ξ is unique with probability 1, then

$$
\xi_n\to_L \xi\ on\ R^p.
$$

6.2 Self-weighted Quantile Estimation

The quantile estimation was first proposed by Koenker and Bassett (1978). It includes the LAE as a special case and has been extensively investigated in the literature, see for examples, Ruppert and Carroll (1980), Bassett and Koenker (1982), Koenker and Bassett (1982), Koenker and D'Orey (1987), and Portnoy and Koenker (1989). In the time series setup, it was studied by Koenker and Zhao (1996) and Mukherjee (1999), and Ling and McAleer (2003).

The self-weighted quantile estimator (SQE) of $\lambda(\tau) \equiv \lambda + (F^{-1}(\tau), 0, \cdots, 0)'$ as

$$\hat{\lambda}(\tau) = \mathrm{argmin}_{\lambda \in R^{p+1}} \sum_{t=1}^{n} \frac{1}{w_t} \rho_\tau (y_t - X'_{t-1}\lambda),$$

where $\rho_\tau(u) = u[\tau - I(u<0)]$, $u \in R$, $\tau \in (0,1)$, $X_t = (1, y_t, \cdots, y_{t-p+1})'$, and $w_t = (1 + \sum_{i=1}^{p} y_{t-i}^2)^{3/2}$. LADe is SQE when $\tau = 1/2$. Define

$$T_n(s,\tau) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{X_{t-1}}{w_t} [I(\varepsilon_t \le F^{-1}(\tau) + s' X_{t-1}/\sqrt{n}) - \tau], \tag{6.13}$$

where $s \in R^{p+1}$. $T_n(s,\tau)$ serves as the score function in the MLE. We can see that the corresponding information-type matrix is bounded.

Theorem 6.3. *If Assumption 2.1 is satisfied and $F(x)$ has a positive density $f(x)$ on $\{x : 0 < F(x) < 1\}$ with $\sup_{x\in R} f(x) < \infty$ and $\sup_{x \in R} f'(x) < \infty$, then*

$$\begin{aligned} \sqrt{n}[\hat{\lambda}_n(\tau) - \lambda(\tau)] &= -\frac{\Sigma^{-1}}{q(\tau)} T_n(0,\tau) + o_p(1) \\ &\longrightarrow_{\mathscr{L}} N\Big(0, \frac{\tau(1-\tau)}{q^2(\tau)} \Sigma^{-1} \Omega \Sigma^{-1}\Big), \end{aligned}$$

where $q(\tau) = f(F^{-1}(\tau))$, $\Sigma = E(X_{t-1}X'_{t-1}/w_t)$ and $\Omega = E(X_{t-1}X'_{t-1}/w_t^2)$.

In what follows, we denote Euclidean norm by $\|\cdot\|$ and a bounded random sequence in probability by $O_p(1)$, and let $\mathscr{F}_t = \sigma\{\varepsilon_t, \varepsilon_{t-1}, \cdots\}$.

Lemma 4.1 *If the assumptions of Theorem 2.1 hold, then it follows that*

$$\begin{aligned} &(i)\ \ \|T_n(\sqrt{n}[\hat{\lambda}_n(\tau) - \lambda(\tau)], \tau)\| = O_p(\frac{1}{\sqrt{n}}), \\ &(ii)\ \ T_n(0,\tau) \longrightarrow_{\mathscr{L}} N(0, \tau(1-\tau)\Omega). \end{aligned}$$

Proof. Since F is continuous, for each t, there exists no constant vector c with $c'c \neq 0$ such that $c'X_t = 0$ almost surely (a.s.). Furthermore, note that $\max_{1 \le t \le n} \|X_{t-1}\|/w_t \le 1$ a.s.. Exactly following the arguments as for Lemma 4.2 in Ruppert and Carroll (1980), we can show that

$$\|T_n(\sqrt{n}[\hat{\lambda}_n(\tau)-\lambda(\tau)],\tau)\| \leq 2(p+1)\max_{1\leq t\leq n}\frac{\|X_{t-1}\|}{\sqrt{n}w_t} = O_p(\frac{1}{\sqrt{n}}),$$

i.e. (i) holds. Since $a_t \equiv (X_{t-1}/w_t)[I(\varepsilon_t \leq F^{-1}(\tau)) - \tau]$ is strictly stationary and ergodic with $E(a_t|\mathscr{F}_{t-1}) = 0$ and $E(a_t a_t') = \tau(1-\tau)\Omega$. (ii) holds by the central limit theorem. This completes the proof. □

Lemma 4.2 *Under the assumptions of Theorem 2.1, for any constant $M \geq 0$,*

$$\sup_{\|s\|\leq M}\|T_n(s,\tau)-T_n(0,\tau)-q(\tau)\Sigma s\| = o_p(1).$$

Proof. Let $g_t(s,u) = (s'X_{t-1} + u|s'X_{t-1}|)/\sqrt{n}$ with $u \in [0,M]$. We define

$$Z_t(s,u) = I[\varepsilon_t \leq x + g_t(s,u)] - I(\varepsilon_t \leq x) - F[x+g_t(s,u)] + F(x),$$

where $x = F^{-1}(\tau)$. By the monotonicity of F and indicator function, it follows that

$$\begin{aligned}|Z_t(s,u)| &\leq I(x-\frac{2M}{\sqrt{n}}\|X_{t-1}\| \leq \varepsilon_t \leq x+\frac{2M}{\sqrt{n}}\|X_{t-1}\|)\\ &\quad +F(x+\frac{2M}{\sqrt{n}}\|X_{t-1}\|) - F(x-\frac{2M}{\sqrt{n}}\|X_{t-1}\|).\end{aligned}$$

Thus, we have

$$\begin{aligned}E[Z_t^2(s,u)|\mathscr{F}_{t-1}] &\leq 4[F(x+\frac{2M}{\sqrt{n}}\|X_{t-1}\|) - F(x-\frac{2M}{\sqrt{n}}\|X_{t-1}\|)]\\ &\leq \frac{16M\|X_{t-1}\|}{\sqrt{n}} f(x+n^{-1/2}\xi_{t-1}^*) \leq \frac{C\|X_{t-1}\|}{\sqrt{n}},\end{aligned}$$

where $-2M\|X_{t-1}\|/\sqrt{n} \leq \xi_{t-1}^* \leq 2M\|X_{t-1}\|/\sqrt{n}$ and C is a constant. Let $\xi_{it}^+ = \max\{y_{t-i},0\}/w_t$ and $\xi_{it}^- = \max\{-y_{t-i},0\}/w_t$. Denote

$$T_{in}^{\pm}(s,\tau,u) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\xi_{it}^{\pm}Z_t(s,u),$$

where $i = 1,\cdots,p+1$. For any $\varepsilon > 0$, since $\xi_{it}^{\pm}Z_t(s,u)$ is a martingale difference in terms of $\mathscr{F}_t$, by Markov's inequality, we have

$$P(\left|T_{in}^{\pm}(s,\tau,u)\right| \geq \varepsilon) \leq \frac{1}{n\varepsilon^2}\sum_{t=1}^{n}E[\xi_{it}^{\pm}Z_t(s,u)]^2 \leq \frac{C^2}{n^2\varepsilon^2}\sum_{t=1}^{n}E\Big(\frac{\|X_{t-1}\|^4}{w_t^2}\Big) \to 0 \quad (6.14)$$

as $n \to \infty$, for each $s \in R^p$ and $u \in R$, where $i = 1,\cdots,p+1$.

Denote $D_M = [-M,M]^{p+1}$. Since D_M is a bounded and closed region of R^{p+1}, for every $\delta > 0$, there is a finite number of open subsets $\Delta_i(\delta)$, $i = 1,\cdots,m$, each with diameter δ, such that $\bigcup_{i=1}^m \Delta_i(\delta) \supset D_M$ and $\tilde{\Delta}_i \equiv \Delta_i(\delta)\cap D_M$ is not empty . Let s_r be any fixed point in $\tilde{\Delta}_r$. Then for any $u \in \tilde{\Delta}_r$, we know that

$$|g_t(s,u) - g_t(s_r,u)| \le \|s - s_r\| \cdot \|X_{t-1}\|/\sqrt{n} \le \delta\|X_{t-1}\|/\sqrt{n},$$

that is, $g_t(s_r, u-\delta) \le g_t(s,u) \le g_t(s_r, u+\delta)$. By the monotonicity of the indicator function, we obtain that

$$T_{in}^{\pm}(s,\tau,0) \le T_{in}^{\pm}(s_r,\tau,\delta) + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\xi_{it}^{\pm}[F(x+g_t(s_r,\delta)) - F(x+g_t(s,0))]$$

and a reverse inequality holds as δ is replaced by $-\delta$.

By the assumption given and the mean value theorem, it follows that

$$\begin{aligned}
&\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\xi_{it}^{\pm}[F(x+g_t(s_r,\pm\delta)) - F(x+g_t(s,0))]\Big| \\
&\le \sup_x |f(x)| \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\xi_{it}^{\pm}|g_t(s_r,\pm\delta) - g_t(s,0)| \\
&\le \frac{2\delta \sup_x |f(x)|}{n}\sum_{t=1}^{n}\frac{\|X_{t-1}\|^2}{w_t} = \delta O_p(1),
\end{aligned} \tag{6.15}$$

where $O_p(1)$ uniformly holds for all $s \in \tilde{\Delta}_r$ and all $r = 1, \cdots, m$. Given any small $\varepsilon > 0$ and $\eta > 0$, by (4.2), there exists a $\delta_\varepsilon > 0$ such that

$$P\{\frac{1}{\sqrt{n}}\sup_r \sup_{s\in\tilde{\Delta}_r}\Big|\sum_{t=1}^{n}[F(x+g_t(s_r,\pm\delta_\varepsilon)) - F(x+g_t(s,0))]\Big| \ge \frac{\varepsilon}{3}\} \le \eta\,. \tag{6.16}$$

For the $\pm\delta_\varepsilon$, by (4.1), it follows that

$$P\{\max_r |T_{in}^{\pm}(s_r,\tau,\pm\delta_\varepsilon)| \ge \frac{\varepsilon}{3}\} \le r\max_r P\{|T_{in}(s_r,\tau,\pm\delta_\varepsilon)| \ge \frac{\varepsilon}{3}\} \le \eta, \tag{6.17}$$

as n is large enough. By (4.3)-(4.4), we know that

$$\begin{aligned}
&P\{\sup_{s\in D_M}|T_{in}^{\pm}(s,\tau,0)| \ge \varepsilon\} \\
&\le P\{\max_r |T_{in}^{\pm}(s_r,\tau,\delta_\varepsilon)| \ge \frac{\varepsilon}{3}\} + P\{\max_r |T_{in}^{\pm}(s_r,\tau,-\delta_\varepsilon)| \ge \frac{\varepsilon}{3}\} \\
&\quad + P\{\frac{1}{\sqrt{n}}\sup_r \sup_{s\in\tilde{\Delta}_r}|\sum_{t=1}^{n}[F(x+g_t(s_r,\pm\delta_\varepsilon)) - F(x+g_t(s,0))]| \ge \frac{\varepsilon}{3}\} \\
&\le 3\eta.
\end{aligned} \tag{6.18}$$

By (4.5), we can show that

$$\sup_{\|s\|\le M}\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{X_{t-1}}{w_t}Z_t(s,0)\Big| = o_p(1). \tag{6.19}$$

Furthermore, by Taylor's expansion and the assumption given, we have

$$\sup_{s \in D_M} \Big| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{X_{t-1}}{w_t} \Big[F(x + \frac{1}{\sqrt{n}} s' X_{t-1}) - F(x) - \frac{f(x)}{\sqrt{n}} s' X_{t-1} \Big] \Big|$$
$$\leq \sup_{s \in D_M} f'(\xi_t^*) \frac{M^2}{n\sqrt{n}} \sum_{t=1}^{n} \frac{\|X_{t-1}\|^3}{w_t} = O_p(\frac{1}{\sqrt{n}}). \tag{6.20}$$

By the ergodic theorem, $\sum_{t=1}^{n}(X_{t-1}X_{t-1}'/w_t)/n = \Sigma + o_p(1)$. Furthermore, by (4.6) and (4.7), we can claim that the conclusion holds. This completes the proof. □

Proof of Theorem 2.1. Denote $\Upsilon_n(\tau) = \sqrt{n}[\hat{\lambda}_n(\tau) - \lambda(\tau)]$. For any $\varepsilon, \eta > 0$, by Lemma 4.1 (i), there exists an integer $n_1 > 0$ such that, when $n > n_1$,

$$P\Big\{ \|T_n(\Upsilon_n(\tau), \tau)\| > \eta \Big\} < \varepsilon.$$

Thus, for a positive constant M, when $n > n_1$,

$$P\{\|\Upsilon_n(\tau)\| \geq M\}$$
$$\leq P\{\|\Upsilon_n(\tau)\| \geq M, \|T_n(\Upsilon_n(\tau), \tau)\| \leq \eta\} + P\{\|T_n(\Upsilon_n(\tau), \tau)\| \geq \eta\}$$
$$\leq P\Big\{ \inf_{\|s_1\| \geq M} \|T_n(s_1, \tau)\| \leq \eta \Big\} + \varepsilon. \tag{6.21}$$

Note that $s_1' T_n(\nu s_1, \tau)$ is a non-decreasing function of ν for any $\tau \in (0,1)$ and $s_1 \in R^{p+1}$. Writing s_1 as $s_1 = \nu s$ with $\nu \geq 1$ and $\|s\| = M$ for any $\|s_1\| \geq M$, by the Cauchy-Schwarz inequality, we have

$$\inf_{\|s\|=M} |s' T_n(s, \tau)| \leq \inf_{\|s\|=M, \nu \geq 1} |s' T_n(\nu s, \tau)| \leq M \inf_{\|s_1\| \geq M} \|T_n(s_1, \tau)\|.$$

Thus, by (4.8),

$$P\{\|\Upsilon_n(\tau)\| \geq M\} \leq P\Big\{ \inf_{\|s\|=M} |s' T_n(s, \tau)| \leq \eta M \Big\} + \varepsilon. \tag{6.22}$$

Denote $R_n(\tau) = \sup_{\|s\|=M} |s'[T_n(s,\tau) - T_n(0,\tau)] - s'\Sigma s q(\tau)|$ and let c_0 be the minimum eigenvalue of Σ. Since

$$|s' T_n(s, \tau)| \geq \inf_{\|s\|=M} [s' \Sigma s q(\tau)] - R_n(\tau) - \sup_{\|s\|=M} |s' T_n(0, \tau)|,$$

by (4.9), it follows that

$$P\Big\{ \|\Upsilon_n(\tau)\| \geq M \Big\}$$
$$\leq P\Big\{ R_n(\tau) \geq \inf_{\|s\|=M} [s' \Sigma s q(\tau)] - \sup_{\|s\|=M} |s' T_n(0, \tau)| - \eta M \Big\} + \varepsilon$$
$$\leq P\Big\{ R_n(\tau) \geq - \sup_{\|s\|=M} |s' T_n(0, \tau)| - \eta M + c_0 M^2 q(\tau)] \Big\} + \varepsilon. \tag{6.23}$$

By Lemma 4.1 (ii), there exists a large constant M_1 and an integer n_2 such that, when $n > n_2$,

$$P(\sup_{\|s\|=M} |s'T_n(0,\tau)| > MM_1) \leq P(\|T_n(0,\tau)\| > M_1) < \varepsilon. \tag{6.24}$$

Thus, by (4.11), when $n > \max\{n_2, n_3\}$,

$$\begin{aligned}
&P\Big\{R_n(\tau) \geq -\sup_{\|s\|=M} |s'T_n(0,\tau)| - \eta M + c_0 M^2 q(\tau)\Big\} \\
&\leq P\Big\{R_n(\tau) \geq -\sup_{\|s\|=M} |s'T_n(0,\tau)| - \eta M + c_0 M^2 q(\tau), \\
&\quad \sup_{\|s\|=M} |s'T_n(0,\tau)| \leq MM_1\Big\} + P\Big(\sup_{\|s\|=M} |s'T_n(0,\tau)| > MM_1\Big) \\
&\leq P\Big\{R_n(\tau) \geq c_0 M^2 q(\tau) - MM_1 - \eta M\Big\} + \varepsilon.
\end{aligned} \tag{6.25}$$

We may choose M large enough such that $c = c_0 M q(\tau) - M_1 - \eta > 0$. For the constant c, by Lemma 4.2, there exists an integer n_3 such that, when $n > n_3$,

$$P\Big\{R_n(\tau) \geq Mc\Big\} \leq P\Big\{\sup_{\|s\|=M} \|[T_n(s,\tau) - T_n(0,\tau)] - q(\tau)\Sigma s]\| \geq c\Big\} < \varepsilon. \tag{6.26}$$

Thus, by (4.10) and (4.12)-(4.13), when $n > \max\{n_1, n_2, n_3\}$, $P\{\| \Upsilon_n(\tau)\| \geq M\} < 4\varepsilon$. Finally, by Lemma 4.1(i) and 4.2, we can show that

$$\sqrt{n}[\hat{\lambda}_n(\tau) - \lambda(\tau)] = -\frac{\Sigma^{-1}}{q(\tau)} T_n(0,\tau) + o_p(1).$$

Furthermore, by Lemma 4.1(ii) and the equation above, the conclusion holds. This completes the proof. □

6.3 Self-weighted LAD for ARMA Models

Assume that the random sample $\{Y_1, \cdots, Y_n\}$ is from model (2.6). With the initial values $\tilde{Y}_0 \equiv \{Y_0, U_{-1}, \cdots\}$, We investigate the following objective function:

$$L_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} w_t |\varepsilon_t(\theta)|, \tag{6.27}$$

where $\varepsilon_t(\theta)$ is defined as (3.4), $w_t = w(y_{t-1}, y_{t-2}, ...) > 0$ and w is a measurable and bounded function on R^{Z_0} with $Z_0 = \{0, 1, 2...\}$, and it satisfies the following condition:

Assumption 6.4 *$E[(w_t + w_t^2)\xi_{\rho t-1}^2] < \infty$ for any $\rho \in (0,1)$, where $\xi_{\rho t-1} = 1 + \sum_{i=0}^{\infty} \rho^i |y_{t-1-i}|$ a.s.*

The minimizer, $\hat{\theta}_n$, of $L_n(\theta)$ on Θ is called the global self-weighted LAD (SLAD) estimator of θ, i.e.,

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

When $E\varepsilon_t^2 < \infty$, we take $w_t = 1$ and in this case, $\hat{\theta}_n$ is the LAD estimator for the finite variance ARMA(p,q) model (**??**). The weights $w_t \neq 1$ are used only when $E\varepsilon_t^2 = \infty$. Obviously, there are a lot of weight functions satisfy Assumption 6.4. Let $2\iota_0$ be the tail index of data $\{y_1, \cdots, y_n\}$. When $\iota_0 > \frac{1}{2}$ (i.e., $E|\varepsilon_t| < \infty$), we can choose the weight function as

$$w_t = \left(\max\left\{ 1, C^{-1} \sum_{k=1}^{\infty} \frac{1}{k^9} |y_{t-k}| I\{|y_{t-k}| > C\} \right\} \right)^{-4},$$

where $C > 0$ is a constant. In practice, it works well when we select C as the 90% quantile of data $\{y_1, \cdots, y_n\}$. When $\iota_0 \in (0, \frac{1}{2}]$ and $q > 0$, the weight function need to be modified as follows:

$$w_t = \left(\max\left\{ 1, C^{-1} \sum_{k=1}^{\infty} \frac{1}{k^{1+8/\iota_1}} |y_{t-k}| I\{|y_{t-k}| > C\} \right\} \right)^{-4}$$

for any $\iota_1 < \iota_0$, see Ling (2007). For the choice of w_t and its motivation, we refer to Ling (2005) and Pan et al. (2007). For instance, the weight functions defined in Pan et al. (2007) satisfy Assumption 3.4 as well. Without this weight, one cannot obtain the asymptotic normality of the estimated parameters when $E\varepsilon_t^2 = \infty$, see Mikosch et al. (1995). We assume that Y_0 is from model (**??**) and hence $\varepsilon_t(\theta_0) = \varepsilon_t$ in this case. For the case that Y_0 is unobservable, we will consider it at the end of this section. The following theorem gives the consistency of $\hat{\theta}_n$ and its proof is given in the Appendix.

Theorem 6.4. *If Assumptions 3.4-6.4 hold and ε_t has a zero median, then*

$$\hat{\theta}_n \to \theta_0 \quad \textit{almost surely (a.s.)} \quad \textit{as } n \to \infty.$$

To study the rate of convergence of $\hat{\theta}_n$, we first reparametrize the objective function as

$$L_n(u) \equiv nH_n(\theta_0 + u) - nH_n(\theta_0),$$

where $u \in \Lambda \equiv \{u : u + \theta_0 \in \Theta\}$. Let $\hat{u}_n = \hat{\theta}_n - \theta_0$. Then, $\hat{u}_n$ is the minimizer of $L_n(u)$ in Λ. Furthermore, by (6.27), we have

$$L_n(u) = \sum_{t=1}^{n} w_t \left[|\varepsilon_t(\theta_0 + u)| - |\varepsilon_t(\theta_0)| \right]. \tag{6.28}$$

Using the identity (**??**), it follows that

$$\begin{aligned} |\varepsilon_t(\theta_0 + u)| - |\varepsilon_t(\theta_0)| = q_t(u)[I(\varepsilon_t > 0) - I(\varepsilon_t < 0)] \\ + 2\int_0^{-q_t(u)} X_t(s)ds, \end{aligned} \tag{6.29}$$

where $X_t(s) = I(\varepsilon_t < s) - I(\varepsilon_t < 0)$,

$$q_t(u) = u' \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta} + \frac{1}{2} u' \frac{\partial^2 \varepsilon_t(\xi^*)}{\partial \theta \partial \theta'} u,$$

and ξ^* lies between θ_0 and $\theta_0 + u$. Using equation (6.29), we can decompose (6.28) as follows:

$$L_n(u) = u' T_n + \sum_{t=1}^{n} E[\xi_t(u) | \mathscr{F}_{t-1}] + \alpha_n(u) + \beta_n(u), \tag{6.30}$$

where

$$\begin{aligned} T_n &= \sum_{t=1}^{n} w_t \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta} [I(\varepsilon_t > 0) - I(\varepsilon_t < 0)], \\ \xi_t(u) &= 2w_t \int_0^{-u' \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta}} X_t(s)ds, \\ \alpha_n(u) &= \sum_{t=1}^{n} \{ \xi_t(u) - E[\xi_t(u) | \mathscr{F}_{t-1}] \}, \\ \beta_n(u) &= \frac{u'}{2} \sum_{t=1}^{n} w_t \frac{\partial^2 \varepsilon_t(\xi^*)}{\partial \theta \partial \theta'} [I(\varepsilon_t > 0) - I(\varepsilon_t < 0)] u \\ &\quad + 2 \sum_{t=1}^{n} w_t \int_{-u' \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta}}^{-q_t(u)} X_t(s)ds. \end{aligned}$$

We next introduce one more assumption and three lemmas. The first lemma follows directly from the central limit theorem. The second and third lemmas give the expansions of $\alpha_n(u)$, $\beta_n(u)$ and $\sum_{t=1}^{n} E[\xi_t(u) | \mathscr{F}_{t-1}]$ and their proofs are given in Section 3. The key technical argument lies in applying the bracketing method from Pollard (1985) and is used in Lemma 6.2.

Assumption 6.5 ε_t *has a zero median with a continuous density function* $g(x)$ *satisfying* $g(0) > 0$ *and* $\sup_{x \in R} g(x) < \infty$.

Assumption 6.5 is weaker than the condition $\sup_x |g'(x)| < \infty$ in Pan et al. (2007). When ε_t is stable with support R and median zero, Assumption 6.5 is satisfied, see Rachev (2003, p.110).

Lemma 6.1. *If Assumptions 3.4-6.5 hold, then*

$$\frac{1}{\sqrt{n}} T_n \to_d N(0, \Omega_0),$$

as $n \to \infty$, where $\Omega_0 = E[w_t^2(\partial \varepsilon_t(\theta_0)/\partial \theta)(\partial \varepsilon_t(\theta_0)/\partial \theta')]$.

Lemma 6.2. *If Assumptions 3.4-6.5 hold, then for any sequence of random variables $\{u_n\}$ such that $u_n = o_p(1)$, it follows that*

$$\alpha_n(u_n) = o_p(\sqrt{n}\|u_n\| + n\|u_n\|^2).$$

Lemma 6.3. *If Assumptions 3.4-6.5 hold, then for any sequence of random variables $\{u_n\}$ such that $u_n = o_p(1)$, it follows that*

$$(i)\ \sum_{t=1}^{n} E\left[\xi_t(u)|\mathscr{F}_{t-1}\right]\Big|_{u=u_n} = (\sqrt{n}u_n)'\,[g(0)\Sigma_0]\,(\sqrt{n}u_n) + o_p(n\|u_n\|^2),$$
$$(ii)\ \beta_n(u_n) = o_p(n\|u_n\|^2),$$

where $\Sigma_0 = E[w_t(\partial \varepsilon_t(\theta_0)/\partial \theta)(\partial \varepsilon_t(\theta_0)/\partial \theta')]$.

We now can state our main results as follows.

Theorem 6.5. *If Assumptions 3.4-6.5 hold, then*

$$(i)\ \sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1),$$
$$(ii)\ \sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N\left(0, [2g(0)]^{-2}\Sigma_0^{-1}\Omega_0\Sigma_0^{-1}\right) \text{ as } n \to \infty.$$

Proof. (i). First, we have $\hat{u}_n = o_p(1)$ by Theorem 6.4. Furthermore, by (6.30) and Lemmamas 6.2-6.3, we have

$$L_n(\hat{u}_n) = \hat{u}_n' T_n + (\sqrt{n}\hat{u}_n)'\,[g(0)\Sigma_0]\,(\sqrt{n}\hat{u}_n) + o_p(\sqrt{n}\|\hat{u}_n\| + n\|\hat{u}_n\|^2). \tag{6.31}$$

Let $\lambda_{min} > 0$ be the minimum eigenvalue of $g(0)\Sigma_0$. Then

$$L_n(\hat{u}_n) \geq -\|\sqrt{n}\hat{u}_n\|\left[\left\|\frac{1}{\sqrt{n}}T_n\right\| + o_p(1)\right] + n\|\hat{u}_n\|^2\left[\lambda_{min} + o_p(1)\right].$$

Note that $L_n(\hat{u}_n) \leq 0$. By the previous equation, it follows that

$$\sqrt{n}\|\hat{u}_n\| \leq [\lambda_{min} - 2\delta + o_p(1)]^{-1}\left[\left\|\frac{1}{\sqrt{n}}T_n\right\| + o_p(1)\right] = O_p(1), \tag{6.32}$$

where the last step holds by Lemmama 6.1. Thus, (i) holds.

(ii). Let $u_n^* = -[2g(0)\Sigma_0]^{-1}T_n/n$. Then, by Lemmama 6.1, we have

$$\sqrt{n}u_n^* \to_d N\left(0, [2g(0)]^{-2}\Sigma_0^{-1}\Omega_0\Sigma_0^{-1}\right) \text{ as } n \to \infty.$$

Hence it is sufficient to show that $\sqrt{n}\hat{u}_n - \sqrt{n}u_n^* = o_p(1)$. By (6.31) and (6.32), we have

$$\begin{aligned} L_n(\hat{u}_n) &= (\sqrt{n}\hat{u}_n)'\frac{1}{\sqrt{n}}T_n + (\sqrt{n}\hat{u}_n)'[g(0)\Sigma_0](\sqrt{n}\hat{u}_n) + o_p(1) \\ &= (\sqrt{n}\hat{u}_n)'[g(0)\Sigma_0](\sqrt{n}\hat{u}_n) - 2(\sqrt{n}\hat{u}_n)'[g(0)\Sigma_0](\sqrt{n}u_n^*) + o_p(1). \end{aligned}$$

Note that (6.31) still holds when $\hat{u}_n$ is replaced by u_n^*. Thus

$$\begin{aligned} L_n(u_n^*) &= (\sqrt{n}u_n^*)'\frac{1}{\sqrt{n}}T_n + (\sqrt{n}u_n^*)'[g(0)\Sigma_0](\sqrt{n}u_n^*) + o_p(1) \\ &= -(\sqrt{n}u_n^*)'[g(0)\Sigma_0](\sqrt{n}u_n^*) + o_p(1). \end{aligned}$$

From the previous two equations it follows that

$$\begin{aligned} L_n(\hat{u}_n) - L_n(u_n^*) &= (\sqrt{n}\hat{u}_n - \sqrt{n}u_n^*)'[g(0)\Sigma_0](\sqrt{n}\hat{u}_n - \sqrt{n}u_n^*) + o_p(1) \\ &\geq \lambda_{min}\|\sqrt{n}\hat{u}_n - \sqrt{n}u_n^*\|^2 + o_p(1). \end{aligned} \tag{6.33}$$

Since $L_n(\hat{u}_n) - L_n(u_n^*) = n[H_n(\hat{u}_n + \theta_0) - H_n(u_n^* + \theta_0)] \leq 0$ a.s., by (6.33), we have $\|\sqrt{n}\hat{u}_n - \sqrt{n}u_n^*\| = o_p(1)$. This completes the proof.

6.4 Simulation Study

To examine the performance of the asymptotic results in finite samples, we consider the AR(1) model,

$$Y_t = \phi_0 + \phi Y_{t-1} + \varepsilon_t.$$

We use the optimal bandwidth b_n given in Silverman (1986, p.40) which is automatically searched from the data. Table 6.1 summarizes the empirical means, empirical standard deviations (SD) and asymptotic standard deviations (AD) of the SLADs of (ϕ_0, ϕ). The ADs are calculated using the estimated covariances in (2.2). Table 6.1 shows that all the biases are very small and all the SDs and ADs are very close, particularly, when $n = 400$. As n is increased from 200 to 400, all the SDs and ADs become smaller.

To give an overall view on the approximation of the limiting distribution to the finite sample distribution, we simulate 27000 replications for the case with $\phi = 0.5$, $\eta_t \sim t_2$ and $n = 400$. Denote $N_{SLADn} = \sqrt{n}[\hat{\phi}_n(0.5) - 05]/\hat{\sigma}_{SLAD}$, where $\hat{\sigma}_{SLAD}$ is the SDs of the SLAD of ϕ. Figure 1 shows the density curves of N_{SLADn} and $N(0,1)$. The density curve of N_{SLADn} is approximated by $f(x_i) \approx \sum_{i=1}^{27000} I(x_{i-1} \leq N_{SLADn} \leq x_i)/(27000b)$ with $x_0 = -6.235$, $x_i = x_{i-1} + b$ and $b = 0.215$. From this figure,

Table 6.1 Means and Standard Deviations of SLAD for AR Models with $\phi_0 = 0$ (1000 replications)

ϕ		n=200 $\hat{\phi}_0$	n=200 $\hat{\phi}$	n=400 $\hat{\phi}_0$	n=400 $\hat{\phi}$	n=200 $\hat{\phi}_0$	n=200 $\hat{\phi}$	n=400 $\hat{\phi}_0$	n=400 $\hat{\phi}$
		$\varepsilon_t \sim$ Cauchy				$\varepsilon_t \sim t_2$			
-0.5	Mean	.002	$-.505$	.001	$-.503$	.003	$-.495$	$-.002$	$-.494$
	SD	.134	.103	.098	.071	.130	.107	.093	.071
	AD	.139	.101	.098	.071	.134	.104	.095	.074
0.5	Mean	-.007	.491	-.008	.495	-.004	.487	-.003	.496
	SD	.136	.107	.094	.073	.135	.105	.092	.075
	AD	.139	.102	.098	.071	.134	.104	.094	.073
0.8	Mean	-.014	.787	-.014	.794	-.002	.779	-.001	.792
	SD	.155	.092	.116	.063	.164	.093	.110	.065
	AD	.165	.085	.117	.060	.163	.088	.115	.062

Table 6.2 Sizes and Powers of Wald-test for Null Hypothesis H_0: $(\phi_0, \phi) = (0, 0.5)$ at Significance Level 5% in AR Models (1000 replications)

$\varepsilon_t \sim$	n=200 Cauchy	n=200 t_2	n=400 Cauchy	n=400 t_2
ϕ_0 ϕ				
-.1 .3	.453	.437	.742	.732
.0 .3	.416	.371	.698	.676
-.1 .4	.196	.189	.321	.328
.0 .4	.152	.152	.232	.212
.0 .5	.066	.066	.060	.057
.0 .6	.150	.156	.228	.209
.1 .6	.199	.214	.350	.339
.0 .7	.416	.407	.699	.661
.1 .7	.472	.465	.772	.742

we can see that the density curve of N_{SLADn} is very close to that of $N(0,1)$. This is consistent with our theoretical results. These simulation results indicate that the SLAD performs very well in the finite samples.

We now investigate the size and power of the statistic W_n. Again, the sample sizes are $n = 200$ and 400 and the number of replications is 1000. Cauchy and t_2 distributions are used. The null hypothesis is H_0: $(\phi_0, \phi) = (0, 0.5)$ and the significance level is 5%. Table 6.2 summarizes the sizes and powers of W_n. From this table, we can see that the sizes are a little large, but they are still acceptable. In particular, when $n = 400$, the sizes are getting close to the nominal significance level. The powers are increased when n becomes large or when the distance between the alternative and the null H_0 becomes large. These simulation results indicate that the Wald test works well in the finite samples and should be useful in practice.

6.5 Proofs of Lemmas 6.2-6.3

In this section, we give the proofs of Lemmamas 6.2 and 6.3. In the rest of this paper, C denotes a universal constant.

PROOF OF LEMMAMA 6.2. A direct calculation gives

$$\xi_t(u) = -2u' w_t \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta} M_t(u),$$

where $M_t(u) = \int_0^1 X_t\left(-u' \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta} s\right) ds$. Thus, we have

$$|\alpha_n(u)| \leq 2\|u\| \sum_{j=1}^{m} \left| w_t \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta_j} \sum_{t=1}^{n} \{M_t(u) - E[M_t(u)|\mathscr{F}_{t-1}]\} \right|.$$

It is sufficient to show that

$$\left| w_t \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta_j} \sum_{t=1}^{n} \{M_t(u_n) - E[M_t(u_n)|\mathscr{F}_{t-1}]\} \right| = o_p(\sqrt{n} + n\|u_n\|) \quad (6.34)$$

for each $1 \leq j \leq m$. Denote $m_t = w_t \partial \varepsilon_t(\theta_0)/\partial \theta_j$ and $f_t(u) = m_t M_t(u)$. We define

$$D_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{f_t(u) - E[f_t(u)|\mathscr{F}_{t-1}]\}.$$

In order to prove (6.34), we only need to show that, for any $\eta \in (0,1)$,

$$\sup_{\|u\| \leq \eta} \frac{|D_n(u)|}{1 + \sqrt{n}\|u\|} = o_p(1). \quad (6.35)$$

Note that $m_t = \max\{m_t, 0\} - \max\{-m_t, 0\}$. To make it simple, we only prove the case when $m_t \geq 0$.

We adopt the method in Lemmama 4 of Pollard (1985). Let $\mathfrak{F} = \{f_t(u) : \|u\| \leq \eta\}$ be a collection of functions indexed by u. We first verify that $\mathfrak{F}$ satisfies the bracketing condition in Pollard (1985, p.304). By $B_r(\zeta)$ denote the open ball around ζ with radius $r > 0$. For any fixed $\varepsilon > 0$ and $0 < \delta \leq \eta$, there is a sequence of open balls $\{B_{\varepsilon\delta/C_1}(u_i)\}_{i=1}^{K_\varepsilon}$ to cover $B_\delta(0)$, where K_ε is an integer less than $c_0\varepsilon^{-m}$ and c_0 is a constant not depending on ε and δ, see Huber (1967, p.227). Here, C_1 is a constant to be selected later. Moreover, we can choose $U_i(\delta) \subseteq B_{\varepsilon\delta/C_1}(u_i)$ such that $\{U_i(\delta)\}_{i=1}^{K_\varepsilon}$ is a partition of $B_\delta(0)$.

For each $u \in U_i(\delta)$, we define the bracketing functions as follows:

$$f_t^{\pm}(u) = m_t \int_0^1 X_t\left(-u' \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta} s \pm \frac{\varepsilon\delta}{C_1} \left\| \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta} \right\| \right) ds.$$

Since the indicator function is nondecreasing and $m_t \geq 0$, we can see that, for any $u \in U_i(\delta)$,

$$f_{t,i}^- \leq f_t(u) \leq f_{t,i}^+,$$

where $f_{t,i}^\pm \triangleq f_t^\pm(u_i)$. Note that $\sup_x g(x) < \infty$. It is straightforward to see that

$$E\left[f_{t,i}^+ - f_{t,i}^- | \mathscr{F}_{t-1}\right] \leq \frac{2\varepsilon\delta}{C_1} \sup_x g(x) w_t \left\| \frac{\partial \varepsilon_t(\theta_0)}{\partial \theta} \right\|^2 \equiv \frac{\varepsilon\delta\Delta_t}{C_1}. \tag{6.36}$$

Setting $C_1 = E(\Delta_t)$, we have

$$E\left[f_{t,i}^+ - f_{t,i}^-\right] = E\left\{E\left[f_{t,i}^+ - f_{t,i}^- | \mathscr{F}_{t-1}\right]\right\} \leq \varepsilon\delta.$$

Thus, the family $\mathfrak{F}$ satisfies the bracketing condition.

Put $\delta_k = 2^{-k}\eta$. Define $B(k) \equiv B_{\delta_k}(0)$, and $A(k)$ to be the annulus $B(k) \setminus B(k+1)$. Fix $\varepsilon > 0$, by the bracketing condition, there exists a partition $\{U_i(\delta_k)\}_{i=1}^{K_\varepsilon}$ of $B(k)$.

We first consider the upper tail of $\frac{D_n(u)}{1+\sqrt{n}\|u\|}$. By (6.36), with $\delta = \delta_k$, we have that for each $u \in U_i(\delta_k)$

$$\begin{aligned} D_n(u) &\leq \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{f_{t,i}^+ - E\left[f_{t,i}^- | \mathscr{F}_{t-1}\right]\right\} \\ &= D_n^+(u_i) + \frac{1}{\sqrt{n}} \sum_{t=1}^n E\left[f_{t,i}^+ - f_{t,i}^- | \mathscr{F}_{t-1}\right] \\ &\leq D_n^+(u_i) + \sqrt{n}\varepsilon\delta_k \left[\frac{1}{nC_1} \sum_{t=1}^n \Delta_t\right], \end{aligned}$$

where

$$D_n^+(u_i) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \{f_{t,i}^+ - E[f_{t,i}^+ | \mathscr{F}_{t-1}]\}.$$

Denote the event

$$E_n = \left\{\frac{1}{nC_1} \sum_{t=1}^n \Delta_t < 2\right\}.$$

On E_n it follows that for $u \in U_i(\delta_k)$,

$$D_n(u) \leq D_n^+(u_i) + 2\sqrt{n}\varepsilon\delta_k. \tag{6.37}$$

On $A(k)$, the divisor $1 + \sqrt{n}\|u\| > \sqrt{n}\delta_{k+1} = \sqrt{n}\delta_k/2$. Thus, by (6.37) and Tchebychev's inequality, it follows that

$$
\begin{aligned}
& P\left(\left\{\sup_{u\in A(k)} \frac{D_n(u)}{1+\sqrt{n}\|u\|} > 6\varepsilon\right\}\cap E_n\right) \\
& \le P\left(\left\{\sup_{u\in A(k)} D_n(u) > 3\sqrt{n}\varepsilon\delta_k\right\}\cap E_n\right) \\
& \le P\left(\left\{\max_{1\le i\le K_\varepsilon}\sup_{u\in U_i(\delta_k)\cap A(k)} D_n(u) > 3\sqrt{n}\varepsilon\delta_k\right\}\cap E_n\right) \\
& \le P\left(\left\{\max_{1\le i\le K_\varepsilon} D_n^+(u_i) > \sqrt{n}\varepsilon\delta_k\right\}\cap E_n\right) \\
& \le K_\varepsilon \max_{1\le i\le K_\varepsilon} P\left(D_n^+(u_i) > \sqrt{n}\varepsilon\delta_k\right) \\
& \le K_\varepsilon \max_{1\le i\le K_\varepsilon} \frac{E[(D_n^+(u_i))^2]}{n\varepsilon^2\delta_k^2}. \qquad (6.38)
\end{aligned}
$$

Note that $\|u_i\| \le \delta_k$ and $m_t^2 \le w_t^2\|\partial\varepsilon_t(\theta_0)/\partial\theta_j\|^2 \le Cw_t^2\xi_{\rho t-1}^2$ for some $\rho\in(0,1)$ by Lemmama 6.4 (ii). By the conditional expectation property and the fact that $|X_t(s)|<1$, we have

$$
\begin{aligned}
E\left[\left(f_{t,i}^+\right)^2\right] &= E\left\{E\left[\left(f_{t,i}^+\right)^2|\mathscr{F}_{t-1}\right]\right\} \\
&\le E\left[m_t^2\int_0^1 E\left[\left|X_t\left(-u_i'\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}s+\frac{\varepsilon\delta_k}{C_1}\left\|\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\right\|\right)\right||\mathscr{F}_{t-1}\right]ds\right] \\
&\le CE\left[\sup_{|x|\le\delta_k C'\xi_{\rho t-1}}|G(x)-G(0)|w_t^2\xi_{\rho t-1}^2\right],
\end{aligned}
$$

where $C'=1+\varepsilon/C_1$. Use that $f_{t,i}^+-E[f_{t,i}^+|\mathscr{F}_{t-1}]$ is a martingale difference sequence and the previous inequality to see that

$$
\begin{aligned}
E[(D_n^+(u_i))^2] &= \frac{1}{n}\sum_{t=1}^n E\left\{f_{t,i}^+-E\left[f_{t,i}^+|\mathscr{F}_{t-1}\right]\right\}^2 \\
&\le \frac{1}{n}\sum_{t=1}^n E\left[\left(f_{t,i}^+\right)^2\right] \\
&\le \frac{C}{n}\sum_{t=1}^n E\left[\sup_{|x|\le\delta_k C'\xi_{\rho t-1}}|G(x)-G(0)|w_t^2\xi_{\rho t-1}^2\right], \\
&\equiv \pi_n(\delta_k). \qquad (6.39)
\end{aligned}
$$

Thus, by (6.38)-(6.39), we have

$$
P\left(\left\{\sup_{u\in A(k)}\frac{D_n(u)}{1+\sqrt{n}\|u\|}>6\varepsilon\right\}\cap E_n\right)\le K_\varepsilon\frac{\pi_n(\delta_k)}{n\varepsilon^2\delta_k^2}.
$$

By a similar argument, we can get the same bound for the lower tail. Thus, we can show that

$$P\left(\left\{\sup_{u\in A(k)}\frac{|D_n(u)|}{1+\sqrt{n}\|u\|}>6\varepsilon\right\}\cap E_n\right)\le 2K_\varepsilon\frac{\pi_n(\delta_k)}{n\varepsilon^2\delta_k^2}. \tag{6.40}$$

By Assumption 6.4 and the dominated convergence theorem, we have

$$E\left[\sup_{|x|\le\delta C'\xi_{\rho t-1}}|G(x)-G(0)|w_t^2\xi_{\rho t-1}^2\right]\to 0$$

as $\delta\to 0$ uniformly in t. Thus, $\pi_n(\delta_k)\to 0$ as $k\to\infty$. Choose k_ε such that

$$2\pi_n(\delta_k)K_\varepsilon/(\varepsilon\eta)^2<\varepsilon$$

for $k\ge k_\varepsilon$. Let k_n be an integer such that $n^{-1/2}\le 2^{-k_n}<2n^{-1/2}$. Split $\{u:\|u\|\le\eta\}$ into two sets $B(k_n+1)$ and $B(k_n+1)^c=\cup_{k=0}^{k_n}A(k)$. By (6.40), since $\pi_n(\delta_k)$ is bounded, we have

$$\begin{aligned}
&P\left(\sup_{u\in B(k_n+1)^c}\frac{|D_n(u)|}{1+\sqrt{n}\|u\|}>6\varepsilon\right)\\
&\le\sum_{k=0}^{k_n}P\left(\left\{\sup_{u\in A(k)}\frac{|D_n(u)|}{1+\sqrt{n}\|u\|}>6\varepsilon\right\}\cap E_n\right)+P(E_n^c)\\
&\le\frac{1}{n}\sum_{k=0}^{k_\varepsilon-1}\frac{CK_\varepsilon}{\varepsilon^2\eta^2}2^{2k+1}+\frac{\varepsilon}{n}\sum_{k=k_\varepsilon}^{k_n}2^{2k}+P(E_n^c)\\
&\le O\left(\frac{1}{n}\right)+4\varepsilon\frac{2^{2k_n}}{n}+P(E_n^c)\\
&\le O\left(\frac{1}{n}\right)+4\varepsilon+P(E_n^c).
\end{aligned}\tag{6.41}$$

Since $1+\sqrt{n}\|u\|>1$ and $\sqrt{n}\delta_{k_n+1}<1$, using a similar argument as for (6.38) together with (6.39), we have

$$\begin{aligned}
&P\left(\left\{\sup_{u\in B(k_n+1)}\frac{D_n(u)}{1+\sqrt{n}\|u\|}>3\varepsilon\right\}\cap E_n\right)\\
&\le P\left(\left\{\max_{1\le i\le K_\varepsilon}D_n^+(u_i)>\varepsilon\right\}\cap E_n\right)\\
&\le\frac{K_\varepsilon\pi_n(\delta_{k_n+1})}{\varepsilon^2}.
\end{aligned}$$

We can get the same bound for the lower tail. Thus, we have

$$
\begin{aligned}
&P\left(\sup_{u\in B(k_n+1)}\frac{|D_n(u)|}{1+\sqrt{n}\|u\|}>3\varepsilon\right)\\
&\quad= P\left(\left\{\sup_{u\in B(k_n+1)}\frac{|D_n(u)|}{1+\sqrt{n}\|u\|}>3\varepsilon\right\}\cap E_n\right)+P(E_n^c)\\
&\quad\le \frac{2K_\varepsilon\pi_n(\delta_{k_n+1})}{\varepsilon^2}+P(E_n^c). \qquad (6.42)
\end{aligned}
$$

Note that $\pi_n(\delta_{k_n+1})\to 0$ as $n\to\infty$. Furthermore, $P(E_n)\to 1$ by the ergodic theorem. Hence

$$P(E_n^c)\to 0 \quad \text{as } n\to\infty.$$

Finally, (6.35) follows from (6.41)-(6.42). This completes the proof. □

PROOF OF LEMMAMA 6.3. (i). By a direct calculation we have

$$
\begin{aligned}
\sum_{t=1}^n E\left[\xi_t(u)\middle|\mathscr{F}_{t-1}\right] &= 2\sum_{t=1}^n w_t\int_0^{-u'\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}}[G(s)-G(0)]\,ds\\
&= 2\sum_{t=1}^n w_t\int_0^{-u'\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}} sg(\zeta^*)ds\\
&= (\sqrt{n}u)'K_{1n}(\sqrt{n}u)+n\|u\|^2K_{2n}(u), \qquad (6.43)
\end{aligned}
$$

where ζ^* lies between 0 and s, and

$$
\begin{aligned}
K_{1n} &= \frac{g(0)}{n}\sum_{t=1}^n w_t\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta'},\\
K_{2n}(u) &= \frac{2}{n\|u\|^2}\sum_{t=1}^n w_t\int_0^{-u'\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}} s[g(\zeta^*)-g(0)]ds.
\end{aligned}
$$

Using the ergodic theorem, it is easy to see that

$$K_{1n}=g(0)\Sigma_0+o_p(1). \qquad (6.44)$$

Furthermore, by Lemmama 6.4 (ii), it is straightforward to see that for any $\eta>0$, there exists a $\rho\in(0,1)$ such that

$$
\begin{aligned}
\sup_{\|u\|\le\eta}|K_{2n}(u)| &\le \sup_{\|u\|\le\eta}\frac{2}{n\|u\|^2}\sum_{t=1}^n w_t\int_{-\left|u'\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\right|}^{\left|u'\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\right|}\left|s\left[g(\zeta^*)-g(0)\right]\right|ds\\
&\le \frac{1}{n}\sum_{t=1}^n\left[\sup_{|s|\le C\eta\xi_{\rho t-1}}|g(s)-g(0)|w_t\xi_{\rho t-1}^2\right].
\end{aligned}
$$

Note that by Assumptions 6.4-6.5, $\sup_{x\in R} g(x) < \infty$ and $E(w_t \xi_{\rho t-1}^2) < \infty$. Then, by the dominated convergence theorem, we have

$$E\left[\sup_{|s|\leq C\eta\xi_{\rho t-1}} |g(s) - g(0)|\, w_t \xi_{\rho t-1}^2\right] \to 0$$

as $\eta \to 0$. Thus, by stationarity of $\{y_t\}$ and Markov's Theorem, for $\forall \varepsilon, \delta > 0$, $\exists \eta_0(\varepsilon) > 0$, such that

$$P\left(\sup_{\|u\|\leq \eta_0} |K_{2n}(u)| > \delta\right) < \frac{\varepsilon}{2} \tag{6.45}$$

for all $n \geq 1$. On the other hand, since $u_n = o_p(1)$, it follows that

$$P(\|u_n\| > \eta_0) < \frac{\varepsilon}{2} \tag{6.46}$$

as n is large enough. By (6.45) and (6.46), for $\forall \varepsilon, \delta > 0$, we have

$$\begin{aligned} P(|K_{2n}(u_n)| > \delta) &\leq P(\{|K_{2n}(u_n)| > \delta\} \cap \{\|u_n\| \leq \eta_0\}) + P(\|u_n\| > \eta_0) \\ &\leq P\left(\sup_{\|u\|\leq \eta_0} |K_{2n}(u)| > \delta\right) + \frac{\varepsilon}{2} \\ &\leq \varepsilon \end{aligned} \tag{6.47}$$

as n is large enough, i.e., $K_{2n}(u_n) = o_p(1)$. Furthermore, combining (6.43)-(6.44), we can see that (i) holds.

(ii). Let $\beta_n(u) \equiv \beta_{1n}(u) + \beta_{2n}(u)$, where

$$\begin{aligned} \beta_{1n}(u) &= \frac{u'}{2} \sum_{t=1}^{n} w_t \frac{\partial^2 \varepsilon_t(\xi^*)}{\partial\theta\partial\theta'} [I(\varepsilon_t > 0) - I(\varepsilon_t < 0)]u, \\ \beta_{2n}(u) &= 2\sum_{t=1}^{n} w_t \int_{-u' \frac{\partial \varepsilon_t(\theta_0)}{\partial\theta}}^{-q_t(u)} X_t(s) ds. \end{aligned}$$

By Lemmama 6.4 (iii) and Assumption 6.4, there exists a $\rho \in (0,1)$ such that

$$E\left[\sup_{\xi^*\in\Lambda} w_t \left|\frac{\partial^2 \varepsilon_t(\xi^*)}{\partial\theta\partial\theta'} [I(\varepsilon_t > 0) - I(\varepsilon_t < 0)]\right|\right] \leq CE(w_t \xi_{\rho t-1}) < \infty.$$

Since ε_t has median 0, the conditional expectation property gives

$$E\left[w_t \frac{\partial^2 \varepsilon_t(\xi^*)}{\partial\theta\partial\theta'} [I(\varepsilon_t > 0) - I(\varepsilon_t < 0)]\right] = 0.$$

Then, by Theorem 3.1 in Ling and McAleer (2003), we have

$$\sup_{\xi^*\in\Lambda}\left|\frac{1}{n}\sum_{t=1}^{n} w_t \frac{\partial^2\varepsilon_t(\xi^*)}{\partial\theta\partial\theta'}[I(\varepsilon_t>0)-I(\varepsilon_t<0)]\right| = o_p(1).$$

Thus, it follows that $\beta_{1n}(u_n) = o_p(n\|u_n\|^2)$.

Next, by a direct calculation, we have

$$\begin{aligned}\frac{\beta_{2n}(u)}{n\|u\|^2} &= \frac{2}{n}\sum_{t=1}^{n} w_t \int_0^{-\frac{1}{2\|u\|^2}u'\frac{\partial^2\varepsilon_t(\xi^*)}{\partial\theta\partial\theta'}u} X_t\left(\|u\|^2 s - u'\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\right) ds \\ &\equiv \frac{1}{n}\sum_{t=1}^{n} J_t(u).\end{aligned}$$

By Lemmama 6.4 (ii) and (iii), we have for any $\eta>0$, there exists a $\rho\in(0,1)$ such that

$$\begin{aligned}\sup_{\|u\|\le\eta}|J_t(u)| &\le 2w_t\int_{-C\xi_{\rho t-1}}^{C\xi_{\rho t-1}}\left\{X_t\left(C\eta^2\xi_{\rho t-1}+C\eta\xi_{\rho t-1}\right)\right. \\ &\quad \left.-X_t\left(-C\eta^2\xi_{\rho t-1}-C\eta\xi_{\rho t-1}\right)\right\} ds \\ &= Cw_t\xi_{\rho t-1}\left\{X_t\left(C\eta^2\xi_{\rho t-1}+C\eta\xi_{\rho t-1}\right)\right. \\ &\quad \left.-X_t\left(-C\eta^2\xi_{\rho t-1}-C\eta\xi_{\rho t-1}\right)\right\}.\end{aligned}$$

Then, by Assumptions 6.4-6.5 and the conditional expectation property, it follows that

$$\begin{aligned}E\left[\sup_{\|u\|\le\eta}|J_t(u)|\right] &\le CE\left[w_t\xi_{\rho t-1}\left\{G\left(C\eta^2\xi_{\rho t-1}+C\eta\xi_{\rho t-1}\right)\right.\right. \\ &\quad \left.\left.-G\left(-C\eta^2\xi_{\rho t-1}-C\eta\xi_{\rho t-1}\right)\right\}\right] \\ &\le C(\eta^2+\eta)\sup_x g(x)E(w_t\xi_{\rho t-1}^2)\to 0\end{aligned}$$

as $\eta\to 0$. Similar to the proof of (6.47), we can show that $\beta_{2n}(u_n) = o_p(n\|u_n\|^2)$. This completes the proof of (ii). □

6.6 Appendix: Proof

In this Appendix, we give the proof of Theorem 6.4. We first give two lemmamas. The first one is directly taken from Ling (2007).

Lemma 6.4. *Let $\xi_{\rho t}$ be defined as in Assumption 6.4. If Assumption 3.4 holds, then there exist constants C and $\rho\in(0,1)$ such that the following holds uniformly in θ:*

$$
\begin{aligned}
&(i) \quad \sup_{\Theta} \|\varepsilon_{t-1}(\theta)\| \leq C\xi_{\rho t-1}, \\
&(ii) \quad \sup_{\Theta} \left\| \frac{\partial \varepsilon_t(\theta)}{\partial \theta} \right\| \leq C\xi_{\rho t-1}, \\
&(iii) \quad \sup_{\Theta} \left\| \frac{\partial^2 \varepsilon_t(\theta)}{\partial \theta \partial \theta'} \right\| \leq C\xi_{\rho t-1}.
\end{aligned}
$$

Lemma 6.5. *For any $\theta^* \in \Theta$, let $B_\eta(\theta^*) = \{\theta \in \Theta : \|\theta - \theta^*\| < \eta\}$ be an open neighborhood of θ^* with radius $\eta > 0$. If Assumptions 3.4-6.5 hold, then*

$$
\begin{aligned}
&(i) \quad E[w_t|\varepsilon_t(\theta)|] \text{ has a unique minimum at } \theta_0, \\
&(ii) \quad E\left[\sup_{\theta \in B_\eta(\theta^*)} w_t|\varepsilon_t(\theta) - \varepsilon_t(\theta^*)|\right] \to 0 \text{ as } \eta \to 0.
\end{aligned}
$$

Proof. Note that ε_t has median 0 and $E|X-a| \geq E|X - median(X)|$ for all r.v. X and real number a. It follows that

$$
\begin{aligned}
E[w_t|\varepsilon_t(\theta)|] &= E\left[w_t E\left(\left|\varepsilon_t(\theta_0) + (\theta - \theta_0)' \frac{\partial \varepsilon_t(\tilde{\theta})}{\partial \theta}\right| \big| \mathscr{F}_{t-1}\right)\right] \\
&\geq E\left[w_t E\left(|\varepsilon_t(\theta_0)| \big| \mathscr{F}_{t-1}\right)\right] = E[w_t|\varepsilon_t(\theta_0)|],
\end{aligned}
$$

where $\tilde{\theta}$ lies between θ and θ_0. The equation in the last inequality holds if and only if $(\theta - \theta_0)' \partial \varepsilon_t(\tilde{\theta})/\partial \theta = 0$ a.s., which holds if and only if $\theta = \theta_0$, see Ling (2007). Thus (i) holds. Furthermore, by Lemmama 6.4 (ii) and Assumption 6.4, there exists a $\rho \in (0,1)$ such that

$$
\begin{aligned}
E\left[\sup_{\theta \in B_\eta(\theta^*)} w_t|\varepsilon_t(\theta) - \varepsilon_t(\theta^*)|\right] &= E\left[\sup_{\theta \in B_\eta(\theta^*)} w_t \left|(\theta - \theta^*)' \frac{\partial \varepsilon_t(\theta^{**})}{\partial \theta}\right|\right] \\
&< C\eta E\left[w_t \xi_{\rho t-1}\right] \to 0 \text{ as } \eta \to 0,
\end{aligned}
$$

where θ^{**} lies between θ and θ^*. Thus, we can see that (ii) holds. This completes the proof.

PROOF OF THEOREM 6.4. We use the method in Huber (1967). Let V be any open neighborhood of θ_0. By Lemmama 6.5 (ii), for any $\theta^* \in V^c = \Theta / V$ and $\varepsilon > 0$, there exists an $\eta_0 > 0$ such that

$$
E\left[\inf_{\theta \in B_{\eta_0}(\theta^*)} w_t|\varepsilon_t(\theta)|\right] \geq E[w_t|\varepsilon_t(\theta^*)|] - \varepsilon. \tag{6.48}
$$

By Assumption 6.4 and Lemmama 6.4, there exists a $\rho \in (0,1)$ such that

$$E\left[\inf_{\theta\in B_{\eta_0}(\theta^*)} w_t|\varepsilon_t(\theta)|\right] \le E[w_t\xi_{\rho t-1}] < \infty.$$

By the ergodic theorem, it follows that

$$\frac{1}{n}\sum_{t=1}^{n}\inf_{\theta\in B_{\eta_0}(\theta^*)} w_t|\varepsilon_t(\theta)| \ge E\left[\inf_{\theta\in B_{\eta_0}(\theta^*)} w_t|\varepsilon_t(\theta)|\right] - \varepsilon \tag{6.49}$$

as n is large enough. Let $\{B_{\eta_0}(\theta_i) : \theta_i \in V^c, i = 1,2,\cdots,k\}$ be a finite covering of V^c. Thus, by (6.48)-(6.49), we have

$$\begin{aligned}\inf_{\theta\in V^c} H_n(\theta) &= \min_{1\le i\le k}\inf_{\theta\in B_{\eta_0}(\theta_i)} H_n(\theta)\\ &\ge \min_{1\le i\le k}\frac{1}{n}\sum_{t=1}^{n}\inf_{\theta\in B_{\eta_0}(\theta_i)} w_t|\varepsilon_t(\theta)|\\ &\ge \min_{1\le i\le k} E\left[\inf_{\theta\in B_{\eta_0}(\theta_i)} w_t|\varepsilon_t(\theta)|\right] - \varepsilon. \end{aligned}\tag{6.50}$$

Note that the infimum on the compact set V^c is attained. For each $\theta_i \in V^c$, from Lemmama 6.5 (i), there exists an $\varepsilon_0 > 0$ such that

$$E\left[\inf_{\theta\in B_{\eta_0}(\theta_i)} w_t|\varepsilon_t(\theta)|\right] \ge E[w_t|\varepsilon_t(\theta_0)|] + 3\varepsilon_0. \tag{6.51}$$

Thus, by (6.50)-(6.51), taking $\varepsilon = \varepsilon_0$, it follows that

$$\inf_{\theta\in V^c} H_n(\theta) \ge E[w_t|\varepsilon_t(\theta_0)|] + 2\varepsilon_0. \tag{6.52}$$

On the other hand, by the ergodic theorem, it follows that

$$\inf_{\theta\in V} H_n(\theta) \le H_n(\theta_0) = \frac{1}{n}\sum_{t=1}^{n} w_t|\varepsilon_t(\theta_0)| \le E[w_t|\varepsilon_t(\theta_0)|] + \varepsilon_0. \tag{6.53}$$

Hence, combining (6.52)-(6.53),

$$\inf_{\theta\in V^c} H_n(\theta) \ge E[w_t|\varepsilon_t(\theta_0)|] + 2\varepsilon_0 > E[w_t|\varepsilon_t(\theta_0)|] + \varepsilon_0 \ge \inf_{\theta\in V} H_n(\theta),$$

which implies that

$$\hat{\theta}_n \in V, \quad a.s. \text{ for all } V, \text{ as } n \text{ is large enough .}$$

Since V is arbitrary, it yields $\hat{\theta}_n \to \theta_0$ a.s. This completes the proof. □

Chapter 7
Threshold Models

7.1 Introduction

Since the TAR model was proposed by Tong (1978), it has attracted much attention and been widely investigated in the literature. Today, it has been extended to one class so-called threshold models in nonlinear time series. The threshold model has many important applications in diverse fields such as economics, finance, hydrology, physics, population dynamics, neural science and among others. A fairly comprehensive survey on the TAR model is available in Tong (1990).

The difficulty in threshold models is to estimate the threshold when it is unknown. The asymptotic theory of the TAR model was established first by Chan (1993) when the autoregressive function is discontinuous and by Chan & Tsay (1998) when the autoregressive function is continuous. Hansen (1997, 2000) also studied the TAR model. Under the assumption that the threshold effect is vanishingly small, he obtained the distribution- and parameter-free limit of the estimated threshold. Seo & Linton (2007) proposed a smoothed LSE for the TAR/regression model and showed that the estimated threshold is asymptotically normal but its convergence rate is less than n and depends on the bandwidth.

In this chapter, we discuss the Section 1 is to investigate the QMLE of the threshold double AR model. Section 2 is to study ....

7.2 Quasi-MLE of Threshold Double AR Models

This section considers the threshold double AR model of order p, defined by

$$Y_t = \begin{cases} \phi_{10} + \sum_{j=1}^{p} \phi_{1j} Y_{t-j} + \eta_t \sqrt{\alpha_{10} + \sum_{j=1}^{p} \alpha_{1j} Y_{t-j}^2}, & \text{if } Y_{t-d} \le r, \\ \phi_{20} + \sum_{j=1}^{p} \phi_{2j} Y_{t-j} + \eta_t \sqrt{\alpha_{20} + \sum_{j=1}^{p} \alpha_{2j} Y_{t-j}^2}, & \text{if } Y_{t-d} > r, \end{cases} \tag{7.1}$$

where ϕ_{ij}, α_{ij}, $i=1,2, j=0,1,...,p$, are the coefficients, r is the threshold and d is a positive integer called the delay parameter, and p is a known nonnegative integer. Model (7.1) is a threshold ARCH-type model, but is different from those studied in Li & Lam (1995) and Li & Li (1996) since the volatility in our model is an immediate regression on the observed process.The probabilistic structure of model (7.1) was studied by Cline (2007) and Cline & Pu (2004). When $\phi_{1i}=\phi_{2i}$ and $\alpha_{1i}=\alpha_{2i}$, model (7.1) reduces to the double AR model and the related work can be found in Ling (2004, 2007) and Ling & Li (2008).

Assume that $\{Y_1,...,Y_n\}$ is a sample from model (7.1). Given the initial values $\{Y_{-p},...,Y_0\}$, the conditional log-quasi-MLF (omitting a constant) is defined as

$$L_n(\theta)=\sum_{t=1}^{n} l_t(\theta),$$

where $\theta=(\lambda^{\tau},r)^{\tau}=(\phi_1^{\tau},\alpha_1^{\tau},\phi_2^{\tau},\alpha_2^{\tau},r)^{\tau}$ is the parameter with $\phi_i=(\phi_{i0},\phi_{i1},...,\phi_{ip})^{\tau}$ and $\alpha_i=(\alpha_{i0},\alpha_{i1},...,\alpha_{ip})^{\tau}$ for $i=1,2$,

$$\begin{aligned} l_t(\theta) &= -\frac{1}{2}\left\{\log(\alpha_1^{\tau}X_{t-1})+\frac{(Y_t-\phi_1^{\tau}\tilde{Y}_{t-1})^2}{\alpha_1^{\tau}X_{t-1}}\right\}I(Y_{t-d}\le r)\\ &\quad -\frac{1}{2}\left\{\log(\alpha_2^{\tau}X_{t-1})+\frac{(Y_t-\phi_2^{\tau}\tilde{Y}_{t-1})^2}{\alpha_2^{\tau}X_{t-1}}\right\}I(Y_{t-d}> r)\end{aligned}$$

with $\tilde{Y}_{t-1}=(1,Y_{t-1},...,Y_{t-p})^{\tau}$ and $X_{t-1}=(1,y_{t-1}^2,...,y_{t-p}^2)^{\tau}$, and $I(\cdot)$ is the indicator function. Assume that d is known and $1\le d\le \max(p,1)$. Let Θ be the parameter space of θ. The maximizer $\widehat{\theta}_n=(\widehat{\lambda}_n^{\tau},\widehat{r}_n)^{\tau}$ of $L_n(\theta)$ on Θ is called a MLE of $\theta_0=(\lambda_0^{\tau},r_0)^{\tau}$. That is, $\widehat{\theta}_n$ is defined by

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

Due to the discontinuity of $L_n(\theta)$ in r, one can take two steps to find $\widehat{\theta}_n$:

Step 1. For each fixed r, maximize $L_n(\theta)$ and get its maximizer $\widehat{\lambda}_n(r)$.
Step 2. Since $L_n^*(r)\equiv L_n(\widehat{\lambda}_n(r),r)$ only takes finite possible values, one can get the maximizer $\widehat{r}_n$ of $L_n^*(r)$ by the enumeration approach and then obtain the estimator $\widehat{\theta}_n=(\widehat{\lambda}_n(\widehat{r}_n)^{\tau},\widehat{r}_n)^{\tau}$.

Generally, there exist infinitely many r such that $L_n(\cdot)$ attains its global maximum. One can choose the smallest r as an estimator of r_0. According to this procedure, $\widehat{\theta}_n$ is the QMLE of θ_0, i.e., $L_n(\widehat{\theta}_n)=\max_{\theta\in\Theta}L_n(\theta)$.

Assumption 7.1 *$\{\eta_t\}$ is a sequence of iid r.v.s with mean* 0 *and variance* 1*, and has a positive and continuous density $h(x)$ on $\mathbb{R}$.*

Assumption 7.2 *The parameter space* $\Theta = \{\theta \in \mathbb{R}^{4p+5} : \phi_1 \neq \phi_2$ *or* $\alpha_1 \neq \alpha_2, \alpha_{ij} > 0, i = 1,2, j = 0,1,...,p\}$ *is compact.*

Theorem 7.1. *Suppose that Assumptions 7.1 and 7.2 hold and* $\{Y_t\}$ *is strictly stationary and ergodic. Then,* $\widehat{\theta}_n \to \theta_0$ *with probability one, as* $n \to \infty$.

There is no any requirement for the moment of Y_t in Theorem 7.1. If we allow some α_{ij}'s to be zeros, then the result still holds when $E(Y_t^2) < \infty$.

Let $Z_t = (Y_t, ..., Y_{t-p+1})^\tau$. Then $\{Z_t\}$ is a Markov chain. Denote its l-step transition probability by $\mathbf{P}^l(z, A)$, where $z \in \mathbb{R}^p$ and A is a Borel set. We introduce two more assumptions as follows.

Assumption 7.3 $\{Z_t\}$ *admits a unique invariant measure* $\Pi(\cdot)$ *such that there exist* $K > 0$ *and* $\rho \in [0,1)$, *for any* $z \in \mathbb{R}^p$ *and any n,* $\|\mathbf{P}^n(z,\cdot) - \Pi(\cdot)\|_{\rm v} \leq K(1 + \|z\|^2)\rho^n$, *where* $\|\cdot\|_{\rm v}$ *and* $\|\cdot\|$ *denote the total variation norm and the Euclidian norm, respectively.*

Assumption 7.4 *There exist nonrandom vectors* $w = (1, w_1, ..., w_p)^\tau$ *with* $w_d = r_0$ *and* $W = (1, W_1, ..., W_p)^\tau$ *with* $W_d = r_0^2$ *such that*

$$\{(\phi_{10} - \phi_{20})^\tau w\}^2 + \{(\alpha_{10} - \alpha_{20})^\tau W\}^2 > 0.$$

Assumption 7.3 is on the V-uniform ergodicity of $\{Y_t\}$, under which $\{Y_t\}$ is strictly stationary if the initial value Z_0 follows the stationary distribution Π. Without loss of generality, in what follows we assume that $Z_0 \sim \Pi$. Cline & Pu (2004) showed that Assumption 7.3 holds if Assumption 7.1 holds with $\sup_{x\in\mathbb{R}}\{(1 + |x|)h(x)\} < \infty$ and

$$\Big\{\sum_{i=1}^{p} \max(|\phi_{1i}|, |\phi_{2i}|)\Big\}^2 + \sum_{i=1}^{p} \max(\alpha_{1i}, \alpha_{2i}) < 1.$$

A weaker and more general condition for the V-uniform ergodicity of model (7.1) is given by Cline & Pu (2004), see also Lu (1998). Assumption 7.4 implies that the mean function or the volatility function is not continuous over the hyperplane $Y_{t-d} = r_0$. It is a necessary condition for the n-convergence rate of $\widehat{r}_n$.

Theorem 7.2. *If Assumptions 7.1-7.4 hold,* θ_0 *is an interior point of* Θ *and* $\kappa_4 = E(\eta_1^4) < \infty$, *then*

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

(ii). $\sqrt{n} \sup\limits_{|r-r_0|\leq B/n} \big\|\widehat{\lambda}_n(r) - \widehat{\lambda}_n(r_0)\big\| = o_p(1)$ *for any fixed constant* $0 < B < \infty$.

Furthermore, it follows that

$$\sqrt{n}(\widehat{\lambda}_n - \lambda_0) = \sqrt{n}(\widehat{\lambda}_n(r_0) - \lambda_0) + o_p(1) \longrightarrow_d N(0, \Omega^{-1}\Sigma\Omega^{-1}) \quad as\ n \to \infty,$$

where $\Omega = \mathrm{diag}(A_1, 0.5B_1, A_2, 0.5B_2)$ *and* $\Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2)$ *with*

$$\Sigma_i = \begin{pmatrix} A_i & \frac{\kappa_3}{2} D_i \\ \frac{\kappa_3}{2} D_i^\tau & \frac{\kappa_4 - 1}{4} B_i \end{pmatrix}, \quad i = 1, 2,$$

where $\kappa_3 = E(\eta_1^3)$ *and, for* $i = 1, 2$,

$$A_i = E\left\{ \frac{\tilde{Y}_{p-1} \tilde{Y}_{p-1}^\tau}{\alpha_{i0}^\tau X_{p-1}} g_i(r_0) \right\}, B_i = E\left\{ \frac{X_{p-1} X_{p-1}^\tau}{(\alpha_{i0}^\tau X_{p-1})^2} g_i(r_0) \right\} \text{ and } D_i = E\left\{ \frac{\tilde{Y}_{p-1} X_{p-1}^\tau}{(\alpha_{i0}^\tau X_{p-1})^{3/2}} g_i(r_0) \right\}$$

with $g_1(r_0) = I(Y_{p-d} \le r_0)$ *and* $g_2(r_0) = I(Y_{p-d} > r_0)$.

When ε_1 is symmetric, then $\kappa_3 = 0$ and

$$\Omega^{-1} \Sigma \Omega^{-1} = \operatorname{diag}\left\{ A_1^{-1}, (\kappa_4 - 1) B_1^{-1}, A_2^{-1}, (\kappa_4 - 1) B_2^{-1} \right\}.$$

When $\varepsilon_1 \sim N(0,1)$, then $\widehat{\theta}_n$ is the MLE of θ_0 and $\Omega^{-1}\Sigma\Omega^{-1} = \Omega^{-1}$.

To describe the limiting distribution of $n(\widehat{r}_n - r_0)$, we consider the limiting behavior of a sequence of normalized profile log-likelihood processes defined by

$$\widetilde{L}_n(z) = -2\left\{ L_n\left(\widehat{\lambda}_n(r_0 + \frac{z}{n}), r_0 + \frac{z}{n}\right) - L_n\left(\widehat{\lambda}_n(r_0), r_0\right) \right\}, \quad z \in \mathbb{R}. \tag{7.2}$$

Using Theorem 7.2 and the Taylor expansion, it is straightforward to show that $\widetilde{L}_n(z)$ can be approximated in $\mathbb{D}(\mathbb{R})$, the space of all cadlag functions on $\mathbb{R}$ being equipped with the Skorokhod metric, by

$$\wp_n(z) = I(z<0) \sum_{t=1}^n \zeta_{1t} I\left(r_0 + \frac{z}{n} < Y_{t-d} \le r_0\right) + I(z \ge 0) \sum_{t=1}^n \zeta_{2t} I\left(r_0 < Y_{t-d} \le r_0 + \frac{z}{n}\right),$$

where

$$\begin{aligned} \zeta_{1t} &= \log \frac{\alpha_{20}^\tau X_{t-1}}{\alpha_{10}^\tau X_{t-1}} + \frac{\left\{ (\phi_{10} - \phi_{20})^\tau \tilde{Y}_{t-1} + \varepsilon_t \sqrt{\alpha_{10}^\tau X_{t-1}} \right\}^2}{\alpha_{20}^\tau X_{t-1}} - \eta_t^2, \\ \zeta_{2t} &= \log \frac{\alpha_{10}^\tau X_{t-1}}{\alpha_{20}^\tau X_{t-1}} + \frac{\left\{ (\phi_{10} - \phi_{20})^\tau \tilde{Y}_{t-1} - \varepsilon_t \sqrt{\alpha_{20}^\tau X_{t-1}} \right\}^2}{\alpha_{10}^\tau X_{t-1}} - \eta_t^2. \end{aligned} \tag{7.3}$$

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 further define a two-sided compound Poisson process $\wp(z)$ as

$$\wp(z) = I(z<0) \sum_{i=1}^{N_1(-z)} \xi_i^{(1)} + I(z \ge 0) \sum_{j=1}^{N_2(z)} \xi_j^{(2)}, \; z \in \mathbb{R}, \tag{7.4}$$

where $\{N_1(z), z \ge 0\}$ and $\{N_2(z), z \ge 0\}$ are two independent Poisson processes with $N_1(0) = N_2(0) = 0$ and with identical jump rate $\pi(r_0)$, where $\pi(\cdot)$ is the density of Y_1. $\{\xi_i^{(1)}\}$ and $\{\xi_i^{(2)}\}$ are independent. $\{\xi_i^{(k)}, i \geq 1\}$ are independent random variables with the same distribution $F_k(\cdot|r_0)$, see Figure 8.1 for its realization of one sample path. Clearly, $\wp(z)$ goes to $+\infty$ as $z \to \pm\infty$ almost surely since $E(\xi_1^{(1)}) > 0$

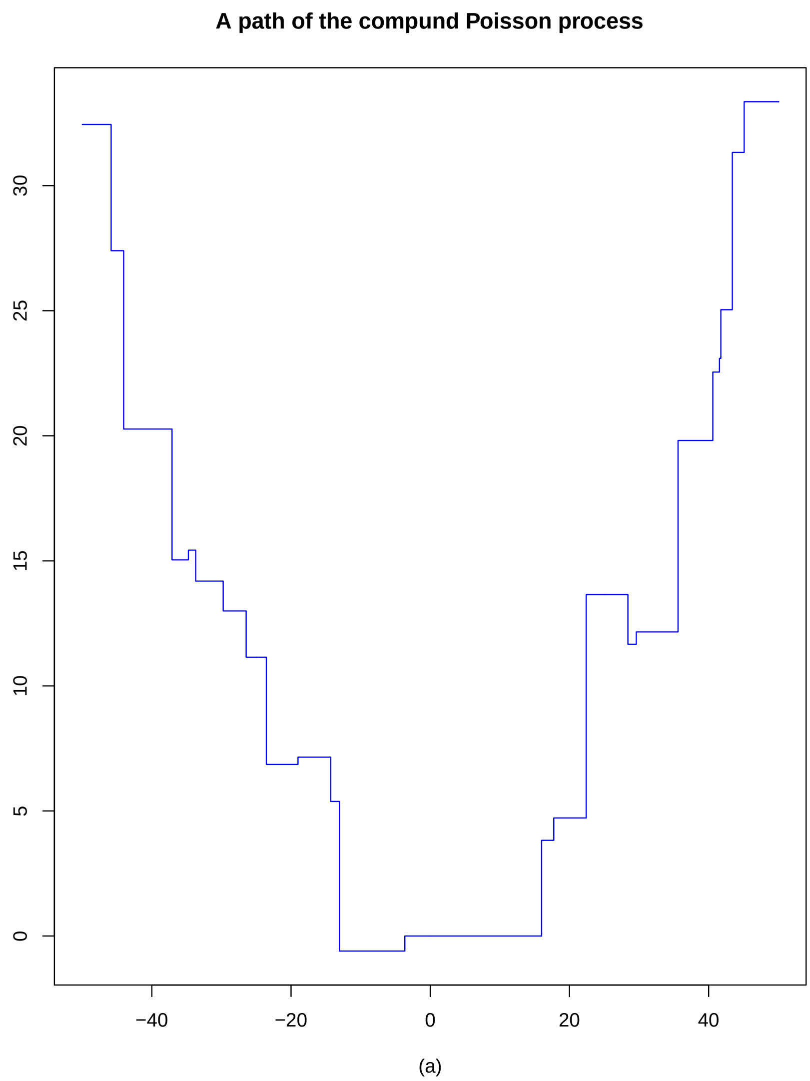


Fig. 7.1 $\theta_0 = (1, -0.6, 1, 0.5, -1, -0.2, 0.5, 0.3, 0)'$ and $\eta_t \sim N(0,1)$.

and $E(\xi_1^{(2)}) > 0$ by Assumption 7.4 and the elementary inequality $\log(1/x) + x -$

$1 > 0$ for $x > 0$ unless $x = 1$. Thus, there exists a unique random interval $[M_-, M_+)$ on which the process $\wp(z)$ attains its global minimum.

Theorem 7.3. *If Assumptions 7.1-7.4 hold, then* $n(\widehat{r}_n - r_0)$ *converges in distribution to* M_-. *Furthermore,* $n(\widehat{r}_n - r_0)$ *is asymptotically independent of* $\sqrt{n}(\widehat{\lambda}_n - \lambda_0)$ *which is* $N(0, \Omega^{-1}\Sigma\Omega^{-1})$ *asymptotically.*

When $\alpha_{10} = \alpha_{20}$ and α_{ij}'s are zeros, $i = 1, 2$, $j = 1, ..., p$, Theorem 7.3 reduces to Theorem 2.2 of Chan (1993) for TAR models. When $\alpha_{1j} = \alpha_{2j}$, $j = 0, 1, ..., p$, model (7.1) reduces to the TAR model with ARCH errors. The corresponding parameter is $\theta = (\lambda^\tau, r)^\tau$ with $\lambda = (\phi_1^\tau, \phi_2^\tau, \alpha^\tau)^\tau$. In this case, Theorem 7.3 still holds with

$$\Omega^{-1}\Sigma\Omega^{-1} = \begin{pmatrix} A_1^{-1} & 0 & \kappa_3 A_1^{-1} D_1 (B_1 + B_2)^{-1} \\ 0 & A_2^{-1} & \kappa_3 A_2^{-1} D_2 (B_1 + B_2)^{-1} \\ \kappa_3 (B_1 + B_2)^{-1} D_1^\tau A_1^{-1} & \kappa_3 (B_1 + B_2)^{-1} D_2^\tau A_2^{-1} & (\kappa_4 - 1)(B_1 + B_2)^{-1} \end{pmatrix},$$

where A_i, B_i and D_i are defined in Theorem 7.2 with replacing α_{i0}'s by α_0. When all $\phi_{ij} = 0$, model (7.1) is a threshold ARCH model. The corresponding parameter is $\theta = (\lambda^\tau, r)^\tau$ with $\lambda = (\alpha_1^\tau, \alpha_2^\tau)^\tau$. In this case, Theorem 3 holds with

$$\Omega^{-1}\Sigma\Omega^{-1} = (\kappa_4 - 1)\mathrm{diag}(B_1^{-1}, B_2^{-1}).$$

When both the mean function and the volatility function are continuous, we conjecture that $\widehat{r}_n$ is $\sqrt{n}$-consistent and asymptotically normal, see Chan & Tsay (1998) for TAR models. In practice, d is unknown and can be estimated consistently by an analogous procedure in Chan (1993). The order p can be determined by the AIC or BIC.

7.3 Numerical Implementation of M_- and Simulation Study

From (7.4), we know that M_- is determined by the jump rate $\pi(r_0)$ and the jump distribution $F_k(\cdot|r_0)$. We can simulate M_- via simulating the two-sided compound Poisson process (7.4) on the interval $[-T, T]$ for $T > 0$ large enough. Modifying Algorithm 6.2 in Cont and Tankov (2004, pp.174) for a one-sided compound Poisson process, the algorithm is as follows.

Algorithm A:
Step A.1. *Generate two independent Poisson random variables* N_1 *and* N_2 *with the same parameter* $\pi(r_0)T$ *which are the total number of jumps on* $[-T, 0]$ *and* $[0, T]$, *respectively.*

Step A.2. *Generate two independent jump time sequences:* $\{U_1, \cdots, U_{N_1}\}$ *and* $\{V_1, \cdots, V_{N_2}\}$*, where* U_i*'s and* V_i*'s are independently and uniformly distributed on* $[-T, 0]$ *and* $[0, T]$*, respectively.*
Step A.3. *Generate two independent jump-size sequences* $\{\xi_1^{(1)}, \cdots, \xi_{N_1}^{(1)}\}$ *and* $\{\xi_1^{(2)}, \cdots, \xi_{N_2}^{(2)}\}$ *from* $F_1(\cdot|r_0)$ *and* $F_2(\cdot|r_0)$*, respectively.*

For $z \in [-T, T]$, the trajectory of (7.4) is given by

$$\wp(z) = I(z < 0) \sum_{i=1}^{N_1} I(U_i > z)\xi_i^{(1)} + I(z \geq 0) \sum_{j=1}^{N_2} I(V_j < z)\xi_j^{(2)}.$$

Then, we take the smallest minimizer of $\wp(z)$ on $[-T, T]$ as one observed value of M_-. Repeating the previous algorithm, we can get a sequence of observations of M_-, which can be used for statistical inferences of r_0.

The key step is to generate the jump-size sequences from $F_k(\cdot|r_0)$ in Step A.3. When $p = d = 1$, $F_k(\cdot|r_0)$ reduces to an unconditional distribution and the sampling is easy from it. For the general case, i.e., $p > 1$ or $d > 1$, note that $E[\pi(r_0|Z_0)] = \pi(r_0)$, where $\pi(x|z)$ is the conditional density of Y_1 given $Z_0 = z$. By the property of conditional expectation and strong law of large numbers, we have

$$\begin{aligned}
F_k(x|r_0) &= \mathbb{P}(\zeta_{k,d+1} \leq x|Y_1 = r_0) \\
&= \int_{\mathbb{R}^p} \mathbb{P}(\zeta_{k,d+1} \leq x|Y_1 = r_0, Z_0 = z) \frac{\pi(r_0|z)}{\pi(r_0)} G(dz) \\
&= \frac{1}{K} \sum_{i=1}^{K} \mathbb{P}(\zeta_{k,d+1} \leq x|Y_1 = r_0, Z_0 = z_i) \frac{\pi(r_0|z_i)}{\pi(r_0)} + o(1) \\
&= \sum_{i=1}^{K} \mathbb{P}(\zeta_{k,d+1} \leq x|Y_1 = r_0, Z_0 = z_i) \frac{\pi(r_0|z_i)}{\sum_{j=1}^{K} \pi(r_0|z_j)} + o(1) \\
&\equiv F_{k,K}(x|r_0, \{z_j\}) + o(1)
\end{aligned}$$

almost surely as $K \to \infty$, uniformly in $x \in \mathbb{R}$ by Theorem 2 in Pollard (1984, page 8), where $G(\cdot)$ is the distribution of $Z_0 = (Y_0, ..., Y_{1-p})^\tau$. From model (7.1), it follows that

$$\pi(x|z) = \frac{1}{\sigma(z, \theta_0)} h\left(\frac{x - \mu(z, \theta_0)}{\sigma(z, \theta_0)}\right)$$

where $z = (z_1, ..., z_p)^\tau$ and

$$\mu(z,\theta) = \{\phi_{10} + \sum_{j=1}^{p} \phi_{1j} z_j\} I(z_d \le r) + \{\phi_{20} + \sum_{j=1}^{p} \phi_{2j} z_j\} I(z_d > r),$$
$$\sigma(z,\theta) = \sqrt{\alpha_{10} + \sum_{j=1}^{p} \alpha_{1j} z_j^2} I(z_d \le r) + \sqrt{\alpha_{20} + \sum_{j=1}^{p} \alpha_{2j} z_j^2} I(z_d > r). \tag{7.5}$$

Let $H(\cdot)$ be the cumulative distribution function of η_t. When θ_0, $\pi(r_0)$, $H(\cdot)$ and $G(\cdot)$ are known, the following algorithm describes how to sample $\xi_1^{(k)}$ from $F_{k,K}(x|r_0, \{z_j\})$.

Algorithm B:
Step B.1. *Draw a sample* $\mathscr{Z} = \{z_1, ..., z_K\}$ *from* $G(\cdot)$.
Step B.2. *For each* $i \in \{1, ..., K\}$*, sample, independently,* $\varepsilon_2, ..., \varepsilon_{d+1}$ *from* $H(\cdot)$ *and generate* $\{Y_2, ..., Y_d\}$ *by iterating model (7.1) with the initial value* $Y_1 = r_0$ *and* $Z_0 = z_i$*. Then calculate* $\zeta_{k,d+1}$ *in (7.3), denoted by* $\eta_{1k}^{(i)}$.
Step B.3. *Sample* U *from the density function* $\mathbb{P}(U = i|\mathscr{Z}) = \pi(r_0|z_i)/[\sum_{l=1}^{K} \pi(r_0|z_l)]$ *for* $i = 1, ..., K$*, independent of all* $\{\varepsilon_2, ..., \varepsilon_{d+1}\}$.
Step B.4. *Obtain* $\xi_1^{(k)} = \eta_{1k}^{(U)}$.

Clearly, by Step 2 and Step B.3, all $\{Y_2, ..., Y_d\}$ is independent of U given $\mathscr{Z}$, and so is $\eta_{1k}^{(i)}$. Denote the conditional measure $\mathbb{P}_{\mathscr{Z}}(\cdot|A) \equiv \mathbb{P}(\cdot|A, \mathscr{Z})$. Thus,

$$\begin{aligned}
\mathbb{P}_{\mathscr{Z}}(\xi_1^{(k)} \le x) &= \sum_{i=1}^{K} \mathbb{P}_{\mathscr{Z}}(\eta_{1k}^{(i)} \le x, U = i) \\
&= \sum_{i=1}^{K} \mathbb{P}_{\mathscr{Z}}(\eta_{1k}^{(i)} \le x) \mathbb{P}_{\mathscr{Z}}(U = i) \\
&= \sum_{i=1}^{K} \mathbb{P}(\zeta_{k,d+1} \le x | Y_1 = r_0, Z_0 = z_i) \frac{\pi(r_0|z_i)}{\sum_{l=1}^{K} \pi(r_0|z_l)} \\
&= F_{k,K}(x|r_0, \{z_j\}).
\end{aligned}$$

Since $F_{k,K}(x|r_0, \{z_j\}) \to F_k(x|r_0)$ almost surely as $K \to \infty$, the distribution of $\xi_1^{(k)}$ given $\mathscr{Z}$ is $F_1(x|r_0)$ asymptotically. Denote a sequence of the two-sided compound Poisson processes by $\{\wp_K(z) : z \in \mathbb{R}\}$ which are determined by the same jump rate $\pi(r_0)$ and jump distributions $F_{k,K}(x|r_0, \{z_j\})$. By Theorem 16 in Pollard (1984, page 134), we have that $\wp_K(z)$ converges weakly to $\wp(z)$ in $\mathbb{D}(\mathbb{R})$. Minimizing the process $\wp_K(z)$, we can get the smallest minimizer $\widetilde{M}_K$. By Theorem 3.1 (on the continuity of the smallest argmax functional) in Seijo and Sen (2011), almost surely we have that $\widetilde{M}_K$ converges weakly to M_- as $K \to \infty$. Summarizing above discussion, we have

$$\lim_{K\to\infty} |\mathbb{P}_{\mathscr{X}}(\widetilde{M}_K \le x) - \mathbb{P}(M_- \le x)| = 0$$

at each x for which $\mathbb{P}(M_- = x) = 0$.

Since $\widetilde{M}_K$ is only relevant to model (7.1) and independent of the estimation of the parameter, without loss of generality, in what follows, we regard $\widetilde{M}_K$ as M_-. In practice, however, since only one sample $\mathfrak{X} \equiv \{Y_1, ..., Y_n\}$ is available given the initial values $\{Y_{-p}, ..., Y_0\}$, we can use it to estimate θ_0 and $\pi(r_0)$ consistently, denoting the estimators as $\widehat{\theta}_n$ and $\widehat{\pi}(\widehat{r}_n)$, respectively, where $\widehat{\pi}(\cdot)$ is the kernel density estimator of $\pi(\cdot)$, and calculate the residuals $\{\widehat{\eta}_t : 1 \le t \le n\}$. Based on the residuals, we can construct the estimators $\widehat{H}(\cdot)$ and $\widehat{h}(\cdot)$ of $H(\cdot)$ and $h(\cdot)$ as follows

$$\widehat{H}(x) = \frac{1}{n}\sum_{t=1}^{n} I(\eta_t^* \le x),$$

where $\eta_t^* = \widehat{\eta}_t - \bar{\hat{\eta}}$ with $\bar{\hat{\eta}} = \frac{1}{n}\sum_{t=1}^{n} \widehat{\eta}_t$, and

$$\widehat{h}(x) = \frac{1}{n}\sum_{t=1}^{n} \frac{1}{\widehat{h}^*_{\text{opt}}} K\Big(\frac{\varepsilon_t^* - x}{\widehat{h}^*_{\text{opt}}}\Big),$$

where $K(x) = (\sqrt{2\pi})^{-1}\exp(-x^2/2)$ is the Gaussian kernel and $\widehat{h}^*_{\text{opt}}$ is the bandwidth, which can be selected 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 $\widehat{h}_{\text{opt}} = 1.06sn^{-1/5}$ is the reference bandwidth selector, and $s, \widehat{\gamma}_3$ and $\widehat{\gamma}_4$ are the sample standard deviation, skewness and kurtosis of the residuals $\{\widehat{\varepsilon}_t : 1 \le t \le n\}$, respectively. Of course, one can use other kernel functions and bandwidths. When $h(\cdot)$ is uniformly continuous, we have $\sup_{x\in\mathbb{R}} |\widehat{h}(x) - h(x)| = o_p(1)$ as $n \to \infty$, see Silverman (1978).

Algorithm C:
Step C.1. *Set* $z_i = (Y_i, ..., Y_{i-p+1})^\tau$ *for* $i = p, ..., n$.
Step C.2. *For each* $i \in \{p, ..., n\}$, *sample, independently,* $\tilde{\varepsilon}_2, ..., \tilde{\varepsilon}_{d+1}$ *from* $\widehat{H}$ *given* $\mathfrak{X}$ *and generate* $\{\tilde{y}_2, ..., \tilde{y}_d\}$ *by iterating model (7.1) with the initial value* $Y_1 = \widehat{r}_n$ *and* $Z_0 = z_i$ *and* θ_0 *being replaced by* $\widehat{\theta}_n$. *Then calculate* $\widetilde{\zeta}_{k,d+1}$, *denoted by* $\widetilde{\eta}_{1k}^{(i)}$, *where*

$$\widetilde{\zeta}_{1,d+1} = \log\frac{\widehat{\alpha}_{2n}^{\tau}X_d^*}{\widehat{\alpha}_{1n}^{\tau}X_d^*} + \frac{\left\{(\widehat{\phi}_{1n}-\widehat{\phi}_{2n})^{\tau}Y_d^* + \tilde{\eta}_{d+1}\sqrt{\widehat{\alpha}_{1n}^{\tau}X_d^*}\right\}^2}{\widehat{\alpha}_{2n}^{\tau}X_d^*} - \tilde{\eta}_{d+1}^2,$$

$$\widetilde{\zeta}_{2,d+1} = \log\frac{\widehat{\alpha}_{1n}^{\tau}X_d^*}{\widehat{\alpha}_{2n}^{\tau}X_d^*} + \frac{\left\{(\widehat{\phi}_{1n}-\widehat{\phi}_{2n})^{\tau}Y_d^* - \tilde{\eta}_{d+1}\sqrt{\widehat{\alpha}_{20}^{\tau}X_d^*}\right\}^2}{\widehat{\alpha}_{1n}^{\tau}X_d^*} - \tilde{\eta}_{d+1}^2,$$

with $Y_d^* = (1, Y_d^*, ..., y_{d+1-p}^*)^{\tau}$, $X_d^* = (1, Y_d^{*2}, ..., y_{d+1-p}^{*2})^{\tau}$ *and*

$$Y_j^* = \begin{cases} \tilde{y}_j, & \textit{if } j \geq 2, \\ \widehat{r}_n, & \textit{if } j = 1, \\ Y_{i+j}, & \textit{if } j \leq 0. \end{cases}$$

Step C.3. *Sample* U *from the density function* $\mathbb{P}_{\mathfrak{X}}(U = i) = \widehat{\pi}(\widehat{r}_n|z_i)/[\sum_{l=p}^{n}\widehat{\pi}(\widehat{r}_n|z_l)]$ *for* $i = p, ..., n$, *independent of all* $\{\tilde{\varepsilon}_2, ..., \tilde{\varepsilon}_{d+1}\}$ *given* $\mathfrak{X}$.
Step C.4. *Obtain* $\widetilde{\xi}_1^{(k)} = \widetilde{\eta}_{1k}^{(U)}$.

Denote $\widehat{M}_n$ obtained by Algorithms A and C as an approximation of M_-. Then, we can show that

$$\lim_{n\to\infty} |\mathbb{P}_{\mathfrak{X}}(\widehat{M}_n \leq x) - \mathbb{P}(M_- \leq x)| = 0$$

at each x for which $\mathbb{P}(M_- = x) = 0$.

To assess the performance of the estimator in finite samples, we consider the TDAR(1) model:

$$Y_t = \begin{cases} 1 - 0.6Y_{t-1} + \varepsilon_t\sqrt{1+0.5Y_{t-1}^2}, & \text{if } Y_{t-1} \leq 0, \\ -1 - 0.2Y_{t-1} + \varepsilon_t\sqrt{0.5+0.3Y_{t-1}^2}, & \text{if } Y_{t-1} > 0. \end{cases} \tag{7.6}$$

Table 7.1 reports the empirical mean, the empirical standard deviation and the asymptotic standard deviation when $\eta_t \sim N(0,1)$ with the sample sizes n and replications 1000. Table7.2 reports the empirical quantiles of M_- at the significance level α when η_t is standard normal, t_5- and double exponential distribution, respectively. Based on the critical values in Table 7.2, Table 7.3 reports the coverage probabilities of r_0.

Figure 1 displays the density of $n(\widehat{r}_n - r_0)$ with sample sizes $n = 400$.

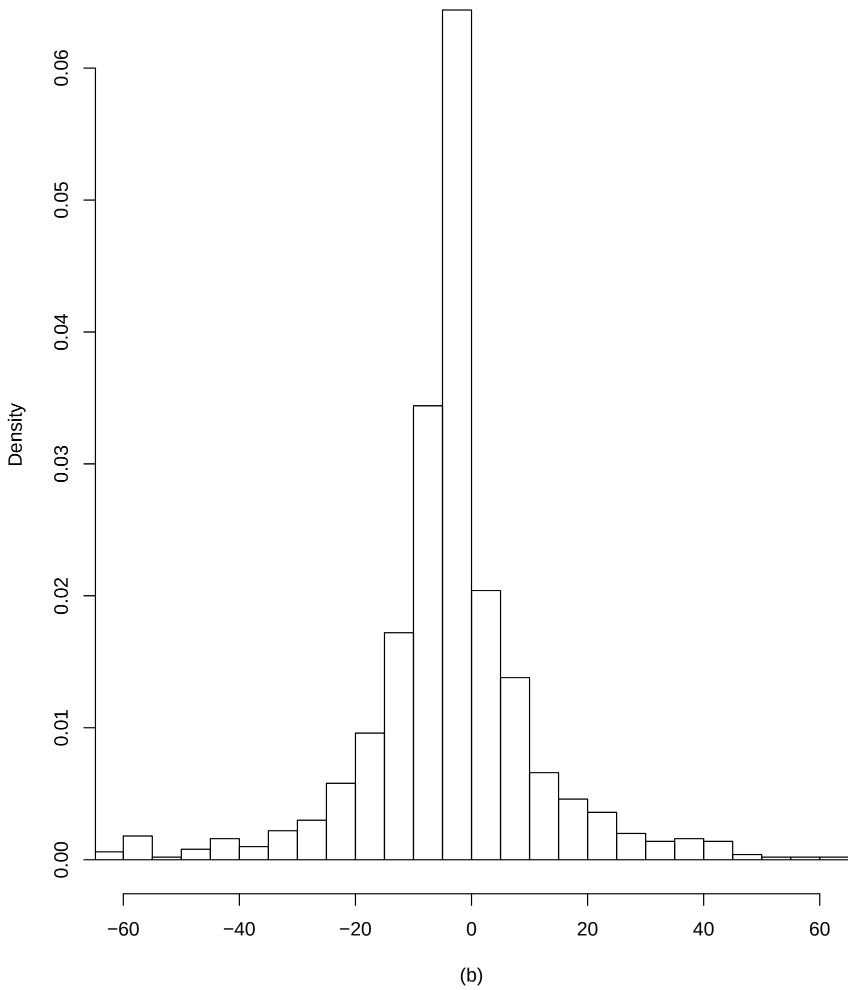


Fig. 7.2 The density of $n(\widehat{r}_n - r_0)$ when $\eta_t \sim N(0,1)$ and $n = 400$.

Table 7.1 Simulation results for model (7.6) with $\theta_0 = (1, -0.6, 1, 0.5, -1, -0.2, 0.5, 0.3, 0)'$ when ε_t is standard normal.

n		ϕ_{10}	ϕ_{11}	α_{10}	α_{11}	ϕ_{20}	ϕ_{21}	α_{20}	α_{21}	r
	EM	1.0477	-0.5741	0.8650	0.4786	-1.0173	-0.1935	0.4180	0.2923	-0.0528
100	ESD	0.3542	0.2547	0.4112	0.2148	0.2555	0.1632	0.2288	0.1082	0.1242
	ASD	0.3203	0.2363	0.3965	0.2116	0.2361	0.1550	0.2182	0.1029	0.1012
	EM	1.0253	-0.5851	0.9398	0.4865	-1.0050	-0.1983	0.4596	0.2939	-0.0250
200	ESD	0.2337	0.1664	0.2931	0.1547	0.1692	0.1086	0.1579	0.0749	0.0548
	ASD	0.2239	0.1670	0.2768	0.1501	0.1639	0.1088	0.1511	0.0725	0.0506
	EM	1.0227	-0.5909	0.9734	0.4988	-1.0135	-0.1970	0.4861	0.2971	-0.0127
400	ESD	0.1605	0.1182	0.1977	0.1069	0.1132	0.0771	0.1088	0.0506	0.0256
	ASD	0.1575	0.1171	0.1951	0.1051	0.1152	0.0764	0.1064	0.0510	0.0253
	EM	1.0042	-0.6006	0.9973	0.4926	-1.0026	-0.1996	0.4946	0.2971	-0.0061
800	ESD	0.1080	0.0811	0.1391	0.0750	0.0830	0.0540	0.0778	0.0377	0.0140
	ASD	0.1110	0.0825	0.1376	0.0741	0.0813	0.0539	0.0751	0.0360	0.0127

EM, empirical mean; ESD, empirical standard deviation; ASD, asymptotic standard deviation.

Table 7.2 Empirical quantiles of M_- (10,000 replications).

α	0.5%	1%	2.5%	5%	95%	97.5%	99%	99.5%
$N(0,1)$	-45.02	-38.20	-30.38	-24.25	5.77	12.50	21.54	28.81
ST(5)	-52.47	-46.91	-37.16	-29.66	8.75	19.23	33.56	46.25
Dexp	-65.14	-56.61	-44.80	-34.44	11.86	22.93	37.78	51.14

$ST(5)$–t_5-distribution and Dexp– double exponential distribution.

Table 7.3 Coverage probabilities.

ε_t	α	100	200	400	800
	0.01	0.979	0.986	0.989	0.984
$N(0,1)$	0.05	0.932	0.940	0.944	0.946
	0.10	0.880	0.893	0.900	0.887
	0.01	0.970	0.980	0.984	0.987
ST(5)	0.05	0.906	0.925	0.934	0.949
	0.10	0.859	0.871	0.884	0.886
	0.01	0.970	0.969	0.987	0.991
Dexp	0.05	0.919	0.922	0.942	0.945
	0.10	0.845	0.878	0.886	0.892

7.4 LSE of Multiple Threshold AR Models

This section considers the m-regime TAR model$(m \geq 2)$ with order p if it satisfies the equation

$$Y_t = \sum_{j=1}^{m} (\mathbf{Y}'_{t-1}\boldsymbol{\beta}_j + \sigma_j \varepsilon_t) I(r_{j-1} < Y_{t-d} \leq r_j), \tag{7.7}$$

where $\mathbf{Y}_{t-1} = (1, Y_{t-1}, ..., Y_{t-p})'$, $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, ..., \beta_{jp})' \in \mathbb{R}^{p+1}$, $\sigma_j > 0, j = 1, ..., m$; $-\infty = r_0 < r_1 < ... < r_m = \infty$ and $I(\cdot)$ is an indicator function. The number m of regimes and the order p of model (7.7) are positive integers. d is a positive integer called the delay parameter. $\{r_1, ..., r_{m-1}\}$ are threshold parameters. The errors

$\{\varepsilon_t\}$ are iid r.v.s with zero mean and unit variance, and ε_t is independent of the past information $\{Y_{t-j} : j \geq 1\}$.

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{r}', d)' = (\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_m', \mathbf{r}', d)' \in \mathbb{R}^{m(p+1)+(m-1)} \times \{1, ..., D_0\}$ and $\mathbf{r} = (r_1, ..., r_{m-1})' \in \mathbb{R}^{m-1}$, where D_0 is a known positive integer. Suppose that a sample $\{Y_1, ..., Y_n\}$ is from model (7.7) with true value $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_{10}', ..., \boldsymbol{\beta}_{m0}', \mathbf{r}_0', d_0)'$. Given the initial values $\{Y_0, ..., Y_{1-p}\}$, the sum of square errors function $L_n(\boldsymbol{\theta})$ is defined as

$$L_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} [Y_t - \mathbb{E}_{\boldsymbol{\theta}}(Y_t | \mathscr{F}_{t-1})]^2,$$

where $\mathscr{F}_t$ is the σ-algebra generated by $\{Y_{1-p}, ..., Y_t\}$. The minimizer $\hat{\boldsymbol{\theta}}_n$ of $L_n(\boldsymbol{\theta})$ is called a LSE of $\boldsymbol{\theta}_0$, that is,

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}).$$

Since $L_n(\boldsymbol{\theta})$ is discontinuous in $\mathbf{r}$ and d, a multi-parameter grid-search algorithm is needed. The way to obtain $\hat{\boldsymbol{\theta}}_n$ is as follows.

- Fix $\mathbf{r} \in \mathbb{R}^{m-1}$ and $d \in \{1, ..., D_0\}$, then minimize $L_n(\boldsymbol{\theta})$ and get its minimizer $\hat{\boldsymbol{\beta}}_n(\mathbf{r}, d)$ and minimum $L_n^*(\mathbf{r}, d) \equiv L_n(\boldsymbol{\theta})|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_n(\mathbf{r},d)}$.
- Since $L_n^*(\mathbf{r}, d)$ only takes finite possible values, one can get the minimizer $(\hat{\mathbf{r}}_n', \hat{d}_n)'$ of $L_n^*(\mathbf{r}, d)$ by the enumeration approach.
- Use a plug-in method, one can finally get $\hat{\boldsymbol{\beta}}_n \equiv \hat{\boldsymbol{\beta}}_n(\hat{\mathbf{r}}_n, \hat{d}_n)$ and $\hat{\boldsymbol{\theta}}_n$.

Generally, $\hat{\mathbf{r}}_n$ is taken as the form $(Y_{(i_1)}, ..., Y_{(i_{m-1})})'$, where $i_1 < \cdots < i_{m-1}$ and $\{Y_{(1)}, ..., Y_{(n)}\}$ is the order statistics of the sample $\{Y_1, ..., Y_n\}$. If $(Y_{(j_1)}, ..., Y_{(j_{m-1})})'$ is an estimator of $\mathbf{r}_0$, then $L_n^*(\mathbf{r}, \hat{d}_n)$ is a constant over the $(m-1)$-dimensional cube $\mathscr{A}$, where

$$\mathscr{A} = \{\mathbf{r} = (r_1, ..., r_{m-1})' : r_i \in [Y_{(j_i)}, Y_{(j_i+1)}), i = 1, ..., m-1\}.$$

Thus, there exist infinitely many $\mathbf{r}$ such that $L_n(\cdot)$ can achieves its global minimum and each $\mathbf{r} \in \mathscr{A}$ can be considered as an estimator of $\mathbf{r}_0$. In this case, we choose $(Y_{(j_1)}, ..., Y_{(j_{m-1})})'$ as a representative of $\mathscr{A}$ and denote it as the estimator of $\mathbf{r}_0$. According to the procedure for obtaining $\hat{\boldsymbol{\theta}}_n$, it is not hard to show that $\hat{\boldsymbol{\theta}}_n$ is the LSE of $\boldsymbol{\theta}_0$.

Let σ_{j0} be the true value of σ_j for $j = 1, ..., m$. Once $\hat{\boldsymbol{\theta}}_n$ is obtained, we then can consistently estimate σ_{j0}^2 by

$$\hat{\sigma}_{jn}^2 = \frac{1}{n_j} \sum_{t=1}^{n} (Y_t - \mathbf{Y}_{t-1}' \hat{\boldsymbol{\beta}}_{jn})^2 I(\hat{r}_{j-1,n} < Y_{t-\hat{d}_n} \leq \hat{r}_{jn}), \tag{7.8}$$

where $n_j = \sum_{t=1}^{n} I(\hat{r}_{j-1,n} < Y_{t-\hat{d}_n} \le \hat{r}_{jn})$.

In order to get the global minimum of $L_n(\cdot)$ with m regimes and sample size n, the required number of calculations is $O(n^{m-1}/(m-1)!)$. When m is large, however, the computational burden becomes substantial, requiring multi-parameter grid-based search over all possible values of all threshold parameters taken together, and hence this algorithm is very time-consuming. For a fixed m, the consumed time soars at an exponential rate as the sample size n increases. This problem is similar to the computational problem arising from multiple change-point models investigated by Bai and Perron (2003, 2006). Tsay (1989) transforms model (7.7) into a change-point model and use the rearranged technique to localize possible positions of threshold parameters. Similarly, using the same rearranged technique, Coakley, Fuertes and Pérez (2003) provides an efficient estimation approach which relies on the computational advantages of QR factorizations of matrices. When m is small, the grid-based search algorithm is an easy way to obtain the global minimum of $L_n(\cdot)$. Throughout this section, we assume that both m and p are known.

Let $\Theta \times \{1,...,D_0\}$ be the parameter space, where $\Theta = \Theta_\beta \times \Theta_r$ is a compact subset of $\mathbb{R}^{m(p+1)} \times \mathscr{R}^{m-1}$ and $\mathscr{R}^{m-1} = \{(r_1,...,r_{m-1}) : -\infty < r_1 < ... < r_{m-1} < \infty\}$. The following result states the strong consistency of $\hat{\boldsymbol{\theta}}_n$.

Theorem 7.4. *Suppose that (i) $\{Y_t\}$ satisfying (7.7) is strictly stationary and ergodic, having finite second moments, (ii) $\boldsymbol{\beta}_{j0} \neq \boldsymbol{\beta}_{j+1,0}$ for $j = 1,...,m-1$, and (iii) ε_1 admits a bounded, continuous and positive density $f_\varepsilon(x)$ on $\mathbb{R}$ with zero mean and unit variance. Then, $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s. as $n \to \infty$ and so are $\hat{\sigma}^2_{jn}$'s in (7.8).*

The condition (ii) in Theorem 7.4 is required to guarantee the identification of $\mathbf{r}_0$. The strong consistency of $\hat{\boldsymbol{\theta}}_n$ holds regardless if the autoregressive function is continuous over all thresholds or not. From Theorem 7.4, we know that $\hat{d}_n$ equals d_0 eventually. Thus, without loss of generality, we assume that the delay parameter d_0 is known for the remainder of this paper and it is deleted from $\boldsymbol{\theta}_0$, i.e., $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0', \mathbf{r}_0')'$, and so is $\hat{d}_n$ from $\hat{\boldsymbol{\theta}}_n$. The parameter space becomes Θ, accordingly, and we write d for d_0 in what follows.

To obtain the convergence rate of $\hat{\mathbf{r}}_n$, the asymptotic normality of $\hat{\boldsymbol{\beta}}_n$ and the limiting distribution of $n(\hat{\mathbf{r}}_n - \mathbf{r}_0)$, we first give four assumptions as follows.

Assumption 7.5 *$\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with $\mathbb{E}\varepsilon_1 = 0$, $\mathbb{E}\varepsilon_1^2 = 1$ and $\mathbb{E}\varepsilon_t^4 < \infty$. ε_1 has a bounded, continuous and positive density $f_\varepsilon(x)$ on $\mathbb{R}$.*

Assumption 7.6 *$\{Y_t\}$ is strictly stationary with $\mathbb{E}Y_t^4 < \infty$.*

Let $\mathbf{Z}_t = (Y_t,...,Y_{t-(p\vee d)+1})'$, where $p \vee d = \max(p,d)$. Then $\{\mathbf{Z}_t\}$ is a Markov chain. Denote its l-step transition probability by $\mathbf{P}^l(\mathbf{z},A)$, where $\mathbf{z} \in \mathbb{R}^{p\vee d}$ and A is a Borel set of $\mathbb{R}^{p\vee d}$.

Assumption 7.7 *$\{\mathbf{Z}_t\}$ admits a unique invariant measure $\Pi(\cdot)$ such that there exist $K > 0$ and $\rho \in [0,1)$, for any $\mathbf{z} \in \mathbb{R}^{p\vee d}$ and any n, $\|\mathbf{P}^n(\mathbf{z},\cdot) - \Pi(\cdot)\|_v \le K(1 + \|\mathbf{z}\|)\rho^n$, where $\|\cdot\|_v$ and $\|\cdot\|$ denote the total variation norm and the Euclidean norm, respectively.*

Under Assumption 7.7, $\{\mathbf{Z}_t\}$ is V-uniformly ergodic with $V(\mathbf{z}) = K(1 + \|\mathbf{z}\|)$, which is stronger than geometric ergodicity. For the concept of V-uniform ergodicity, see Meyn and Tweedie (1993). If Assumption 7.5 holds and $\max_{1\le i\le m}\sum_{j=1}^{p}|\beta_{ij}| < 1$, then Assumption 7.7 holds and $\mathbb{E}Y_t^4 < \infty$, see Chan (1989) and Chan and Tong (1985). If the initial value $\mathbf{Z}_0$ is from the distribution $\Pi(\cdot)$, then Assumption 7.7 implies that $\{Y_t\}$ is strictly stationary.

Assumption 7.8 *There exist nonrandom vectors* $\mathbf{w}_i^* = (1, w_{i1}, ..., w_{ip})'$ *with* $w_{id} = r_{i0}$ *such that* $(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{i+1,0})'\mathbf{w}_i^* \neq 0$ *for* $i = 1, ..., m-1$.

In Assumption 7.8, w_{id} may not be a component of $\mathbf{w}_i^*$ if $d > p$. In this case, Assumption 7.8 is equivalent to the conditions $\|\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{i+1,0}\| > 0$ for $i = 1, ..., m-1$. The latter is necessary and sufficient for the identification of all thresholds. When $p = d = 1$, Assumption 7.8 implies that the autoregressive mean function is discontinuous at all thresholds $\{r_1, ..., r_{m-1}\}$. Assumption 7.8 in the general case implies that $\|\mathbf{Y}_{t-1}'(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{i+1,0})\|$ is bigger than a positive constant with a positive probability and plays a key role in obtaining the n-convergence rate of $\hat{\mathbf{r}}_n$ and its limiting distribution.

Theorem 7.5. *If Assumptions 7.5-7.8 hold, then*

$$\text{(i).}\quad n\|\hat{\mathbf{r}}_n - \mathbf{r}_0\| = O_p(1);$$

$$\text{(ii).}\quad \sqrt{n}\sup_{\|\mathbf{r}-\mathbf{r}_0\|<B/n}\|\hat{\boldsymbol{\beta}}_n(\mathbf{r}) - \hat{\boldsymbol{\beta}}_n(\mathbf{r}_0)\| = o_p(1)\quad \text{for any fixed } B \in (0, +\infty).$$

Furthermore,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \sqrt{n}(\hat{\boldsymbol{\beta}}_n(\mathbf{r}_0) - \boldsymbol{\beta}_0) + o_p(1) \to_d \mathcal{N}(0, \Sigma) \quad \text{as } n \to \infty,$$

where '$\to_d$*' denotes convergence in distribution and* $\Sigma = \mathrm{diag}(\sigma_{10}^2\Sigma_1, ..., \sigma_{m0}^2\Sigma_m)$ *with*

$$\Sigma_j^{-1} = \mathbb{E}[\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}' I(r_{j-1,0} < Y_{t-d} \le r_{j0})], \quad j = 1, ..., m.$$

From Theorem 7.5(i), we know that the convergence rate of $\hat{\mathbf{r}}_n$ is n. To study the limiting distribution of $n(\hat{\mathbf{r}}_n - \mathbf{r}_0)$, we consider the following profile sum of squares errors function:

$$\widetilde{L}_n(\mathbf{s}) = L_n\Big(\hat{\boldsymbol{\beta}}_n(\mathbf{r}_0 + \frac{\mathbf{s}}{n}), \mathbf{r}_0 + \frac{\mathbf{s}}{n}\Big) - L_n\Big(\hat{\boldsymbol{\beta}}_n(\mathbf{r}_0), \mathbf{r}_0\Big), \quad \mathbf{s} \in \mathbb{R}^{m-1}. \tag{7.9}$$

Using Theorem 7.5 and Taylor's expansion, we can show that $\widetilde{L}_n(\mathbf{s})$ can be approximated in the function space $\mathbb{D}(\mathbb{R}^{m-1})$ (defined in the proof of Theorem 7.6) by

$$
\begin{aligned}
\wp_n(\mathbf{s}) &= L_n(\boldsymbol{\beta}_0, \mathbf{r}_0 + \frac{\mathbf{s}}{n}) - L_n(\boldsymbol{\beta}_0, \mathbf{r}_0) \\
&= \sum_{i=1}^{m-1} \sum_{t=1}^{n} \Big[\big\{ [\mathbf{Y}_{t-1}'(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{i+1,0})]^2 + 2\sigma_{i0}\varepsilon_t [\mathbf{Y}_{t-1}'(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{i+1,0})] \big\} \\
&\qquad\qquad \times I\Big(r_{i0} + \frac{s_i}{n} < Y_{t-d} \le r_{i0} \Big) I(s_i < 0) \\
&\qquad + \big\{ [\mathbf{Y}_{t-1}'(\boldsymbol{\beta}_{i+1,0} - \boldsymbol{\beta}_{i0})]^2 + 2\sigma_{i+1,0}\varepsilon_t [\mathbf{Y}_{t-1}'(\boldsymbol{\beta}_{i+1,0} - \boldsymbol{\beta}_{i0})] \big\} \\
&\qquad\qquad \times I\Big(r_{i0} < Y_{t-d} \le r_{i0} + \frac{s_i}{n} \Big) I(s_i \ge 0) \Big] \\
&= \sum_{i=1}^{m-1} \sum_{t=1}^{n} \Big[\xi_t^{(i,i+1)} I\Big(r_{i0} + \frac{s_i}{n} < Y_{t-d} \le r_{i0} \Big) I(s_i < 0) \\
&\qquad + \xi_t^{(i+1,i)} I\Big(r_{i0} < Y_{t-d} \le r_{i0} + \frac{s_i}{n} \Big) I(s_i \ge 0) \Big],
\end{aligned}
$$

where

$$
\xi_t^{(i,j)} = \big[\mathbf{Y}_{t-1}'(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{j0}) \big]^2 + 2\sigma_{i0}\varepsilon_t \mathbf{Y}_{t-1}'(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_{j0}), \quad i, j = 1, ..., m. \tag{7.10}
$$

Let $F_{(i,j)}(\cdot|r)$ be the conditional distribution function of $\xi_{d+1}^{(i,j)}$ given $Y_1 = r$. We first define $(m-1)$ independent one-dimensional two-sided compound Poisson processes $\{\mathscr{P}_j(z), z \in \mathbb{R}\}$ as

$$
\mathscr{P}_j(z) = I(z < 0) \sum_{k=1}^{N_1^{(j)}(-z)} Y_k^{(j,j+1)} + I(z \ge 0) \sum_{k=1}^{N_2^{(j)}(z)} Z_k^{(j+1,j)}, \quad z \in \mathbb{R} \tag{7.11}
$$

for $j = 1, ..., m-1$, where $\{N_1^{(j)}(z), z \ge 0\}$ and $\{N_2^{(j)}(z), z \ge 0\}$ are two independent Poisson processes with $N_1^{(j)}(0) = N_2^{(j)}(0) = 0$ a.s. and with the same jump rate $\pi(r_{j0})$, where $\pi(\cdot)$ is the density function of Y_1. $\{Y_k^{(j,j+1)} : k \ge 1\}$ are i.i.d. random variables with the distribution $F_{(j,j+1)}(\cdot|r_{j0})$, and $\{Z_k^{(j+1,j)} : k \ge 1\}$ are i.i.d. random variables with the distribution $F_{(j+1,j)}(\cdot|r_{j0})$. $\{Y_k^{(j,j+1)} : k \ge 1\}$ and $\{Z_k^{(j+1,j)} : k \ge 1\}$ are mutually independent. Here, we work with the left continuous version for $N_1^{(j)}(\cdot)$ and the right continuous version for $N_2^{(j)}(\cdot)$ for $j = 1, ..., m-1$.

We further define a spatial compound Poisson process $\wp(\mathbf{s})$ as follows,

$$
\wp(\mathbf{s}) = \sum_{j=1}^{m-1} \mathscr{P}_j(s_j), \quad \mathbf{s} = (s_1, ..., s_{m-1})' \in \mathbb{R}^{m-1}. \tag{7.12}
$$

Clearly, $\wp(\mathbf{s})$ goes to $+\infty$ a.s. when $\|\mathbf{s}\| \to \infty$ since $\mathbb{E}Y_1^{(i,i+1)} = \mathbb{E}Z_1^{(i+1,i)} > 0$ by Assumption 7.8 for $i = 1, ..., m-1$. Therefore, there exists a unique random $(m-1)$-dimensional cube $[\mathbf{M}_-, \mathbf{M}_+) \equiv [M_-^{(1)}, M_+^{(1)}) \times \cdots \times [M_-^{(m-1)}, M_+^{(m-1)})$ on which the process $\wp(\mathbf{s})$ attains its global minimum a.s. That is,

$$[\mathbf{M}_-, \mathbf{M}_+) = \arg \min_{\mathbf{s} \in \mathbb{R}^{m-1}} \wp(\mathbf{s}).$$

From (7.12), the minimization above is equivalent to

$$[M_-^{(j)}, M_+^{(j)}) = \arg\min_{z \in \mathbb{R}} \mathscr{P}_j(z), \quad j = 1, ..., m-1.$$

Note that the processes $\{\mathscr{P}_j(z) : j = 1, ..., m-1\}$ are independent, so are $\{M_-^{(j)} : j = 1, ..., m-1\}$. Now, we can state our another result as follows.

Theorem 7.6. *If Assumptions 7.5-7.8 hold, then $n(\hat{\mathbf{r}}_n - \mathbf{r}_0)$ converges weakly to $\mathbf{M}_-$ and its components are asymptotically independent as $n \to \infty$. Furthermore, $n(\hat{\mathbf{r}}_n - \mathbf{r}_0)$ is asymptotically independent of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ which is always asymptotically normal.*

When $m = 2$, Theorem 7.6 reduces to Theorem 2.2 of Chan (1993). The limit distribution of M_- does not have a closed form and depends on the nuisance parameters and the distribution of ε_t. In next section, we will describe how to simulate $\mathbf{M}_-$ via a numerical approach.

When m is known and p_i is unknown, we can use the AIC below to determine the order in each regime,

$$\text{AIC}(\{p_i\}) = \sum_{j=1}^{m} [n_j \log(\hat{\sigma}_{jn}^2) + 2(p_j + 1)], \tag{7.13}$$

where p_j is the order in the jth regime. See Tsay (1998). More information criteria as model selection tools for nonlinear threshold models, see Kapetanios (2001), in which the author established the consistency of lag selection and compared the small sample performance among different criteria. For the choice of m, Gonzalo and Pitarakis (2002) proposed a sequential model selection approach and considered its weak consistency under some conditions for model (7.7) with all σ_i's being equal. For general threshold models, it seems that the literature does not offer any formal methodology for selecting the number of regimes. More work should be needed in the future.

7.5 Proofs of Theorems

Proof (of Theorem 7.1). Let $\beta(\theta) = E\{\varphi_t(\theta) - \varphi_t(\theta_0)\}$. For any given open neighborhood V of $\theta_0 \in \Theta$ and any $\theta \in V^c \cap \Theta$, a conditional argument yields that

$$-2\beta(\theta) = E\{K_{1t} I(y_{t-d} \le r_0) + K_{2t} I(r_0 < y_{t-d} \le r) + K_{3t} I(y_{t-d} > r)\},$$

where

$$K_{1t} = \log\frac{\alpha_1^{\mathrm{T}}X_{t-1}}{\alpha_{10}^{\mathrm{T}}X_{t-1}} + \frac{\alpha_{10}^{\mathrm{T}}X_{t-1}}{\alpha_1^{\mathrm{T}}X_{t-1}} - 1 + \frac{\{(\phi_{10}-\phi_1)^{\mathrm{T}}Y_{t-1}\}^2}{\alpha_1^{\mathrm{T}}X_{t-1}},$$
$$K_{2t} = \log\frac{\alpha_1^{\mathrm{T}}X_{t-1}}{\alpha_{20}^{\mathrm{T}}X_{t-1}} + \frac{\alpha_{20}^{\mathrm{T}}X_{t-1}}{\alpha_1^{\mathrm{T}}X_{t-1}} - 1 + \frac{\{(\phi_{20}-\phi_1)^{\mathrm{T}}Y_{t-1}\}^2}{\alpha_1^{\mathrm{T}}X_{t-1}},$$
$$K_{3t} = \log\frac{\alpha_2^{\mathrm{T}}X_{t-1}}{\alpha_{20}^{\mathrm{T}}X_{t-1}} + \frac{\alpha_{20}^{\mathrm{T}}X_{t-1}}{\alpha_2^{\mathrm{T}}X_{t-1}} - 1 + \frac{\{(\phi_{20}-\phi_2)^{\mathrm{T}}Y_{t-1}\}^2}{\alpha_2^{\mathrm{T}}X_{t-1}}.$$

Observe that all $K_{it} \geq 0$ almost surely by the elementary inequality $\log(1/x)+x-1>0$ for $x>0$ unless $x=1$. Hence, $\beta(\theta)<0$. The remainder is similar to that of Theorem 1 in Li et al. (2010).

Proof (of Theorem 7.2). (i). We only prove the case $p=1$. When $p>1$, using the technique in Chan (1993) (last paragraph in page 529), the proof would go through with a minor modification. Since $\widehat{\theta}_n$ is strongly consistent, we restrict the parameter space to a neighborhood $V_\delta = \{\theta \in \Theta : \|\lambda-\lambda_0\|<\delta,\ |r-r_0|<\delta\}$ of θ_0 for some $0<\delta<1$ to be determined later. Then, it suffices to prove that there exist constants $B>0$ and $\gamma>0$ such that, for any $\varepsilon>0$,

$$\mathrm{pr}\Big(\sup_{\substack{B/n<|r-r_0|\leq\delta\\ \theta\in V_\delta}} \frac{L_n(\lambda,r)-L_n(\lambda,r_0)}{nG(|r-r_0|)} < -\gamma\Big) > 1-\varepsilon, \tag{7.14}$$

as n is large enough, where $G(u)=\mathrm{pr}(r_0<y_0\leq r_0+u)$. Writing $r=r_0+u$ for some $u\geq 0$. By a calculation, it follows that

$$\begin{aligned}\frac{2\{L_n(\lambda,r)-L_n(\lambda,r_0)\}}{nG(u)} &= \frac{-1}{nG(u)}\sum_{t=1}^{n}\zeta_{2t}I(r_0<y_{t-1}\leq r_0+u)+O_p(\sqrt{\delta})\\ &= -K_4\frac{G_n(u)}{G(u)}+K_5\frac{\sum_{t=1}^{n}\varepsilon_t I(r_0<y_{t-1}\leq r_0+u)}{nG(u)}\\ &\quad +K_6\frac{\sum_{t=1}^{n}(\varepsilon_t^2-1)I(r_0<y_{t-1}\leq r_0+u)}{nG(u)}+O_p(\sqrt{\delta}),\end{aligned}$$

where $G_n(u)=\sum_{t=1}^{n}I(r_0<y_{t-1}\leq r_0+u)/n$,

$$K_4 = \log\frac{\alpha_{10}^{\mathrm{T}}X}{\alpha_{20}^{\mathrm{T}}X}+\frac{\alpha_{20}^{\mathrm{T}}X}{\alpha_{10}^{\mathrm{T}}X}-1+\frac{\{(\phi_{20}-\phi_{10})^{\mathrm{T}}Y\}^2}{\alpha_{10}^{\mathrm{T}}X},$$
$$K_5 = \frac{2\{(\phi_{10}-\phi_{20})^{\mathrm{T}}Y\}\sqrt{\alpha_{20}^{\mathrm{T}}X}}{\alpha_{10}^{\mathrm{T}}X} \quad \text{and} \quad K_6 = \frac{(\alpha_{10}-\alpha_{20})^{\mathrm{T}}X}{\alpha_{10}^{\mathrm{T}}X}$$

with $Y=(1,r_0)^{\mathrm{T}}$ and $X=(1,r_0^2)^{\mathrm{T}}$. Similar to Claim 2 in Chan (1993), for any $\varepsilon>0$ and $\eta>0$, there exists a positive constant B such that as n is large enough

$$\mathrm{pr}\Big(\sup_{B/n<u\le\delta}\Big|\frac{G_n(u)}{G(u)}-1\Big|<\eta\Big)>1-\varepsilon,$$

$$\mathrm{pr}\Big(\sup_{B/n<u\le\delta}\Big|\frac{\sum_{t=1}^{n}\varepsilon_t I(r_0<y_{t-1}\le r_0+u)}{nG(u)}\Big|<\eta\Big)>1-\varepsilon,$$

$$\mathrm{pr}\Big(\sup_{B/n<u\le\delta}\Big|\frac{\sum_{t=1}^{n}(\varepsilon_t^2-1)I(r_0<y_{t-1}\le r_0+u)}{nG(u)}\Big|<\eta\Big)>1-\varepsilon.$$

Note that $K_4>0$ by Assumption 7.4. Choosing δ small enough and $\gamma=K_4/4$, (7.14) holds and so does (i).

The proof of (ii) is trivial and hence it is omitted.

Proof (of Theorem 7.3). Without loss of generality, we assume that ζ_{it}, defined in (7.3), is bounded. Otherwise, we can truncate it using the technique in Li et al. (2010) and consider a new process made up of the truncated random variables. Consider the weak convergence of the process $\wp_n(z)$ on the interval $[0,T]$. The tightness of $\wp_n(z)$ can be easily shown by Theorem 5 in Kushner (1984, page 32). The key step is to describe convergence of finite dimensional distributions. To do this, for any $0\le z_1\le z_2<z_3\le z_4\le T$ and for any constants c_1 and c_2, the linear combination of the increments of $\wp_n(z)$ is

$$S_n\equiv c_1\{\wp_n(z_2)-\wp_n(z_1)\}+c_2\{\wp_n(z_4)-\wp_n(z_3)\}=\sum_{t=1}^{n}J_t^{\varepsilon},$$

where $J_t^{\varepsilon}=\zeta_{2t}\{c_1I_t(z_1,z_2)+c_2I_t(z_3,z_4)\}$, $\varepsilon=1/n$ and $I_t(u,v)=I(r_0+u\varepsilon<y_{t-1}\le r_0+v\varepsilon)$. We first verify Assumptions A.1-A.3 in Li et al. (2010) for J_t^{ε}. By Assumption 7.3, it follows that

$$\lim_{n\to\infty}\varepsilon^{-1}P_k^{\varepsilon}(J_n^{\varepsilon}\ne 0)=\pi(r_0)\{(z_2-z_1)+(z_4-z_3)\}. \tag{7.15}$$

By Assumption 7.3 again, for any Borel set B, it follows that

$$Q^*(B)=\lim_{n\to\infty}\mathrm{pr}(J_n^{\varepsilon}\in B|J_n^{\varepsilon}\ne 0)=wQ_1^*(B)+(1-w)Q_2^*(B), \tag{7.16}$$

where $w=(z_2-z_1)/\{(z_2-z_1)+(z_4-z_3)\}$ and $Q_i^*(B)=\mathrm{pr}(c_i\zeta_{2t}\in B)$, $i=1,2$. By a conditional argument, for any $f\in\widehat{C}_0^2$, a space of functions with compact support and continuous second derivative, and a scalar x,

$$\begin{aligned}E_k^{\varepsilon}\{f(x+J_n^{\varepsilon})-f(x)|J_n^{\varepsilon}\ne 0\}&=E\{f(x+J_n^{\varepsilon})-f(x)|J_n^{\varepsilon}\ne 0\}\\&\to\int\{f(x+u)-f(x)\}Q^*(du),\end{aligned} \tag{7.17}$$

as $n\to\infty$. By (7.15)-(7.17), Assumptions A.1-A.3 in Li et al. (2010) hold. Furthermore, by their Theorem A.1, we have that S_n converges weakly to a compound Poisson random variable J with jump rate $\pi(r_0)\{(z_2-z_1)+(z_4-z_3)\}$ and the jump distribution Q^*. The characteristic function $f_J(t)$ of J is equal to that of

$c_1\{\wp(z_2)-\wp(z_1)\}+c_2\{\wp(z_4)-\wp(z_3)\}$, where

$$\wp(z)=\sum_{i=1}^{N(z)}\xi_i^{(2)}, \quad z\in[0,\infty),$$

and $\{N(z), z\in[0,\infty)\}$ is a Poisson process with jump rate $\pi(r_0)$ and $\{\xi_i^{(2)}\}$ is independent random variables with the same distribution $F_2(\cdot|r_0)$. Thus, $\widetilde{L}_n(z)$, defined in (7.2), converges weakly to $\wp(z)$ as $n\to\infty$. The remainder of the proof is similar to that of Theorem 2 in Chan (1993).

Proof (of Theorem **??***).* From Algorithm C, we have

$$\mathrm{pr}_{\mathfrak{X}}(\widetilde{\xi}_1^{(k)}\le x)=\sum_{i=1}^{K}\mathrm{pr}_{\mathfrak{X}}(\tilde{\zeta}_{k,d+1}\le x|y_1=\widehat{r}_n, Z_0=z_i)\frac{\widehat{\pi}(\widehat{r}_n|z_i)}{\sum_{l=1}^{K}\widehat{\pi}(\widehat{r}_n|z_l)}.$$

Since $h(x)$ is uniformly continuous, $\sup_{x\in\mathbb{R}}|\widehat{h}(x)-h(x)|=o_p(1)$, see Silverman (1978). Thus, $\widehat{\pi}(\widehat{r}_n|z_i)=\pi(r_0|z_i)+o_p(1)$. By Theorem 16 in Pollard (1984, page 134) and Lemma A.3 (on the continuity of the smallest argmax functional) in Seijo and Sen (2010), it suffices to prove

$$\mathrm{pr}_{\mathfrak{X}}(\tilde{\zeta}_{k,d+1}\le x|y_1=\widehat{r}_n, Z_0=z_i)=\mathrm{pr}(\zeta_{k,d+1}\le x|y_1=r_0, Z_0=z_i)+o_p(1). \quad (7.18)$$

First of all, let $\hat{z}_i=(\widehat{r}_n, y_i, ..., y_{i-p+2})^{\mathrm{T}}$ and $\tilde{z}_i=(r_0, y_i, ..., y_{i-p+2})^{\mathrm{T}}$. By a simple calculation, we have

$$\begin{aligned}&\sup_{x\in\mathbb{R}}|\mathrm{pr}_{\mathfrak{X}}(\tilde{y}_2\le x|y_1=\widehat{r}_n, Z_0=z_i)-\mathrm{pr}(y_2\le x|y_1=r_0, Z_0=z_i)|\\ &\le\sup_{x\in\mathbb{R}}|\widehat{H}(x)-H(x)|+\{\sup_{x\in\mathbb{R}}h(x)\}\frac{|\mu(\tilde{z}_i,\theta_0)-\mu(\hat{z}_i,\theta_n)|}{\sigma(\hat{z}_i,\theta_n)}+\sup_{x\in\mathbb{R}}\Big|H\Big(\frac{\sigma(\hat{z}_i,\theta_n)}{\sigma(\tilde{z}_i,\theta_0)}x\Big)-H(x)\Big|,\end{aligned}$$

where $\mu(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ are defined in (7.5). Using a similar technique in Koul and Ling (2003), we can show that $\sup_{x\in\mathbb{R}}|\widehat{H}(x)-H(x)|=o_p(1)$. Using Theorem 7.1, we can show that $\mu(\hat{z}_i,\theta_n)\to\mu(\tilde{z}_i,\theta_0)$ and $\sigma(\hat{z}_i,\theta_n)\to\sigma(\tilde{z}_i,\theta_0)$ almost surely. Since the density function $h(x)$ is uniformly continuous, $\sup_{x\in\mathbb{R}}h(x)<\infty$. Thus, the second term above goes to zero almost surely. By the monotonicity and uniform continuity of $H(x)$, we can show the third term tends to zero almost surely. Thus,

$$\mathrm{pr}_{\mathfrak{X}}(\tilde{y}_2\le x|y_1=\widehat{r}_n, Z_0=z_i)=\mathrm{pr}(y_2\le x|y_1=r_0, Z_0=z_i)+o_p(1).$$

Let $\widetilde{H}_{[k]}(\cdot)$ be the conditional distribution of $\widetilde{\mathscr{Y}_k}\equiv(\tilde{y}_k, ..., \tilde{y}_2)^{\mathrm{T}}$ given $y_1=\widehat{r}_n$, $Z_0=z_i$ and $\mathfrak{X}$, and $H_{[k]}(\cdot)$ be the conditional distribution of $\mathscr{Y}_k\equiv(y_k, ..., y_2)^{\mathrm{T}}$ given $y_1=r_0$ and $Z_0=z_i$. By the induction over k, we have $\widetilde{H}_{[d]}=H_{[d]}+o_p(1)$ as $n\to\infty$. Note that $\tilde{\varepsilon}_{d+1}$ is independent of $\widetilde{\mathscr{Y}_d}$ given $y_1=\widehat{r}_n$, $Z_0=z_i$ and $\mathfrak{X}$ and ε_{d+1} is independent of $\mathscr{Y}_d$ given $y_1=r_0$ and $Z_0=z_i$. By the continuous mapping theorem, (7.18) holds. Thus, the result holds. □

7.6 Appendix: Weak Convergence of a Pure Jump Process

Let $\{X_k^\varepsilon, k \geq 0\}$, indexed by ε, denote a discrete parameter process generated by:

$$X_{k+1}^\varepsilon = X_k^\varepsilon + J_{k+1}^\varepsilon$$

where the initial value is X_0^ε and $\{J_k^\varepsilon, k \geq 1\}$ is a sequence of jumps. Define the piecewise constant *interpolated process* $x^\varepsilon(t)$ for $t \in [0,1]$ by

$$x^\varepsilon(t) = X_j^\varepsilon, \qquad t \in [j\varepsilon, (j+1)\varepsilon) \quad \text{for} \quad j = 0, 1, ..., [1/\varepsilon] - 1,$$

and

$$x^\varepsilon(t) = X_{[1/\varepsilon]}^\varepsilon, \qquad t \in [[1/\varepsilon]\varepsilon,\ 1],$$

where $[1/\varepsilon]$ denotes the integer part of $1/\varepsilon$. What we need is the weak convergence of the interpolated sequence $\{x^\varepsilon(\cdot)\}$. When the limiting process of $\{x^\varepsilon(\cdot)\}$ is an ordinary differential equation or diffusion process, Kushner (1984) gave the detailed and rigorous demonstration through two different methods: the *perturbed test function* method and the *direct-averaging* method. However, when the limit is a pure jump process with J_m^ε being a Markov chain, only an outline is presented. Here, we generalize his result for J_m^ε being measurable in terms of $\mathscr{G}_m = \sigma\{X_i^\varepsilon, i \leq m\}$. Clearly, this result is of interest by itself and can be applied to many other nonlinear time series models. Let $\widehat{C}_0^2$ be a space of functions with compact support and continuous second derivative and let $\mathbb{P}_m^\varepsilon$ and $\mathbb{E}_m^\varepsilon$ be the conditional probability and conditional expectation on $\mathscr{G}_m$, respectively. We first give the following assumptions:

Assumption 7.9 *For each $\varepsilon > 0$, $\{J_k^\varepsilon\}$ is strictly stationary and there exists a constant $\lambda \in (0, +\infty)$ such that*

$$\lim_{\varepsilon \to 0} \lim_{m \to \infty} \mathbb{P}_k^\varepsilon (J_m^\varepsilon \neq 0)/\varepsilon = \lambda.$$

Assumption 7.10 *There exists a random variable U such that $\mathbb{P}(J_k^\varepsilon \in B | J_k^\varepsilon \neq 0) \to \mathbb{P}(U \in B)$ as $\varepsilon \to 0$ for any Borel set $B \in \mathscr{B}(\mathbb{R})$.*

Assumption 7.11 *For any $f \in \widehat{C}_0^2$ and x is a scalar,*

$$\lim_{\varepsilon \to 0} \lim_{m \to \infty} \mathbb{E}_k^\varepsilon \{f(x + J_m^\varepsilon) - f(x) | J_m^\varepsilon \neq 0\} = \mathbb{E}\{f(x+U) - f(x)\}.$$

Assumption 7.12 *There is a positive $M < \infty$ such that $|J_k^\varepsilon| \leq M$ for each $k \geq 1$.*

Assumptions 7.9 and 7.10 characterize the jump rate and the distribution of the jump size in the limiting process, respectively. Assumption 7.11 is a sufficient condition for the average used in the *direct-averaging* method. Assumption 7.12 requires the jumps to be bounded. This is a technical condition. In most applications, the jumps are not bounded in general. We can use the truncated technique to deal with the jumps and consider the truncated process. For some details, see the proof of Theorem 7.3. Based on the above assumptions, we have the following theorem.

Theorem 7.7. *Suppose Assumptions 7.9-7.12 hold. If* $X_0^\varepsilon \Longrightarrow x_0$*, then* $x^\varepsilon(t) \Longrightarrow x(t)$ *in* $D[0,1]$ *and* $x(t) = J(t) + x_0$*, where* $J(t)$ *is a compound Poisson process with jump rate* λ *and jump distribution* $\mathbb{Q}(\cdot)$ *induced by* U *at time* t *and* $J(0) = 0$.

Proof. Let n_ε be an integer satisfying $n_\varepsilon \to \infty$ and $\delta_\varepsilon = \varepsilon n_\varepsilon \to 0$ as $\varepsilon \to 0$. For any fixed function $f(\cdot) \in \widehat{C}_0^2$, define the piecewise constant function

$$\widetilde{A}^\varepsilon f(t) = \frac{1}{\varepsilon n_\varepsilon} \sum_{j=ln_\varepsilon}^{(l+1)n_\varepsilon - 1} \mathbb{E}_{ln_\varepsilon}^\varepsilon (f(X_{j+1}^\varepsilon) - f(X_j^\varepsilon))$$

for $t \in [l\delta_\varepsilon, (l+1)\delta_\varepsilon)$. Clearly, it follows that

$$\mathbb{E}_k^\varepsilon\{f(X_{m+1}^\varepsilon) - f(X_m^\varepsilon)\} = \mathbb{P}_k^\varepsilon(J_m^\varepsilon \neq 0)\mathbb{E}_k^\varepsilon\{f(X_m^\varepsilon + J_m^\varepsilon) - f(X_m^\varepsilon)|J_m^\varepsilon \neq 0\}.$$

By Assumption 7.9, we have

$$\begin{aligned}\widetilde{A}^\varepsilon f(t) &= \frac{1}{\varepsilon n_\varepsilon} \sum_{j=ln_\varepsilon}^{(l+1)n_\varepsilon - 1} \mathbb{E}_{ln_\varepsilon}^\varepsilon \{f(X_{j+1}^\varepsilon) - f(X_j^\varepsilon)\} \\ &= \frac{\lambda}{n_\varepsilon} \sum_{j=ln_\varepsilon}^{(l+1)n_\varepsilon - 1} \mathbb{E}_{ln_\varepsilon}^\varepsilon \{f(X_j^\varepsilon + J_j^\varepsilon) - f(X_j^\varepsilon)|J_j^\varepsilon \neq 0\} + o(1).\end{aligned}$$

Let $\widehat{A}^\varepsilon(t) = \int_0^t \widetilde{A}^\varepsilon f(s)ds$. By the boundedness of λ and J_m^ε in Assumptions 7.9 and 7.12, it follows that $\{(x^\varepsilon(t), \widehat{A}^\varepsilon(t))\}$ is tight in $D^2[0,1]$. In fact, for the tightness of $x^\varepsilon(t)$, see Kushner (1984, the last paragraph on page 32). The tightness of $\widehat{A}^\varepsilon(t)$ is implied by the boundness of $\widetilde{A}^\varepsilon f(t)$ due to $f \in \widehat{C}_0^2$. Since it is sufficient to work with an arbitrary weakly convergent subsequence also indexed by ε, without loss of generality, suppose that $(x^\varepsilon(t), \widehat{A}^\varepsilon(t)) \Longrightarrow (x(t), \widehat{A}(t))$ in $D^2[0,1]$. By means of the *Skorohod imbedding Theorem* in Kushner (1984, page 29), we assume that $(x^\varepsilon(t), \widehat{A}^\varepsilon(t))$ converges to $(x(t), \widehat{A}(t))$ a.s.

Let $\mathscr{C} = \{s \in [0,1] : x(t)$ is continuous at $s\}$. Then for any $s \in \mathscr{C}$, there exists an integer l_ε such that $s \in [l_\varepsilon \delta_\varepsilon, (l_\varepsilon + 1)\delta_\varepsilon)$. Let $m_\varepsilon = l_\varepsilon n_\varepsilon$. Then, for $f(\cdot) \in \widehat{C}_0^2$, by Assumptions 7.10 -7.11, it follows that

$$\begin{aligned}
\widetilde{A}^{\varepsilon} f(s) &= \lambda\Big\{\frac{1}{n_\varepsilon}\sum_{j=m_\varepsilon}^{m_\varepsilon+n_\varepsilon-1}\mathbb{E}^{\varepsilon}_{l_\varepsilon n_\varepsilon}\{f(X^{\varepsilon}_j+J^{\varepsilon}_j)-f(x(s)+J^{\varepsilon}_j)|J^{\varepsilon}_j\neq 0\}\Big\}\\
&\quad+\lambda\Big\{\frac{1}{n_\varepsilon}\sum_{j=m_\varepsilon}^{m_\varepsilon+n_\varepsilon-1}\mathbb{E}^{\varepsilon}_{l_\varepsilon n_\varepsilon}\{f(x(s))-f(X^{\varepsilon}_j)|J^{\varepsilon}_j\neq 0\}\Big\}\\
&\quad+\lambda\Big\{\frac{1}{n_\varepsilon}\sum_{j=m_\varepsilon}^{m_\varepsilon+n_\varepsilon-1}\mathbb{E}^{\varepsilon}_{l_\varepsilon n_\varepsilon}\{f(x(s)+J^{\varepsilon}_j)-f(x(s))|J^{\varepsilon}_j\neq 0)\Big\}+o(1)\\
&\to \lambda\mathbb{E}\{f(x(s)+U)-f(x(s))\}\\
&= \lambda\int[f(x(s)+u)-f(x(s))]\,\mathbb{Q}(du)\equiv Af(x(s)).
\end{aligned}$$

Thus, $\widehat{A}(t)=\int_0^t Af(x(s))ds$. For arbitrary k,t,s with $s_1<s_2<\cdots<s_k<t<t+s\le T$ and any bounded and continuous function $g(\cdot)$, by Taylor's expansion, it follows that

$$\mathbb{E}\Big\{g(x^{\varepsilon}(s_j)),j\le m)\times\Big[f(x^{\varepsilon}(t+s))-f(x^{\varepsilon}(t))-\int_t^{t+s}\widetilde{A}^{\varepsilon}f(u)du\Big]\Big\}=\triangle_\varepsilon,$$

where $\triangle_\varepsilon\to 0$ as $\varepsilon\to 0$. Hence,

$$\mathbb{E}\Big\{g(x(s_j)),j\le k)\Big[f(x(t+s))-f(x(t))-(\widehat{A}(t+s)-\widehat{A}(t))\Big]\Big\}=0,$$

which implies that $x(\cdot)$ solves the martingale problem for the operator A and the initial condition x_0. That is,

$$f(x(t))-\int_0^t Af(x(s))ds \quad \text{is a martingale for the operator } A.$$

Then $x^{\varepsilon}(t)\Longrightarrow x(t)=J(t)+x_0$ in $D[0,1]$, where $J(t)$ is a compound Poisson process with jump rate λ and the jump distribution $\mathbb{Q}(\cdot)$ and $J(0)=0$. ■