On maximum likelihood estimators for a threshold autoregression\textsuperscript{1}

Lianfen Qian

Department of Mathematical Sciences, Florida Atlantic University, Boca Raton, FL 33431, USA

Received 12 March 1997; received in revised form 6 March 1998

Abstract

For a stationary ergodic self-exciting threshold autoregressive model with single threshold parameter, Chan (1993) obtained the consistency and limiting distribution of the least-squares estimator for the underlying true parameters. In this paper, we derive the similar results for the maximum likelihood estimators of the same model under some regularity conditions on the error density, not necessarily Gaussian. © 1998 Elsevier Science B.V. All rights reserved.

AMS classification: primary 62M10; secondary 62A10, 62G05

Keywords: Threshold estimator; Coefficient parameter estimator; \( n \)-Consistency; Asymptotic normality and compound Poisson process

1. Introduction

In the past five decades, linear time-series models have dominated the development of the time-series analysis. But many applied fields such as electronics, oceanography, hydrology, ecology, marine engineering, medical engineering, solar astrophysics, and physics have revealed a common connecting theme: piecewise linearity. Tong (1977) mentioned the usefulness of a time-series that is piecewise linear in the past variables and in parameters. Later, Tong (1978a,b,1980) developed these models further in a systematic way for modeling of discrete time-series data. He argued that various phenomena such as limit cycles, jump resonance, harmonic distortion, modulation effects and chaos can be modeled by discrete time-series that are piecewise linear. He called these models the self-exciting threshold autoregressive (SETAR) models. See Tong (1983,1990) for a comprehensive introduction to general SETAR models.

\textsuperscript{1}Research was partly supported by the NSF grant: DMS-94 02904.
This paper is concerned with the large sample behaviors of maximum likelihood estimators in a two regime SETAR model, called SETAR(2;p,p), defined as follows:

\[ X_i = h(X_{i-1}, \theta) + \varepsilon_i, \quad i \geq 1 \]  

(1)

for some \( \theta = (\theta'_0, \theta'_1, r, d)' \in \mathbb{R}^{2p+3} \times \{1, 2, \ldots, p\} \), where \( X_{i-1} = (X_{i-1}, \ldots, X_{i-p})' \), \( \theta_j = (\theta_{0j}, \theta_{1j}, \ldots, \theta_{pj})' \in \mathbb{R}^{p+1}, j = 1, 2 \) and for \( x \in \mathbb{R}^p \),

\[ h(x, \theta) = \left( \theta_{01} + \sum_{k=1}^{p} \theta_{k1} x_k \right) I(x_d \leq r) + \left( \theta_{02} + \sum_{k=1}^{p} \theta_{k2} x_k \right) I(x_d > r). \]

The errors \( \{\varepsilon_i\} \) in Eq. (1) are independent and identically distributed random variables with mean zero, finite nonzero variance and \( \varepsilon_i \) is independent of \( X_{i-1}, X_{i-2}, \ldots, i \geq 1 \). The parameter \( r \), the location of the change of the autoregressive function \( h \), is called the threshold. The time delay \( d \) is called the delay parameter.

In this paper, we assume that the time-series in model (1) is stationary and ergodic. Detailed discussions about the stationarity and ergodicity can be found in Chan et al. (CPTW)(1985), Chan and Tong (1985).

For the case of the threshold having only finite number of possible values and assuming Gaussian errors, Tong (1983) constructed maximum likelihood estimators of the unknown parameters by using Akaike Information Criterion (Akaike, 1973). If the threshold \( r \) is known, CPTW (1985) obtained the consistency and asymptotic normality property of the least-squares estimators of the coefficient parameter \( \hat{\theta}_c = (\hat{\theta}'_0, \hat{\theta}'_1)' \) under some regularity conditions for \( p = 1 \). But, in practice, the threshold parameter \( r \) is unknown and can take infinitely many values in \( \mathbb{R} \). In this case, Petruccelli (1986) proved that the conditional least-squares estimator (CLSE) of \( \theta \) is strongly consistent for SETAR(2;1,1) model.

The present paper is motivated by Chan (1993), who developed the strong consistency and limiting distribution of the CLSE in a SETAR(2;p,p) model (1). It derives the asymptotics of maximum likelihood estimator (MLE) of the underlying parameter \( \theta \) in model (1), when the errors have a density \( f \), not necessarily Gaussian. Unlike the popular AR model, the likelihood function of SETAR(2;p,p) model is not continuous in the threshold parameter in general. Thus, the routine method of computing maximum likelihood estimator cannot be adopted. Instead, Section 2 discusses the maximum likelihood method to obtain the MLE \( \hat{\theta}_c = (\hat{\theta}'_0, \hat{\theta}'_1, \hat{\theta}'_2)' \) of \( \theta = (\theta'_0, r, d)' \). Section 3 describes assumptions and obtains the strong consistency of the MLE of the true parameter \( \theta \).

Section 4 shows the \( n \)-consistency of the threshold estimator under the assumption of discontinuity of \( h \) at \( r \). Section 5 obtains the uniform asymptotic normality of the coefficient parameter estimator \( \hat{\theta}_c \) over a bounded interval and some more byproduct results. In Section 6, as a consequence of the \( n \)-consistency of \( \hat{\theta}_c \), a suitably normalized log-likelihood sequence of processes \( \{\hat{\ell}_n\} \) is shown to be approximated by a sequence of simpler processes which describe the log-likelihood under known coefficient parameter \( \theta \). Through the latter processes, we obtain the limiting distribution of the standardized maximum likelihood estimator as the left endpoint of a random interval on which a superposition of independent compound Poisson processes attains a minimum.
Notation. Throughout the paper, the symbol $\theta$ is the fixed unknown underlying parameter, the function $f$ is the p.d.f. of $\varepsilon_1$ and $F$ denotes the distribution function corresponding to $f$. The expectation under $\theta$ is denoted by $E$. Weak convergence is denoted by $\Rightarrow$. A sequence (random) goes to zero (in probability) is denoted by $o(1)(o_p(1))$ while $O(1)(O_p(1))$ means that it is bounded (in probability). The multivariate normal distribution with mean zero and covariance matrix $\Gamma$ is denoted by $N(0,\Gamma)$. Let $\mathcal{R}$ be the real line $(-\infty, \infty)$, and $\hat{\mathcal{R}} = \mathcal{R} \cup \{-\infty, \infty\}$, then the compactness of the set $\hat{\mathcal{R}}$ is under the metric $d(\cdot, \cdot)$ defined by $d(x, y) = |\arctan x - \arctan y|$. A function $\phi$ satisfies the Lip(1) if $\forall x, y \in \mathcal{R}, \exists \alpha > 0$, such that $|\phi(x) - \phi(y)| \leq L|x - y|$. For any event $A$, the complement event of $A$ is denoted by $A^c$ and the indicator function is denoted by $I(A)$. Throughout, the capital letter $C$, the symbols $\gamma_i, i = 1, 2, \ldots$ stand for absolute constants and they can have different values in different places. The notation $x'y$ stands for the inner product of vectors $x$ and $y$. For any matrix $M = (m_{ij}), \|M\| = \sum_{i,j} |m_{ij}|$, $\det(M)$ and $\text{adj}(M)$ stand for the determinant and adjoint matrix of $M$, respectively. Vectors of dimension more than one are denoted by bold face letters. The index $i$ in the summation varies from 1 to $n$ unless specified otherwise.

2. The maximum likelihood estimation

We begin with the definition of the maximum likelihood estimators of the unknown underlying parameter $\theta$ in model (1). Throughout we assume that the true parameter $\theta$ is an interior point of the parameter space $\mathcal{R}_{p+2} \times \hat{\mathcal{R}} \times \{1, 2, \ldots, p\}$. There exists a compact subset $K$ of $\mathcal{R}_{p+2}$ such that $\theta$ is an interior point of $K \times \hat{\mathcal{R}} \times \{1, 2, \ldots, p\}$.

Denote $\Omega = K \times \hat{\mathcal{R}} \times \{1, 2, \ldots, p\}$, then $\Omega$ is a compact set. Let $\vartheta = (\vartheta', \beta', s, q)'$ be any point in $\Omega$. Let $X_i = (X_{i1}, \ldots, X_{i-\rho+1})'$, then $\{X_i\}$ is a Markov chain. Let $g_{\vartheta}(X_0)$ be the initial density of $X_0$ under $\vartheta$, $f$ be the density function of $\varepsilon_1$, then the one step transition densities, starting with $X_0$, is $f(X_i - h(X_{i-1}, \vartheta))$, $i \geq 1$. If one observes $(X_{0i}, \ldots, X_n)$, then the likelihood function under $\vartheta$ is $\prod_{i=1}^n f(X_i - h(X_{i-1}, \vartheta))g_{\vartheta}(X_0)$. Let $\hat{\vartheta}_n = (\hat{\vartheta}_{cn}, \hat{r}_n, \hat{d}_n)'$ be any measurable function of $(X'_{0i}, X_{1i}, \ldots, X_{ni})$ from $\mathcal{R}_{p+1}$ to $\Omega$ such that $\hat{\vartheta}_n$ maximizes the conditional likelihood function

$$L_n(\vartheta) := \prod_{i=1}^n f(X_i - h(X_{i-1}, \vartheta))$$ over $\Omega$.

Write $\vartheta = (\vartheta', s, q)'$, $\theta = (\theta', r, d)'$. Because of the behavior of the threshold parameter $r$ in the likelihood function, the maximizing algorithm will be taken in the following fashion:

Step 1: For fixed $s \in \hat{\mathcal{R}}, q \in \{1, 2, \ldots, p\}$, denote $L_{\text{max}}(\vartheta_c) = L_n(\vartheta_c, s, q) = L_n(\vartheta)$. Let $\vartheta_{cn}(s, q) \in K$ be any value satisfying the following equation:

$$\vartheta_{cn}(s, q) = \arg \max_{\vartheta_c \in K} L_{\text{max}}(\vartheta_c).$$
**Step 2:** Consider the profile conditional likelihood function \((s, q) \rightarrow L_n(\theta_{cn}(s, q), s, q)\).

Note that \(L_n(\theta_{cn}(s, q), s, q)\) has only finite number of possible values. Let \(\hat{r}_n, \hat{d}_n\) be the smallest values satisfying the following equation:

\[
L_n(\theta_{cn}(\hat{r}_n, \hat{d}_n), \hat{r}_n, \hat{d}_n) = \max_{s \in A, q \in \{1, \ldots, p\}} L_n(\theta_{cn}(s, q), s, q),
\]

and substitute \(\hat{r}_n, \hat{d}_n\) into \(\theta_{cn}(s, q)\) to get \(\hat{\theta}_{cn} = \theta_{cn}(\hat{r}_n, \hat{d}_n)\). Then

\[
\hat{\theta}_n = (\hat{\theta}_{cn}', \hat{r}_n, \hat{d}_n)' \quad \text{is a maximum likelihood estimator of } \theta.
\]

To see Eq. (2), for any \(\theta = (\theta_{cn}', s, q)' \in \Omega\), by the definitions of \(\hat{\theta}_{cn}\) and \(\hat{r}_n, \hat{d}_n\), we have

\[
L_n(\hat{\theta}_{cn}, \hat{r}_n, \hat{d}_n) = L_n(\theta_{cn}(\hat{r}_n, \hat{d}_n), \hat{r}_n, \hat{d}_n) \geq L_n(\theta_{cn}(s, q), s, q) \geq L_n(\theta)
\]

and, hence,

\[
L_n(\hat{\theta}_{cn}, \hat{r}_n, \hat{d}_n) = \sup_{\theta \in \Omega} L_n(\theta).
\]

This means that \(\hat{\theta}_n = (\hat{\theta}_n', \hat{r}_n, \hat{d}_n)'\) is a MLE of \(\theta\).

### 3. Assumptions and strong consistency

We begin this section by listing the needed assumptions on the density \(f\) of the error \(\varepsilon_1\) and the underlying process.

(C1) \(f\) is absolutely continuous and positive everywhere on \(\mathcal{R}\). With the a.e. derivative \(\hat{f}\), let \(\phi = f / \hat{f}\) and \(I(f) = \int \phi^2(x) f(x) \, dx < \infty\).

(C2) \(\phi\) is Lip(1).

(C3) \(\phi\) is differentiable and the derivative \(\hat{\phi}\) is Lip(1).

(C4) \(E|\varepsilon_1|^4 < \infty\).

To derive the \(n\)-consistency and the limiting distribution of the threshold estimator, we need to make the following model assumptions:

(M1) The threshold \(r\) in \(\mathcal{R}\) is the discontinuity point of \(h\), or equivalently, there exists \(Z^x = (1, x_0, \ldots, x_{1-d})'\) such that \((\theta_1 - \theta_2)'Z^x \neq 0, x_{1-d} = r\).

(M2) \(\{X_k\}\) admits a unique measure \(\pi\) such that \(\exists K, \rho < 1, \forall x \in \mathcal{R}^p, \forall k \geq 1, ||P^k(x, \cdot) - \pi(\cdot)|| \leq K \rho^k(1 + |x|), \) where \(|| \cdot ||\) and \(\cdot\) denote the total variation norm and the Euclidean norm, respectively.

**Remark 1.** Examples of density satisfying (C1)–(C4) include normal density, logistic density, and the \(t\)-density with degrees of freedom \(m, m > 4\). From the invariant equation \(g_\theta(x) = \int f(x - h(y, \theta)) g_\theta(y) \, dy, x \in \mathcal{R}^p\), condition (C1) implies that \(g_\theta\) is bounded away from 0 and \(\infty\) over compact sets. It is the minimal requirement for obtaining asymptotically efficient estimators of the coefficient parameters under the known threshold, see Koul and Schick (1995). The finiteness of \(I(f)\) implies the boundedness.
of \( f \) (Koul, 1992, p. 52). \( \text{Lip}(1) \) and the differentiability of \( \phi \) imply the boundedness of \( \phi \).

**Remark 2.** For the detailed discussion on (M2), see Remark B(i) in Chan (1993).

Throughout in the following proofs, we use the fact that \( E|\epsilon_i|^k < \infty \) implies \( E|X_0|^k < \infty \), for \( k = 2, 3, 4 \), as proved by Chan et al. (1985).

We are going to show the strong consistency of the MLE \( \hat{\theta}_n \). To this effect, let \( l_n \) be the conditional log-likelihood ratio:

\[
l_n(\theta) = \frac{1}{n} \sum \ln \frac{f(X_i - h(X_{i-1}, \theta))}{f(X_i - h(X_{i-1}, \theta))}, \quad \theta \in \Omega,
\]

and denote

\[
\psi(X_{i-1}, \epsilon_i, \theta) = \ln \frac{f(\epsilon_i + h(X_{i-1}, \theta) - h(X_{i-1}, \theta))}{f(\epsilon_i)}, \quad 1 \leq i \leq n. \tag{3}
\]

Note that

\[
l_n(\theta) = \frac{1}{n} \sum \psi(X_{i-1}, \epsilon_i, \theta), \quad \theta \in \Omega.
\]

Write \( \theta = (\theta', s, q)' \in \Omega \) and \( h(x, \theta) = h_{sq}(x, \theta), \ x \in \mathbb{R}^p \). Let \( Z = (1, x')' \), and

\[
\hat{h}_{sq}(x) = (\partial / \partial \theta)(h_{sq}(x, \theta)) = (Z' I(x_q \leq s), Z' I(x_q > s))', \quad s \in \mathbb{R}, \ x \in \mathbb{R}^p, \ q \in \{1, \ldots, p\}.
\]

Then,

\[
|\hat{h}_{sq}(x)| \leq \sqrt{1 + |x|^2}. \tag{4}
\]

Observe that for any \( x \in \mathbb{R}^p \),

\[
h(x, \theta) = \partial h_{sq}(x), \tag{5}
\]

so that

\[
|h(x, \theta)| \leq |\partial h_{sq}(x)| \leq |\theta_c| \sqrt{1 + |x|^2}. \tag{6}
\]

Furthermore, for any \( s \in \mathbb{R}, \ t \in \mathbb{R}, \ q \in \{1, \ldots, p\}, \)

\[
|\hat{h}_{sq}(x) - \hat{h}_{sq}(x)| \leq \sqrt{2(1 + |x|^2)I(s \wedge t \leq x_q \leq s \vee t)} \leq \sqrt{2(1 + |x|^2)I(|x_q - t| \leq |s - t|)}. \tag{7}
\]

We are ready to state the first main result now.

**Theorem 1.** Suppose that \( \{X_i\} \) in model (1) is stationary and ergodic, the conditions (C1) and (C2) hold. Then,

\[
\hat{\theta}_n \overset{a.s.}{\to} \theta \quad \text{as} \ n \to \infty \quad \text{(under } \theta). \]
Before proving Theorem 1, we need the following lemma. Let \( U_\theta \) denote any open neighborhood of \( \theta \).

**Lemma 1.** Under the assumptions of Theorem 1, for any \( \theta \in \Omega \) and its open neighborhood \( U_\theta \), \( E \sup_{\theta' \in U_\theta} |\psi(X_0, \varepsilon_1, \theta^*) - \psi(X_0, \varepsilon_1, \theta) - \delta(X_0, \theta^{*})| \rightarrow 0 \), as \( U_\theta \) shrinks to \( \theta \).

**Proof.** Define

\[ U_\theta(\eta) = \{ \theta^* = (\theta^*_c, s^*, q) \in \Omega : |\theta^*_c - \theta_c| < \eta, \quad d(s^*, s) < \eta \}, \eta > 0. \]

It suffices to show that

\[ E \sup_{\theta^* \in U_\theta(\eta)} |\psi(X_0, \varepsilon_1, \theta^*) - \psi(X_0, \varepsilon_1, \theta)| \rightarrow 0 \quad \text{as } \eta \rightarrow 0. \quad (8) \]

Let \( \varepsilon_1(\theta) = X_1 - h(X_0, \theta) \) and \( \delta(X, \theta^*) = h(x, \theta) - h(x, \theta^*) \), \( x \in \mathcal{P} \). For any \( \theta = (\theta^*_c, s, q) \), recall that for \( x \in \mathcal{P} \),

\[ h_{sq}(x, \theta_c) = h(x, \theta_c, s, q) = h(x, \theta), \quad (9) \]

and rewrite \( \delta(x, \theta^*) = h_{sq}(x, \theta_c) - h_{sq}(x, \theta^*_c) \). For any \( x \in \mathcal{P} \), by Eqs. (4) and (5),

\[ |h_{sq}(x, \theta_c) - h_{sq}(x, \theta^*_c)| \leq |\theta_c - \theta^*_c| \sqrt{1 + |x^2|}, \quad (10) \]

and by Eqs. (5) and (7),

\[ |h_{sq}(x, \theta^*_c) - h_{sq}(x, \theta^*_c)| \leq |\theta^*_c| \sqrt{2(1 + |x^2|)} I(s \wedge s^{*} < x_q \leq s \vee s^{*}) \]

\[ \leq |\theta^*_c| \sqrt{2(1 + |x^2|)} I(|x_q - s| \leq |s^* - s|). \quad (11) \]

Thus for \( \theta^* \in U_\theta(\eta) \) and for \( s \in \mathcal{R}, x \in \mathcal{P} \),

\[ |\delta(x, \theta^*)| \leq |h_{sq}(x, \theta_c) - h_{sq}(x, \theta^*_c)| + |h_{sq}(x, \theta^*_c) - h_{sq}(x, \theta^*_c)| \leq |\theta_c - \theta^*_c| \sqrt{2(1 + |x^2|)} \| I(s \wedge s^{*} < x_q \leq s \vee s^{*}) \| \quad (12) \]

\[ \leq |\theta^*_c| \sqrt{2(1 + |x^2|)} I(|x_q - s| \leq |s^* - s|) + \eta \sqrt{1 + |x^2|} \equiv d(x, \eta) \quad \text{(say)}, \]

where \( s_0(\eta) \) is such that \( d(s_0(\eta), s) = \eta \).

Note that

\[ \varphi(\varepsilon_1(\theta)) = [\varphi(\varepsilon_1 + h(X_0, \theta) - h(X_0, \theta))] + \varphi(\varepsilon_1). \]

Condition (C1) and (C2), \( E|X_0|^2 < \infty \) and Eq. (6) imply that there exists a constant \( L \) such that for any \( \theta \in \Omega \),

\[ E\varphi^2(\varepsilon_1(\theta)) \leq 2E[\varphi^2(\varepsilon_1) + L|h(X_0, \theta) - h(X_0, \theta)|^2] \]

\[ \leq 2E(F) + 4L(\theta_c^2 + |\theta_c^*|^2)E(1 + |X_0|^2) < \infty. \quad (14) \]
The absolute continuity of \( \ln f \), which follows from (C1), and (13) imply that
\[
|\psi(X_0, \epsilon_1, \theta^0) - \psi(X_0, \epsilon_1, \theta)| \leq \int_{-d(X_0, \eta)}^{d(X_0, \eta)} |\varphi(\epsilon_1(\theta) + v)| \, dv. \tag{15}
\]
Thus, Eqs. (14), (15) and Cauchy–Schwarz inequality imply that the LHS of Eq. (8) is
\[
E f[|\psi(X_0, \epsilon_1, \theta^0) - \psi(X_0, \epsilon_1, \theta)|] \leq 2(E \varphi^2(\epsilon_1(\theta)))^{1/2}(E \Delta^2(X_0, \theta))^{1/2} + LE\Delta^2(X_0, \theta).
\]
Moreover,
\[
E \Delta^2(X_0, \theta) = E(1 + X_0^2)(\sqrt{2}d|\theta|/\sqrt{1 + |x|^2}) \leq (\sqrt{2}d|\theta|/\sqrt{1 + |x|^2}) \leq E (X_0^2) / \sqrt{1 + |x|^2} \equiv \Delta_1(x, \eta),
\]
where \( d(s_0(\eta), \infty) = \eta \). Again,
\[
E \Delta^2(X_0, \theta) \to 0 \quad \text{as} \quad \eta \to 0.
\]
Thus, the proof goes through for \( s = \infty \). The proof is similar in the case \( s = -\infty \), except one replaces \( I(X_{1-q} > s^*) \) by \( I(X_{1-q} < s^*) \). Therefore Lemma 1 is proved.

**Proof of Theorem 1.** We will adopt Huber’s method (1967). Let \( z(\theta) = E \psi(X_0, \epsilon_1, \theta) \) for \( \theta \in \Omega \). Conditions (C1) and (C2), the mean value theorem, the independence of \( \epsilon_1 \) and \( X_0 \) and Cauchy–Schwarz inequality imply that \( E |\psi(X_0, \epsilon_1, \theta)| < \infty \). Thus, \( z \) is a well-defined finite function from \( \Omega \) to \( R \). Note that \( z(\theta) = 0 \) and \( \ln x < x - 1 \), unless \( x = 1 \). For any given open neighborhood \( V \) of \( \theta \) in \( \Omega \) and any \( \theta \in V^c = \Omega \setminus V \), a conditional argument yields that
\[
z(\theta) = E \ln \frac{f(\epsilon_1 + h(X_0, \theta) - h(X_0, \theta))}{f(\epsilon_1)}
= E \left\{ E \left[ \ln \frac{f(\epsilon_1 + h(X_0, \theta) - h(X_0, \theta))}{f(\epsilon_1)} \bigg| X_0 \right] \right\}
< E \left\{ \int [f(y + h(X_0, \theta) - h(X_0, \theta)) - f(y)] \, dy \right\} = 0.
\]
By Lemma 1, \( z \) is continuous and, hence, the compactness of \( V^c \) implies that there exists \( \theta_0 \in V^c \), such that
\[
\sup_{\theta \in V^c} z(\theta) = z(\theta_0) < 0.
\]
Let \( \delta_0 = -\alpha(\vtheta_0)/3 \). For any \( \vtheta \in V_c \), by Lemma 1 again, there exists \( \eta_0 > 0 \), such that

\[
E \sup_{\vtheta^* \in U_\eta(\eta_0)} \psi(X_0, e_1, \vtheta^*) \leq E\psi(X_0, e_1, \vtheta) + \delta_0 \leq \alpha(\vtheta_0) + \delta_0 = -2\delta_0. \tag{17}
\]

Again, the compactness of \( V_c \) implies that there exists a finite number \( M \) of \( U_{\vtheta_j}(\eta_0) \), \( \vtheta_j \in V_c, j = 1, 2, \ldots, M \) such that \( \bigcup_{j=1}^M U_{\vtheta_j}(\eta_0) = V_c \). Then by the ergodic theorem and Eq. (17), there exists an \( n_0 \) such that for any \( n \geq n_0, 1 \leq j \leq M \),

\[
\sup_{\vtheta^* \in U_{\vtheta_j}(\eta_0)} l_n(\vtheta^*) \leq \frac{1}{n} \sum_{\vtheta^* \in U_{\vtheta_j}(\eta_0)} \psi(X_{j-1}, e_1, \vtheta^*) \leq E \sup_{\vtheta^* \in U_{\vtheta_j}(\eta_0)} \psi(X_0, e_1, \vtheta^*) + \delta_0 \leq -\delta_0 \text{ a.s.}
\]

But

\[
\sup_{\vtheta \in V} l_n(\vtheta) \geq l_n(\vtheta_0) = 0.
\]

Therefore, for any neighborhood \( V \) of \( \vtheta \) in \( \Omega \), \( \exists n_0 \), s.t. for all \( n \geq n_0 \),

\[
\sup_{\vtheta \in \Omega \cap V} l_n(\vtheta^*) \leq \max_{1 \leq j \leq M} \sup_{\vtheta \in U_{\vtheta_j}(\eta_0)} l_n(\vtheta^*) \leq -\delta_0 < 0 \leq \sup_{\vtheta \in V} l_n(\vtheta).
\]

This implies that

\[
\hat{\vtheta}_n \in V, \text{ a.s. } \forall V \text{ and } \forall n \geq n_0.
\]

By the arbitrary of \( V \), \( \hat{\vtheta}_n \) goes to \( \vtheta \) almost surely. \( \square \)

**Remark 3.** In fact, we proved that any maximum likelihood estimator by our algorithm is consistent. In order to get the limiting distribution of \( \hat{r}_n \), we shall use the one with the smallest \( \hat{r}_n, \hat{d}_n \).

**Remark 4.** Under conditions (C1)–(C2), from Theorem 1, \( \hat{d}_n = d \) eventually. Thus, we assume the delay parameter \( d \) is known for the rest of the paper as Chan (1993).

### 4. \( n \)-consistency of the threshold estimator

From now on we will invoke conditions (M1) and (M2). The discontinuity of \( h \) at \( r \) will give a stronger result about the estimator \( \hat{r}_n \) of the threshold \( r \), i.e., the \( n \)-consistency of \( \hat{r}_n \).

**Theorem 2.** Suppose conditions (C1)–(C4), (M1) and (M2) hold, then

\[
|\hat{r}_n - r| = O_p(1).
\]

First assume \( p = d = 1 \). We will begin with some notation. Let \( J : \mathbb{R}^2 \to \mathbb{R} \) and

\[
p(x) = E J(x, e_1), \quad p_1(x) = E |J(x, e_1)|, \quad p_2(x) = E J^2(x, e_1), \quad x \in \mathbb{R}, \tag{18}
\]
For \( u \geq 0 \), define

\[
G(u) = EI(r < X_0 \leq r + u), \quad G_n(u) = \frac{1}{n} \sum I(r < X_{i-1} \leq r + u),
\]

and

\[
R_n(u) = \frac{1}{n} \sum J(X_{i-1}, \varepsilon_i) I(r < X_{i-1} \leq r + u),
\]

\[
r_n(u) = \frac{1}{n} \sum p(X_{i-1}) I(r < X_{i-1} \leq r + u).
\]

Also, let \( J_c(X_{i-1}, \varepsilon_i) = J(X_{i-1}, \varepsilon_i) - p(X_{i-1}) \). For \( u_2 \geq u_1 \geq 0 \), let

\[
\tilde{R}_n(u_1, u_2) = \frac{1}{n} \sum J_c(X_{i-1}, \varepsilon_i) I(r + u_1 < X_{i-1} \leq r + u_2),
\]

\[
\tilde{R}(u_1, u_2) = E\tilde{R}_n(u_1, u_2).
\]

**Lemma 2.** Suppose that \((C1)\) holds, then there exists constants \( 0 < m \leq M \leq \infty \) and \( 0 < C < \infty \) independent of \( n \). For any \( 0 < \delta < 1 \) and \( \forall u, u_1, u_2 \in [0, \delta], \forall n \),

\[
u m \leq G(u) \leq Mu, \tag{19}
\]

\[
\text{Var}(I(r < X_0 \leq r + u)) \leq CG(u), \tag{20}
\]

\[
\text{Var}(nG_n(u)) \leq nCG(u). \tag{21}
\]


**Remark 5.** The above lemma generalizes a similar result proved in Chan (1993) for the case of \( J(x, y) \equiv xy \). Many details of the proof of the above lemma are similar to those in Chan (1993), but there are some differences. For the convenience of a reader we give a proof of Eq. (24) which exemplifies these similarities and differences.

Expanding the left-hand side of Eq. (24),

\[
\text{Var}(\tilde{R}_n(u_1, u_2)) = \frac{1}{n} \text{Var}(\tilde{J}(X_0, \varepsilon_1) | I(u_1 < X_0 \leq u_2))
\]

\[
+ \frac{1}{n^2} \sum_{k \neq j} \text{Cov}(\tilde{J}(X_{k-1}, \varepsilon_k) | I(u_1 < X_{k-1} \leq u_2), \tilde{J}(X_{j-1}, \varepsilon_j) | I(u_1 < X_{j-1} \leq u_2)).
\]
Note that \( |p(x)| \leq p_1(x) \). By the continuity of \( p_1 \), the property of Markov chain and Remark 2, for any \( k \geq 2 
abla \),
\[
|E[|J^*(X_{k-1}, e_k)|I(u_1 < X_{k-1} \leq u_2)]|X_1, X_0] - E[|J^*(X_0, e_1)|I(u_1 < X_0 \leq u_2)]
\]
\[= \left| \int_{u_1}^{u_2} \int |J^*(x, y)| \, dF(y) I(u_1 < X_{k-1} \leq u_2)|X_1 \right|
\]
\[\leq \int_{u_1}^{u_2} \left( \int |J^*(x, y)| \, dF(y) \right) \left( |P^{k-2}(X_1, dx) - \pi(dx)| \right)
\]
\[\leq 2 \sup_{x \in [0, 1]} p_1(x) \|P^{k-2}(X_1, \cdot) - \pi(\cdot)\|_n
\]
\[\leq C \rho^{k-2}(1 + |X_1|) \leq C \rho^{k-2}(1 + |h(\theta, X_0)| + |\varepsilon_1|). \tag{26}
\]

The continuity of \( p_2 \) and \( p_1(x) \leq p_2^{1/2}(x) \), \( E\varepsilon_1^2 < \infty \), the Cauchy–Schwarz inequality imply that
\[
E[|J^*(X_0, e_1)\varepsilon_1|I(u_1 < X_0 \leq u_2)]
\]
\[= E \{E[|J^*(X_0, e_1)\varepsilon_1|I(u_1 < X_0 \leq u_2)]\}
\]
\[\leq 2 \sup_{x \in [0, 1]} p_2(x) E(\varepsilon_1^2) (G(u_2) - G(u_1)). \tag{27}
\]

Then the definition of the autoregressive function \( h \), the Markov property of \( \{X_i\} \), Eqs. (26), (27), Remark 2 and the conditioning argument yield that
\[
|\text{Cov}(|J^*(X_0, e_1)|I(u_1 < X_0 \leq u_2), |J^*(X_{k-1}, e_k)|I(u_1 < X_{k-1} \leq u_2))|
\]
\[= |E \{|J^*(X_0, e_1)|I(u_1 < X_0 \leq u_2)
\]
\[
\times (E[|J^*(X_{k-1}, e_k)|I(u_1 < X_{k-1} \leq u_2)]|X_1, X_0] - E[|J^*(X_0, e_1)|I(u_1 < X_0 \leq u_2)])
\]
\[\leq C \rho^{k-2} E|J^*(X_0, e_1)|I(u_1 < X_0 \leq u_2)(1 + |h(\theta, X_0)| + |\varepsilon_1|)
\]
\[\leq C \rho^{k-2}(G(u_2) - G(u_1)).
\]

Therefore, Eq. (24) follows from the stationarity of \( \{X_i\} \) and the fact \( \sum_{k \neq j} \rho^{k-j} = O(n) \).

**Proposition 1.** Suppose that (C1) holds and the functions \( p_1 \) and \( p_2 \) in Eq. (18) are continuous. Then, for each \( \varepsilon > 0, \eta > 0 \), there is a constant \( B < \infty \), \( \forall \delta > 0 \) and \( \forall n \geq [B/\delta] + 1 \),
\[
P \left( \sup_{\delta < u < \theta} \left| G_n(u)/G(u) - 1 \right| < \eta \right) > 1 - \varepsilon. \tag{28}
\]
\[ P \left( \sup_{B/\delta < u \leq \delta} \left| \frac{R_n(u) - r_n(u)}{G(u)} \right| < \eta \right) > 1 - \varepsilon. \] (29)

**Note.** Condition (C1) is for Eq. (28), the continuity of \( p_1 \) and \( p_2 \) is for Eq. (29).

**Proof of Proposition 1.** It is similar to that of the Claim 2 in Chan (1993) but it uses Lemmas 2 and 3. For detail proof, see Qian (1996).

Now, let
\[ J(X_{i-1}, e_i) \equiv \psi(X_{i-1}, e_i) \equiv \ln \frac{f(e_i + a + \beta X_{i-1})}{f(e_i)} , \]
where \( a = \theta_{02} - \theta_{01}, \beta = \theta_{12} - \theta_{11} \). The functions \( p, p_1 \) and \( p_2 \) are defined correspondingly. For \( u > 0 \), define
\[ D_n(u) = \frac{1}{n} \sum \psi(X_{i-1}, e_i) I(r < X_{i-1} \leq r + u), \]
\[ d_n(u) = \frac{1}{n} \sum p(X_{i-1}) I(r < X_{i-1} \leq r + u). \]

**Corollary 1.** Suppose that (C1)–(C2) hold. Then, for each \( \varepsilon > 0 \) and \( \eta > 0 \), there is a constant \( B < \infty \) such that \( \forall 0 < \delta < 1 \) and \( \forall n \geq [B/\delta] + 1, \)
\[ P \left( \sup_{B/\delta < u \leq \delta} \left| \frac{D_n(u) - d_n(u)}{G(u)} \right| < \eta \right) > 1 - \varepsilon. \] (30)

**Proof.** The continuity of \( p_1, p_2 \) can be derived from conditions (C1) and (C2). Thus, Eq. (30) follows from (29) immediately.

Before proving Theorem 2, we need some more notation. Rewrite the \( \psi \) of Eq. (3) as
\[ \psi(X_{i-1}, e_i, t, s) = \ln \frac{f(X_i - h_s(X_{i-1}, t))}{f(e_i)}, \quad t \in \mathcal{A}, \quad s \in \mathcal{R}, \]
where \( h_s \) is defined in Eq. (9) when \( d \) is known. Let
\[ \tilde{\psi}(X_{i-1}, e_i, t, s) = \psi(X_{i-1}, e_i, t, s) - \psi(X_{i-1}, e_i, t, r), \quad t \in \mathcal{A}, \quad s \in \mathcal{R}, \ 1 \leq i \leq n. \]
Then, for \( 1 \leq i \leq n, \ t \in \mathcal{A}, \ s \in \mathcal{R}, \)
\[ \tilde{\psi}(X_{i-1}, e_i, t, s) \equiv \frac{\partial}{\partial t} (\tilde{\psi}(X_{i-1}, e_i, t, s)) \]
\[ = - \varphi(X_i - h_s(X_{i-1}, t)) \hat{h}_s(X_{i-1}) - \varphi(X_i - h_s(X_{i-1}, t)) \hat{h}_s(X_{i-1}) \]
\[ = - \left[ \varphi(X_i - h_s(X_{i-1}, t)) - \varphi(X_i - h_s(X_{i-1}, t)) \right] \hat{h}_s(X_{i-1}) \]
\[ - \varphi(X_i - h_s(X_{i-1}, t)) \left[ \hat{h}_s(X_{i-1}) - \hat{h}_s(X_{i-1}) \right]. \]

Now, we are ready to prove Theorem 2.
Proof of Theorem 2. The idea is similar to Chan’s but the details are different. Since \( \hat{\theta}_n \) is strongly consistent by Theorem 1, without loss of generality, we can assume that \( d \) is known and the parameter space can be restricted to a neighborhood of \( \theta \), say,

\[
\Omega(\delta) = \{ \theta \in \Omega : |\theta_e - \theta_e| < \delta, |s - r| < \delta \},
\]

for some \( 0 < \delta < 1 \) to be determined later. Then, it suffices to show that \( l_n(\theta_e, s) - l_n(\theta_e, r) \) is negative for \( n|s - r| \) large enough. More specifically, we shall show that for every \( \varepsilon > 0 \), there is a \( B > 0 \) and \( \gamma > 0 \), \( 0 < \delta < 1 \) and \( n_0 \) such that for any \( n \geq n_0 \),

\[
P \left( \sup_{|s-r| \leq \delta, \theta \in \Omega(\delta)} \frac{[l_n(\theta_e, s) - l_n(\theta_e, r)]}{G(|s - r|)} < -\gamma \right) > 1 - 2\varepsilon. \tag{31}
\]

For any \( \theta = (\theta_e', s') \), denote \( l_n(\theta_e') = l_n(\theta_e, s') = l_n(\theta) \). Now, decompose \( l_n(\theta_e, s) - l_n(\theta_e, r) \) into two terms as follows:

\[
l_n(\theta_e, s) - l_n(\theta_e, r) = [l_n(\theta_e) - l_n(\theta_e)] + [l_n(\theta_e) - l_n(\theta_e)] + [l_n(\theta_e) - l_n(\theta_e)]
\]

\[
= l_n(\theta) + l_n(s) \quad \text{(say)}.
\]

We shall prove that there exists a \( \delta \) small enough such that

\[
\sup_{|s-r| \leq \delta, \theta \in \Omega(\delta)} \left| \frac{l_n(\theta)}{G(|s - r|)} \right| = o_p(1), \tag{32}
\]

and

\[
P \left( \sup_{|s-r| \leq \delta} \frac{l_n^2(s)}{G(|s - r|)} < -2\gamma \right) > 1 - \varepsilon. \tag{33}
\]

We will prove the case \( s > r \) only and write \( s = r + u \) for some \( u > 0 \). For the case \( s < r \), the proof will be exactly the same. To prove Eq. (32), by using the absolute continuity of \( \psi \),

\[
l_n(\theta) \equiv \frac{1}{n} \sum \left[ \zeta(X_{i-1}, e_i, \theta_e, s) - \zeta(X_{i-1}, e_i, \theta_e, s) \right]
\]

\[
= \frac{1}{n} \sum \int_0^1 \zeta'(X_{i-1}, e_i, \theta_e + v(\theta_e - \theta_e), s)(\theta_e - \theta_e) \ dv. \tag{34}
\]

By the Lip(1) of \( \varphi \), Eqs. (10), (11), (4) and (7) imply that there exists a constant \( L \), for \( 1 \leq i \leq n, t \in \mathcal{R}^d, s \in \mathcal{R}^d \),

\[
|\zeta(X_{i-1}, e_i, t, s)| \leq L(h(X_{i-1}, t) - h(X_{i-1}, t)) \sqrt{1 + X_{i-1}^2}
\]

\[
+ \left( |\varphi(e_i)| + L|h(X_{i-1}, \theta_e) - h(X_{i-1}, t)| \right) \sqrt{1 + X_{i-1}^2} I(s \wedge r < X_{i-1} \leq s \vee r)
\]

\[
\leq (L|r| \sqrt{1 + X_{i-1}^2} + |\varphi(e_i)| + L|\theta_e - t| \sqrt{1 + X_{i-1}^2})
\]

\[
\times \sqrt{1 + X_{i-1}^2} I(s \wedge r < X_{i-1} \leq s \vee r).
\]
Thus, for any $\vartheta \in \Omega(\delta)$, by Eqs. (34) and (35), there exists $0 < C < \infty$, such that
\[
\left| \frac{L_n(\vartheta)}{G(u)} \right| \leq C\delta \left[ \frac{G_n(u)}{G(u)} + \left| \frac{R_n(u) - r_n(u)}{G(u)} \right| \right],
\]
with $J(x, y) = |\varphi(y)|$ in the definitions of $R_n$ and $r_n$. Thus, Eq. (32) follows from the Proposition 1 by choosing $\delta > 0$ sufficiently small.

To prove Eq. (33), recall that $\theta = \theta_{02} - \theta_{01}, \beta = \theta_{12} - \theta_{11}$ and let $J = \psi$ in the definition of $p$ in Eq. (18). Then $l_n^2(s)$ can be decomposed as follows:
\[
l_n^2(s) = \frac{l_n^2(r + u)}{G(u)} = \frac{1}{n} \sum \left\{ \ln \left( \frac{f(x_i + a + \beta X_{i-1})}{f(x_i)} \right) - p(a + \beta X_{i-1}) \right\} I(r < X_{i-1} \leq r + u)
+ \frac{1}{n} \sum [p(a + \beta X_{i-1}) - p(a + \beta r)]
\times I(r < X_{i-1} \leq r + u) + p(a + \beta r) G_n(u)
\equiv [D_n(u) - d_n(u)] + E_n(u) + I_n(u).
\]

Then
\[
\sup_{0 < u \leq \delta} \frac{l_n^2(r + u)}{G(u)} \leq \sup_{0 < u \leq \delta} \left\{ \left| \frac{D_n(u) - d_n(u)}{G(u)} \right| + \left| \frac{E_n(u)}{G(u)} \right| + \left| \frac{G_n(u)}{G(u)} - 1 \right| \right\} + p(a + \beta r).
\]

Note that $p(a + \beta r) = \int \ln [f(x + a + \beta r)/f(x)] dF(x) < 0$ by the condition (M1). Fix an $\eta > 0$ such that $\gamma = [-p(a + \beta r) - \eta(2 + |p(a + \beta r)|)]/2 > 0$. Note that
\[
\sup_{B/|n < u \leq \delta} \left| \frac{E_n(u)}{G(u)} \right| \leq \sup_{B/|n < u \leq \delta} \left\{ \left| \frac{D_n(u) - d_n(u)}{G(u)} \right| + \left| \frac{G_n(u)}{G(u)} - 1 \right| \right\} + p(a + \beta r).
\]

By the Proposition 1 and the continuity of $p$ which follows from the conditions (C1)–(C2), we can see that $\sup_{B/|n < u \leq \delta} \left| \frac{E_n(u)}{G(u)} \right|$ goes to zero in probability for sufficiently
small \( \delta > 0 \). Let

\[
\mathcal{A} = \left\{ \sup_{B_n < u \leq \delta} \left[ \frac{D_n(u) - d_n(u)}{G(u)} + \frac{E_n(u)}{G(u)} + \frac{G_n(u)}{G(u)} - 1 \right] \left| p(a + br) \right| \right\} < \eta (2 + \left| p(a + br) \right|) \right\},
\]

\[
\mathcal{B} = \left\{ \sup_{B_n < u \leq \delta} \left| \frac{D_n(u) - d_n(u)}{G(u)} \right| < \eta \right\},
\]

\[
\mathcal{C} = \left\{ \sup_{B_n < u \leq \delta} \left| \frac{E_n(u)}{G(u)} \right| < \eta \right\},
\]

\[
\mathcal{D} = \left\{ \sup_{B_n < u \leq \delta} \left| \frac{G_n(u)}{G(u)} - 1 \right| < \eta \right\},
\]

\[
\mathcal{E} = \left\{ \sup_{B_n < u \leq \delta} \left| \frac{\ln^2(r + u)}{G(u)} \right| < -\gamma \right\}.
\]

Observe that \( \mathcal{B} \cap \mathcal{C} \cap \mathcal{D} \subset \mathcal{A} \) and Eq. (36) implies \( \mathcal{A} \subset \mathcal{E} \). Hence, by Eqs. (28), (30), (37) and choosing \( \delta > 0 \) sufficiently small,

\[
P(\mathcal{E}) \geq P(\mathcal{A}) \geq P(\mathcal{B} \cap \mathcal{C} \cap \mathcal{D}) > 1 - \epsilon.
\]

This completes the proof of Eq. (33).

Thus, \( \forall \epsilon > 0 \), there exists \( \gamma > 0 \) and \( \delta > 0 \) sufficiently small such that with Eqs. (32) and (33),

\[
P \left( \sup_{B_n < |s-r| \leq \delta} \left[ \frac{L_n(\theta, s) - L_n(\theta, r)}{G(|s-r|)} \right] < -\gamma \right)
\]

\[
\geq P \left( \sup_{B_n < |s-r| \leq \delta} \left| \frac{I_1^n(\theta)}{G(|s-r|)} \right| < \gamma, \right.
\]

\[
\sup_{B_n < |s-r| \leq \delta} \left| \frac{I_2^n(s)}{G(|s-r|)} \right| < -2\gamma \right) \geq 1 - P \left( \sup_{B_n < |s-r| \leq \delta} \left| \frac{I_1^n(\theta)}{G(|s-r|)} \right| \geq \gamma \right) \]
This ends the proof of Eq. (31) and hence of Theorem 2 in the case of \( p = d = 1 \).

For the case of \( p > 1 \) and \( 1 \leq d \leq p \), a suitably modified version of the above proof goes through. The main modification is located in proving Eq. (33). By condition (M1), as Chan’s discussion, there exists \( \gamma > 0 \) such that \(|(\theta_2 - \theta_1)'Z| > 0 \) for all \( Z \) satisfying \( |Z - Z^*| \leq \gamma \). In the decomposition of Eq. (35), we shall use \( Z_i^* \), the vector obtained by substituting \( r \) in the \((d + 1)\)th coordinate of \( Z_i \), in the second term and the third term can be replaced by \( p((\theta_2 - \theta_1)'Z_i^*)I(|Z_i^* - Z^*| \leq \gamma) \) which is strictly negative by condition (M1). This completes the proof of Proposition 1.

**Remark 6.** For the case of \( p = d = 1 \), we only need the finite third moment of the error \( \varepsilon_1 \).

5. **Asymptotic normality**

We now consider the limiting distribution of \( \hat{\theta}_n \). Recall that

\[
\psi(X_{i-1}, e_i, \theta) = \ln \frac{f(e_i + h(X_{i-1}, \theta) - h(X_{i-1}, \theta))}{f(e_i)}, \quad \theta \in \Omega,
\]

and the log likelihood ratio function is

\[
l_n(\theta) = \frac{1}{n} \sum \psi(X_{i-1}, e_i, \theta), \quad \theta \in \Omega.
\]

In the definition of the MLE \( \hat{\theta}_n \), the first \( 2p + 2 \) components of the parameter point in \( \Omega \) is treated separately from the last component and we have proved that \( \hat{\theta}_n \) is \( n \)-consistent. Thus, we need some results of \( \hat{\theta}_n(s) \) uniformly in \( s \) in the interval \([r - B/n, r + B/n]\) for some \( B, 0 < B < \infty \) which is given in the following theorem.

**Theorem 3.** Suppose that (C1) and (C2) hold. For any \( 0 < B < \infty \),

\[
\sup_{|s-r| \leq B/n} |\hat{\theta}_n(s) - \theta_c| = o_P(1).
\]

First of all, we need an analogue of the Lemma 1. Recall that \( \theta = (\theta', s)' \in \Omega \) and \( l_n(\theta_c) = l_n(\theta_c, s) \). Now, write

\[
\psi(X_{i-1}, e_i, \theta_c) = \psi(X_{i-1}, e_i, \theta_c, s).
\]

Let \( \eta > 0 \), define

\[
U_{\theta_c}(\eta) = \{ \theta^* \in K : |\theta^* - \theta_c| < \eta \}.
\]
Lemma 4. Under conditions (C1) and (C2), for any \( \vartheta_c \in K \) and its neighborhood \( U_{\vartheta}(\eta) \) in \( K \),

\[
E \sup_{s \in \mathcal{S}, \vartheta^*_c \in U_{\vartheta}(\eta)} |\psi_s(X_0, \varepsilon_1, \vartheta^*_c) - \psi_s(X_0, \varepsilon_1, \vartheta_c)| \to 0 \quad \text{as } \eta \to 0. \tag{38}
\]

Proof. It is similar to that of Lemma 1, see Qian (1996). □

Proof of Theorem 3. The following argument is an analogue of the one used in that of Theorem 1, except here we deal with the component \( \vartheta_c \) of \( \vartheta \) for all \( s \in \mathcal{S} \). Let \( \varkappa_s(\vartheta_c) = \varkappa(\vartheta_c, s) \). By the definition of the function \( \varkappa \), for any open neighborhood \( V_c \) of \( \vartheta_c \) in \( K \) and any \( \vartheta_c \in V_c \), \( s \in \mathcal{S}, \varkappa(\vartheta) < 0 \). Similar proof as that of Theorem 1, there exists \( \delta > 0 \) and \( n_1 \) such that for any \( n > n_1 \)

\[
\sup_{s \in \mathcal{S}, \vartheta_c \in K} l_m(\vartheta^*_c) < -\delta_01, \quad \text{a.s.} \tag{39}
\]

But,

\[
\sup_{s \in \mathcal{S}, \vartheta_c \in K} l_m(\vartheta_c) \geq \sup_{s \in \mathcal{S}} l_m(\vartheta_c). \tag{40}
\]

Taylor’s expansion of \( \ln f \) at \( e_l \) yields that \( \exists \gamma, |\gamma| < 1 \), such that,

\[
l_m(\vartheta_c) = \frac{1}{n} \sum \varphi(e_i) + \gamma(h_s(X_{i-1}, \vartheta_c) - h_s(X_{i-1}, \vartheta_c)) \\
\times [h_s(X_i, \vartheta_c) - h_s(X_{i-1}, \vartheta_c)].
\]

Then for any \( B, 0 < B < \infty \), the Lip(1) condition of \( \varphi \) and Eq. (12) and routine nature of the proof, yield that

\[
\sup_{|s-r| \leq B/n} |l_m(\vartheta_c)| \leq \frac{1}{n} \left[ \sum |\varphi(e_i)| + L|\vartheta_c| \sqrt{1 + |X_{i-1}|^2} \right] |\theta_c| \\
\times \sqrt{1 + |X_{i-1}|^2} I(|X_i - d - r| \leq B/n).
\]

Take expectation of both sides to obtain

\[
E \left( \sup_{|s-r| \leq B/n} |l_m(\vartheta_c)| \right) = O(n^{-1}).
\]

Thus, for any \( \varepsilon > 0 \), there exists \( n_2 \), such that as \( n > n_2, \forall |s-r| \leq B/n \),

\[
P \left( \inf_{|s-r| \leq B/n} l_m(\vartheta_c) > -\delta_01 \right) > 1 - \varepsilon.
\]

This together with Eqs. (39) and (40), \( \exists n_0 = n_1 \lor n_2 \) such that \( \forall n > n_0 \),

\[
P \left( \sup_{s \in \mathcal{S}, \vartheta^*_c \in K \setminus V_c} l_m(\vartheta^*_c) < \inf_{|s-r| \leq B/n, \vartheta_c \in V_c} \sup_{|s-r| \leq B/n} l_m(\vartheta_c) \right) > 1 - \varepsilon.
\]
Let

\[ A_z = \left\{ \sup_{x \in \mathcal{B}} \sup_{\theta^* \in \mathcal{K} \setminus V_c} l_{n}(\theta^*_c) \leq \inf_{|x - r| \leq B/n} \sup_{\theta \in V_c} l_{n}(\theta) \right\}. \]

Then, on, \( A_z \), \( \forall n \geq n_0 \),

\[ \theta_{cn}(s) \in V_c, \quad \forall |x - r| \leq B/n. \]

Thus, by the arbitrary of \( V_c \),

\[ \sup_{|x - r| \leq B/n} |\theta_{cn}(s) - \theta_c| = o_p(1). \]

This ends the proof of Theorem 3.

Before stating the next theorem, recall that \( l_{n}(\theta_c) = l_n(\theta_c, s) \), and for \( x \in \mathcal{R}^p \),

\[ h_t(x) = \frac{\partial}{\partial \theta^*}(h_t(x, \theta_c)) = (Z'1(x_d \leq s), Z'1(x_d > s))'. \]

Let \( u_i(\theta_c) = -\varphi(X_i - h_t(X_{i-1}, \theta_c))h_t(X_{i-1}) \), \( 1 \leq i \leq n \). Denote \( A_c(x) = \hat{h}_t(x)(\hat{h}_t(x))' \) and \( \Gamma = \text{cov}(\varepsilon_i)A_t(X_0) \).

**Lemma 5.** Suppose that \( 0 < I(f) < \infty \) and \( E|X_0|^2 < \infty \), then

\[ \frac{1}{\sqrt{n}} \sum_{j \leq i} u_i(\theta_c) \Rightarrow N(0, \Gamma). \]  

**Proof.** Note that with \( \mathcal{F}_{i-1} = \sigma\{X_j, j \leq i\} \),

\[ E_\theta[u_i(\theta_c)|\mathcal{F}_{i-1}] = E_\theta[\varphi(\varepsilon_i)\hat{h}_t(X_{i-1})|\mathcal{F}_{i-1}] = \hat{h}_t(X_{i-1})E_\theta(-\varphi(\varepsilon_i)) = 0 \quad \text{a.s.} \]

Therefore, for any vector \( v \in \mathcal{R}^{2p+2} \), by the finite Fisher information of \( f, v'\sum u_i(\theta_c) \) is a zero mean square integrable martingale. By ergodic theorem,

\[ \frac{1}{n} \sum_{i \leq j} v'[E[u_i(\theta_c)(u_i(\theta_c))']|\mathcal{F}_{i-1}]v = v'I(f) \frac{1}{n} \sum_{i \leq j} A_t(X_{i-1})v \to v'\Gamma v \quad \text{a.s.} \]  

Thus, the martingale central limiting theorem of Hall and Heday (1980) shows that the sum \( n^{-1/2}\sum u_i(\theta_c) \) converges weakly to the normal distribution with mean zero and variance \( v'\Gamma v \) for all \( v \in \mathcal{R}^{2p+2} \). Thus, \( n^{-1/2}\sum u_i(\theta_c) \) converges weakly to the multivariate normal distribution with mean vector zero and covariance matrix \( \Gamma \).

**Theorem 4.** Suppose that conditions (C1)–(C4) hold. Then for any \( B, 0 < B < \infty \),

\[ \sup_{|x - r| \leq B/n} \sqrt{n}(\theta_{cn}(s) - \theta_c) \Rightarrow N(0, \Gamma^{-1}). \]

As a consequence, for any \( B, 0 < B < \infty \),

\[ \sup_{|x - r| \leq B/n} \sqrt{n}(\theta_{cn}(s) - \theta_c) = o_p(1). \]
Remark 7. This result is a surprise. It holds for the supremum over the bounded interval of rate $n^{-1}$ from the threshold $r$. It is much stronger than the one in Tong (1990) which only has the asymptotic normality result for the Gaussian errors.

Proof of Theorem 4. The proof is routine but lengthy in nature. We shall just sketch the proof, for details, see Qian (1996). Note that for any $s \in \mathcal{R}$,

$$\frac{\partial}{\partial \theta_c} (l_{n}(\theta_c)) = \sum u_n(\theta_c).$$

Consider the Taylor’s expansion of $(\partial/\partial \theta_c) (l_{n}(\theta_c))$ at $\theta_c$:

$$\frac{\partial}{\partial \theta_c} l_{n}(\theta_c) = \sum u_n(\theta_c) + J_n(\theta^*_\mathrm{CM})(\theta_c - \theta_c),$$

(43)

where $\theta^*_\mathrm{CM} = \theta_c + \gamma_c(\theta_c - \theta_c)$, $\gamma_c$ is a function of $(X^*_0, \ldots, X_n, \theta_c, s)$, $|\gamma_c| < 1$ and

$$J_n(t) = \sum \frac{\partial}{\partial t} u_n(t) = \sum \phi(X_i - h_i(X_{i-1}, t) A_i(X_{i-1}), \ t \in \mathcal{R}^{p+2}.$$

Then, the definition of $\theta^*_{CM}(s)$ and Eq. (43) yield that

$$\frac{1}{\sqrt{n}} \sum u_n(\theta_c) + \frac{J_n(\theta^*_\mathrm{CM})}{\sqrt{n}}(\theta^*_{CM}(s) - \theta_c) = 0.$$

(44)

It can be verified that for any $B, \ 0 < B < \infty$,

$$\sup_{|s-r| \leq B/n} \left| \frac{J_n(\theta^*_\mathrm{CM})}{\sqrt{n}}(\theta^*_{CM}(s) - \theta_c) \right| = o_P(1).$$

(45)

and

$$E \left[ \sup_{|s-r| \leq B/n} \left| \frac{1}{\sqrt{n}} \sum [u_n(\theta_c) - u_{i'}(\theta_c)] \right| \right] = O(n^{-1/2}).$$

(46)

Thus, it follows from Eqs. (44)–(46),

$$\sup_{|s-r| \leq B/n} \left| \sqrt{n}(\theta^*_{CM}(s) - \theta_c) + \Gamma^{-1}n^{-1/2}\sum u_n(\theta_c) \right| = o_P(1).$$

By Lemma 5, Theorem 4 is proved. \hfill \Box

As a consequence of Theorems 3 and 4, we have the following uniform convergence rate of $\theta^*_{CM}(\cdot)$.

Theorem 5. Suppose that (C1)–(C4) hold, then for any $B, \ 0 < B < \infty$,

$$\sup_{|s-r| \leq B/n} |\theta^*_{CM}(s) - \theta^*_{CM}(r)| = o_P(n^{-1/2}).$$

Proof. See Qian (1996). \hfill \Box
6. Limiting distribution of the threshold estimator

In this section, we first discuss the limiting behavior for a sequence of normalized profile log-likelihood processes. Then we obtain the limiting distribution of the standardized maximum likelihood estimator of the threshold parameter.

Recall that
\[ l_n(\theta_c, s) = \frac{1}{n} \sum \ln \frac{f(X_i - h_{s}(X_{i-1}, \theta_c))}{f(\theta_i)}, \quad (\theta_c', s') \in \Omega. \]

For \( t \in \mathbb{R} \), a sequence of normalized profile log-likelihood processes is
\[ \hat{l}_n(t) = -\frac{2}{n} \left[ \ln \left( l_n(\theta_c + t/n, \theta_c + t/n) - l_n(\theta_c, \theta_c) \right) \right]. \]

Observe that in view of Theorem 3, \( \theta_c + t/n \) is an approximation of \( \theta_c \) uniformly in \( t \) over bounded sets. Thus, a natural candidate for the approximation of \( \hat{l}_n \) is \( \hat{l}_n \) defined as follows: For \( t \in \mathbb{R} \),
\[ \hat{l}_n(t) = -\frac{2}{n} \left[ \ln \left( l_n(\theta_c + t/n, \theta_c + t/n) - l_n(\theta_c, \theta_c) \right) \right] \]
\[ = -\frac{2}{n} \sum \ln \frac{f(X_i - h_{t/n}(X_{i-1}, \theta_c))}{f(\theta_i)}. \]

The following theorem states that the process \( \hat{l}_n \) is a nice approximation of \( \hat{l}_n \).

**Theorem 6.** Suppose that (C1)–(C4) hold. Then for any \( B, 0 < B < \infty \),
\[ \sup_{|t| \leq B} |\hat{l}_n(t) - \hat{l}_n(t)| = o_P(1). \]

**Proof.** Without loss of generality, assume \( r = 0 \). Decompose the concerned process in the following way:
\[ -\frac{1}{2} [\hat{l}_n(t) - \hat{l}_n(t)] = -\frac{1}{2} \left[ \hat{l}_n(t) - \sum \ln \frac{f(X_i - h_{t/n}(X_{i-1}, \theta_c))}{f(\theta_i)} \right] \]
\[ = \sum \left[ \ln \frac{f(X_i - h_{t/n}(X_{i-1}, \theta_c(t/n)))}{f(X_i - h_0(X_{i-1}, \theta_c(t/n)))} - \ln \frac{f(X_i - h_{t/n}(X_{i-1}, \theta_c))}{f(X_i - h_0(X_{i-1}, \theta_c))} \right] \]
\[ + \sum \ln \frac{f(X_i - h_0(X_{i-1}, \theta_c(t/n)))}{f(X_i - h_0(X_{i-1}, \theta_c(0)))} \]
\[ \equiv \hat{l}_{1n}(t) + \hat{l}_{2n}(t) \quad (\text{say}). \]

It suffices to show that \( \forall B < \infty \),
\[ \sup_{|t| \leq B} |\hat{l}_{1n}(t)| = o_P(1), \tag{48} \]
and
\[ \sup_{|t| \leq B} |\hat{l}_{2n}(t)| = o_P(1). \tag{49} \]
Actually, we shall prove a slightly stronger result than (48). To state this stronger result, recall that for $1 \leq i \leq n$, 
$$\varepsilon_i(\theta, s) = \varepsilon_i(\theta) = X_i - h_i(X_{i-1}, \theta_i), \ (\theta'_i, s') \in \Omega.$$
Denote 
$$p_{\text{nat}}(t) = \ln \frac{f(\varepsilon_i(t, t/n))}{f(\varepsilon_i(t, 0))}; \ R^{p+2} \rightarrow \ R, \ 1 \leq i \leq n.$$
Then, for any $i$, $1 \leq i \leq n$,
$$\hat{p}_{\text{nat}}(t) = -\varphi(\varepsilon_i(t, t/n)) \hat{h}_i(\theta_{i-1}) + \varphi(\varepsilon_i(t, 0)) \hat{h}_0(\theta_{i-1})$$
$$= [-\varphi(\varepsilon_i(t, t/n)) + \varphi(\varepsilon_i(t, 0))] \hat{h}_i(\theta_{i-1})$$
$$+ [-\varphi(\varepsilon_i(t, 0)) + \varphi(\varepsilon_i)] [\hat{h}_i(\theta_{i-1}) - \hat{h}_0(\theta_{i-1})]$$
$$- \varphi(\varepsilon_i) [\hat{h}_i(\theta_{i-1}) - \hat{h}_0(\theta_{i-1})].$$
(50)
Note that $\hat{h}_i(t) = \sum [p_{\text{nat}}(\theta_{i-1}(t/n)) - p_{\text{nat}}(\theta_i)]$. From the consequence of Theorem 5:
$$\sup_{|t| < B} |\sqrt{n}(\theta_{i-1}(t/n) - \theta_c)| = O_p(1),$$
the stronger result that will be proved is that for any $0 < C < \infty$,
$$E \left\{ \sup_{|t| < B, |t - \theta_c| < C} |\sum [p_{\text{nat}}(t) - p_{\text{nat}}(\theta_c)] \right\} = O(n^{-1/2}).$$
(51)
Denote $\theta_{cs}(t) = \theta_c + s(t - \theta_c)$ and recall that $\varphi = \hat{f}/f$. By the absolute continuity of $\ln f$,
$$p_{\text{nat}}(t) - p_{\text{nat}}(\theta_c) = \int_0^1 \hat{p}_{\text{nat}}(\theta_{cs}(t))(t - \theta_c) \, ds \equiv s_{1it}(t) + s_{2it}(t) + s_{3it}(t) \text{ (say),}$$
where, by Eq. (50),
$$s_{1it}(t) = \int_0^1 [-\varphi(\varepsilon_i(\theta_{cs}(t), t/n)) + \varphi(\varepsilon_i(\theta_{cs}(t), 0))]$$
$$\times \hat{h}_i(\theta_{i-1})(t - \theta_c) \, ds,$$
$$s_{2it}(t) = \int_0^1 [-\varphi(\varepsilon_i(\theta_{cs}(t), 0)) + \varphi(\varepsilon_i)]$$
$$\times [\hat{h}_i(\theta_{i-1}) - \hat{h}_0(\theta_{i-1})](t - \theta_c) \, ds,$$
$$s_{3it}(t) = -\varphi(\varepsilon_i)[\hat{h}_i(\theta_{i-1}) - \hat{h}_0(\theta_{i-1})](t - \theta_c).$$
Now, $\varphi$ being Lip(1) and Eq. (12) imply that uniformly in all $t, i$ and $s$,
$$|\varphi(\varepsilon_i(\theta_{cs}(t), t/n)) - \varphi(\varepsilon_i(\theta_{cs}(t), 0))|$$
$$\leq L|\hat{h}_i(\theta_{cs}(t), s) - \hat{h}_0(\theta_{cs}(t))|$$
$$\leq L|\theta_{cs}(t)| \sqrt{1 + |X_{i-1}|^2} I(|X_{i-1}| \leq |t|/n)$$
$$\leq L(|\theta_c| + |t - \theta_c|) \sqrt{1 + |X_{i-1}|^2} I(|X_{i-1}| \leq |t|/n).$$
Therefore, by Eqs. (11) and (4),

\[
\sup_{|t| \leq B, |\theta - \mu| \leq C} |s_1(t)| = \sup_{|t| \leq B, |\theta - \mu| \leq C} \left| \sum_{j=1}^{n} X_i \right| = O(n^{-1/2}).
\]

Thus, the stationarity of \( \{X_i\} \) and Remark 1 yield that

\[
E \left\{ \sup_{|t| \leq B, |\theta - \mu| \leq C} \left| \sum_{l=1}^{n} s_{1t}(t) \right| \right\} = O(n^{-1/2}). \tag{52}
\]

For \( s_{20}(t) \) and \( s_{30}(t) \), the proofs are analogous to that of Eq. (52) and the followings hold:

\[
E \left\{ \sup_{|t| \leq B, |\theta - \mu| \leq C} \left| \sum_{l=1}^{n} s_{2t}(t) \right| \right\} = O(n^{-1}). \tag{53}
\]

\[
E \left\{ \sup_{|t| \leq B, |\theta - \mu| \leq C} \left| \sum_{l=1}^{n} s_{3t}(t) \right| \right\} = O(n^{-1/2}). \tag{54}
\]

Therefore, Eqs. (52)–(54) imply Eq. (51) and, hence, Eq. (48).

Now, it remains to prove Eq. (49). To prove Eq. (49), the result of Theorem 5 will be needed. Recall that

\[
\hat{\theta}_n(t) = \sum_{j=1}^{n} \ln \frac{f(X_i - h_0(X_{i-1}, \theta(t/n)))}{f(X_i - h_0(X_{i-1}, \theta(0)))}.
\]

For any \( t \in \mathcal{R} \), denote \( \theta_{cn}(t) = \theta_{cn}(0) + t(\theta_{cn}(t/n) - \theta_{cn}(0)) \). Then the absolutely continuity of \( \ln f \) gives

\[
\hat{\theta}_n(t) = \sum_{j=1}^{n} \int_{0}^{1} \left[ -\varphi(X_i - h_0(X_{i-1}, \theta_{cn}(t))) \hat{h}_0(X_{i-1}) \right] (\theta_{cn}(t/n) - \theta_{cn}(0)) \, dt
\]

\[
= \sum_{j=1}^{n} \int_{0}^{1} \left[ -\varphi(X_i - h_0(X_{i-1}, \theta_{cn}(t))) + \varphi(\varepsilon_i) \right] \, dt \hat{h}_0(X_{i-1})
\]

\[
\times (\theta_{cn}(t/n) - \theta_{cn}(0)) - \sum_{j=1}^{n} \varphi(\varepsilon_i) \hat{h}_0(X_{i-1})(\theta_{cn}(t/n) - \theta_{cn}(0))
\]

Thus, the Lip(1) of \( \varphi \), Theorem 5, Eqs. (10) and (4) imply

\[
\sup_{|t| \leq B} \{|\text{The first term of } \hat{\theta}_n(t)|\} \leq \sup_{|t| \leq B} L|\theta_{cn}(t/n) - \theta_{cn}(0)||\hat{h}_0(X_{i-1})|
\]

\[
\times \sum_{j=1}^{n} \int_{0}^{1} |h_0(X_{i-1}, \theta_{cn}(t)) - h_0(X_{i-1}, \theta_{c})| \, dt
\]
\[
\leq \mathop{O_p}(n^{-1/2}) \int_0^1 \sup_{|t| \leq B} |\varphi_{cn}(t) - \varphi_c| \, dt \sum (1 + |X_{i-1}|^2)
\]
\[
= \mathop{O_p}(n^{-1/2}) \mathop{O_p}(n^{-1/2}) \mathop{O_p}(n) = \mathop{O_p}(1).
\]
For the second term of \( \tilde{I}_{2n}(t) \), since \( \sum \varphi(c_i) \tilde{h}_0(X_{i-1}) \) is a zero mean square integrable martingale process,
\[
|\sum \varphi(c_i) \tilde{h}_0(X_{i-1})| = \mathop{O_p}(n^{1/2}).
\]
But Theorem 5 implies
\[
\sup_{|t| \leq B} |\varphi_{cn}(t/n) - \varphi_{cn}(0)| = \mathop{O_p}(n^{-1/2}).
\]
Therefore, Eq. (49) follows from the above two equations and Eq. (55). Hence, the proof of Theorem 6 is completed.

Thus, the limiting process of the sequence \( \{\tilde{I}_n(t), t \in \mathcal{R} \} \) is the same as that of the process \( \hat{I}_n \) by Theorem 6.

To describe the limiting process of \( \tilde{I}_n \), let \( \{\tilde{I}^{(1)}(t), t \geq 0\} \) and \( \{\tilde{I}^{(2)}(t), t \geq 0\} \) be two independent compound Poisson processes, both with rate \( g(r) \), the marginal density of \( X_0 \) evaluated at the threshold \( r \) under the true parameter \( \theta \), such that the following hold: \( \tilde{I}^{(1)}(0) = \tilde{I}^{(2)}(0) = 0 \), a.s.; the distribution of jumps being given by the conditional distribution of \( \zeta_1 = -2 \ln[f(c_1 + (\theta_2 - \theta_1)^T \mathbf{Z}_1)/f(c_1)] \) given \( X_{i-d} = r^+ \) and the conditional distribution of \( \zeta_2 = -2 \ln[f(c_1 - (\theta_2 - \theta_1)^T \mathbf{Z}_1)/f(c_1)] \) given \( X_{i-d} = r^- \), respectively, where \( \mathbf{Z}_i = (1,X_{i-1},\ldots,X_{i-p})^T \). The former conditional distribution is the limiting conditional distribution of \( \zeta_1 \) given \( r < X_{i-d} < r + \delta \) as \( \delta \downarrow 0 \) and the latter is that of \( \zeta_2 \) given \( r - \delta < X_{i-d} < r \) as \( \delta \downarrow 0 \). Note that in the Gaussian error case, this reduces to the results of Chan (1993). Let \( v_1 \) and \( v_2 \) be the induced measures of \( \zeta_1 \) and \( \zeta_2 \), respectively.

**Theorem 7.** Suppose (C1) and (C2) hold. Then \( \{\tilde{I}_n(t), t \geq 0\}, \{\tilde{I}_n(t), t \geq 0\} \) converges weakly to \( \{\tilde{I}^{(1)}(t), t \geq 0\}, \{\tilde{I}^{(2)}(t), t \geq 0\} \) in \( D[0, \infty) \times D[0, \infty) \), the product space being equipped with the product Skorokhod metric. (See Kushner, 1984, pp. 29–33.)

By an immediate corollary to a theorem of Whitt (1980), it is enough to prove the weak convergence on \( D[0,T] \times D[0,T] \) for every \( T > 0 \). Furthermore, it suffices to treat the case \( t \geq 0 \), the opposite case being analogous.

Now for \( 0 \leq t \leq T \), we rewrite
\[
\tilde{I}_n(t) = -2 \sum \ln \frac{f(c_i + (\theta_2 - \theta_1)^T \mathbf{Z}_i)}{f(c_i)} I(r < X_{i-d} \leq r + t/n)
\]
we need to show that
\[
\tilde{I}_n \Rightarrow \tilde{I}^{(1)} \quad \text{for every } T > 0.
\]
We shall adopt the modern semimartingale method (Jacod and Mémin, 1980, Theorem (5.4), p. 244) used by Pflug (1983, Section 2) to prove the weak convergence. The method mentioned above is different from the one, directing averaging method from Kushner (1984), used by Chan (1993).

Now, we give the detail modification of the proof of Pflug (1983) to the present case for stationary process. For every large $\alpha > 0$, define the semimartingale with bounded jumps as follows:

$$M_n(t) = -\sum_{X_{i-1} \in A_n(t)} \ln\left(\frac{f(e_i + (\theta_2 - \theta_1)'Z)}{f(e_i)}\right)^{2a} \text{ for large } \alpha > 0,$$

where $x^b = \text{sgn}(x) \min(|x|, b)$ and $A_n(t) = (r, r + t/n]$. Since $\tilde{I}_n$ is a pure jump process in $D[0, \infty)$, the similar argument as in Chan, it follows the tightness of the process $\tilde{I}_n$.

To prove that $M_n(t)$ converges to a pure jumps process with independent increments, by a theorem of Jacod and Memin (1980) (Theorem (5.4), p. 244) or the results of modern semimartingale theory (see also, Jacod and Shiryaev, 1987, Ch. 8), it is sufficient to consider the triplet of local characteristics of $M_n(t)$. Since $M_n(t)$ is a pure jumps process, it is characterized by its random jump measure. We shall show that the compensating measure of this measure converges in probability to that of a compound Poisson process with rate $g(r)$ and jump measure $v_1^a$, where $v_1^a$ is obtained from $v_1$ by putting the mass outside from $[-2a, 2a]$ at the endpoints of this interval.

Let

$$\psi^a(y, z) = -\ln\left(\frac{f(y + (\theta_2 - \theta_1)'z)}{f(y)}\right) I\left(\left|\ln\left(\frac{f(y + (\theta_2 - \theta_1)'z)}{f(y)}\right)\right| \leq 2a\right).$$

and

$$h(y, z, u) = \begin{cases} z_{d+1} - r & \text{if } r < z_{d+1} \text{ and } \psi^a(y, z) \leq u, \\ 1 & \text{otherwise.} \end{cases}$$

Then the compensating measure of the jump measure $\sum_{s \leq t} I(\Delta M(s) \in (-\infty, u])$ is (see Jacod and Shiryaev, 1987, p. 99)

$$\Pi_n(t, u) \equiv \Pi_n(\omega_n(0, t], (-\infty, u]) = \sum_{s \leq t} \ln\left(1 - H\left(h(e_i, Z_i, u) \wedge \frac{t}{n}, u\right)\right),$$

We show that for every $t \in [-T, T]$ and every $u \in \mathcal{R}$ with $v_1^a(\{u\}) = 0$, $\Pi_n(\omega_n(0, t], (-\infty, u])$ converges in probability to the constant $\mu(t, u) \equiv g(r)v_1^a(-\infty, u]$.

Note that for $t < 1$, $H(t, u)$ is the distribution function of $h(y, z, u)$ under $P$. Hence if $n > T$, by the stationary property,

$$E\Pi_n(\omega_n(0, t], (-\infty, u]) = nE \ln\left(1 - H\left(h(e_1, Z_1, u) \wedge \frac{t}{n}, u\right)\right) = n \ln(1 - H(t/n, u)(1 - H(t/n, u))$$
\[ + \int_0^{t/n} \ln(1 - H(y,u)) \, dH(y,u) \]
\[ = nH(t/n,u). \]

Since \( H(t/n,u) = P\{\psi^a(t/n, Z_1) \leq u | A_n(t/n)\} P\{A_n(t/n)\} \), it follows from the assumption (M2):
\[
\mu_n(t,u) \equiv nH(t/n,u) \to g(r)v'_1(-\infty, u] \equiv \mu(t,u), \tag{58}
\]
if \( u \) is a continuity point of \( v'_1(-\infty, \cdot] \). By condition (M2), it can be verified that
\[ E(\Pi_n(t,u) - \mu(t,u))^2 \to 0. \tag{59} \]

To see Eq. (59),
\[ E(\Pi_n(t,u) - \mu(t,u))^2 = E(\Pi_n(t,u) - \mu_n(t,u) + \mu_n(t,u) - \mu(t,u))^2 \]
\[ \leq 2[E(\Pi_n(t,u) - \mu_n(t,u))^2 + (\mu_n(t,u) - \mu(t,u))^2]. \]

By Eq. (58), the second term above goes to zero. Now, we prove that the first term above goes to zero also. For fixed \( t \) and \( u \), denote \( \psi_i \equiv \psi_i(t,u) = \ln(1 - H(\rho_i, Z_i,u) \wedge t/n,u) \), then \( \Pi_n(t,u) = \sum \psi_i(t,u) \) and
\[
E(\Pi_n(t,u) - \mu_n(t,u))^2 = n\Var(\psi_1(t,u)) + \sum_{i \neq j} \Cov(\psi_i(t,u), \psi_j(t,u)).
\]

By the integration by parts,
\[ \Var(\psi_1(t,u)) = 2[\ln(1 - H(t/n))(1 - H(t/n,u)) + H(t/n,u)], \]
and by condition (M2) and conditioning argument,
\[
|\Cov(\psi_i(t,u), \psi_j(t,u))| = |E\{(\psi_i(t,u) - E\psi_i(t,u))(\psi_j(t,u) - E\psi_j(t,u))\}| \leq E(\psi_i(t,u) - E\psi_i(t,u)) \]
\[ \times \left| \int \left| \int \ln(1 - H(h(y, Z_i,u) \wedge t/n,u)) \, dF(y) \right| \right| \]
\[ \times |P_j(Z_1) \, d\pi - \pi(\, d\pi)| \leq C\rho^j E|\psi_i - E\psi_i| \sqrt{\Var(\psi_i)} \]
\[ \leq C\rho^j [\Var(\psi_i) + \sqrt{\Var(\psi_i)}|E\psi_i|]. \]

By Eq. (58), it can be verified that \( n\Var(\psi_i) \to 0 \), and with the fact \( \sum_{i < j} \rho^{j-i} = O(n) \),
\[ \sum_{i \neq j} \Cov(\psi_i(t,u), \psi_j(t,u)) \to 0. \]

Hence Eq. (59) is proved, i.e., the convergence in probability of the compensating measure is shown. Thus, we have established \( M_n \Rightarrow \tilde{\iota}_1^{(1)} \), where \( \tilde{\iota}_1^{(1)} \) is the compound Poisson process with rate \( g(r) \) and jump measure \( v'_1 \).
Note that both jump measures \( v_1, v_2 \) have positive means and are absolutely continuous. The rest part of the proof is analogous to that in Pflug (1983) except we use
\[
C_a = \left\{ x \in D[0, \infty) \mid \sup_{s \leq T} x(s) \leq -a \text{ or } \sup_{s > t_x} x(s) \leq a/2 \right\},
\]
where \( t_x = \inf \{ s : x(s) \geq a \} \). Therefore Theorem 7 is proved.

Denote the compound Poisson process by
\[
\{ \tilde{I}(t) : t \in \mathcal{R} \} = \{ \tilde{I}^{(1)}(-t)I(t \leq 0) + \tilde{I}^{(2)}(t)I(t \geq 0), t \in \mathcal{R} \}.
\]
Then the two random walks associated with the compound Poisson processes tend to \(+\infty\), a.s. Hence there exists an unique random interval \([M_-, M_+]\) on which the process \( \{ \tilde{I}(t) : t \in \mathcal{R} \} \) attains its global minimum almost surely.

Now, we are ready to state the limiting distribution theorem for the threshold parameter.

**Theorem 8.** Suppose that conditions (C1)–(C4), (M2) and (M2) hold. Then, \( n(\hat{r}_n - r) \) converges weakly to \( M_- \). Furthermore, \( n(\hat{r}_n - r) \) is asymptotically independent of \( \sqrt{n}(\hat{\theta}_n - \theta) \) and the latter is asymptotically normal with a distribution same as that for the case when \( r \) is known.

**Proof.** It is the same as that of Theorem 2 in Chan (1993).

**Acknowledgements**

I would sincerely like to express my deep gratitude to my thesis advisor, Professor Hira L. Koul, for all his expert guidance, numerous useful suggestions and continuous encouragement. I also thank an anonymous referee for many useful comments and suggestions.

**References**


