This is to certify that the

dissertation entitled

Parameter estimation in non-linear time series: Random coefficient autoregressive and self-exciting threshold models

presented by

Lianfen Qian

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Statistics

Department of Statistics and Probability

HKoul
Major professor

Date June 17, 1996

PARAMETER ESTIMATION IN NON-LINEAR TIME SERIES: RANDOM COEFFICIENT AUTOREGRESSIVE AND SELF-EXCITING THRESHOLD MODELS

By

Lianfen Qian

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1996

ABSTRACT

PARAMETER ESTIMATION IN NONLINEAR TIME SERIES: RANDOM COEFFICIENT AUTOREGRESSIVE AND SELF-EXCITING THRESHOLD MODELS

By

Lianfen Qian

This dissertation studies the parameter estimation in two nonlinear time series models: *Random coefficient autoregressive* and *self-exciting threshold autoregressive* models.

For the random coefficient autoregressive model of order p (RCAR(p)), we discuss a class of minimum distance (MD) estimators for the true unknown parameters. These estimators are defined via certain weighted empiricals as in Koul (1986). The class of estimators considered includes the least absolute deviation estimator and an analogue of the Hodges-Lehmann estimator. The dissertation contains a proof of the asymptotic normality of these estimators and a simulation study. It is observed that RCAR(2) model with the Hodges-Lehmann type estimator fits the Canadian lynx data at least as well as with the least square estimator.

For the first order stationary ergodic *self-exciting threshold autoregressive* model with single threshold parameter, we show that the maximum likelihood estimators of the underlying true parameters are strongly consistent under some regularity conditions on the error density. Then, we prove that the maximum likelihood estimator of the threshold parameter is $n-$consistent if the threshold parameter is the discontinuity point of the autoregressive function. Further, we derive the asymptotic normality of the estimators of the coefficient parameters. We also obtain a simple approximation of a sequence of normalized log-likelihood processes, hence prove the tightness of the sequence of normalized log-likelihood processes.

To: **Jenny Yao ——— my threshold daughter**

Qingchuan Yao — my beloved husband

ACKNOWLEDGMENTS

I want to express my deep gratitude to Professor Hira L. Koul, my thesis advisor, for all his continuous encouragement, expert guidance, generous support and extreme patience during the preparation of this dissertation. His love of statistics and devotion to research have served as the main source of inspiration to my research.

I also like to thank Professors Mandrekar, Gilliland and Shapiro for serving on my thesis committee. My special thanks go to Professor Mandrekar for his carefully reading my dissertation and making positive suggestions that improve the presentation, Professor Gilliland for training me working for the statistical consulting service, Professor Shapiro for preparing me the strongly mathematical background.

I appreciate the support of the Department of Statistics and Probability in the last three years for pursuing my Ph.D. in Statistics. I would like to thank all the faculty members and staff in the department for all of their help during my graduate work at Michigan State University. Finally, I want to express my deep thanks to my husband, Qingchuan, for his continuous support, encouragement and patience. Major portion of this research was supported by the **NSF Grant DMS 94-02904**.

Contents

List of Tables

List of Figures

Introduction

Nonlinear time series analysis has achieved a rapid development in the past two decades. The two main factors expediting nonlinear time series model building are: Essentially complete theory of linear time series analysis and some complicated dynamics phenomena that can not be modeled by linear time series models. Many different types of nonlinear time series models have been studied in literature (see Priestly (1980), Pagan (1980), Nicholls and Quinn (1982) and Tong (1983, 1990). In this dissertation, we will focus on the parameter estimations of two nonlinear time series models: The *random coefficient autoregressive* and the *self-exciting threshold autoregressive* models.

The first part of this dissertation is concerned with the *random coefficient autoregressive* model of order p ($RCAR(p)$) in which one observes $\{X_i, i \in \mathbf{Z}\}$ satisfying

$$X_i = (\boldsymbol{\theta} + \boldsymbol{Z}_i)^T \boldsymbol{Y}_{i-1} + \epsilon_i, \quad i \in \mathcal{Z}, \tag{0.1}$$

for some $\boldsymbol{\theta} \in \mathcal{R}^p$, where $\{\epsilon_i,\ i \in \mathcal{Z}\}$ and $\{\boldsymbol{Z}_i,\ i \in \mathcal{Z}\}$ are independent sequences of independent identically distributed random vectors with respective distribution functions F and G. Here $\mathbf{Y}_0 := (X_0, \ldots, X_{1-p})^T$ is an observable random vector independent of $\{\epsilon_i\}$, $\boldsymbol{Y}_{i-1} := (X_{i-1}, \ldots, X_{i-p})^T$, $\boldsymbol{Z}_i := (Z_{i1}, \ldots, Z_{ip})^T$, $p \geq 1$ is a known integer and $\mathcal{Z}$ denotes the set of all integers. For the importance of these models in time series analysis, see the Lecture Notes by Nicholls and Quinn (1982) and the monograph by Tong (1990). RCAR(p) models include the well known AR(p) models (take $\boldsymbol{Z}_i$ to be degenerate at $\mathbf{0}$).

The problem of interest here is to estimate the unknown parameter $\boldsymbol{\theta}$ based on $\{\mathbf{Y}_0, X_1, ...X_n\}$. We will study the Minimum Distance (MD) estimators $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ which are based on minimizing certain types of distance functions $M_g(\cdot)$, called *dispersion*, related to the data and the parameter, for measurable function $\boldsymbol{g}$ from $\mathcal{R}^p$ to $\mathcal{R}^p$. The importance of this methodology in linear models is discussed in Koul (1992, Chapters 5 and 7). These estimators have many desirable properties, including consistency, robustness against outliers in the error, efficiency and asymptotic normality in AR models.

Koul (1992) discussed the asymptotic behavior of the estimators $\hat{\boldsymbol{\theta}}$ under the AR(p) setup. We obtain the asymptotic normality of $\hat{\boldsymbol{\theta}}$ under the RCAR(p) setup. The method of proof is similar to that of Koul (1992) which requires obtaining the asymptotic uniform quadraticity of $M_g(\boldsymbol{t})$ and showing $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = O_P(1)$.

This part of the material is organized as follows. Assumptions and statements of main theorems appear in Section 1.2 while proofs appear in Section 1.3. Section 1.4 contains a simulation study and an application to the Canadian lynx data. The simulation study shows that the Hodges-Lehmann type estimator is as good as the Least Square (LS) estimator and Huber estimator (HE). For the additive effects outliers model, Dhar (1990, 1991) established the robustness of the MD estimator. Later, Dhar (1993) working on the AR(p) model showed through simulation that the MD estimator, with $H(x) = x$ and $\boldsymbol{g}(\mathbf{y}) = \mathbf{y}$, has the smallest absolute bias even for the small sample size $n = 10$ and the smallest mean square error for $n = 50$ and 100, under the logistic error distribution.

As an application, we fit the RCAR(2) model with MD estimators to the annual trappings of the Canadian lynx over the years 1821-1934. The result shows that this model provides an acceptable alternative to the more widely adopted class of AR models.

A *self-exciting threshold autoregressive* model is a piecewise linear model. It is fitted by different linear autoregressive functions in both past variables and parameters for different subsets of data. Tong (1977) first mentioned the usefulness of these models. Later, Tong (1978a, 1978b, 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 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.

The second part of the dissertation is concerned with the large sample behavior of maximum likelihood estimators in a special SETAR model, called SETAR(2;1,1), defined as follows:

$$X_i = h(X_{i-1}, \boldsymbol{\theta}) + \epsilon_i, \quad i \geq 1, \tag{0.2}$$

for some $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, r)^T \in \mathcal{R}^5$, where $\boldsymbol{\theta}_1 = (a_0, a_1, b_0, b_1)^T \in \mathcal{R}^4$ and for any $x \in \mathcal{R}$,

$$h(x, \boldsymbol{\theta}) = (a_0 + a_1 x) I(x \leq r) + (b_0 + b_1 x) I(x > r).$$

Here, the errors $\{\epsilon_i\}$ are independent and identically distributed random variables with mean zero, finite nonzero variance and ϵ_1 is independent of X_0. The parameter r, the location of the change of the autoregressive function h, is called the *threshold*.

Define a region Θ of parameters as follows.

$$\Theta = \{\boldsymbol{\vartheta} = (\alpha_0, \alpha_1, \beta_0, \beta_1, s)^T \in \mathcal{R}^5 : \ \alpha_1 < 1, \beta_1 < 1, \ \alpha_1 \, \beta_1 < 1\}. \tag{0.3}$$

Petruccelli and Woolford (PW)(1984) proved that the model (0.2) with $a_0 = b_0 = 0$, $r = 0$ is ergodic if and only if $\boldsymbol{\theta} \in \Theta$. Note that Θ is much wider compared to the region of stationarity of AR(1) model. Chan, Petruccelli, Tong and Woolford (CPTW)(1985) continued PW's work and found some other sufficient conditions on the parameters for $\{X_i\}$ in model (0.2) to be ergodic. Note that the process $\{X_i\}$ defined in (0.2) is a Markov chain. From ergodicity, one can readily obtain stationarity if the measure induced by the initial distribution of the Markov chain is the same as the invariant measure of ergodicity. Since this part of the dissertation discusses the asymptotic properties of the maximum likelihood estimators, it will be assumed that the initial measure is equal to the invariant measure. That is, we will work with stationary and ergodic SETAR (2;1,1) model.

For the case of the *threshold* r having only finite number of possible values and assuming Gaussian errors, Tong (1983) constructed a maximum likelihood estimator of the unknown parameters using Akaike Information Criterion (1973). If the *threshold* r is known, CPTW (1985) obtained the consistency and asymptotic normality property of the least-square estimators of the coefficient parameter $\boldsymbol{\theta}_1$ under some regularity conditions. But in practice, the *threshold* parameter r is unknown and can take infinitely many values in $\mathcal{R}$. In this case, Petruccelli (1986) proved that the conditional least-square estimator (CLSE) of $\boldsymbol{\theta}$ is strongly consistent for the SETAR(2;1,1) model. Chan (1993) developed the strong consistency of the same CLSE in a general SETAR model. Furthermore, he claimed that he obtained the limiting distribution of the CLSE of the *threshold* under some regularity conditions on the errors.

We derive the asymptotics of a maximum likelihood estimator (MLE) of the underlying parameter $\boldsymbol{\theta}$ in model (0.2), when the errors have a density f, not necessarily to be Gaussian. Unlike the popular AR model, the likelihood function of the SETAR(2;1,1) model is not differentiable with respect to the parameters. Actually, it is not continuous in the *threshold* parameter in general. Thus the routine method of computing maximum likelihood estimator can not be adopted. Instead, in Chapter 2, Section 2.1 discusses a profile maximum likelihood method to obtain the MLE $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\theta}}_{1n}^T, \hat{r}_n)^T$ of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, r)^T$. Section 2.2 states assumptions for latter use. In Section 2.3, Theorem 2.3.1 shows that the MLE is strongly consistent. If the threshold parameter r is a discontinuity point of the autoregressive function h, then the maximum likelihood estimator $\hat{r}_n$ is not only consistent, but also n-consistent as shown in Theorem 2.4.1, i.e. $|n(\hat{r}_n - r)|$ is bounded in probability. In Chapter 3, we develop the asymptotic normality of the coefficient parameter estimator $\hat{\boldsymbol{\theta}}_{1n}$ and some more byproduct results. In Chapter 4, as a consequence of the n-consistency of $\hat{r}_n$, the suitably normalized log-likelihood sequence of processes $\{\hat{l}_n\}$ (see section 4.1) is shown to be approximated by a sequence of simpler processes which describe the log-likelihood under known coefficient parameter $\boldsymbol{\theta}_1$. Through the latter processes, the tightness of $\{\hat{l}_n\}$ is derived. It is expected that this result will be useful in obtaining the limiting distribution of the standardized maximum likelihood estimator of the threshold parameter.

Notation. Throughout this dissertation, the symbol $\boldsymbol{\theta}$ is the fixed unknown underlying parameter, the function f is the $p.d.f.$ of ϵ_1 and F denotes the distribution function corresponding to f. The expectation under $\boldsymbol{\theta}$ 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 Γ is denoted by $N(0,\Gamma)$. Let $\mathcal{R}$ be the real line $(-\infty,\infty)$, and $\bar{\mathcal{R}} = \mathcal{R} \cup \{-\infty,\infty\}$, then the compactness of the set $\bar{\mathcal{R}}$ is under the metric $d(\cdot,\cdot)$ defined by $d(x,y) = |\arctan x - \arctan y|$. A function φ satisfies the Lip (1) if $\forall\ x,\ y \in \mathcal{R}$, $\exists\ L \geq 0$, such that

$$|\varphi(x) - \varphi(y)| \leq L|x-y|.$$

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 γ_i, $i = 1,2,\ldots$ stand for absolute constants and they can have different values in different places. The notation $\boldsymbol{x}^T\boldsymbol{y}$ stands for the inner product of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. For any matrix $M = (m_{ij})$, $\|M\| = \sum_{i,j} |m_{ij}|$, M^T stands for the transpose of M, $det(M)$ and $adj(M)$ stand for the determinant, adjoint matrix of M, respectively. Vectors of dimension more than one are denoted by bold face letters. The index $\boldsymbol{i}$ in the summation varies from 1 to $\boldsymbol{n}$ unless specified otherwise.

Part I

Random Coefficient Autoregressive Model

Chapter 1

Minimum distance estimation

1.1 Introduction

This part of the dissertation considers the random coefficient autoregressive model of order p ($RCAR(p)$) in which one observes $\{X_i, i \in \mathcal{Z}\}$ satisfying

$$X_i = (\boldsymbol{\theta} + \boldsymbol{Z}_i)^T \boldsymbol{Y}_{i-1} + \epsilon_i, \quad i \in \mathcal{Z}, \tag{1.1}$$

for some unknown $\boldsymbol{\theta} \in \mathcal{R}^p$ and for independent sequences $\{\epsilon_i, i \in \mathcal{Z}\}$ and $\{\boldsymbol{Z}_i, i \in \mathcal{Z}\}$ of independent identically distributed random vectors with distribution functions F and G, respectively. Also, it is assumed that $E\epsilon_1 = 0$ and $E\epsilon_1^2 = \sigma_F^2 > 0$, $E\boldsymbol{Z}_1 = 0$ and $E\boldsymbol{Z}_1\boldsymbol{Z}_1^T = \Sigma \geq 0$. Here, $\mathbf{Y}_0 := (X_0, \ldots, X_{1-p})^T$ is an observable random vector independent of $\{\epsilon_i\}$, $\boldsymbol{Y}_{i-1} := (X_{i-1}, \ldots, X_{i-p})^T, \boldsymbol{Z}_i := (Z_{i1}, \ldots, Z_{ip})^T$, $p \geq 1$ is a known integer and $\mathcal{Z}$ denotes the set of all integers. This model includes the well known AR(p) model (take $\boldsymbol{Z}_i$ to be degenerate at $\mathbf{0}$). For the importance of RCAR(p) models in time series analysis, see the Lecture Notes by Nicholls and Quinn (1982) and the monograph by Tong (1990) .

The problem of interest here is to estimate the unknown parameter $\boldsymbol{\theta}$ in model (1.1) based on $\{\mathbf{Y}_0, X_1, ...X_n\}$. We study Minimum Distance (MD) estimators of $\boldsymbol{\theta}$ which are based on minimizing some types of distance functions, called *dispersion*, related to the data and the parameter. The importance of this methodology in linear models can be found in Koul (1992, Chapters 5 and 7). These estimators have many desirable properties, including consistency, robustness against outlier in the error, efficiency and asymptotic normality in AR models.

To describe these estimators, let $\boldsymbol{g} = (g_1, ..., g_p)^T$ be a measurable function from $\mathcal{R}^p$ to $\mathcal{R}^p$, and $|\cdot|$ be the Euclidean norm. For a given nondecreasing right continuous function H on $\mathcal{R}$, define the dispersion function, for $\boldsymbol{u} \in \mathcal{R}^p$,

$$M_g(\boldsymbol{u}) = \int |n^{-1/2} \sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{Y}_{i-1})\{I(X_i - \boldsymbol{u}^T \boldsymbol{Y}_{i-1} \le y) - I(-X_i + \boldsymbol{u}^T \boldsymbol{Y}_{i-1} < y)\}|^2 dH(y)$$

and a class of MD estimators of $\boldsymbol{\theta}$, one for each $\boldsymbol{g}$ and H, to be

$$\hat{\boldsymbol{\theta}} := argmin\{M_g(\boldsymbol{u}); \boldsymbol{u} \in \mathcal{R}^p\}.$$

Here $I(A)$ is the indicator function of the event A. The existence of the MD estimator $\hat{\boldsymbol{\theta}}$ follows from Dhar (1993).

Note that if we take $H(x) = x$, $\boldsymbol{g}(\mathbf{y}) = \mathbf{y}$, then $\hat{\boldsymbol{\theta}}$ is the Hodges-Lehmann type estimator. If we denote $U_i := \boldsymbol{Z}_i^T \boldsymbol{Y}_{i-1} + \epsilon_i$, then the RCAR(p) model becomes $X_i = \boldsymbol{\theta}^T \boldsymbol{Y}_{i-1} + U_i$ and M_g is essentially the same as the K_g^+ of Koul (1992) with ϵ_i there replaced by U_i.

Koul (1992) discussed the asymptotic behavior of the estimators $\hat{\boldsymbol{\theta}}$ under the AR(p) setup. A simulation study of Dhar (1993) shows that many of these MD estimators outperform the least square estimator in an AR(p) model with asymmetric error. In this paper, we obtain the asymptotic normality of $\hat{\boldsymbol{\theta}}$ under the RCAR(p) setup. The method of proof is similar to that of Koul (1992) which requires obtaining the asymptotic uniform quadraticity of $M_g(\mathbf{t})$ and showing $|\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})| = O_P(1)$.

The material is organized as follows. Assumptions and statements of main theorems appear in section 1.2 while proofs appear in Section 1.3. Section 1.4 contains a simulation study and an application to the Canadian lynx data. The simulation study shows that the Hodges-Lehmann type estimator is at least as good as the least square (LS) estimator and the Huber estimator in the sense of having smaller biase and mean squared error. For the annual trappings data of the Canadian lynx over the years 1821-1934, it is observed that the RCAR(2) model with $\boldsymbol{\theta}$ estimated by $\hat{\boldsymbol{\theta}}$ provides an acceptable alternative to the more widely adapted class of AR models.

1.2 Assumptions and Theorems

Throughout this part of the dissertation, we assume that $\{X_i\}$ is strictly stationary and ergodic satisfying model (1.1). Sufficient conditions for this to happen are discussed in Theorems 2.1, 2.7 and Corollary 2.3.2 of Nicholls and Quinn (1982). In particular, when $p = 1$, then the following two assumptions imply the strictly stationarity and ergodicity of $\{X_i\}$.

(i) $\{\epsilon_i, i \in \mathcal{Z}\}$ and $\{Z_i, i \in \mathcal{Z}\}$ have mean zeros and finite variances σ_F^2 and σ_G^2, respectively.

(ii) $\theta^2 + \sigma_G^2 < 1$.

Now let γ be a measurable function from $\mathcal{R}^p$ to $\mathcal{R}$, $y, a \in \mathcal{R}, \boldsymbol{t} \in \mathcal{R}^p$. Define

$$p_i(y; \boldsymbol{t}, a) = \int F(y - \mathbf{z}^T \boldsymbol{Y}_{i-1} + n^{-1/2}(\boldsymbol{t}^T \boldsymbol{Y}_{i-1} + a|\boldsymbol{Y}_{i-1}|))dG(\mathbf{z}),$$

$$V(y; \boldsymbol{t}, a) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma(\boldsymbol{Y}_{i-1}) I(U_i \le y + n^{-1/2}(\boldsymbol{t}^T \boldsymbol{Y}_{i-1} + a|\boldsymbol{Y}_{i-1}|)),$$

$$v(y; \boldsymbol{t}, a) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma(\boldsymbol{Y}_{i-1}) p_i(y; \boldsymbol{t}, a),$$

$$W(y; \boldsymbol{t}, a) = V(y; \boldsymbol{t}, a) - v(y; \boldsymbol{t}, a).$$

Write $p_i(y, \boldsymbol{t})$, $W(y, \boldsymbol{t})$, $v(y, \boldsymbol{t})$, $V(y, \boldsymbol{t})$ for $p_i(y; \boldsymbol{t}, 0)$, $W(y; \boldsymbol{t}, 0)$, $v(y; \boldsymbol{t}, 0)$, $V(y; \boldsymbol{t}, 0)$, respectively. Also define

$$K_\gamma(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t}) = \int [V(y, \boldsymbol{t}) + V(-y, \boldsymbol{t}) - n^{-1/2} \sum_{i=1}^{n} \gamma(\boldsymbol{Y}_{i-1})]^2 dH(y), \ \ \boldsymbol{t} \in \mathcal{R}^p.$$

Observe that choosing $\gamma(\boldsymbol{Y}_{i-1}) = g_j(\boldsymbol{Y}_{i-1})$ in $K_\gamma(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t})$ gives the jth summand in $M_g(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t})$.

We now state the conditions required for the asymptotic uniform quadraticity of $K_\gamma(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t})$ in $\boldsymbol{t}$. In the sequel, all expectations are taken under the true parameter $\boldsymbol{\theta}$. Moreover, by the symmetry around 0 of the distribution G, we mean $dG(\boldsymbol{z}) = (-1)^p dG(-\boldsymbol{z})$.

(F) Functions F, G and H are symmetric around 0 and F has a Lebesgue density f.

(C1) $E\, \gamma^2(\boldsymbol{Y}_0) < \infty$.

(C2) For any $\boldsymbol{t} \in \mathcal{R}^p$ and $a \in \mathcal{R}$,

$$\int E\ \gamma^2(\boldsymbol{Y}_0)|p_1(y;\boldsymbol{t},a) - p_1(y;\boldsymbol{0},0)|dH(y) = o(1).$$

(C3) There exists a constant k, $0 < k < \infty$, such that $\forall\ \delta > 0, \forall\ \boldsymbol{t} \in \mathcal{R}^p$

$$\liminf_n P\Big(\int \Big[n^{-1/2}\sum_{i=1}^n \gamma^{\pm}(\boldsymbol{Y}_{i-1})(p_i(y;\boldsymbol{t},\delta) - p_i(y;\boldsymbol{t},-\delta))\Big]^2 dH(y) \le k\delta^2\Big) = 1,$$

where $\gamma^+ = \max(\gamma, 0), \gamma^- = \gamma^+ - \gamma$.

(C4) For every $\boldsymbol{t} \in \mathcal{R}^p$,

$$\int \Big\{n^{-1/2}\sum_{i=1}^n \gamma(\boldsymbol{Y}_{i-1})[p_i(y,\boldsymbol{t}) - p_i(y,\boldsymbol{0})] - \mathcal{A}(y)\boldsymbol{t}/2\Big\}^2 dH(y) = o_P(1).$$

where $\mathcal{A}(y) = 2E\ \gamma(\boldsymbol{Y}_0)\boldsymbol{Y}_0^T \int f(y - \mathbf{z}^T\boldsymbol{Y}_0)dG(\mathbf{z})$.

(C5) $\int E\ \gamma^2(\boldsymbol{Y}_0)F(y - \boldsymbol{Z}_1^T\boldsymbol{Y}_0)(1 - F(y - \boldsymbol{Z}_1^T\boldsymbol{Y}_0))dH(y) < \infty$.

We state two more conditions required for the asymptotic normality of the estimator $\hat{\boldsymbol{\theta}}$: Let $\mathcal{B}(y) = E\mathbf{g}(\boldsymbol{Y}_0)\boldsymbol{Y}_0^T \int f(y - \mathbf{z}^T\boldsymbol{Y}_0)dG(\mathbf{z})$, $y \in \mathcal{R}$.

(C6) The matrix $\mathcal{B}(y)$ is nonnegative definite for each $y \in \mathcal{R}$, $\int \mathcal{B}(y)dH(y)$ and $\int \mathcal{B}^T(y)\mathcal{B}(y)dH(y)$ are positive definite $p \times p$ matrices.

(C7) Either

$$\boldsymbol{s}^T\boldsymbol{g}(\boldsymbol{Y}_{i-1})\boldsymbol{Y}_{i-1}^T\boldsymbol{s} \ge 0, \quad \forall\ 1 \le i \le n,\ \forall\ \boldsymbol{s} \in \mathcal{R}^p,\ |\boldsymbol{s}| = 1, a.s.$$

or

$$\boldsymbol{s}^T\boldsymbol{g}(\boldsymbol{Y}_{i-1})\boldsymbol{Y}_{i-1}^T\boldsymbol{s} \le 0, \quad \forall\ 1 \le i \le n,\ \forall\ \boldsymbol{s} \in \mathcal{R}^p,\ |\boldsymbol{s}| = 1, a.s.$$

Remark 1 The above conditions are assumed so that the desired asymptotic uniform quadraticity and asymptotic normality of $\hat{\boldsymbol{\theta}}$ are achievable. In the case of the AR(p) model, i.e., when $\boldsymbol{Z}_i \equiv 0$, the above conditions (C1)-(C4) correspond to the conditions (7.4.7a), (7.4.8)-(7.4.10) of Koul (1992) and the condition (C5) is implied by (7.4.7a) and (5.5.69) of Koul (1992). For the Huber type estimator, i.e., for $\boldsymbol{y} \in \mathcal{R}^p$, $\boldsymbol{g}(\boldsymbol{y}) = \boldsymbol{y}I(|\boldsymbol{y}| \le c) + k\frac{\boldsymbol{y}}{|\boldsymbol{y}|}I(|\boldsymbol{y}| > c)$, for some positive constants k and c, the conditions (C7) is a *priori* satisfied. If H is a finite measure, $\gamma(\boldsymbol{y}) = y_j$, the

jth component of $\boldsymbol{y}$, $j = 1, ..., p$, $\boldsymbol{g}(\boldsymbol{y}) = \boldsymbol{y}$ and F has uniformly continuous density, having positive integral with respect to the measure induced by H, then all of the above conditions (C1)-(C7) are implied by the strict stationarity and ergodicity of $\{X_i\}$.

Now we are ready to state our main theorems.

Theorem 1.2.1 *Suppose that conditions (F), (C1)-(C5) hold. Then,* $\forall\ 0 < B < \infty$,

$$\sup_{|\boldsymbol{t}| \leq B} |K_\gamma(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t}) - \hat{K}_\gamma(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t})| = o_P(1),$$

where

$$\hat{K}_\gamma(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t}) = \int [V(y, \mathbf{0}) + V(-y, \mathbf{0}) - n^{-1/2} \sum_{i=1}^{n} \gamma(\boldsymbol{Y}_{i-1}) + \mathcal{A}(y)\ \boldsymbol{t}]^2 dH(y).$$

Upon using this result p times, the jth time with

$$\gamma(\boldsymbol{Y}_{i-1}) = g_j(\boldsymbol{Y}_{i-1}), \quad j = 1, ..., p, \tag{1.2}$$

we obtain the required asymptotic uniform quadraticity of the dispersion $M_g(\boldsymbol{u})$. For stating the desired results, we need to clarify the conditions (C1)-(C5) when γ is as in the equation (1.2). Condition (C1) is now equal to

(C1g) $Eg_j^2(\boldsymbol{Y}_0) < \infty$ for all $j = 1, ..., p$.

Similarly, condition (C2) is equal to $\forall\ \boldsymbol{t} \in \mathcal{R}^p$, $a \in \mathcal{R}$, $1 \leq j \leq p$, $||\mathbf{t}|| \leq B < \infty$,

(C2g) $\int Eg_j^2(\boldsymbol{Y}_0)|p_1(y; \boldsymbol{t}, a) - p_1(y; \mathbf{0}, 0)| dH(y) = o(1)$.

Let (C3g) stand for the condition (C3) after $\gamma^{\pm}(\boldsymbol{Y}_{i-1})$ is replaced by $g_j^{\pm}(\boldsymbol{Y}_{i-1}), 1 \leq j \leq p$, in condition (C3), $1 \leq i \leq n$. Interpret (C4g), (C5g) similarly.

Corollary 1.2.1 *Suppose that conditions (F), (C1g)-(C5g) hold. Then*

$$\sup_{|\boldsymbol{t}| \leq B} |M_g(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t}) - \hat{M}_g(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t})| = o_P(1), \tag{1.3}$$

where

$$\hat{M}_g(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t}) = \int |\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{Y}_{i-1})\{I(U_i \leq y) - I(-U_i < y)\} + \mathcal{B}(y)\boldsymbol{t}|^2 dH(y).$$

Theorem 1.2.2 *In addition to the assumptions of Corollary 1.2.1, assume that the conditions (C6) and (C7) hold. Then*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \Longrightarrow N(\mathbf{0}, \Gamma), \tag{1.4}$$

where

$$\Gamma := V^{-1} E(\psi(-U_1) - \psi(U_1^-))\boldsymbol{g}(\boldsymbol{Y}_0)\boldsymbol{g}^T(\boldsymbol{Y}_0)(\psi(-U_1) - \psi(U_1^-))^T V^{-1},$$

$$\psi(y) := \int I(x \le y)\mathcal{B}^T(x)dH(x), \quad \psi(y^-) := \int I(x < y)\mathcal{B}^T(x)dH(x),$$

$$V := \int \mathcal{B}^T(y)\mathcal{B}(y)dH(y).$$

Remark 2 *Least Absolute Deviation Estimator (l.a.d.).* If we choose $\boldsymbol{g}(\mathbf{y}) = \mathbf{y}$ and $H(\cdot)$ degenerates at 0, then $\hat{\boldsymbol{\theta}}$ is the *l.a.d.* estimator, v.i.z.

$$\hat{\boldsymbol{\theta}}_{lad} = argmin\{|n^{-1/2}\sum \boldsymbol{Y}_{i-1} sign(X_i - \boldsymbol{t}^T\boldsymbol{Y}_{i-1})|^2,\ \boldsymbol{t} \in \mathcal{R}^p\}.$$

Because of the importance of the *l.a.d.* estimator, we summarize all the conditions on f for the case $p = 1$. All conditions (F) and (C1)-(C7) are satisfied when G is symmetric around zero, F has a uniformly continuous and even density f, and $EX_0^2 f(Z_1X_0) > 0$. Therefore, Theorem 1.2.2 implies that $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \Rightarrow N(0, \hat{\sigma}_{lad}^2)$, where

$$\begin{aligned} \hat{\sigma}_{lad}^2 &= \frac{EX_0^2\mathcal{A}^2(0)[I(-U_1 \le 0) - I(U_1 > 0)]^2}{\mathcal{A}^4(0)} \\ &= \frac{EX_0^2}{\mathcal{A}^2(0)} = \frac{E(X_0^2)}{4(EX_0^2 f(Z_1X_0))^2}. \end{aligned}$$

1.3 Proofs

Notation For any measurable functions f and g from $\mathcal{R}^{p+1}$ to $\mathcal{R}$, define

$$|f\boldsymbol{s} - g\boldsymbol{u}|_H^2 := \int [f(y, \boldsymbol{s}) - g(y, \boldsymbol{u})]^2 dH(y),$$

where $\boldsymbol{u}, \boldsymbol{s} \in \mathcal{R}^p$. In the following, $W^{\pm}, v^{\pm}$ stand for W, v with γ replaced by $\gamma^{\pm}$.

The proof of Theorem 1.2.1 is similar to that of Theorem 7.4.1 of Koul (1992) and is facilitated by the following two Lemmas.

Lemma 1.3.1 *Suppose that the RCAR(p) model (1.1) holds. Then the followings hold.*

(A). The condition (C2) implies that $\forall\ 0 < B < \infty$,

$$E\int [W^{\pm}(y;\boldsymbol{t},a) - W^{\pm}(y;\boldsymbol{t},0)]^2 dH(y) = 0(1), \forall\ |\boldsymbol{t}| \le B, a \in \mathcal{R}. \tag{1.5}$$

(B). The condition (C3) implies that $\forall\ |\boldsymbol{t}| \le B$, $\forall\ 0 < B < \infty$,

$$\liminf_n P(\sup_{|\boldsymbol{s}-\boldsymbol{t}|\le B} |v^{\pm}(y;\boldsymbol{s}) - v^{\pm}(y;\boldsymbol{t})|^2 \le k\delta^2) = 1, \tag{1.6}$$

where k and δ are as in (C3).

(C). The conditions (C1), (C3) and (C4) imply that $\forall\ 0 < B < \infty$,

$$\sup_{|\boldsymbol{t}|\le B} \int [v(y,\boldsymbol{t}) - v(y;\mathbf{0}) - \mathcal{A}(y)\ \boldsymbol{t}/2]^2 dH(y) = o_P(1). \tag{1.7}$$

Lemma 1.3.2 *Suppose that the conditions (C2) and (C3) hold, then* $\forall\ 0 < B < \infty$,

$$\sup_{|\boldsymbol{t}|\le B} \int [W^{\pm}(y,\boldsymbol{t}) - W^{\pm}(y,\mathbf{0})]^2 dH(y) = o_P(1), \tag{1.8}$$

$$\sup_{|\boldsymbol{t}|\le B} \int [W(y,\boldsymbol{t}) - W(y,\mathbf{0})]^2 dH(y) = o_P(1). \tag{1.9}$$

The proofs of the above lemmas are similar to those of Lemmas 7.4.2 and 7.4.3 of Koul (1992) with the following modification: Replace the $\sigma-$fields used there by $\mathcal{F}_i = \sigma\{\epsilon_j, \boldsymbol{Z}_j, \mathbf{Y}_0, j \le i\}$, $i \ge 1$ and the linear term there by $\mathcal{A}\ \boldsymbol{t}/2$.

Proof of Corollary 1.2.1. Note that the jth summand in M_g is the same as that in K_γ when the function γ is replaced by g_j in (1.2). Therefore (1.3) follows from Theorem 1.2.1 easily. □

Before proving the Theorem 1.2.2, we need the following three lemmas.

Lemma 1.3.3 *Let $\{u_1, u_2, ...\}$ be a stationary ergodic stochastic process such that $E\{u_1^2\}$ is finite and $E(u_i|u_1, ..., u_{i-1}) = 0, \forall\ i \ge 1$, with probability one. Then the distribution of $n^{-1/2}\sum_{i=1}^n u_i$ approaches the normal distribution with mean zero and variance Eu_1^2.*

Proof. See Billingsley (1961).

Lemma 1.3.4 *Suppose that the assumptions of Theorem 1.2.2 hold. Then*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = O_P(1).$$

Proof. It suffices to prove that, for every $\epsilon > 0$, $\exists\ B > 0$ and integer $N_\epsilon \ni$

$$P(\inf_{|n^{1/2}(\boldsymbol{u}-\boldsymbol{\theta})|>B} M_g(\boldsymbol{u}) \geq M_g(\boldsymbol{\theta})) > 1 - \epsilon, \quad \forall\ n > N_\epsilon, \tag{1.10}$$

because then

$$P(|\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})| > B) \leq P(\inf_{|n^{1/2}(\boldsymbol{u}-\boldsymbol{\theta})|>B} M_g(\boldsymbol{u}) < M_g(\boldsymbol{\theta})) < \epsilon, \quad \forall\ n > N_\epsilon.$$

The following lemma gives (1.10). □

Lemma 1.3.5 *For any $\epsilon > 0, \exists\ B$ (depending on ϵ) and N_ϵ, $0 < B < \infty$, such that*

$$P(\inf_{|\boldsymbol{t}|>B} M_g(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t}) \geq M_g(\boldsymbol{\theta})) > 1 - \epsilon, \quad \forall\, n \geq N_\epsilon, \tag{1.11}$$

$$P(\inf_{|\boldsymbol{t}|>B} \hat{M}_g(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t}) \geq M_g(\boldsymbol{\theta})) > 1 - \epsilon, \quad \forall\, n \geq N_\epsilon. \tag{1.12}$$

Proof. Write $V_j(y,\boldsymbol{t})$, $W_j(y,\boldsymbol{t})$ for $V(y,\boldsymbol{t})$, $W(y,\boldsymbol{t})$, respectively, when γ replaced by g_j in (2) at jth time. Put $V_g(y,\boldsymbol{t}) := (V_1(y,\boldsymbol{t}),...,V_p(y,\boldsymbol{t}))^T$ and $W_g(y,\boldsymbol{t}) := (W_1(y,\boldsymbol{t}),...,W_p(y,\boldsymbol{t}))^T$. Note that the measure generated by H being σ-finite, there exists a partition $\{A_i\}$ of $\mathcal{R}$ such that $0 < \int_{A_i} dH < \infty, i = 1,2,....$ Let $h = \sum_{i=1}^{\infty} I_{A_i}$, then $0 < |h|_H^2 = \int h^2 dH < \infty$. For $\boldsymbol{t} \in \mathcal{R}^p$, define

$$N(\boldsymbol{t}) := \int [V_g(y,\boldsymbol{t}) - V_g(-y,\boldsymbol{t})]h(y)dH(y),$$

$$\hat{N}(\boldsymbol{t}) := \int [V_g(y,\boldsymbol{0}) - V_g(-y,\boldsymbol{0}) + \mathcal{B}(y)\boldsymbol{t}]h(y)dH(y).$$

Then, for any $\boldsymbol{s} \in \mathcal{R}^p, |\boldsymbol{s}| = 1$, by the C-S inequality,

$$\begin{aligned} |\boldsymbol{s}^T N(\boldsymbol{t})|^2 &= \Big|\int \boldsymbol{s}^T[V_g(y,\boldsymbol{t}) - V_g(-y,\boldsymbol{t})]h(y)dH(y)\Big|^2 \\ &\leq \int \left(\boldsymbol{s}^T[V_g(y,\boldsymbol{t}) - V_g(-y,\boldsymbol{t})]\right)^2 dH(y)|h|_H^2 \\ &\leq M_g(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t})|h|_H^2. \end{aligned}$$

For a $0 < B < \infty$, $\boldsymbol{t} \in \mathcal{R}^p$, write $\boldsymbol{t} = \boldsymbol{s}u, |\boldsymbol{s}| = 1$, then

$$\inf_{|u|>B,|\boldsymbol{s}|=1} M_g(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{s}u) \geq \inf_{|u|>B,|\boldsymbol{s}|=1} \frac{|\boldsymbol{s}^T N(\boldsymbol{t})|^2}{|h|_H^2}.$$

Similarly,

$$\inf_{|u|>B,|\boldsymbol{s}|=1} \hat{M}_g(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{s}u) \geq \inf_{|u|>B,|\boldsymbol{s}|=1} \frac{|\boldsymbol{s}^T \hat{N}(\boldsymbol{s}u)|^2}{|h|_H^2}.$$

Note that, by the condition (C5),

$$E\int W^2(y,\mathbf{0})dH(y) = E\gamma^2(\mathbf{Y}_0)\int F(y - \boldsymbol{Z}_1^T\mathbf{Y}_0)(1 - F(y - \boldsymbol{Z}_1^T\mathbf{Y}_0))dH(y) < \infty.$$

It follows easily that $M_g(\boldsymbol{\theta})$ is bounded in probability. That is, $\forall\ \epsilon > 0, \exists\ M_\epsilon < \infty$ such that

$$P(M_g(\boldsymbol{\theta}) \leq M_\epsilon) \geq 1 - \epsilon/2, \quad \text{for all } n \geq 1. \tag{1.13}$$

Thus it suffices to prove that

$$P(\inf_{|u|>B,|\boldsymbol{s}|=1} \frac{|\boldsymbol{s}^T N(\boldsymbol{s}u)|^2}{|h|_H^2} \geq M_\epsilon) > 1 - \epsilon, \tag{1.14}$$

$$P(\inf_{|u|>B,|\boldsymbol{s}|=1} \frac{|\boldsymbol{s}^T \hat{N}(\boldsymbol{s}u)|^2}{|h|_H^2} \geq M_\epsilon) > 1 - \epsilon. \tag{1.15}$$

But, $\forall\ \boldsymbol{t} \in \mathcal{R}^p$,

$$\begin{aligned}
&|N(\boldsymbol{t}) - \hat{N}(\boldsymbol{t})|^2 \\
\leq\ & \int |V_g(y,\boldsymbol{t}) - V_g(y,\mathbf{0}) - V_g(-y,\boldsymbol{t}) + V_g(-y,\mathbf{0}) - \mathcal{B}(y)\boldsymbol{t}|^2 dH(y)|h|_H^2 \\
\leq\ & 2|h|_H^2\left[|W_{g\boldsymbol{t}} - W_{g0} + W_{-g\boldsymbol{t}} + W_{-g0}|_H^2 + |v_{g\boldsymbol{t}} - v_{g0} + v_{-g\boldsymbol{t}} + v_{-g0} - \mathcal{B}\boldsymbol{t}|_H^2\right].
\end{aligned}$$

By using Lemma 1.3.1(C) p times and Lemma 1.3.2, we obtain that $\forall\ 0 < B < \infty$,

$$\sup_{|\boldsymbol{t}|\leq B} |N(\boldsymbol{t}) - \hat{N}(\boldsymbol{t})|^2 = o_P(1).$$

Now rewrite $\hat{N}(\boldsymbol{t})$,

$$\begin{aligned}
\hat{N}(\boldsymbol{t}) &= \int [V_g(y,\mathbf{0}) - V_g(-y,\mathbf{0})]h(y)dH(y) + \int \mathcal{B}(y)h(y)dH(y)\ \boldsymbol{t} \\
&:= \hat{N}_1 + \hat{N}_2\ \boldsymbol{t}.
\end{aligned}$$

By the C-S inequality,

$$|\hat{N}_1|^2 \leq M_g(\boldsymbol{\theta})|h|_H^2.$$

Therefore, by (1.13), there exists $b = M_\epsilon^{1/2}|h|_H$, such that,

$$P(|\hat{N}_1| \le b) \ge 1 - \epsilon. \tag{1.16}$$

Denote $\alpha_s = \boldsymbol{s}^T \int \mathcal{B}(y)h(y)dH(y)\boldsymbol{s} = \boldsymbol{s}^T \int \mathcal{B}(y)dH(y)\boldsymbol{s}$, $\alpha = \inf_{|s|=1}\{\alpha_s\}$. Then by the condition (C6), $\alpha > 0$. The rest of the proof is similar to that of Lemma 5.5.4. of Koul (1992). □

Proof of Theorem 1.2.2. Expand $\hat{M}_g(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t})$ in $\boldsymbol{t}$,

$$\begin{aligned}
&\hat{M}_g(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{t}) \\
&= \hat{M}_g(\boldsymbol{\theta}) + 2\,\boldsymbol{t}^T \frac{1}{\sqrt{n}} \sum_{i=1}^n \int \{I(U_i \le y) - I(-U_i < y)\}\mathcal{B}^T(y)\boldsymbol{g}(\boldsymbol{Y}_{i-1})dH(y) \\
&\qquad\qquad\qquad + \boldsymbol{t}^T \int \mathcal{B}^T(y)\mathcal{B}(y)dH(y)\,\boldsymbol{t}.
\end{aligned}$$

Let $\tilde{\boldsymbol{\theta}} := argmin\{\hat{M}_g(\boldsymbol{u}) : \boldsymbol{u} \in \mathcal{R}^p\}$. Then by the same proof as that of the Theorem 5.4.1 of Koul (1992), we have

$$\Big|(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})^T \int \mathcal{B}^T(y)\mathcal{B}(y)dH(y)(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})\Big| = o_P(1).$$

Therefore it is enough to prove

$$\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \Rightarrow N(0, \Gamma). \tag{1.17}$$

But $\tilde{\boldsymbol{\theta}}$ must satisfy the following equation

$$\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = (\int \mathcal{B}^T(y)\mathcal{B}(y)dH(y))^{-1}\frac{\mathbf{S}_n}{\sqrt{n}}, \tag{1.18}$$

where

$$\boldsymbol{S}_n := \sum_{i=1}^n (\psi(-U_i) - \psi(U_i^-))\boldsymbol{g}(\boldsymbol{Y}_{i-1}).$$

Since $\{X_i\}$ is strictly stationary and ergodic, so are $(\psi(-U_i) - \psi(U_i^-))\boldsymbol{g}(\boldsymbol{Y}_{i-1})$. Furthermore, by the condition (F), U_1 is a continuous random variable and given Y_0, the conditional distribution of U_1 is the same as that of $-U_1$. Thus

$$E[\psi(-U_1) - \psi(U_1^-)|\mathcal{F}_0] = 0 \tag{1.19}$$

Now, for any p-component vector $\boldsymbol{\beta}$,

$$E\{\boldsymbol{\beta}^T(\boldsymbol{\psi}(-U_i) - \boldsymbol{\psi}(U_i^-))\boldsymbol{g}(\boldsymbol{Y}_{i-1})\boldsymbol{g}^T(\boldsymbol{Y}_{i-1})(\boldsymbol{\psi}(-U_i) - \boldsymbol{\psi}(U_i-))^T\boldsymbol{\beta}\}$$
$$= \boldsymbol{\beta}^T E\{(\boldsymbol{\psi}(-U_1) - \boldsymbol{\psi}(U_1^-))\boldsymbol{g}(\mathbf{Y}_0)\boldsymbol{g}^T(\mathbf{Y}_0)(\boldsymbol{\psi}(-U_1) - \boldsymbol{\psi}(U_1^-))^T\}\boldsymbol{\beta}.$$

This expectation exists if $Eg_j^2(\boldsymbol{Y}_0) < \infty$, $j = 1, ..., p$ and condition (C6) holds. Then $E\{\boldsymbol{\beta}^T(\boldsymbol{\psi}(-U_i) - \boldsymbol{\psi}(U_i^-))\boldsymbol{g}(\boldsymbol{Y}_{i-1})|\mathcal{F}_{i-1}\} = 0$ follows from (1.19) and stationarity of U_i. An application of Lemma 1.3.3 shows that $n^{-1/2}\sum_{i=1}^n \boldsymbol{\beta}^T(\boldsymbol{\psi}(-U_i) - \boldsymbol{\psi}(U_i^-))\boldsymbol{g}(\boldsymbol{Y}_{i-1})$ converges weakly to the normal distribution with mean zero and variance

$$\boldsymbol{\beta}^T E\{(\boldsymbol{\psi}(-U_1) - \boldsymbol{\psi}(U_1^-))\boldsymbol{g}(\mathbf{Y}_0)\boldsymbol{g}^T(\mathbf{Y}_0)(\boldsymbol{\psi}(-U_1) - \boldsymbol{\psi}(U_1^-))^T\}\boldsymbol{\beta},$$

for all $\boldsymbol{\beta} \in \mathcal{R}^p$. U_i Thus, by the Cramer-Wold device, $\boldsymbol{S}_n$ converges weakly to the multivariate normal distribution with mean vector zero and covariance matrix $E\{(\boldsymbol{\psi}(-U_1) - \boldsymbol{\psi}(U_1^-))\boldsymbol{g}(\boldsymbol{Y}_0)\boldsymbol{g}^T(\boldsymbol{Y}_0)(\boldsymbol{\psi}(-U_1) - \boldsymbol{\psi}(U_1^-))^T\}$. Hence $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ converges in distribution to the normal distribution with mean vector zero and covariance matrix

$$V^{-1}E\{(\boldsymbol{\psi}(-U_1) - \boldsymbol{\psi}(U_1^-))\boldsymbol{g}(\mathbf{Y}_0)\boldsymbol{g}^T(\mathbf{Y}_0)(\boldsymbol{\psi}(-U_1) - \boldsymbol{\psi}(U_1^-))^T\}V^{-1}.$$

This ends the proof of Theorem 1.2.2. □

1.4 Simulation results

In this section, we investigate the performance of the MD estimators under the RCAR(1) model for finite samples. A simulation study (100 replications) was performed for samples of size $n = 20$, $n = 50$ and $n = 100$. The samples were generated by the RCAR(1) model,

$$X_i = (\theta + Z_i)X_{i-1} + \epsilon_i$$

with the true parameter $\theta = .5$, and different error distributions and normal distibuted random coefficients.

The comparison is made among LS, Huber and MD estimators. The MD estimators considered for comparison under the RCAR(1) model are as follows. The function H, g are taken to be the identity function (i.e., $H(y) = g(y) = y, y \in \mathcal{R}$). The Huber function ϕ is defined as follows.

$$\phi(x) = \begin{cases} x, & \text{if } |x| \leq c \\ c \,\text{sign}(x), & \text{if } |x| > c. \end{cases} \tag{1.20}$$

with c estimated by $S_x = \text{Median}|X_i|$.

The random coefficient distribution G considered here is a normal distribution with mean zero and variance .25; This enables to have $\theta^2 + \sigma_G^2 < 1$. The error distributions considered are the following.

(a) F is the standard normal distribution,

(b) F is the double exponential(1) distribution,

(c) F is the logistic (1,1) .

The RCAR(1) process is generated as follows:

In case (a),

1. generate a vector $< w(1), w(2) >^T$, where the $w(i)$'s are successivly generated by a standard normal random number generator. So they are independent.

2. Repeat step 1 $(n + 200)$ times where n is the sample size desired. Let the $w(1)$, $.5w(2)$ generated in the mth time be $\epsilon_m, Z_m, m = 1, ..., n + 200$, respectively. Then $< \epsilon_1, ..., \epsilon_{n+200} >^T$ and $< Z_1, ..., Z_{n+200} >^T$ will theoretically be independent, have zero means and $E\epsilon_i^2 = 1$, and $EZ_i^2 = .25$.

3. Calculate

$$X_i = (\theta + Z_i)X_{i-1} + \epsilon_i,$$

where X_0 is generated by the normal distribution with mean zero and variance 2. Then ignore the first 200 X values produced. This enables $\{X_i\}$ to reach an equilibrium since we assume $\{X_i\}$ is stable.

In the case (b) (or (c)), we independently generate $w(1), w(2)$ from double exponential(1) (or logistic (1,1)) and standard normal distribution, respectively. Then do the same thing as in case (a) for steps 2 and 3.

The LS estimator is computed using the formula $\sum_{i=1}^{n} X_i X_{i-1} / \sum_{i=1}^{n} X_{i-1}^2$. The Huber estimator is the solution of

$$T(u) := \sum_{i=1}^{n} X_{i-1}\phi(X_i - uX_{i-1}) = 0. \tag{1.21}$$

with ϕ as given in (1.20).

For the Huber estimator, S_x =Median $|X_i - uX_{i-1}|$ is first computed using the LS estimator $\hat{\theta}_{ls}$, then S_x is again computed by using the zero of (1.21). This iterative procedure is terminated when the absolute difference between the consecutive zeros of (1.21) is less than 10^{-6}.

To compute the MD estimator, we minimize the dispersion $M_g(u)$ over $[-1, 1]$ and the minimizer is denoted by θ_{md}. Table 1.1 contains the simulation results of the averages (Mean) and the mean squared errors (MSE) of the estimators for the true parameter $\theta = 0.5$ in the RCAR(1) model with 100 replicates and sample size n. Notice that the MD estimator computed in Table 1.1 is a local minimization. According to the paper by Dhar (1993, Lemma 1.1), in the case $H(x) = x$, the minimizer of function $M_g(u)$ can only be one or a convex combination of a pair of elements from the set

$$D = \left\{ \begin{array}{l} (X_i - X_j)/(X_{i-1} - X_{j-1}), (X_i + X_j)/(X_{i-1} + X_{j-1}) : \\ X_{i-1} \neq X_{j-1}, X_{i-1} \neq -X_{j-1}, 1 \leq i, j \leq n. \end{array} \right\}.$$

Thus the global minimizer can be computed through comparing $(\theta_d, M_g(\theta_d))$ and pairs $(u, M_g(u))$ for $u \in D$ starting with $(\theta_{md}, M_g(\theta_{md}))$.

From Table 1.1, we observe that the Huber estimator has the biggest MSEs except in the case of Dexp(1) error and $n = 50$ which could be caused by the computing accuracy. Most of the biases and the MSEs of MD estimator are less than these of LSEs. Also, the estimated standard deviation of the averages of the estimates can be computed by $SV/\sqrt{100} = SV/10$, where SV is the sample variance which is related to MSE by the formula $MSE = (n-1)SV/n + (Mean - \theta)^2$. Again, we observe that all estimators are under estimating. For the sample of size $n = 100$, the MD estimators with $H(x) = g(x) = x$ are between LSE and Huber estimator. The simulation study was done by Mathematica.

Table 1.1: Simulation results

	Error distribution					
	N(0, 1)		Logistic (1,1)		Dexp(1)	
Estimator	Mean	MSE	Mean	MSE	Mean	MSE
	n=20					
LS	.437970	.055660	.432906	.058949	.439949	.068925
Huber	.452448	.061695	.447024	.066406	.447882	.075071
MD	.441372	.055152	.433595	.061095	.437264	.076194
	n=50					
LS	0.469420	.020073	.444754	.025691	.477704	.020230
Huber	0.463232	.026582	.458541	.030475	.479045	.026283
MD	0.473069	.020896	.454640	.0263490	.476869	.031124
	n=100					
LS	.475662	.013030	.486802	.010064	.474512	.013748
Huber	.488874	.014368	.498575	.13089	.487110	.013204
MD	.483978	.012277	.494234	.11173	.483249	.012016

Table 1.2: The estimators of the RCAR(2) model for lynx data

Estimators	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\Sigma}_{11}$	$\hat{\Sigma}_{12}$	$\hat{\Sigma}_{22}$	$\hat{\sigma}_F^2$
LS	1.3844	-.7479	.0770	-.0694	.0821	.0364
ML	1.4274	-.8073	.0664	-.0489	.0839	.0300
MD	1.3932	-.7495	.0764	-.0706	.0845	.0367

An applied example

We now fit a second order autoregressive random coefficient model

$$X_i^* = (\theta_1 + Z_{1i})X_{i-1}^* + (\theta_2 + Z_{2i})X_{i-2}^* + \epsilon_i$$

to the classical Canadian lynx data. Here $X_i^* = X_i - \bar{X}$, $\bar{X}$ is the average value of the X_i's, and X_i is the log_{10} of the ith data. We took the first one hundred observations. The MD estimators $\hat{\theta}_1$, $\hat{\theta}_2$ of θ_1, θ_2 are 1.3932, -.749498. Let $\boldsymbol{Z}_1 = (Z_{11}, Z_{21})'$ and

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix} = E\boldsymbol{Z}_1\boldsymbol{Z}_1^T.$$

To estimate the covariance matrix Σ of $\boldsymbol{Z}_1$ and the variance σ_F^2 of ϵ_1, substitute the MD estimator $\hat{\boldsymbol{\theta}}$ into (3.2.4) and (3.2.5) of Nicholls and Quinn. The estimators $\hat{\Sigma}_{11}$, $\hat{\Sigma}_{12}$, $\hat{\Sigma}_{22}$ and $\hat{\sigma}_F^2$ obtained thus are .076433, -.070556, .084552 and .03668. The comparison of the LSE , MLE of Nicholls and Quinn's and the MD estimator is given in Table 1.2.

From Table 1.2, we can see that the MD estimator performs at least as well as the LSE. Also, notice that the ML estimator of Nicholls and Quinn has the smallest estimated variance $\hat{\sigma}_F^2$ and the smallest norm of the estimated covariance matrix of $\boldsymbol{Z}_1$. The three dimensional graph of the dispersion $M_g(\boldsymbol{u})$, $\boldsymbol{u} \in \mathcal{R}^2$ is in Figure 1.1.

The zeros of the characteristic polynomial $(1 - 1.3932z + .749498z^2)$ are $1.15509\exp\{\pm i2\pi/9.88329\}$, and so by using RCAR(2) model, it exhibits a period of 9.88329 cycle which is close to the result of Moran(1953).

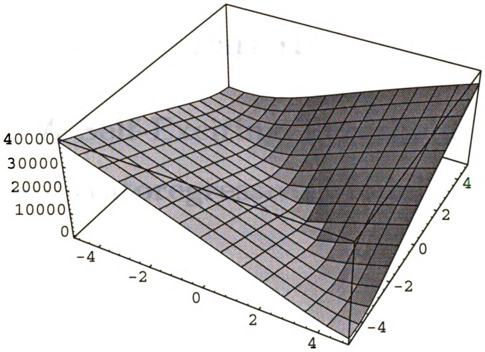

Figure 1.1: The graph of the dispersion $M_g(\boldsymbol{u})$

Part II

A Self-Exciting Threshold Autoregressive Model

Chapter 2

Definitions, assumptions and consistency

2.1 The profile maximum likelihood estimation

Recall that the SETAR(2;1,1) model is defined by

$$X_i = h(X_{i-1}, \boldsymbol{\theta}) + \epsilon_i, \qquad i \geq 1, \tag{2.1}$$

for some $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, r)^T \in \mathcal{R}^5$, where $\boldsymbol{\theta}_1 = (a_0, a_1, b_0, b_1)^T \in \mathcal{R}^4$ and for any $x \in \mathcal{R}$,

$$h(x, \boldsymbol{\theta}) = (a_0 + a_1 x) I(x \leq r) + (b_0 + b_1 x) I(x > r).$$

Here, the errors $\{\epsilon_i\}$ are independent and identically distributed random variables with mean zero, finite nonzero variance and ϵ_1 is independent of X_0.

We begin with the definition of the maximum likelihood estimators of the unknown underlying parameter $\boldsymbol{\theta}$ in model (2.1). Assume $\boldsymbol{\theta}$ is an interior point of Θ defined in (0.3). Note that Θ is an open subset of $\mathcal{R}^5$. There exists a compact subset K of $\mathcal{R}^4$ such that $\boldsymbol{\theta}$ is an interior point of $K \times \bar{\mathcal{R}}$.

Denote $\Omega = K \times \bar{\mathcal{R}}$, then Ω is a compact set. Let $\boldsymbol{\vartheta} = (\alpha_0, \alpha_1, \beta_0, \beta_1, s)^T$ be any point in Ω. Note that $\{X_i\}$ in model (2.1) forms a Markov chain. Let $g_{\boldsymbol{\vartheta}}(X_0)$ be the initial density of X_0 under $\boldsymbol{\vartheta}$, f be the density function of ϵ_1, then the one step transition densities, starting with X_{i-1}, is $f(X_i - h(X_{i-1}, \boldsymbol{\vartheta}))$, $i \geq 1$. If one observes

$(X_0, \cdots, X_n)$, then the likelihood function under $\boldsymbol{\vartheta}$ is $\Pi_1^n f(X_i - h(X_{i-1}, \boldsymbol{\vartheta})) g_{\boldsymbol{\vartheta}}(X_0)$. Let $\hat{\boldsymbol{\theta}}_n = (\hat{a}_{0n}, \hat{a}_{1n}, \hat{b}_{0n}, \hat{b}_{1n}, \hat{r}_n)^T$ be any measurable function of $(X_0, X_1, ..., X_n)$ from $\mathcal{R}^{n+1}$ to Ω such that $\hat{\boldsymbol{\theta}}_n$ maximizes the conditional likelihood function

$$L_n(\boldsymbol{\vartheta}) := \Pi_1^n f(X_i - h(X_{i-1}, \boldsymbol{\vartheta})), \text{ over } \Omega.$$

Write $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1^T, s)^T$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, r)^T$. 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 $s \in \bar{\mathcal{R}}$, denote $L_{ns}(\boldsymbol{\vartheta}_1) = L_n(\boldsymbol{\vartheta}_1, s) = L_n(\boldsymbol{\vartheta})$. Let $\boldsymbol{\vartheta}_{1n}(s) \in K$ be any value satisfying the following equation:

$$\boldsymbol{\vartheta}_{1n}(s) = argmax_{\boldsymbol{\vartheta}_1 \in K} L_{ns}(\boldsymbol{\vartheta}_1).$$

Step 2. Consider the profile conditional likelihood function $s \to L_n(\boldsymbol{\vartheta}_{1n}(s), s)$. Note that $L_n(\boldsymbol{\vartheta}_{1n}(s), s)$ has only finite number of possible values. Let $\hat{r}_n$ be the any value satisfying the following equation

$$\hat{r}_n = argmax_{s \in \mathcal{R}} L_n(\boldsymbol{\vartheta}_{1n}(s), s),$$

and substitute $\hat{r}_n$ into $\boldsymbol{\vartheta}_{1n}(s)$ to get

$$\hat{\boldsymbol{\theta}}_{1n} = \boldsymbol{\vartheta}_{1n}(\hat{r}_n).$$

Then

$$\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\theta}}_{1n}^T, \hat{r}_n)^T \text{ is a maximum likelihood estimator of } \boldsymbol{\theta}. \tag{2.2}$$

To see (2.2), for any $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1^T, s)^T \in \Omega$, by the definitions of $\hat{\boldsymbol{\theta}}_{1n}$ and $\hat{r}_n$, we have

$$L_n(\hat{\boldsymbol{\theta}}_{1n}, \hat{r}_n) = L_n(\boldsymbol{\vartheta}_{1n}(\hat{r}_n), \hat{r}_n) \geq L_n(\boldsymbol{\vartheta}_{1n}(s), s) \geq L_n(\boldsymbol{\vartheta}),$$

and hence,

$$L_n(\hat{\boldsymbol{\theta}}_{1n}, \hat{r}_n) = \sup_{\boldsymbol{\vartheta} \in \Omega} L_n(\boldsymbol{\vartheta}).$$

This means that $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\theta}}_{1n}^T, \hat{r}_n)^T$ is a MLE of $\boldsymbol{\theta}$.

2.2 Assumptions

In this section, we are listing assumptions and some examples. The following assumptions on the density f of ϵ_1 will be used in the following chapters of part II.

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

(C2) φ is Lip(1).

(C3) φ is differentiable and the derivative φ' is Lip(1).

(C4) $E|\epsilon_1|^3 < \infty$.

To derive the n-consistency, we need to assume the following:

(**M**) The threshold r in $\mathcal{R}$ is the discontinuity point of h, or equivalently,

$$b_0 - a_0 + r(b_1 - a_1) \neq 0, \;\; r \in \mathcal{R}.$$

Remark 1. From the invariant equation $g_{\boldsymbol{\theta}}(y) = \int f(y - h(x, \boldsymbol{\theta})) g_{\boldsymbol{\theta}}(x)\, dx,\; y \in \mathcal{R}$, the condition (C1) implies that $g_{\boldsymbol{\theta}}$ is bounded away from 0 and ∞ over compact sets. It is the minimal requirement for obtaining asymptotically efficient estimators of the coefficient parameters. See Koul and Schick (1995).

Remark 2. For the Markov chain $\{X_i\}$ in model (2.1), denote its k-step transition probability by $P^k(x, B)$ where $x \in \mathcal{R}$ and B is a Borel set. By the discussions in Chan and Tong (1985) and Chan (1989), the condition (C1) and $\boldsymbol{\theta} \in \Theta$ imply that $\{X_i\}$ admits an unique invariant measure $G_{\boldsymbol{\theta}}(\cdot)$ such that $\exists\; C,\; \rho < 1,\; \forall\; x \in \mathcal{R},\; \forall\; k \geq 1,\; \|P^k(x, \cdot) - G_{\boldsymbol{\theta}}(\cdot)\|_{tv} \leq C(1 + |x|)\rho^k$, where $\|\cdot\|_{tv}$ and $|\cdot|$ denote the total variation norm and the Euclidean norm, respectively.

Remark 3. The Lip(1) and the differentiability of φ imply the boundedness of φ'.

Remark 4. When discussing the CLSE, Chan (1993) assumed the finiteness of fourth moment of ϵ_1 while we need the finiteness of the third moment of ϵ_1 only.

Examples satisfying the conditions (C1)-(C4).

Example 1. If f is the standard normal density function, then (C1)-(C4) hold.

Example 2. If f is the logistic density, i.e. $f(x) \equiv F(x)(1 - F(x))$, then (C1)-(C4) hold.

Example 3. Let $f(x) = c(m)(1 + x^2/m)^{-(m+1)/2}$, $-\infty < x < \infty$, where m is a positive integer and $c(m)$ is a constant such that f is a density function. Then, $E\epsilon_1^2 = m/(m-2) < \infty$, $m > 2$ and

$$f'(x) = c(m)\Big(-\frac{m+1}{2}\Big)\Big(1 + \frac{x^2}{m}\Big)^{-\frac{m+1}{2}-1}\frac{2x}{m}. \tag{2.3}$$

Thus,

$$\varphi = f'/f = -\frac{m+1}{2}\frac{2x/m}{1 + x^2/m},$$

and

$$|f'(x)| \le \frac{m+1}{2} f(x). \tag{2.4}$$

Hence (C1) holds. By (2.4),

$$\frac{f'^2(x)}{f(x)} \le \Big(\frac{m+1}{2}\Big)^2 f(x).$$

This implies that $I(f) < \infty$. The Lip(1) of φ holds because of

$$|\varphi'| = \left| -\frac{m+1}{m}\Big(1 - \frac{x^2}{m}\Big)\Big/\Big(1 + \frac{x^2}{m}\Big)^2 \right| \le \frac{m+1}{m}.$$

Hence (C2) holds. Note that

$$\varphi'' = 2\frac{m+1}{m^2}x\Big(3 - \frac{x^2}{m}\Big)\Big/\Big(1 + \frac{x^2}{m}\Big)^3,$$

and

$$|\varphi''| \le 6\frac{m+1}{m}.$$

So (C3) holds. Furthermore, for $m > 6$, $E|\epsilon_1|^3 < \infty$ which implies (C4).

Throughout in the following proofs, we use the fact that $E|\epsilon_1|^k < \infty$ implies $E|X_0|^k < \infty$, for $k = 2, 3$, as proved by Chan, Petruccelli, Tong and Woolford (1985).

2.3 Strong consistency of the MLE

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

$$l_n(\boldsymbol{\vartheta}) = \frac{1}{n}\sum \ln \frac{f(X_i - h(X_{i-1}, \boldsymbol{\vartheta}))}{f(X_i - h(X_{i-1}, \boldsymbol{\theta}))} \tag{2.5}$$

and denote

$$\psi(X_{i-1}, \epsilon_i, \boldsymbol{\vartheta}) = \ln \frac{f(\epsilon_i + h(X_{i-1}, \boldsymbol{\theta}) - h(X_{i-1}, \boldsymbol{\vartheta}))}{f(\epsilon_i)}, \; 1 \le i \le n. \tag{2.6}$$

Note that

$$l_n(\boldsymbol{\vartheta}) = \frac{1}{n} \sum \psi(X_{i-1}, \epsilon_i, \boldsymbol{\vartheta}). \tag{2.7}$$

Write $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1^T, s)^T \in \Omega$ and $h(x, \boldsymbol{\vartheta}) = h_s(x, \boldsymbol{\vartheta}_1)$. Let

$$\begin{aligned} \dot{h}_s(x) &= (\partial/\partial\boldsymbol{\vartheta}_1)(h_s(x, \boldsymbol{\vartheta}_1)) \\ &= (I(x \le s), xI(x \le s), I(x > s), xI(x > s))^T, \; s \in \bar{\mathcal{R}}, \; x \in \mathcal{R}. \end{aligned}$$

Observe that for any $x \in \mathcal{R}$,

$$h(x, \boldsymbol{\vartheta}) = \boldsymbol{\vartheta}_1^T \dot{h}_s(x). \tag{2.8}$$

Also,

$$|\dot{h}_s(x)| \le \sqrt{1 + x^2}, \tag{2.9}$$

and for any $s \in \bar{\mathcal{R}}$, $t \in \bar{\mathcal{R}}$,

$$\begin{aligned} |\dot{h}_s(x) - \dot{h}_t(x)| &\le \sqrt{2(1 + x^2)} I(s \wedge t < x \le s \vee t) & (2.10) \\ &\le \sqrt{2(1 + x^2)} I(|x - t| \le |s - t|) & (2.11) \end{aligned}$$

Thus, by (2.8) and (2.9),

$$|h(x, \boldsymbol{\vartheta})| \le |\boldsymbol{\vartheta}_1| \sqrt{1 + x^2}. \tag{2.12}$$

Recall that $\boldsymbol{\theta} \in \Theta$ means the stationarity and ergodicity of underlying process $\{X_i\}$. Throughout, we will work on the stationary and ergodic process $\{X_i\}$.

Theorem 2.3.1 *Suppose that the conditions (C1) and (C2) hold. Then,*

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}, \quad \textit{as} \quad n \to \infty, \qquad (\textit{under } \boldsymbol{\theta}). \tag{2.13}$$

Before proving Theorem 2.3.1, we need the following lemma. Let $U_{\boldsymbol{\vartheta}}$ denote any open neighborhood of $\boldsymbol{\vartheta}$.

Lemma 2.3.1 *Under the assumptions of Theorem 2.3.1, for any $\boldsymbol{\vartheta} \in \Omega$ and its open neighborhood $U_{\boldsymbol{\vartheta}}$,*

$$E \sup_{\boldsymbol{\vartheta}^* \in U_{\boldsymbol{\vartheta}}} |\psi(X_0, \epsilon_1, \boldsymbol{\vartheta}^*) - \psi(X_0, \epsilon_1, \boldsymbol{\vartheta})| \to 0, \quad \textit{as } U_{\boldsymbol{\vartheta}} \textit{ shrinks to } \boldsymbol{\vartheta}. \tag{2.14}$$

Proof. Define

$$U_{\boldsymbol{\vartheta}}(\eta) = \{\boldsymbol{\vartheta}^* = (\boldsymbol{\vartheta}_1^{*T}, s^*)^T \in \Omega : |\boldsymbol{\vartheta}_1^* - \boldsymbol{\vartheta}_1| < \eta,\ d(s^*, s) < \eta\},\ \eta > 0.$$

It suffices to show that

$$E \sup_{\boldsymbol{\vartheta}^* \in U_{\boldsymbol{\vartheta}}(\eta)} |\psi(X_0, \epsilon_1, \boldsymbol{\vartheta}^*) - \psi(X_0, \epsilon_1, \boldsymbol{\vartheta})| \to 0, \quad \text{as } \eta \to 0. \tag{2.15}$$

Let $\epsilon_1(\boldsymbol{\vartheta}) = X_1 - h(X_0, \boldsymbol{\vartheta})$ and $\delta(X_0, \boldsymbol{\vartheta}^*) = h(X_0, \boldsymbol{\vartheta}) - h(X_0, \boldsymbol{\vartheta}^*)$. For any $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1^T, s)^T$, recall

$$h_s(X_0, \boldsymbol{\vartheta}_1) = h(X_0, \boldsymbol{\vartheta}_1, s) = h(X_0, \boldsymbol{\vartheta}), \tag{2.16}$$

and rewrite

$$\delta(X_0, \boldsymbol{\vartheta}^*) = h_s(X_0, \boldsymbol{\vartheta}_1) - h_{s^*}(X_0, \boldsymbol{\vartheta}_1^*).$$

For any $x \in \mathcal{R}$, by (2.8) and (2.9),

$$|h_s(x, \boldsymbol{\vartheta}_1) - h_s(x, \boldsymbol{\vartheta}_1^*)| \le |\boldsymbol{\vartheta}_1 - \boldsymbol{\vartheta}_1^*|\sqrt{1 + x^2}, \tag{2.17}$$

and by (2.10) and (2.11),

$$\begin{aligned} |h_s(x, \boldsymbol{\vartheta}_1^*) - h_{s^*}(x, \boldsymbol{\vartheta}_1^*)| &\le |\boldsymbol{\vartheta}_1^*|\sqrt{2(1 + x^2)}\ I(s \wedge s^* < x \le s \vee s^*) && (2.18)\\ &\le |\boldsymbol{\vartheta}_1^*|\sqrt{2(1 + x^2)}\ I(|x - s| \le |s^* - s|). && (2.19) \end{aligned}$$

Thus on $U_{\boldsymbol{\vartheta}}(\eta)$ and for $s \in \mathcal{R}$,

$$\begin{aligned} |\delta(X_0, \boldsymbol{\vartheta}^*)| &\le |h_s(X_0, \boldsymbol{\vartheta}_1) - h_{s^*}(X_0, \boldsymbol{\vartheta}_1)| + |h_{s^*}(X_0, \boldsymbol{\vartheta}_1) - h_{s^*}(X_0, \boldsymbol{\vartheta}_1^*)| \\ &\le [\sqrt{2}|\boldsymbol{\vartheta}_1| I(|X_0 - s| \le |s^* - s|) + |\boldsymbol{\vartheta}_1 - \boldsymbol{\vartheta}_1^*|]\sqrt{(1 + X_0^2)} \\ &\le [\sqrt{2}|\boldsymbol{\vartheta}_1| I(|X_0 - s| \le |s_0(\eta) - s|) + \eta]\sqrt{(1 + X_0^2)} \\ &\equiv \Delta(\eta, X_0),\ (\textit{say}), && (2.20) \end{aligned}$$

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

Note that

$$\varphi(\epsilon_1(\boldsymbol{\vartheta})) = [\varphi(\epsilon_1 + h(X_0, \boldsymbol{\theta}) - h(X_0, \boldsymbol{\vartheta})) - \varphi(\epsilon_1)] + \varphi(\epsilon_1). \tag{2.21}$$

Condition (C2), $EX_0^2 < \infty$ and (2.12) imply that there exists a constant L such that for any $\boldsymbol{\vartheta} \in \Omega$,

$$\begin{aligned} E\varphi^2(\epsilon_1(\boldsymbol{\vartheta})) &\leq 2E\left[\varphi^2(\epsilon_1) + L|h(X_0, \boldsymbol{\theta}) - h(X_0, \boldsymbol{\vartheta})|^2\right] \\ &\leq 2I(f) + 4L(|\boldsymbol{\theta}_1|^2 + |\boldsymbol{\vartheta}_1|^2)E(1 + X_0^2) < \infty. \end{aligned} \tag{2.22}$$

The absolute continuity of $\ln f$, which follows from (C1), and (2.20) imply that

$$|\psi(X_0, \epsilon_1, \boldsymbol{\vartheta}^*) - \psi(X_0, \epsilon_1, \boldsymbol{\vartheta})| \leq \int_{-\Delta(\eta, X_0)}^{\Delta(\eta, X_0)} |\varphi(\epsilon_1(\boldsymbol{\vartheta}) + v)| \, dv. \tag{2.23}$$

Thus, (2.22), (2.23) and Cauchy-Schwarz inequality imply that the

$$\begin{aligned} \text{LHS of (2.15)} &\leq E\{[|2\varphi(\epsilon_1(\boldsymbol{\vartheta}))| + L\Delta(\eta, X_0)]\Delta(\eta, X_0)\} \\ &\leq 2(E\varphi^2(\epsilon_1(\boldsymbol{\vartheta})))^{1/2}(E\Delta^2(\eta, X_0))^{1/2} + LE\Delta^2(\eta, X_0). \end{aligned} \tag{2.24}$$

Moreover,

$$E\Delta^2(\eta, X_0) = E(1 + X_0^2)\left(\sqrt{2}|\boldsymbol{\vartheta}_1|I(|X_0 - s| \leq |s_0(\eta) - s|) + \eta\right)^2 \to 0, \tag{2.25}$$

as $\eta \to 0$. Therefore, the finiteness of $I(f)$ and (2.25) imply (2.15) for any $s \in \mathcal{R}$.

In the case $s = \infty$, similar to (2.20),

$$\begin{aligned} |\delta(X_0, \boldsymbol{\vartheta}^*)| &\leq \left(\sqrt{2}|\boldsymbol{\vartheta}_1|I(X_0 > s^*) + \eta\right)\sqrt{(1 + X_0^2)} \\ &\leq \left(\sqrt{2}|\boldsymbol{\vartheta}_1|I(X_0 > s_0(\eta)) + \eta\right)\sqrt{1 + X_0^2} \\ &\equiv \Delta_1(\eta, X_0) \end{aligned}$$

where $d(s_0(\eta), \infty) = \eta$. Again,

$$E\Delta_1^2(\eta, X_0) \to 0, \text{ as } \eta \to 0.$$

Thus the proof goes through for $s = \infty$. The proof is similar in the case $s = -\infty$, except one replaces $I(X_0 > s^*)$ by $I(X_0 < s^*)$. Therefore Lemma 2.3.1 is proved. □

Proof of Theorem 2.3.1. Let $\alpha(\boldsymbol{\vartheta}) = E\psi(X_0, \epsilon_1, \boldsymbol{\vartheta})$ for $\boldsymbol{\vartheta} \in \Omega$. The conditions (C1) and (C2), the mean value theorem, the independence of ϵ_1 and X_0 and Cauchy-Schwarz inequality imply that $E|\psi(X_0, \epsilon_1, \boldsymbol{\vartheta})| < \infty$. Thus α is a well defined finite function from Ω to $\mathcal{R}$. Note that $\alpha(\boldsymbol{\theta}) = 0$ and $\ln x < x - 1$, unless $x = 1$. For any given open neighborhood V of $\boldsymbol{\theta}$ in Ω and any $\boldsymbol{\vartheta} \in V^c = \Omega \setminus V$, an conditional argument yields that

$$\begin{aligned}
\alpha(\boldsymbol{\vartheta}) &= E \ln \frac{f(\epsilon_1 + h(X_0, \boldsymbol{\theta}) - h(X_0, \boldsymbol{\vartheta}))}{f(\epsilon_1)} \qquad (2.26)\\
&= E\left\{ E\left[\ln \frac{f(\epsilon_1 + h(X_0, \boldsymbol{\theta}) - h(X_0, \boldsymbol{\vartheta}))}{f(\epsilon_1)} \Big| X_0 \right] \right\}\\
&< E\left\{ \int [f(y + h(X_0, \boldsymbol{\theta}) - h(X_0, \boldsymbol{\vartheta})) - f(y)]\, dy \right\} = 0
\end{aligned}$$

By Lemma 2.3.1, α is continuous and hence the compactness of V^c implies that there exists $\boldsymbol{\vartheta}_0 \in V^c$, such that

$$\sup_{\boldsymbol{\vartheta} \in V^c} \alpha(\boldsymbol{\vartheta}) = \alpha(\boldsymbol{\vartheta}_0) < 0.$$

Let $\delta_0 = -\alpha(\boldsymbol{\vartheta}_0)/3$. For any $\boldsymbol{\vartheta} \in V^c$, by Lemma 2.3.1 again, there exists $\eta_0 > 0$, such that

$$E \sup_{\boldsymbol{\vartheta}^* \in U_{\boldsymbol{\vartheta}}(\eta_0)} \psi(X_0, \epsilon_1, \boldsymbol{\vartheta}^*) \leq E\psi(X_0, \epsilon_1, \boldsymbol{\vartheta}) + \delta_0 \leq \alpha(\boldsymbol{\vartheta}_0) + \delta_0 = -2\delta_0. \qquad (2.27)$$

Again, the compactness of V^c implies that there exists a finite number M of $U_{\boldsymbol{\vartheta}_j}(\eta_0)$, $\boldsymbol{\vartheta}_j \in V^c$, $j = 1, 2, ..., M$ such that $\bigcup_1^M U_{\boldsymbol{\vartheta}_j}(\eta_0) = V^c$. Then by the ergodic theorem and (2.27), there exists a n_0 such that for any $n \geq n_0$, $1 \leq j \leq M$,

$$\begin{aligned}
\sup_{\boldsymbol{\vartheta}^* \in U_{\boldsymbol{\vartheta}_j}(\eta_0)} l_n(\boldsymbol{\vartheta}^*) &\leq \frac{1}{n} \sum \sup_{\boldsymbol{\vartheta}^* \in U_{\boldsymbol{\vartheta}_j}(\eta_0)} \psi(X_{i-1}, \epsilon_i, \boldsymbol{\vartheta}^*)\\
&\leq E \sup_{\boldsymbol{\vartheta}^* \in U_{\boldsymbol{\vartheta}_j}(\eta_0)} \psi(X_0, \epsilon_1, \boldsymbol{\vartheta}^*) + \delta_0 \leq -\delta_0, \quad a.\ s.
\end{aligned}$$

But,

$$\sup_{\boldsymbol{\vartheta} \in V} l_n(\boldsymbol{\vartheta}) \geq l_n(\boldsymbol{\theta}) = 0. \qquad (2.28)$$

Therefore, for any neighborhood V of $\boldsymbol{\theta}$ in Ω, $\exists\ n_0$, s.t. for all $n \geq n_0$,

$$\sup_{\boldsymbol{\vartheta}^* \in \Omega \setminus V} l_n(\boldsymbol{\vartheta}^*) \leq \max_{1 \leq j \leq M} \sup_{U_{\boldsymbol{\vartheta}_j}(\eta_0)} l_n(\boldsymbol{\vartheta}^*) \leq -\delta_0 < 0 \leq \sup_{\boldsymbol{\vartheta} \in V} l_n(\boldsymbol{\vartheta}).$$

This implies that

$$\hat{\boldsymbol{\theta}}_n \in V, \quad a.\ s.\ \forall\ V \text{and}\ \forall\ n \geq n_0.$$

By the arbitrary of V, $\hat{\boldsymbol{\theta}}_n$ goes to $\boldsymbol{\theta}$ almost surely. □

2.4 n-consistency of the threshold estimator

From now on we will invoke the condition (M). 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 2.4.1 *Suppose conditions (C1)-(C3) and (M) hold, then*

$$|n(\hat{r}_n - r)| = O_P(1).$$

The proof of Theorem 2.4.1 is technical and lengthy but interesting. We will begin with some notation. Let $J : \mathcal{R}^2 \to \mathcal{R}$ and

$$p(x) = EJ(x, \epsilon_1),\ p_1(x) = E|J(x, \epsilon_1)|,\ p_2(x) = EJ^2(x, \epsilon_1), x \in \mathcal{R}, \tag{2.29}$$

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

$$R_n(u) = \frac{1}{n}\sum J(X_{i-1}, \epsilon_i) I(r < X_{i-1} \leq r + u).$$

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

Also, let $J^c(X_{i-1}, \epsilon_i) = J(X_{i-1}, \epsilon_i) - p(X_{i-1})$. For $u_2 \geq u_1 \geq 0$, let

$$\tilde{R}_n(u_1, u_2) = \frac{1}{n}\sum |J^c(X_{i-1}, \epsilon_i)| I(r + u_1 < X_{i-1} \leq r + u_2),$$

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

Lemma 2.4.1 *Suppose that (C1) holds, then there exists constants* $0 < m \leq M < \infty$ *and* $0 < C < \infty$ *independent of* n. *For any* $0 < \delta < 1$ *and* $\forall\ u,\ u_1,\ u_2 \in [0, \delta]$, $\forall\ n$,

$$mu \leq G(u) \leq Mu, \tag{2.30}$$

$$Var(I(r < X_0 \leq r + u)) \leq CG(u), \tag{2.31}$$

$$Var(nG_n(u)) \leq nCG(u). \tag{2.32}$$

Proof. The facts (2.30) and (2.31) follow from the assumption (C1) and Remark 1. To prove (2.32), without loss of generality, assume $r = 0$. Note that

$$Var(nG_n(u)) = nVar(I(0 < X_0 \leq u)) + \sum_{k \neq j} Cov(I(0 < X_{k-1} \leq u), I(0 < X_{j-1} \leq u)).$$

But, Remark 2 and a conditioning argument yield that for any $k \geq 2$ and $u \leq \delta$,

$$\begin{aligned}
&|Cov(I(0 < X_0 \leq u), I(0 < X_{k-1} \leq u))| \\
= \ & |EI(0 < X_0 \leq u)[I(0 < X_{k-1} \leq u) - G(u)]| \\
\leq \ & EI(0 < X_0 \leq u)|[P^k(X_0, (0, u]) - G(u)]| \leq C\rho^{k-1}G(u).
\end{aligned}$$

Thus (2.32) follows from the stationarity of $\{X_i\}$ and the fact $\sum_{k \neq j} \rho^{k-j} = O(n)$. □

Lemma 2.4.2 *Suppose that the functions* p_1 *and* p_2 *are continuous over* $\mathcal{R}$. *Then, there exists constant* $0 < C < \infty$ *independent of* n *such that* $\forall\ 0 < \delta < 1$ *and* $\forall\ u, u_1, u_2 \in [0, \delta],\ u_2 \geq u_1,\ \forall\ n$,

$$\tilde{R}(u_1, u_2) \leq C(G(u_2) - G(u_1)). \tag{2.33}$$

$$Var(|J^c(X_0, \epsilon_1)| I(r + u_1 < X_0 \leq r + u_2)) \leq C(G(u_2) - G(u_1)). \tag{2.34}$$

$$Var(\tilde{R}_n(u_1, u_2)) \leq C(G(u_2) - G(u_1))/n. \tag{2.35}$$

$$Var(R_n(u) - r_n(u)) \leq CG(u)/n. \tag{2.36}$$

Proof. Without loss of generality, assume $r = 0$ again. The continuity of p_1 implies that

$$\begin{aligned}
\tilde{R}(u_1, u_2) \ &= \ E\Big\{I(u_1 < X_0 \leq u_2)E\Big[|J^c(X_0, \epsilon_1)|\ \Big|X_0\Big]\Big\} \\
&\leq \ 2 \sup_{x \in [0,1]} p_1(x) E\ I(u_1 < X_0 \leq u_2) \\
&\leq \ RHS\ of\ (2.33).
\end{aligned}$$

The continuity of p_2 and the Cauchy-Schwarz inequality imply that the

$$\begin{aligned} LHS\ (2.34) &\le E|J^c(X_0,\epsilon_1)|^2 I(u_1 < X_0 \le u_2) \\ &= E\{I(u_1 < X_0 \le u_2) E[|J^c(X_0,\epsilon_1)|^2 \big| X_0]\} \\ &\le 4 \sup_{x\in[0,1]} p_2(x) EI(u_1 < X_0 \le u_2) \\ &\le RHS\ of\ (2.34). \end{aligned} \tag{2.37}$$

Next, we shall verify (2.35). The argument is similar to the proof of (2.32). Expanding the left hand side of (2.35),

$$\begin{aligned} &Var(n\tilde{R}_n(u_1,u_2)) = nVar(|J^c(X_0,\epsilon_1)|I(u_1 < X_0 \le u_2)) \\ &\quad + \sum_{k\ne j} Cov\Big(|J^c(X_{k-1},\epsilon_k)|I(u_1 < X_{k-1} \le u_2), |J^c(X_{j-1},\epsilon_j)|I(u_1 < X_{j-1} \le u_2)\Big). \end{aligned}$$

Note that $|p(x)| \le p_1(x)$. By the continuity of p_1, the property of Markov chain and Remark 2, for any $k \ge 2$,

$$\begin{aligned} &\Big|E\Big[|J^c(X_{k-1},\epsilon_k)|I(u_1 < X_{k-1} \le u_2)\Big|X_1, X_0\Big] - E|J^c(X_0,\epsilon_1)|I(u_1 < X_0 \le u_2)\Big| \\ =\ &\Big|E\Big[\int |J^c(X_{k-1},y)|\, dF(y) I(u_1 < X_{k-1} \le u_2)\Big|X_1\Big] - \int_{u_1}^{u_2}\int |J^c(x,y)|\, dF(y) dG_{\boldsymbol{\theta}}(x)\Big| \\ \le\ &\int_{u_1}^{u_2} \Big[\int |J^c(x,y)|\, dF(y)\Big] \Big(\big|P^{k-2}(X_1,dx) - G_{\boldsymbol{\theta}}(dx)\big|\Big) \\ \le\ &2 \sup_{x\in[0,1]} p_1(x) \|P^{k-2}(X_1,\cdot) - G_{\boldsymbol{\theta}}(\cdot)\|_{tv} \\ \le\ &C\rho^{k-2}(1+|X_1|) \le C\rho^{k-2}(1+|h(\boldsymbol{\theta},X_0)|+|\epsilon_1|). \end{aligned} \tag{2.38}$$

The continuity of p_2 and $p_1^2(x) \le p_2(x)$, $E\epsilon_1^2 < \infty$, the Cauchy-Schwarz inequality and an argument like (2.37) imply that

$$\begin{aligned} &E\left[|J^c(X_0,\epsilon_1)\epsilon_1|I(u_1 < X_0 \le u_2)\right] \\ =\ &E\Big\{E\left[|J^c(X_0,\epsilon_1)\epsilon_1|\Big|X_0\right] I(u_1 < X_0 \le u_2)\Big\} \\ \le\ &2\sqrt{\sup_{x\in[0,1]} p_2(x) E(\epsilon_1)^2}\ (G(u_2) - G(u_1)). \end{aligned} \tag{2.39}$$

Then the definition of the autoregressive function h, the Markov property of $\{X_i\}$, (2.38), (2.39) and Remark 2 yield that

$$\Big|Cov\Big(|J^c(X_0,\epsilon_1)|I(u_1 < X_0 \le u_2), |J^c(X_{k-1},\epsilon_k)|I(u_1 < X_{k-1} \le u_2)\Big)\Big|$$

$$= \Big|E\Big\{|J^c(X_0,\epsilon_1)|I(u_1 < X_0 \le u_2) \times$$
$$\Big(E\Big[|J^c(X_{k-1},\epsilon_k)|I(u_1 < X_{k-1} \le u_2)\Big|X_1, X_0\Big] - E|J^c(X_0,\epsilon_1)|I(u_1 < X_0 \le u_2)\Big)\Big\}\Big|$$
$$\le C\rho^{k-2} E|J^c(X_0,\epsilon_1)|I(u_1 < X_0 \le u_2)\Big(1 + |h(\boldsymbol{\theta}, X_0)| + |\epsilon_1|\Big)$$
$$\le C\rho^{k-2}(G(u_2) - G(u_1)),$$

Therefore (2.35) follows from the stationarity of $\{X_i\}$ and the fact $\sum_{k\neq j} \rho^{k-j} = O(n)$. The proof of (2.36) follows from the property of the square integrable martingale $R_n(u) - r_n(u)$, for fixed $u > 0$. That is,

$$Var(n(R_n(u) - r_n(u))) = nVar(J^c(X_0,\epsilon_1)I(u_1 < X_0 \le u_2)).$$

This completes the proof of Lemma 2.4.2.

Proposition 2.4.1 *Suppose that (C1) holds and the functions p_1 and p_2 are continuous. Then, for each $\epsilon > 0$, $\eta > 0$, there is a constant $B < \infty$, $\forall\ 0 < \delta < 1$ and $\forall\ n \ge [B/\delta] + 1$,*

$$P\left(\sup_{B/n<u\le\delta} |G_n(u)/G(u) - 1| < \eta\right) > 1 - \epsilon, \tag{2.40}$$

$$P\left(\sup_{B/n<u\le\delta} \left|\frac{R_n(u) - r_n(u)}{G(u)}\right| < \eta\right) > 1 - \epsilon, \tag{2.41}$$

Note. The condition (C1) is for (2.40), the continuity of p_1 and p_2 is for (2.41).

Proof of Proposition 2.4.1. For any $B > 0$ and $0 < \delta < 1$, choose a partition of the interval $(B/n, \delta]$ as follows: Fix a $b > 1$ and let M_0 be the greatest integer less than or equal to $\ln(n\delta/B)/\ln b$. Note that

$$(B/n,\delta] = \bigcup_{i=0}^{M_0} I_i, \quad I_i = (b^i B/n, b^{i+1}B/n],\ i = 0, ..., M_0 - 1,\ I_{M_0} = (b^{M_0}B/n, \delta].$$

Then (2.30) and (2.32) of Lemma 2.4.1 imply that $\forall\ \eta_1 > 0$,

$$\begin{aligned} &P(\sup_i |G_n(b^iB/n)/G(b^iB/n) - 1| \ge \eta_1) \\ &\le \sum_i Var(G_n(b^iB/n))/(\eta_1^2 G^2(b^iB/n)) \\ &\le C\sum_i 1/(m\eta_1^2 Bb^i) = C/(m\eta_1^2 B(1 - b^{-1})). \end{aligned} \tag{2.42}$$

For $0 < x \le y \le bx \le \delta$ with $|G_n(x)/G(x) - 1| < \eta_1$ and $|G_n(bx)/G(bx) - 1| < \eta_1$, we have,

$$\begin{aligned}(1-\eta_1)G(x)/G(bx) - 1 &\le G_n(x)/G(bx) - 1 \le G_n(y)/G(y) - 1 \\ &\le G_n(bx)/G(x) - 1 \le G(bx)/G(x)(1+\eta_1) - 1. \quad (2.43)\end{aligned}$$

The strictly increasing property of G and Dini theorem imply that $\forall\, \eta > 0$, one can choose $\eta_1 > 0$ and $b > 1$ sufficiently small such that

$$\max_{0\le i\le M_0} \left|\frac{G(b^iB/n)}{G(b^{i+1}B/n)}(1-\eta_1) - 1\right| \vee \left|\frac{G(b^{i+1}B/n)}{G(b^iB/n)}(1+\eta_1) - 1\right| < \eta. \qquad (2.44)$$

Now let

$$A_n = \left\{\max_{0\le i\le M_0}\left|\frac{G_n(b^iB/n)}{G(b^iB/n)} - 1\right| < \eta_1, \max_{0\le i\le M_0-1}\left|\frac{G_n(b^{i+1}B/n)}{G(b^{i+1}B/n)} - 1\right| < \eta_1\right\}.$$

Then on A_n, (2.43) and (2.44) imply that

$$\begin{aligned}&\sup_{B/n<u\le\delta} |G_n(u)/G(u) - 1| = \max_{0\le i\le M_0}\sup_{u\in I_i}|G_n(u)/G(u) - 1| \\ &\le \max_{0\le i\le M_0}\left\{\left|\frac{G(b^iB/n)}{G(b^{i+1}B/n)}(1-\eta_1) - 1\right| \vee \left|\frac{G(b^{i+1}B/n)}{G(b^iB/n)}(1+\eta_1) - 1\right|\right\} \\ &< \eta.\end{aligned}$$

And then by (2.42), choosing B sufficiently large and $n_0 = [B/\delta] + 1$, for any $n \ge n_0$,

$$P(\sup_{B/n<u\le\delta}|G_n(u)/G(u) - 1| \ge \eta) \le P(A_n^c) < \epsilon. \qquad (2.45)$$

Hence (2.40) holds. □

To prove (2.41), for $b^iB/n < u \le b^{i+1}B/n$, $i \ge 0$,

$$\begin{aligned}|R_n(u) - r_n(u)| &\le |R_n(u) - r_n(u) - (R_n(b^iB/n) - r_n(b^iB/n)| \\ &\quad + |R_n(b^iB/n) - r_n(b^iB/n)| \\ &\le \tilde{R}_n(b^iB/n, u) + |R_n(b^iB/n) - r_n(b^iB/n)|.\end{aligned}$$

Thus, by the increasing property of $\tilde{R}_n$ and G,

$$\begin{aligned}&\sup_{\frac{b^iB}{n}<u\le\frac{b^{i+1}B}{n}}\left|\frac{R_n(u) - r_n(u)}{G(u)}\right| \\ &\le \frac{\tilde{R}_n(b^iB/n, b^{i+1}B/n)}{G(b^iB/n)} + \frac{|R_n(b^iB/n) - r_n(b^iB/n)|}{G(b^iB/n)}. \qquad (2.46)\end{aligned}$$

Also, note that (2.30) of Lemma 2.4.1 and (2.36) of Lemma 2.4.2 imply that for any $\eta_1 > 0$,

$$\begin{aligned}
&P\left(\sup_i \left|\frac{R_n(b^iB/n) - r_n(b^iB/n)}{G(b^iB/n)}\right| \geq \eta_1\right) \\
&\leq \sum_i Var(R_n(b^iB/n) - r_n(b^iB/n))/(\eta_1^2 G^2(b^iB/n)) \\
&\leq C\sum_i 1/(m\eta_1^2 B b^i) = C/(m\eta_1^2 B(1-b^{-1})),
\end{aligned} \tag{2.47}$$

and (2.30) and (2.35) yield that

$$\begin{aligned}
&P\left(\sup_i |\tilde{R}_n(b^iB/n, b^{i+1}B/n)/G(b^iB/n) - \tilde{R}(b^iB/n, b^{i+1}B/n)/G(b^iB/n)| \geq \eta_1\right) \\
&\leq \sum_i Var(\tilde{R}_n(b^iB/n, b^{i+1}B/n))/(\eta_1^2 G^2(b^iB/n)) \\
&\leq C(b-1)/(\eta_1^2 m^2 B)\sum_i 1/b^i = Cb/(\eta_1^2 m^2 B).
\end{aligned} \tag{2.48}$$

By (2.30) and (2.33),

$$\sup_i |\tilde{R}(b^iB/n, b^{i+1}B/n)/G(b^iB/n)| \leq C(b^iB(b-1)/n)/(mb^iB/n) \leq C(b-1)/m. \tag{2.49}$$

Thus for any $\epsilon > 0$, $\eta > 0$, one can choose $\eta_1 > 0$ and $b > 1$ sufficiently small such that $\eta_1 + C(b-1)/m < \eta$ and then choose sufficiently large B such that

$$B > \frac{2C}{m\eta_1^2(1-b^{-1})\epsilon} \vee \frac{2b}{m^2\eta_1^2\epsilon}.$$

Then, by choosing $n_0 = [B/\delta] + 1$, (2.46)-(2.49) imply that for any $n \geq n_0$,

$$P\left(\sup_{B/n<u\leq\delta}\left|\frac{R_n(u) - r_n(u)}{G(u)}\right| \geq \eta\right) < \epsilon. \tag{2.50}$$

This completes the proof of Proposition 2.4.1. □

Now let

$$J(X_{i-1}, \epsilon_i) \equiv \psi(X_{i-1}, \epsilon_i) \equiv \ln\frac{f(\epsilon_i + a + \beta X_{i-1})}{f(\epsilon_i)},$$

where $a = b_0 - a_0$, $\beta = b_1 - a_1$. The functions p, p_1 and p_2 are defined correspondingly. For $u \geq 0$, define

$$D_n(u) = \frac{1}{n}\sum \psi(X_{i-1}, \epsilon_i) I(r < X_{i-1} \leq r + u).$$

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

Corollary 2.4.1 *Suppose that (C1)-(C2) hold. Then, for each $\epsilon > 0$ and $\eta > 0$, there is a constant $B < \infty$ such that $\forall\ 0 < \delta < 1$ and $\forall\ n \geq [B/\delta] + 1$,*

$$P\left(\sup_{B/n<u\leq\delta}\left|\frac{D_n(u) - d_n(u)}{G(u)}\right| < \eta\right) > 1 - \epsilon, \tag{2.51}$$

Proof. The continuity of p_1, p_2 can be derived from the conditions (C1) and (C2) (See Appendix). Thus (2.51) follows (2.41) immediately. □

Before proving Theorem 2.4.1, we need some more notation. Write

$$\tilde{\psi}(X_{i-1}, \epsilon_i, \boldsymbol{t}, s) = \ln\frac{f(X_i - h_s(X_{i-1}, \boldsymbol{t}))}{f(\epsilon_i)}, \ \boldsymbol{t} \in \mathcal{R}^4, \ s \in \bar{\mathcal{R}},$$

where h_s is defined in (2.16). Let

$$\xi(X_{i-1}, \epsilon_i, \boldsymbol{t}, s) = \tilde{\psi}(X_{i-1}, \epsilon_i, \boldsymbol{t}, s) - \tilde{\psi}(X_{i-1}, \epsilon_i, \boldsymbol{t}, r), \ \boldsymbol{t} \in \mathcal{R}^4, \ s \in \bar{\mathcal{R}}, \ 1 \leq i \leq n.$$

Then, for $1 \leq i \leq n$, $\boldsymbol{t} \in \mathcal{R}^4$, $s \in \bar{\mathcal{R}}$,

$$\begin{aligned}
\dot{\xi}(X_{i-1}, \epsilon_i, \boldsymbol{t}, s) &\equiv \frac{\partial}{\partial \boldsymbol{t}}(\xi(X_{i-1}, \epsilon_i, \boldsymbol{t}, s)) \\
&= -\varphi(X_i - h_s(X_{i-1}, \boldsymbol{t}))\dot{h}_s(X_{i-1}) - \varphi(X_i - h_r(X_{i-1}, \boldsymbol{t}))\dot{h}_r(X_{i-1}) \\
&= -\left[\varphi(X_i - h_s(X_{i-1}, \boldsymbol{t})) - \varphi(X_i - h_r(X_{i-1}, \boldsymbol{t}))\right]\dot{h}_s(X_{i-1}) \\
&\quad -\varphi(X_i - h_r(X_{i-1}, \boldsymbol{t}))\left[\dot{h}_s(X_{i-1}) - \dot{h}_r(X_{i-1})\right].
\end{aligned}$$

Now we are ready to prove Theorem 2.4.1.

Proof of Theorem 2.4.1. Since $\hat{\boldsymbol{\theta}}_n$ is strongly consistent by Theorem 2.3.1, without loss of generality, the parameter space can be restricted to a neighborhood of $\boldsymbol{\theta}$, say,

$$\Omega(\delta) = \{\boldsymbol{\vartheta} \in \Omega : |\boldsymbol{\vartheta}_1 - \boldsymbol{\theta}_1| < \delta, |s - r| < \delta\},$$

for some $0 < \delta < 1$ to be determined later. Then, it suffices to show that $l_n(\boldsymbol{\vartheta}_1, s) - l_n(\boldsymbol{\vartheta}_1, r)$ is negative for $n|s - r|$ large enough. More specifically, we shall show that for every $\epsilon > 0$, there is a $B > 0$ and $\gamma > 0$, $1 > \delta > 0$ and n_0 such that for any $n \geq n_0$,

$$P\left(\sup_{B/n<|s-r|\leq\delta, \boldsymbol{\vartheta}\in\Omega(\delta)}\frac{[l_n(\boldsymbol{\vartheta}_1, s) - l_n(\boldsymbol{\vartheta}_1, r)]}{G(|s - r|)} < -\gamma\right) > 1 - 2\epsilon. \tag{2.52}$$

For any $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1^T, s)^T$, denote $l_{ns}(\boldsymbol{\vartheta}_1) = l_n(\boldsymbol{\vartheta}_1, s) = l_n(\boldsymbol{\vartheta})$. Now, decompose $l_n(\boldsymbol{\vartheta}_1, s) - l_n(\boldsymbol{\vartheta}_1, r)$ into two terms as follows:

$$\begin{aligned} l_n(\boldsymbol{\vartheta}_1, s) - l_n(\boldsymbol{\vartheta}_1, r) &= [l_{ns}(\boldsymbol{\vartheta}_1) - l_{nr}(\boldsymbol{\vartheta}_1) - (l_{ns}(\boldsymbol{\theta}_1) - l_{nr}(\boldsymbol{\theta}_1))] + [l_{ns}(\boldsymbol{\theta}_1) - l_{nr}(\boldsymbol{\theta}_1)] \\ &\equiv l_n^1(\boldsymbol{\vartheta}) + l_n^2(s), \ (say). \end{aligned}$$

We shall prove that there exists a δ small enough such that

$$\sup_{B/n<|s-r|\le\delta, \boldsymbol{\vartheta}\in\Omega(\delta)} \left| \frac{l_n^1(\boldsymbol{\vartheta})}{G(|s-r|)} \right| = o_P(1), \tag{2.53}$$

and

$$P\left(\sup_{B/n<|s-r|\le\delta} \frac{l_n^2(s)}{G(|s-r|)} < -2\gamma \right) > 1 - \epsilon. \tag{2.54}$$

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 (2.53), by using the absolute continuity of ψ,

$$\begin{aligned} l_n^1(\boldsymbol{\vartheta}) &\equiv \frac{1}{n}\sum [\xi(X_{i-1}, \epsilon_i, \boldsymbol{\vartheta}_1, s) - \xi(X_{i-1}, \epsilon_i, \boldsymbol{\theta}_1, s)] \\ &= \frac{1}{n}\sum \int_0^1 \dot{\xi}^T(X_{i-1}, \epsilon_i, \boldsymbol{\theta}_1 + v(\boldsymbol{\vartheta}_1 - \boldsymbol{\theta}_1), s)(\boldsymbol{\vartheta}_1 - \boldsymbol{\theta}_1)\, dv. \end{aligned} \tag{2.55}$$

By the Lip(1) of φ, (2.17), (2.18), (2.9), (2.10) and (2.11) imply that there exists a constant L, for $1 \le i \le n$, $\boldsymbol{t} \in \mathcal{R}^4$, $s \in \bar{\mathcal{R}}$,

$$\begin{aligned} & |\dot{\xi}(X_{i-1}, \epsilon_i, \boldsymbol{t}, s)| \\ \le\ & L\left|h_s(X_{i-1}, \boldsymbol{t}) - h_r(X_{i-1}, \boldsymbol{t})\right|\sqrt{1 + X_{i-1}^2} \\ & + \left[|\varphi(\epsilon_i)| + L|h_r(X_{i-1}, \boldsymbol{\theta}_1) - h_r(X_{i-1}, \boldsymbol{t})|\right]\sqrt{1 + X_{i-1}^2}\ I(s \wedge r < X_{i-1} \le s \vee r) \\ \le\ & \left(L|\boldsymbol{t}|\sqrt{1 + X_{i-1}^2} + \left[|\varphi(\epsilon_i)| + L|\boldsymbol{\theta}_1 - \boldsymbol{t}|\sqrt{1 + X_{i-1}^2}\right]\right) \\ & \times\sqrt{1 + X_{i-1}^2}\, I(s \wedge r < X_{i-1} \le s \vee r). \end{aligned} \tag{2.56}$$

Thus, for any $\boldsymbol{\vartheta} \in \Omega(\delta)$, by (2.55) and (2.56), there exists $0 < C < \infty$, such that

$$\left| \frac{l_n^1(\boldsymbol{\vartheta})}{G(u)} \right| \le C\delta \left[\left| \frac{G_n(u)}{G(u)} \right| + \left| \frac{R_n(u) - r_n(u)}{G(u)} \right| \right],$$

with $J(x, y) = |\varphi(y)|$ in the definitions of R_n and r_n. Thus, (2.53) follows from the Proposition 2.4.1 by choosing $\delta > 0$ sufficiently small.

To prove (2.54), recall that $a = b_0 - a_0, \beta = b_1 - a_1$ and let $J = \psi$ in the definition of p in (2.29). Then $l_n^2(s)$ can be decomposed as follows:

$$\begin{aligned} l_n^2(s) &\equiv l_n^2(r+u) = \frac{1}{n}\sum \psi(X_{i-1}, \epsilon_i) I(r < X_{i-1} \le r+u) \\ &= \frac{1}{n}\sum \left[\ln \frac{f(\epsilon_i + a + \beta X_{i-1})}{f(\epsilon_i)} - p(a+\beta X_{i-1})\right] I(r < X_{i-1} \le r+u) \\ &\quad + \frac{1}{n}\sum [\, p(a+\beta X_{i-1}) - p(a+\beta r)] I(r < X_{i-1} \le r+u) + p(a+\beta r) G_n(u) \\ &\equiv [D_n(u) - d_n(u)] + E_n(u) + I_n(u) \end{aligned}$$

Then

$$\sup_{\frac{B}{n}<u\le\delta} \frac{l_n^2(r+u)}{G(u)} \le \sup_{\frac{B}{n}<u\le\delta} \left\{ \left|\frac{D_n(u)-d_n(u)}{G(u)}\right| + \left|\frac{E_n(u)}{G(u)}\right| + \left|\frac{G_n(u)}{G(u)} - 1\right| |p(a+\beta r)| \right\} + p(a+\beta r). \tag{2.57}$$

Note that $p(a+\beta r) = \int \ln[f(x+a+\beta r)/f(x)]dF(x) < 0$ by the condition (M). Fix an $\eta > 0$ such that $\gamma = [-p(a+\beta r) - \eta(2 + |p(a+\beta r)|)]/2 > 0$. Note that

$$\begin{aligned} &\sup_{B/n<u\le\delta} \left|\frac{E_n(u)}{G(u)}\right| \\ \le\ & \sup_{B/n<u\le\delta} \frac{1}{n}\sum |p(a+\beta X_{i-1}) - p(a+\beta r)| \frac{I(r < X_{i-1} \le r+u)}{G(u)} \\ \le\ & \sup_{x\in[0,\delta]} |p(a+\beta r+\beta x) - p(a+\beta r)| \left[\left|\frac{G_n(u)}{G(u)} - 1\right| + 1\right] \end{aligned} \tag{2.58}$$

By the Proposition 2.4.1 and the continuity of p (see Appendix), we can see that $\sup_{B/n<u\le\delta} \left|\frac{E_n(u)}{G(u)}\right|$ goes to zero in probability for sufficiently small $\delta > 0$. Let

$$\mathcal{A} = \left\{ \sup_{B/n<u\le\delta} \left[\left|\frac{D_n(u)-d_n(u)}{G(u)}\right| + \left|\frac{E_n(u)}{G(u)}\right| + \left|\frac{G_n(u)}{G(u)} - 1\right| |p(a+\beta r)|\right] < \eta(2 + |p(a+\beta r)|)\right\},$$

$$\mathcal{B} = \left\{ \sup_{B/n<u\le\delta} \left|\frac{D_n(u)-d_n(u)}{G(u)}\right| < \eta \right\},$$

$$\mathcal{C} = \left\{ \sup_{B/n<u\le\delta} \left|\frac{E_n(u)}{G(u)}\right| < \eta \right\},$$

$$\mathcal{D} = \left\{ \sup_{B/n<u\le\delta} \left|\frac{G_n(u)}{G(u)} - 1\right| < \eta \right\},$$

$$\mathcal{E} = \left\{ \sup_{B/n<u\le\delta} \frac{l_n^2(r+u)}{G(u)} < -2\gamma \right\}.$$

Observe that $\mathcal{B} \cap \mathcal{C} \cap \mathcal{D} \subset \mathcal{A}$ and (2.57) implies $\mathcal{A} \subset \mathcal{E}$. Hence, by (2.40), (2.51) and choosing $\delta > 0$ sufficiently small,

$$P(\mathcal{E}) \geq P(\mathcal{A}) \geq P(\mathcal{B} \cap \mathcal{C} \cap \mathcal{D}) > 1 - \epsilon. \tag{2.59}$$

This completes the proof of (2.54).

Thus, $\forall\, \epsilon > 0$, there exists $\gamma > 0$ and $\delta > 0$ sufficiently small such that with (2.53) and (2.54),

$$\begin{aligned}
& P\Big(\sup_{\frac{B}{n} < |s-r| \leq \delta, \boldsymbol{\vartheta} \in \Omega(\delta)} \frac{[l_n(\boldsymbol{\vartheta}_1, s) - l_n(\boldsymbol{\vartheta}_1, r)]}{G(|s-r|)} < -\gamma \Big) \\
\geq\ & P\left(\sup_{\frac{B}{n} < |s-r| \leq \delta, \boldsymbol{\vartheta} \in \Omega(\delta)} \left| \frac{l_n^1(\boldsymbol{\vartheta})}{G(|s-r|)} \right| < \gamma, \sup_{\frac{B}{n} < |s-r| \leq \delta} \frac{l_n^2(s)}{G(|s-r|)} < -2\gamma \right) \\
\geq\ & 1 - P\left(\sup_{\frac{B}{n} < |s-r| \leq \delta, \boldsymbol{\vartheta} \in \Omega(\delta)} \left| \frac{l_n^1(\boldsymbol{\vartheta})}{G(|s-r|)} \right| \geq \gamma \right) - P\left(\sup_{\frac{B}{n} < |s-r| \leq \delta} \frac{l_n^2(s)}{G(|s-r|)} \geq -2\gamma \right) \\
>\ & 1 - 2\epsilon.
\end{aligned}$$

This ends the proof of (2.52) and hence of Theorem 2.4.1. $\square$

Chapter 3

Limiting distribution of $\hat{\boldsymbol{\theta}}_{1n}$

We now consider the limiting distribution of $\hat{\boldsymbol{\theta}}_{1n}$. Recall that

$$\psi(X_{i-1}, \epsilon_i, \boldsymbol{\vartheta}) = \ln \frac{f(\epsilon_i + h(X_{i-1}, \boldsymbol{\theta}) - h(X_{i-1}, \boldsymbol{\vartheta}))}{f(\epsilon_i)}, \ \boldsymbol{\vartheta} \in \Omega,$$

and the log likelihood ratio function is

$$l_n(\boldsymbol{\vartheta}) = \frac{1}{n} \sum \psi(X_{i-1}, \epsilon_i, \boldsymbol{\vartheta}), \ \boldsymbol{\vartheta} \in \Omega.$$

In the definition of the MLE $\hat{\boldsymbol{\theta}}_n$, the first four components of the parameter point in Ω is treated separately from the last component and we have proved that $\hat{r}_n$ is $n-$consistent. Thus, we need some results of $\boldsymbol{\vartheta}_{1n}(s)$ 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 3.1.1.

3.1 Uniform consistency

Theorem 3.1.1 *Suppose that (C1) and (C2) hold. For any* $0 < B < \infty$,

$$\sup_{|s-r| \le B/n} |\boldsymbol{\vartheta}_{1n}(s) - \boldsymbol{\theta}_1| = o_P(1).$$

First of all, we need an analogue of the Lemma 2.3.1. Recall that $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1^T, s)^T$ and $l_{ns}(\boldsymbol{\vartheta}_1) = l_n(\boldsymbol{\vartheta}_1, s)$. Now write

$$\psi_s(X_{i-1}, \epsilon_i, \boldsymbol{\vartheta}_1) = \psi(X_{i-1}, \epsilon_i, \boldsymbol{\vartheta}_1, s).$$

Let $\eta > 0$, define

$$U_{\vartheta_1}(\eta) = \{\vartheta_1^* \in K : |\vartheta_1^* - \vartheta_1| < \eta\}.$$

Lemma 3.1.1 *Under the conditions (C1) and (C2), for any $\vartheta_1 \in K$ and its neighborhood $U_{\vartheta_1}(\eta)$ in K,*

$$E \sup_{s \in \bar{\mathcal{R}}, \vartheta_1^* \in U_{\vartheta_1}(\eta)} |\psi_s(X_0, \epsilon_1, \vartheta_1^*) - \psi_s(X_0, \epsilon_1, \vartheta_1)| \to 0, \quad as\ \eta \to 0. \tag{3.1}$$

Proof. Let $\delta_s(X_0, \vartheta_1^*) = h_s(X_0, \vartheta_1^*) - h_s(X_0, \vartheta_1)$. Observe that by (2.17) on $U_{\vartheta_1}(\eta)$,

$$|\delta_s(X_0, \vartheta_1^*)| \le |\vartheta_1^* - \vartheta_1| \sqrt{1 + X_0^2} \le \eta \sqrt{1 + X_0^2}. \tag{3.2}$$

By the absolute continuity of $\ln f$, which follows from (C1),

$$|\psi_s(X_0, \epsilon_1, \vartheta_1^*) - \psi_s(X_0, \epsilon_1, \vartheta_1)| \le \int_{-\eta\sqrt{1+X_0^2}}^{\eta\sqrt{1+X_0^2}} |\varphi(\epsilon_1(\vartheta) + v)|\, dv. \tag{3.3}$$

The condition (C2), $EX_0^2 < \infty$ and (2.12) imply that

$$E \sup_{s \in \bar{\mathcal{R}}} \varphi^2(\epsilon_1(\vartheta_1, s)) < \infty. \tag{3.4}$$

Therefore, (3.3)-(3.4) and the finiteness of the second moment of X_0 yield that the

$$\begin{aligned} \text{LHS of (3.1)} &\le \eta E \left[\sup_{s \in \bar{\mathcal{R}}} 2|\varphi(\epsilon_1(\vartheta))| \sqrt{1 + X_0^2} + LE(1 + X_0^2)\eta^2 \right] \\ &\le 2\eta \left[E \sup_{s \in \bar{\mathcal{R}}} \varphi^2(\epsilon_1(\vartheta)) \right]^{1/2} [E(1 + X_0^2)]^{1/2} + LE(1 + X_0^2)\eta^2 \\ &\to 0, \text{ as } \eta \to 0, \end{aligned}$$

thereby completing the proof of Lemma 3.1.1. □

Proof of Theorem 3.1.1. The following argument is similar to the one used in the proof of Theorem 2.3.1, except here we deal with the component ϑ_1 of ϑ for all $s \in \bar{\mathcal{R}}$. Let $\alpha_s(\vartheta_1) = \alpha(\vartheta_1, s)$. By the definition of the function α, for any open neighborhood V_1 of θ_1 in K and any $\vartheta_1 \in V_1^c$, $s \in \bar{\mathcal{R}}$, $\alpha(\vartheta) < 0$.

Given $V_1 \subset K$, by the continuity of α and compactness of $(K \setminus V_1) \times \bar{\mathcal{R}}$, there exists a $\vartheta_{01} \in (K \setminus V_1) \times \bar{\mathcal{R}}$, such that

$$\sup_{\vartheta \in (K \setminus V_1) \times \bar{\mathcal{R}}} \alpha(\vartheta) = \alpha(\vartheta_{01}) < 0.$$

Let $\delta_{01} = -\alpha(\boldsymbol{\vartheta}_{01})/3 > 0$. By using Lemma 3.1.1, there exists an $\eta_{01} > 0$,

$$E \sup_{s\in\bar{\mathcal{R}}} \sup_{\boldsymbol{\vartheta}_1^*\in U_{\boldsymbol{\vartheta}_1}(\eta_{01})} \psi_s(X_0, \epsilon_1, \boldsymbol{\vartheta}_1^*) \le E\psi_s(X_0, \epsilon_1, \boldsymbol{\vartheta}_1) + \delta_{01} \le \alpha(\boldsymbol{\vartheta}_{01}) + \delta_{01} = -2\delta_{01}. \quad (3.5)$$

By the compactness of $K \setminus V_1$ again, there exists a finite number M_1 of $U_{\boldsymbol{\vartheta}_{1j}}(\eta_{01})$, $\boldsymbol{\vartheta}_{1j} \in K \setminus V_1$, $j = 1, 2, ..., M_1$ such that $\bigcup_1^{M_1} U_{\boldsymbol{\vartheta}_{1j}}(\eta_{01}) = K \setminus V_1$. Then by the ergodic theorem and (3.5), there exists an n_1 such that for any $n \ge n_1$, $1 \le j \le M_1$,

$$\begin{aligned}
\sup_{s\in\bar{\mathcal{R}}} \sup_{\boldsymbol{\vartheta}_1^*\in U_{\boldsymbol{\vartheta}_{1j}}(\eta_{01})} l_{ns}(\boldsymbol{\vartheta}_1^*) &\le \frac{1}{n}\sum \sup_{s\in\bar{\mathcal{R}}} \sup_{\boldsymbol{\vartheta}_1^*\in U_{\boldsymbol{\vartheta}_{1j}}(\eta_{01})} \psi_s(X_{i-1}, \epsilon_i, \boldsymbol{\vartheta}_1^*) \\
&\le E \sup_{s\in\bar{\mathcal{R}}} \sup_{\boldsymbol{\vartheta}_1^*\in U_{\boldsymbol{\vartheta}_{1j}}(\eta_{01})} \psi_s(X_0, \epsilon_1, \boldsymbol{\vartheta}_1^*) + \delta_{01} \\
&\le -\delta_{01}, \ a.s.,
\end{aligned}$$

which implies that

$$\sup_{s\in\bar{\mathcal{R}}} \sup_{\boldsymbol{\vartheta}_1^*\in K\setminus V_1} l_{ns}(\boldsymbol{\vartheta}_1^*) \le \sup_{s\in\bar{\mathcal{R}}} \max_{1\le j\le M_1} \sup_{\boldsymbol{\vartheta}_1^*\in U_{\boldsymbol{\vartheta}_{1j}}(\eta_{01})} l_{ns}(\boldsymbol{\vartheta}_1^*) \le -\delta_{01}, \ a.s. \quad (3.6)$$

But,

$$\sup_{s\in\bar{\mathcal{R}}} \sup_{\boldsymbol{\vartheta}_1\in V_1} l_{ns}(\boldsymbol{\vartheta}_1) \ge \sup_{s\in\bar{\mathcal{R}}} l_{ns}(\boldsymbol{\theta}_1). \quad (3.7)$$

Taylor's expansion of $\ln f$ at ϵ_i yields that $\exists\ \gamma,\ |\gamma| < 1$, such that,

$$l_{ns}(\boldsymbol{\theta}_1) = \frac{1}{n}\sum \varphi(\epsilon_i + \gamma(h_r(X_{i-1}, \boldsymbol{\theta}_1) - h_s(X_{i-1}, \boldsymbol{\theta}_1)))[h_r(X_{i-1}, \boldsymbol{\theta}_1) - h_s(X_{i-1}, \boldsymbol{\theta}_1)].$$

Then the Lip(1) condition of φ and (2.19) imply that

$$\begin{aligned}
l_{ns}(\boldsymbol{\theta}_1) &\le \frac{1}{n}\sum \left[|\varphi(\epsilon_i)| + L|h_r(X_{i-1}, \boldsymbol{\theta}_1) - h_s(X_{i-1}, \boldsymbol{\theta}_1)|\right] \\
&\qquad \times \left[|\boldsymbol{\theta}_1|\sqrt{1 + X_{i-1}^2}\, I(|X_{i-1} - r| \le |s - r|)\right] \\
&\le \frac{1}{n}\sum \left[|\varphi(\epsilon_i)| + L|\boldsymbol{\theta}_1|\sqrt{1 + X_{i-1}^2}\right] |\boldsymbol{\theta}_1|\sqrt{1 + X_{i-1}^2}\, I(|X_{i-1} - r| \le |s - r|). \quad (3.8)
\end{aligned}$$

Therefore, for any B, $0 < B < \infty$, (3.8) yields that

$$\begin{aligned}
&\sup_{|s-r|\le B/n} |l_{ns}(\boldsymbol{\theta}_1)| \\
&\le \frac{1}{n}\sum \left[|\varphi(\epsilon_i)| + L|\boldsymbol{\theta}_1|\sqrt{1 + X_{i-1}^2}\right] |\boldsymbol{\theta}_1|\sqrt{1 + X_{i-1}^2}\, I(|X_{i-1} - r| \le B/n) \\
&\le \frac{1}{n}\sum \left[|\varphi(\epsilon_i)| + L|\boldsymbol{\theta}_1|\sqrt{1 + (|r| + B/n)^2}\right] |\boldsymbol{\theta}_1| \\
&\qquad \times \sqrt{1 + (|r| + B/n)^2}\, I(|X_{i-1} - r| \le B/n).
\end{aligned}$$

Take expectation both sides to obtain

$$E\Big(\sup_{|s-r|\le B/n} |l_{ns}(\boldsymbol{\theta}_1)|\Big) = O(n^{-1}).$$

Thus, for any $\epsilon > 0$, there exists n_2, such that as $n \ge n_2$, $\forall\ |s-r| \le B/n$,

$$P(\inf_{|s-r|\le B/n} l_{ns}(\boldsymbol{\theta}_1) > -\delta_0) > 1-\epsilon.$$

By (3.6), (3.7) and the above, there exists a $n_0 = n_1 \vee n_2$ such that $\forall\ n \ge n_0$,

$$P(\sup_{s\in\mathcal{R}} \sup_{\boldsymbol{\vartheta}_1^*\in K\setminus V_1} l_{ns}(\boldsymbol{\vartheta}_1^*) < \inf_{|s-r|\le B/n} \sup_{\boldsymbol{\vartheta}_1\in V_1} l_{ns}(\boldsymbol{\vartheta}_1)) > 1-\epsilon. \tag{3.9}$$

Let

$$A_\epsilon = \Big\{\sup_{s\in\mathcal{R}} \sup_{\boldsymbol{\vartheta}_1^*\in K\setminus V_1} l_{ns}(\boldsymbol{\vartheta}_1^*) < \inf_{|s-r|\le B/n} \sup_{\boldsymbol{\vartheta}_1\in V_1} l_{ns}(\boldsymbol{\vartheta}_1)\Big\}.$$

Then, on A_ϵ,

$$\boldsymbol{\vartheta}_{1n}(s) \in V_1, \quad \forall\ |s-r| \le B/n,\ \forall\ n \ge n_0.$$

Thus, by the arbitrary of V_1,

$$\sup_{|s-r|\le B/n} |\boldsymbol{\vartheta}_{1n}(s) - \boldsymbol{\theta}_1| = o_P(1).$$

This ends the proof of Theorem 3.1.1. □

3.2 Asymptotic normality of $\hat{\boldsymbol{\theta}}_{1n}$

Before stating the next theorem, recall that $l_{ns}(\boldsymbol{\vartheta}_1) = l_n(\boldsymbol{\vartheta}_1, s)$, and for $x \in \mathcal{R}$,

$$\dot{h}_s(x) = \frac{\partial}{\partial\boldsymbol{\vartheta}_1}(h_s(x,\boldsymbol{\vartheta}_1)) = (I(x\le s), xI(x\le s), I(x>s), xI(x>s))^T.$$

Let

$$u_{is}(\boldsymbol{\vartheta}_1) = -\varphi(X_i - h_s(X_{i-1},\boldsymbol{\vartheta}_1))\dot{h}_s(X_{i-1}),\ 1\le i\le n.$$

Denote $A_s(x) = \dot{h}_s(x)(\dot{h}_s(x))^T$.

Lemma 3.2.1 *Suppose that* $0 < I(f) < \infty$ *and* $EX_0^2 < \infty$, *then*

$$\frac{1}{\sqrt{n}}\sum u_{ir}(\boldsymbol{\theta}_1) \Longrightarrow N(0,\Gamma), \tag{3.10}$$

where

$$\Gamma = E\ \varphi^2(\epsilon_1)A_r(X_0).$$

Proof. Note that with $\mathcal{F}_i = \sigma\{X_j, j \le i\}$,

$$E_{\boldsymbol{\theta}}\left[u_{ir}(\boldsymbol{\theta}_1)|\mathcal{F}_{i-1}\right] = E_{\boldsymbol{\theta}}\left[-\varphi(\epsilon_i)\dot{h}_r(X_{i-1})|\mathcal{F}_{i-1}\right] = \dot{h}_r(X_{i-1})E_{\boldsymbol{\theta}}(-\varphi(\epsilon_i)) = 0, \quad a.\ s.$$

Therefore, for any vector $v \in \mathcal{R}^4$, by the finite Fisher information of f, $v^T\sum u_{ir}(\boldsymbol{\theta}_1)$ is a zero mean square integrable martingale. By ergodic theorem,

$$\begin{aligned}
& \frac{1}{n}\sum v^T\left\{E\left[u_{ir}(\boldsymbol{\theta}_1)(u_{ir}(\boldsymbol{\theta}_1))^T\Big|\mathcal{F}_{i-1}\right]\right\}v \\
= \;& v^T I(f)\frac{1}{n}\sum A_r(X_{i-1})v \\
\to \;& v^T\Gamma v,\ a.s.
\end{aligned}$$

Thus, the martingale central limiting theorem of Hall and Hedye (1980) shows that the sum $n^{-1/2}\sum v^T u_{ir}(\boldsymbol{\theta}_1)$ converges weakly to the normal distribution with mean zero and variance $v^T\Gamma v$ for all $v \in \mathcal{R}^4$. Thus, $n^{-1/2}\sum u_{ir}(\boldsymbol{\theta}_1)$ converges weakly to the multivariate normal distribution with mean vector zero and covariance matrix Γ. $\square$

Theorem 3.2.1 *Suppose that conditions (C1)-(C4) hold. Then for any* B, $0 < B < \infty$,

$$\sup_{|s-r|\le B/n} \sqrt{n}(\boldsymbol{\vartheta}_{1n}(s) - \boldsymbol{\theta}_1) \Longrightarrow N(0,\Gamma^{-1}). \tag{3.11}$$

As a consequence, for any B, $0 < B < \infty$,

$$\sup_{|s-r|\le B/n} |\sqrt{n}(\boldsymbol{\vartheta}_{1n}(s) - \boldsymbol{\theta}_1)| = O_P(1). \tag{3.12}$$

Proof. Note that for any $s \in \bar{\mathcal{R}}$,

$$\frac{\partial}{\partial\boldsymbol{\vartheta}_1}(l_{ns}(\boldsymbol{\vartheta}_1)) = \sum u_{is}(\boldsymbol{\vartheta}_1).$$

Consider the Taylor's expansion of $(\partial/\partial\boldsymbol{\vartheta}_1)(l_{ns}(\boldsymbol{\vartheta}_1))$ at $\boldsymbol{\theta}_1$:

$$\frac{\partial}{\partial\boldsymbol{\vartheta}_1}l_{ns}(\boldsymbol{\vartheta}_1)=\sum u_{is}(\boldsymbol{\theta}_1)+J_{ns}(\boldsymbol{\theta}^*_{1ns})(\boldsymbol{\vartheta}_1-\boldsymbol{\theta}_1), \tag{3.13}$$

where $\boldsymbol{\theta}^*_{1ns}=\boldsymbol{\theta}_1+\gamma_1(\boldsymbol{\vartheta}_1-\boldsymbol{\theta}_1)$, γ_1 is a function of $(X_0,..,X_n,\boldsymbol{\theta}_1,s)$, $|\gamma_1|<1$ and

$$J_{ns}(\boldsymbol{t})=\sum\frac{\partial}{\partial\boldsymbol{t}}u_{is}(\boldsymbol{t})=\sum\varphi'(X_i-h_s(X_{i-1},\boldsymbol{t}))A_s(X_{i-1}),\ \boldsymbol{t}\in\mathcal{R}^4.$$

Then, the definition of $\boldsymbol{\vartheta}_{1n}(s)$ and (3.13) yield that

$$\frac{1}{\sqrt{n}}\sum u_{is}(\boldsymbol{\theta}_1)+\frac{J_{ns}(\boldsymbol{\theta}^*_{1ns})}{n}\sqrt{n}(\boldsymbol{\vartheta}_{1n}(s)-\boldsymbol{\theta}_1)=0. \tag{3.14}$$

For any matrix $M=(m_{ij})$, define $\|M\|=\sum|m_{ij}|$. Then for any finite number of matrices $\{M_i\}$ and finite number of real numbers a_i, we have

$$\left\|\sum a_iM_i\right\|\leq\sum|a_i|\|M_i\|. \tag{3.15}$$

By the definition of A_s,

$$\|A_s(x)\|=(1+|x|)^2,\ \text{for all } s\in\bar{\mathcal{R}} \text{ and } x\in\mathcal{R}. \tag{3.16}$$

First, we prove the following: For any B, $0<B<\infty$,

$$\sup_{|s-r|\leq B/n}\left\|\frac{J_{ns}(\boldsymbol{\theta}^*_{1ns})}{n}+\Gamma\right\|=o_P(1). \tag{3.17}$$

For any s, such that $|s-r|\leq B/n$,

$$\begin{aligned}
J_{ns}(\boldsymbol{\theta}^*_{1ns}) &= \sum\varphi'(X_i-h_s(X_{i-1},\boldsymbol{\theta}^*_{1ns}))A_s(X_{i-1})\\
&= \sum\Big[\varphi'(X_i-h_s(X_{i-1},\boldsymbol{\theta}^*_{1ns}))-\varphi'(X_i-h_s(X_{i-1},\boldsymbol{\theta}_1))\Big]A_s(X_{i-1})\\
&\quad+\sum\Big[\varphi'(X_i-h_s(X_{i-1},\boldsymbol{\theta}_1))-\varphi'(\epsilon_i)\Big]A_s(X_{i-1})\\
&\quad+\sum\varphi'(\epsilon_i)A_s(X_{i-1})\\
&\equiv J_{1ns}(\boldsymbol{\theta}^*_{1ns})+J_{2ns}+J_{3ns},\ (say).
\end{aligned} \tag{3.18}$$

For the first term in (3.18), the Lip(1) of φ', (3.15), (3.16) , (2.16) and (2.17) imply that

$$\begin{aligned}
\|J_{1ns}(\boldsymbol{\theta}^*_{1ns})\| &\leq \sum\Big|\varphi'(X_i-h_s(X_{i-1},\boldsymbol{\theta}^*_{1ns}))-\varphi'(X_i-h_s(X_{i-1},\boldsymbol{\theta}_1))\Big|\ \|A_s(X_{i-1})\|\\
&\leq L\sum|h_s(X_{i-1},\boldsymbol{\theta}^*_{1ns})-h_s(X_{i-1},\boldsymbol{\theta}_1)|\ (1+|X_{i-1}|)^2\\
&\leq L\sum|\boldsymbol{\vartheta}_{1n}(s)-\boldsymbol{\theta}_1|(1+X^2_{i-1})^{1/2}\ (1+|X_{i-1}|)^2\\
&\leq L\sup_{|s-r|\leq B/n}|\boldsymbol{\vartheta}_{1n}(s)-\boldsymbol{\theta}_1|\sum(1+|X_{i-1}|)^3.
\end{aligned} \tag{3.19}$$

Thus (3.19), Theorem 3.1.1, the ergodicity of $\{X_i\}$ and the finiteness of the third moment of X_0 yield that

$$\sup_{|s-r|\le B/n}\left\|\frac{J_{1ns}(\boldsymbol{\theta}^*_{1ns})}{n}\right\| \le L \sup_{|s-r|\le B/n} |\boldsymbol{\vartheta}_{1n}(s)-\boldsymbol{\theta}_1|\frac{1}{n}\sum(1+|X_{i-1}|)^3 = o_P(1). \quad (3.20)$$

Recall that $\epsilon_i(\boldsymbol{\vartheta}_1, s) = X_i - h_s(X_{i-1}, \boldsymbol{\vartheta}_1)$. For J_{2ns}, again the Lip(1) of φ', (3.15), (3.16) and (2.19) imply that

$$\begin{aligned}
\sup_{|s-r|\le B/n}\|J_{2ns}\| &\le \sup_{|s-r|\le B/n}\sum\left|\varphi'(\epsilon_i(\boldsymbol{\theta}_1,s))-\varphi'(\epsilon_i)\right|\|A_s(X_{i-1})\| \\
&\le L|\boldsymbol{\theta}_1|\sum(1+X_{i-1}^2)^{1/2}I(|X_{i-1}-r|\le B/n)(1+|X_{i-1}|)^2 \\
&\le L|\boldsymbol{\theta}_1|\sum(1+|r|+B/n)^3 I(|X_{i-1}-r|\le B/n).
\end{aligned}$$

The above inequality, the stationarity of $\{X_i\}$ and Remark 2 imply that

$$E\sup_{|s-r|\le B/n}\left\|\frac{J_{2ns}}{n}\right\| = O(n^{-1}). \quad (3.21)$$

For J_{3ns}, observe that for all $x\in\mathcal{R}$,

$$\|A_s(x)-A_r(x)\| = 2(1+|x|)^2 I(s\wedge r < x \le s\vee r) \le 2(1+|x|)^2\ I(|x-r|\le|s-r|).$$

This implies that

$$\begin{aligned}
& E\sup_{|s-r|\le B/n}\|A_s(X_0)-A_r(X_0)\| \\
\le\ & 2E(1+|X_0|)^2 I(|X_0-r|\le B/n) \\
\le\ & 2(1+|r|+B/n)^2 EI(|X_0-r|\le B/n) = O(n^{-1}). \qquad (3.22)
\end{aligned}$$

Also, by (3.15),

$$\begin{aligned}
& \sup_{|s-r|\le B/n}\left\|\frac{J_{3ns}}{n}+\Gamma\right\| \\
=\ & \sup_{|s-r|\le B/n}\left\|\frac{1}{n}\sum\varphi'(\epsilon_i)\big[A_s(X_{i-1})-A_r(X_{i-1})\big]\right\| + \left\|\frac{1}{n}\sum\varphi'(\epsilon_i)A_r(X_{i-1})+\Gamma\right\| \\
\equiv\ & term_1 + term_2,\ (say).
\end{aligned}$$

Thus (3.22), the stationarity of $\{X_i\}$ and the independence between ϵ_i and X_{i-1} imply that

$$\begin{aligned}
E(term_1) &\le n^{-1}E\Big(n|\varphi'(\epsilon_1)|\sup_{|s-r|\le B/n}\big\|A_s(X_0)-A_r(X_0)\big\|\Big) \\
&= E|\varphi'(\epsilon_1)|O(n^{-1}) = O(n^{-1}),
\end{aligned}$$

and by the ergodic theorem, $term_2 \to 0$, $a.s.$ Thus,

$$\sup_{|s-r|\le B/n} \left\| \frac{J_{3ns}}{n} + \Gamma \right\| = o_P(1). \tag{3.23}$$

Therefore, (3.20)-(3.23) and Markov inequality yield (3.17).

Next, we are going to show that for any B, $0 < B < \infty$,

$$E\left[\sup_{|s-r|\le B/n} \left| \frac{1}{\sqrt{n}} \sum [u_{is}(\boldsymbol{\theta}_1) - u_{ir}(\boldsymbol{\theta}_1)] \right| \right] = O(n^{-1/2}). \tag{3.24}$$

To prove (3.24),

$$\begin{aligned}
&\sum [u_{is}(\boldsymbol{\theta}_1) - u_{ir}(\boldsymbol{\theta}_1)] \\
&= \sum [-\varphi(X_i - h_s(X_{i-1}, \boldsymbol{\theta}_1))\dot{h}_s(X_{i-1}) + \varphi(\epsilon_i)\dot{h}_r(X_{i-1})] \\
&= \sum [-\varphi(\epsilon_i + h_r(X_{i-1}, \boldsymbol{\theta}_1) - h_s(X_{i-1}, \boldsymbol{\theta}_1)) + \varphi(\epsilon_i)]\dot{h}_s(X_{i-1}) \\
&\quad - \sum \varphi(\epsilon_i)[\dot{h}_s(X_{i-1}) - \dot{h}_r(X_{i-1})] \\
&\equiv I_{1ns} - I_{2ns}, \ (say).
\end{aligned}$$

For any s, $|s - r| \le B/n$, the triangle inequality, the Lip(1) of φ, (2.19) and (1.20) imply that

$$\begin{aligned}
|I_{1ns}| &\le L \sum |h_s(X_{i-1}, \boldsymbol{\theta}_1) - h_r(X_{i-1}, \boldsymbol{\theta}_1)| \, |\dot{h}_s(X_{i-1})| \\
&\le L|\boldsymbol{\theta}_1| \sum (1 + X_{i-1}^2) I(|X_{i-1} - r| \le |s - r|) \\
&\le L|\boldsymbol{\theta}_1| \sum (1 + (|r| + B/n)^2) I(|X_{i-1} - r| \le B/n).
\end{aligned}$$

Thus, by the stationarity of $\{X_i\}$ and Remark 2,

$$E\Big(\sup_{|s-r|\le B/n} \left| \frac{I_{1ns}}{\sqrt{n}} \right| \Big) = O(n^{-1/2}). \tag{3.25}$$

Similarly, the independent of ϵ_i and X_{i-1}, the stationarity of $\{X_i\}$, the integrability of φ and (2.19) imply that

$$E\Big(\sup_{|s-r|\le B/n} \left| \frac{I_{2ns}}{\sqrt{n}} \right| \Big) = O(n^{-1/2}). \tag{3.26}$$

Therefore (3.25) and (3.26) yield (3.24).

Note that

$$EI(X_0 \le r) EX_0^2 I(X_0 \le r) > (EX_0 I(X_0 \le r))^2,$$

and

$$EI(X_0 > r)EX_0^2 I(X_0 > r) > (EX_0 I(X_0 > r))^2.$$

Thus $EA_r(X_0)$ is positive definite matrix and so is Γ.

Denote $\Gamma_{ns} = -J_{ns}(\boldsymbol{\theta}^*_{1ns})/n$. By (3.17) and the positive definiteness of Γ, Γ_{ns} is positive definite eventually for every s, $|s-r| \leq B/n$, and

$$\sup_{|s-r|\leq B/n} |det(\Gamma_{ns}) - det(\Gamma)| = o_P(1). \tag{3.27}$$

By the Cramer's rule,

$$(\Gamma_{ns})^{-1} = \frac{adj(\Gamma_{ns})}{det(\Gamma_{ns})}.$$

Then the continuity property of the $adj(\Gamma_{ns})$ in the components of Γ_{ns} and (3.15) yield that

$$\begin{aligned} &\sup_{|s-r|\leq B/n} \|\Gamma_{ns}^{-1} - \Gamma^{-1}\| = \left\|\frac{adj(\Gamma_{ns})}{det(\Gamma_{ns})} - \frac{adj(\Gamma)}{det(\Gamma)}\right\| \\ \leq & \left\|\frac{adj(\Gamma_{ns}) - adj(\Gamma)}{det(\Gamma_{ns})}\right\| + \left\|adj(\Gamma)\right\|\left|\frac{det(\Gamma_{ns}) - det(\Gamma)}{det(\Gamma_{ns})det(\Gamma)}\right| = o_P(1). \end{aligned}$$

That is,

$$\sup_{|s-r|\leq B/n} \|\Gamma_{ns}^{-1} - \Gamma^{-1}\| = o_P(1). \tag{3.28}$$

It follows from (3.14), for any s, $|s-r| \leq B/n$,

$$\begin{aligned} & \sqrt{n}(\boldsymbol{\vartheta}_{1n}(s) - \boldsymbol{\theta}_1) + \Gamma^{-1}n^{-1/2}\sum u_{ir}(\boldsymbol{\theta}_1) \\ = & -\left(\Gamma_{ns}^{-1} - \Gamma^{-1}\right)\left(n^{-1/2}\sum u_{is}(\boldsymbol{\theta}_1) - n^{-1/2}\sum u_{ir}(\boldsymbol{\theta}_1)\right) \\ & -\Gamma^{-1}\left(n^{-1/2}\sum u_{is}(\boldsymbol{\theta}_1) - n^{-1/2}\sum u_{ir}(\boldsymbol{\theta}_1)\right) \\ & -\left(\Gamma_{ns}^{-1} - \Gamma^{-1}\right)n^{-1/2}\sum u_{ir}(\boldsymbol{\theta}_1). \end{aligned} \tag{3.29}$$

Thus, by (3.24), (3.28) and Lemma 3.2.1,

$$\sup_{|s-r|\leq B/n} \left|\sqrt{n}(\boldsymbol{\vartheta}_{1n}(s) - \boldsymbol{\theta}_1) + \Gamma^{-1}n^{-1/2}\sum u_{ir}(\boldsymbol{\theta}_1)\right| = o_P(1).$$

Again by Lemma 3.2.1, Theorem 3.2.1 is proved. □

As a corollary of Theorem 3.1.1 and 3.2.1, we have the following uniform convergence rate of $\boldsymbol{\vartheta}_{1n}(\cdot)$.

Theorem 3.2.2 *Suppose that (C1)-(C4) hold, then for any B, $0 < B < \infty$,*

$$\sup_{|s-r|\le B/n} |\boldsymbol{\vartheta}_{1n}(s) - \boldsymbol{\vartheta}_{1n}(r)| = o_P(n^{-1/2}).$$

Proof. Consider the Taylor's expansions of $(d/d\boldsymbol{\vartheta}_1)(l_{ns}(\boldsymbol{\vartheta}_1)$ and $(d/d\boldsymbol{\vartheta}_1)(l_{nr}(\boldsymbol{\vartheta}_1)$ at $\boldsymbol{\theta}_1$ and evaluate them at $\boldsymbol{\vartheta}_{1n}(s)$ and $\boldsymbol{\vartheta}_{1n}(r)$, respectively. We have

$$\begin{aligned} 0 &= \sum u_{is}(\boldsymbol{\theta}_1) + J_{ns}(\boldsymbol{\theta}^*_{1ns})(\boldsymbol{\vartheta}_{1n}(s) - \boldsymbol{\theta}_1) \\ &= \sum u_{ir}(\boldsymbol{\theta}_1) + J_{nr}(\boldsymbol{\theta}^*_{1nr})(\boldsymbol{\vartheta}_{1n}(r) - \boldsymbol{\theta}_1). \end{aligned}$$

Hence,

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum [u_{is}(\boldsymbol{\theta}_1) - u_{ir}(\boldsymbol{\theta}_1)] = \Gamma\sqrt{n}[\boldsymbol{\vartheta}_{1n}(s) - \boldsymbol{\vartheta}_{1n}(r)] \\ &+\left[\Gamma_{ns} - \Gamma\right]\sqrt{n}(\boldsymbol{\vartheta}_{1n}(s) - \boldsymbol{\theta}_1) - \left[\Gamma_{nr} - \Gamma\right]\sqrt{n}(\boldsymbol{\vartheta}_{1n}(r) - \boldsymbol{\theta}_1). \end{aligned}$$

Therefore, Theorem 3.2.1, (3.17) and (3.24) imply that

$$\sup_{|s-r|\le B/n} \left|\sqrt{n}(\boldsymbol{\vartheta}_{1n}(s) - \boldsymbol{\vartheta}_{1n}(r))\right| = o_P(1).$$

This completes the proof. $\square$

Chapter 4

Some asymptotic results on log-likelihood process

In this chapter, we discuss some asymptotic results for a sequence of normalized profile log-likelihood processes. It is expected that these results will be useful in obtaining the limiting distribution of the standardized maximum likelihood estimator of the threshold parameter.

Recall that

$$l_n(\boldsymbol{\vartheta}_1, s) = \frac{1}{n}\sum \ln \frac{f(X_i - h_s(X_{i-1}, \boldsymbol{\vartheta}_1))}{f(\epsilon_i)}, \ (\boldsymbol{\vartheta}_1^T, s)^T \in \Omega.$$

For $z \in \mathcal{R}$, a sequence of normalized profile log-likelihood processes is

$$\hat{l}_n(z) = -2n[l_n(\boldsymbol{\vartheta}_{1n}(r + z/n), r + z/n) - l_n(\boldsymbol{\vartheta}_{1n}(r), r)].$$

Observe that in view of Theorem 3.1.1, $\boldsymbol{\vartheta}_{1n}(r + z/n)$ is an approximation of $\boldsymbol{\theta}_1$ uniformly in z over bounded sets. Thus a natural candidate for the approximation of $\hat{l}_n$ is $\tilde{l}_n$ defined as follows: For $z \in \mathcal{R}$,

$$\tilde{l}_n(z) = -2n[l_n(\boldsymbol{\theta}_1, r + z/n) - l_n(\boldsymbol{\theta}_1, r)] = -2\sum \ln \frac{f(X_i - h_{r+z/n}(X_{i-1}, \boldsymbol{\theta}_1))}{f(\epsilon_i)}. \quad (4.1)$$

4.1 An approximation $\tilde{l}_n$ of the normalized profile log-likelihood process $\hat{l}_n$

Theorem 4.1.1 *Suppose that (C1)-(C4) hold. Then for any B, $0 < B < \infty$,*

$$\sup_{|z|\le B}\left|\hat{l}_n(z) - \tilde{l}_n(z)\right| = o_P(1).$$

Proof. Without loss of generality, assume $r = 0$. Decompose the concerned process in the following way,

$$\begin{aligned}
-\frac{1}{2}\left[\hat{l}_n(z) - \tilde{l}_n(z)\right] &= -\frac{1}{2}\left[\hat{l}_n(z) - \sum \ln\frac{f(X_i - h_{z/n}(X_{i-1},\boldsymbol{\theta}_1))}{f(\epsilon_i)}\right] \\
&= \sum\left[\ln\frac{f(X_i - h_{z/n}(X_{i-1},\boldsymbol{\vartheta}_{1n}(z/n)))}{f(X_i - h_0(X_{i-1},\boldsymbol{\vartheta}_{1n}(z/n)))} - \ln\frac{f(X_i - h_{z/n}(X_{i-1},\boldsymbol{\theta}_1))}{f(X_i - h_0(X_{i-1},\boldsymbol{\theta}_1))}\right] \\
&\quad + \sum \ln\frac{f(X_i - h_0(X_{i-1},\boldsymbol{\vartheta}_{1n}(z/n)))}{f(X_i - h_0(X_{i-1},\boldsymbol{\vartheta}_{1n}(0)))} \\
&\equiv \hat{l}_{1n}(z) + \hat{l}_{2n}(z),\ (say).
\end{aligned}$$

It suffices to show that $\forall\, B < \infty$,

$$\sup_{|z|\le B}|\hat{l}_{1n}(z)| = o_P(1), \tag{4.2}$$

and

$$\sup_{|z|\le B}|\hat{l}_{2n}(z)| = o_P(1). \tag{4.3}$$

Actually, we shall prove a slightly stronger result than (4.2). To state this stronger result, recall that for $1 \le i \le n$,

$$\epsilon_i(\boldsymbol{\vartheta}_1, s) = \epsilon_i(\boldsymbol{\vartheta}) = X_i - h_s(X_{i-1},\boldsymbol{\vartheta}_1),\ (\boldsymbol{\vartheta}_1^T, s)^T \in \Omega.$$

Denote

$$p_{nzi}(\boldsymbol{t}) = \ln\frac{f(\epsilon_i(\boldsymbol{t}, z/n))}{f(\epsilon_i(\boldsymbol{t}, 0))} : \mathcal{R}^4 \to \mathcal{R},\ 1 \le i \le n.$$

Then, for any i, $1 \le i \le n$,

$$\begin{aligned}
\dot{p}_{nzi}(\boldsymbol{t}) &= -\varphi(\epsilon_i(\boldsymbol{t}, z/n))\dot{h}_{z/n}(X_{i-1}) + \varphi(\epsilon_i(\boldsymbol{t},0))\dot{h}_0(X_{i-1}) \\
&= \Big[-\varphi(\epsilon_i(\boldsymbol{t}, z/n)) + \varphi(\epsilon_i(\boldsymbol{t},0))\Big]\dot{h}_{z/n}(X_{i-1}) \\
&\quad + \Big[-\varphi(\epsilon_i(\boldsymbol{t},0)) + \varphi(\epsilon_i)\Big]\Big[\dot{h}_{z/n}(X_{i-1}) - \dot{h}_0(X_{i-1})\Big] \\
&\quad - \varphi(\epsilon_i)\Big[\dot{h}_{z/n}(X_{i-1}) - \dot{h}_0(X_{i-1})\Big].
\end{aligned} \tag{4.4}$$

Note that

$$\hat{l}_{1n}(z) = \sum [p_{nzi}(\boldsymbol{\vartheta}_{1n}(z/n)) - p_{nzi}(\boldsymbol{\theta}_1)].$$

From (3.12), $\sup_{|z|\le B} |\sqrt{n}(\boldsymbol{\vartheta}_{1n}(z/n) - \boldsymbol{\theta}_1)| = O_P(1)$. The stronger result that will be proved is that for any $0 < C < \infty$,

$$E\Big\{ \sup_{|z|\le B, \sqrt{n}|\boldsymbol{t}-\boldsymbol{\theta}_1|\le C} \Big|\sum [p_{nzi}(\boldsymbol{t}) - p_{nzi}(\boldsymbol{\theta}_1)]\Big| \Big\} = O(n^{-1/2}). \tag{4.5}$$

Denote $\boldsymbol{\theta}_{1s}(\boldsymbol{t}) = \boldsymbol{\theta}_1 + s(\boldsymbol{t} - \boldsymbol{\theta}_1)$ and recall that $\varphi = f'/f$. By the absolute continuity of $\ln f$,

$$\begin{aligned} p_{nzi}(\boldsymbol{t}) - p_{nzi}(\boldsymbol{\theta}_1) &= \int_0^1 \dot{p}_{nzi}^T(\boldsymbol{\theta}_{1s}(\boldsymbol{t}))\,(\boldsymbol{t} - \boldsymbol{\theta}_1)\,ds \\ &\equiv s_{1iz}(\boldsymbol{t}) + s_{2iz}(\boldsymbol{t}) + s_{3iz}(\boldsymbol{t}),\ (say), \end{aligned}$$

where, by (4.4),

$$s_{1iz}(\boldsymbol{t}) = \int_0^1 [-\varphi(\epsilon_i(\boldsymbol{\theta}_{1s}(\boldsymbol{t}), z/n)) + \varphi(\epsilon_i(\boldsymbol{\theta}_{1s}(\boldsymbol{t}), 0))]\, \dot{h}_{z/n}^T(X_{i-1})(\boldsymbol{t} - \boldsymbol{\theta}_1)\,ds,$$

$$s_{2iz}(\boldsymbol{t}) = \int_0^1 [-\varphi(\epsilon_i(\boldsymbol{\theta}_{1s}(\boldsymbol{t}), 0)) + \varphi(\epsilon_i)]\left[\dot{h}_{z/n}(X_{i-1}) - \dot{h}_0(X_{i-1})\right]^T (\boldsymbol{t} - \boldsymbol{\theta}_1)\,ds,$$

$$s_{3iz}(\boldsymbol{t}) = -\varphi(\epsilon_i)\left[\dot{h}_{z/n}(X_{i-1}) - \dot{h}_0(X_{i-1})\right]^T\ (\boldsymbol{t} - \boldsymbol{\theta}_1).$$

The Lip(1) of φ and (2.19) imply that uniformly in all $\boldsymbol{t}$, i and s,

$$\begin{aligned} &|\varphi(\epsilon_i(\boldsymbol{\theta}_{1s}(\boldsymbol{t}), z/n)) - \varphi(\epsilon_i(\boldsymbol{\theta}_{1s}(\boldsymbol{t}), 0))| \\ &\le L|h_{z/n}(X_{i-1}, \boldsymbol{\theta}_{1s}(\boldsymbol{t})) - h_0(X_{i-1}, \boldsymbol{\theta}_{1s}(\boldsymbol{t}))| \\ &\le L|\boldsymbol{\theta}_{1s}(\boldsymbol{t})|\sqrt{1 + X_{i-1}^2}\ I(|X_{i-1}| \le |z|/n) \\ &\le L(|\boldsymbol{\theta}_1| + |\boldsymbol{t} - \boldsymbol{\theta}_1|)\sqrt{1 + X_{i-1}^2}\ I(|X_{i-1}| \le |z|/n). \end{aligned}$$

Therefore, by (2.18) and (2.9),

$$\begin{aligned} &\sup_{|z|\le B, |\sqrt{n}(\boldsymbol{t}-\boldsymbol{\theta}_1)|\le C} |s_{1iz}(\boldsymbol{t})| \\ \le\ & CLn^{-1/2}(|\boldsymbol{\theta}_1| + Cn^{-1/2})\sqrt{(1 + X_{i-1}^2)}\ \sup_{|z|\le B} I(|X_{i-1}| \le |z|/n)\ |\dot{h}_{z/n}(X_{i-1})| \\ \le\ & CLn^{-1/2}\left[|\boldsymbol{\theta}_1| + Cn^{-1/2}\right](1 + (B/n)^2) I(|X_{i-1}| \le B/n). \end{aligned}$$

Thus, the stationarity of $\{X_i\}$ and Remark 2 yield that

$$E\left\{\sup_{|z|\le B,|\sqrt{n}(\boldsymbol{t}-\boldsymbol{\theta}_1)|\le C}\left|\sum s_{1iz}(\boldsymbol{t})\right|\right\}=O(n^{-1/2}). \tag{4.6}$$

For $s_{2iz}(\boldsymbol{t})$, the Lip(1) of φ, (2.17) and (2.9) imply that for all $\boldsymbol{t}$ and i,

$$\begin{aligned}
&\sup_{|z|\le B,|\sqrt{n}(\boldsymbol{t}-\boldsymbol{\theta}_1)|\le C}|s_{2iz}(\boldsymbol{t})|\\
\le&\sup_{|z|\le B,|\sqrt{n}(\boldsymbol{t}-\boldsymbol{\theta}_1)|\le C}\int_0^1|\varphi(\epsilon_i(\boldsymbol{\theta}_{1s}(\boldsymbol{t}),0))-\varphi(\epsilon_i)|ds\;|\dot{h}_{z/n}(X_{i-1})-\dot{h}_0(X_{i-1})||\boldsymbol{t}-\boldsymbol{\theta}_1|\\
\le& L\sup_{|\sqrt{n}(\boldsymbol{t}-\boldsymbol{\theta}_1)|\le C}|\boldsymbol{t}-\boldsymbol{\theta}_1|^2(1+(B/n)^2)I(|X_{i-1}|\le B/n)\\
\le& LC^2n^{-1}(1+(B/n)^2)I(|X_{i-1}|\le B/n).
\end{aligned}$$

This together with the stationarity of $\{X_i\}$ and the boundedness of $g_{\boldsymbol{\theta}}$ on compact set implies that

$$E\left\{\sup_{|z|\le B,|\sqrt{n}(\boldsymbol{t}-\boldsymbol{\theta}_1)|\le C}\left|\sum s_{2iz}(\boldsymbol{t})\right|\right\}=O(n^{-1}). \tag{4.7}$$

To $s_{3iz}(\boldsymbol{t})$, by (2.10),

$$\begin{aligned}
&\sup_{|z|\le B,|\sqrt{n}(\boldsymbol{t}-\boldsymbol{\theta}_1)|\le C}|s_{3iz}(\boldsymbol{t})|\\
\le& |\varphi(\epsilon_i)|\sup_{|z|\le B,|\sqrt{n}(\boldsymbol{t}-\boldsymbol{\theta}_1)|\le C}\left[|\dot{h}_{z/n}(X_{i-1})-\dot{h}_0(X_{i-1})||\boldsymbol{t}-\boldsymbol{\theta}_1|\right]\\
\le& C\,n^{-1/2}|\varphi(\epsilon_i)|\sqrt{1+(B/n)^2}I(|X_{i-1}|\le B/n).
\end{aligned}$$

The above, the integrability of φ and Remark 2 yield that

$$E\left\{\sup_{|z|\le B,|\sqrt{n}(\boldsymbol{t}-\boldsymbol{\theta}_1)|\le C}\left|\sum s_{3iz}(\boldsymbol{t})\right|\right\}=O(n^{-1/2}). \tag{4.8}$$

Therefore, (4.6)-(4.8) imply (4.5) and hence (4.2).

Now it remains to prove (4.3). To prove (4.3), the result of Theorem 3.2.2 will be needed. Recall that

$$\hat{l}_{2n}(z)=\sum\ln\frac{f(X_i-h_0(X_{i-1},\boldsymbol{\vartheta}_{1n}(z/n)))}{f(X_i-h_0(X_{i-1},\boldsymbol{\vartheta}_{1n}(0)))}.$$

For any $t\in\mathcal{R}$, denote

$$\boldsymbol{\vartheta}_{1nt}(z)=\boldsymbol{\vartheta}_{1n}(0)+t(\boldsymbol{\vartheta}_{1n}(z/n)-\boldsymbol{\vartheta}_{1n}(0)).$$

Then the absolutely continuity of $\ln f$ gives

$$\begin{aligned}\hat{l}_{2n}(z) &= \sum \int_0^1 \left[-\varphi(X_i - h_0(X_{i-1}, \boldsymbol{\vartheta}_{1nt}(z)))\dot{h}_0^T(X_{i-1})\right](\boldsymbol{\vartheta}_{1n}(z/n) - \boldsymbol{\vartheta}_{1n}(0))\, dt \\ &= \sum \int_0^1 [-\varphi(X_i - h_0(X_{i-1}, \boldsymbol{\vartheta}_{1nt}(z))) + \varphi(\epsilon_i)]\, dt\, \dot{h}_0^T(X_{i-1})(\boldsymbol{\vartheta}_{1n}(z/n) - \boldsymbol{\vartheta}_{1n}(0)) \\ &\quad - \sum \varphi(\epsilon_i)\dot{h}_0^T(X_{i-1})(\boldsymbol{\vartheta}_{1n}(z/n) - \boldsymbol{\vartheta}_{1n}(0))\end{aligned}$$

Thus, the Lip(1) of φ, Theorem 3.2.2, (2.17) and (2.9) imply

$$\begin{aligned}&\sup_{|z|\le B}\{|\text{The first term of } \hat{l}_{2n}(z)|\} \\ \le\ &\sup_{|z|\le B} L|\boldsymbol{\vartheta}_{1n}(z/n) - \boldsymbol{\vartheta}_{1n}(0)||\dot{h}_0(X_{i-1})|\sum \int_0^1 |h_0(X_{i-1}, \boldsymbol{\vartheta}_{1nt}(z)) - h_0(X_{i-1}, \boldsymbol{\theta}_1)|\, dt \\ \le\ &o_P\left(n^{-1/2}\right)\int_0^1 \sup_{|z|\le B}|\boldsymbol{\vartheta}_{1nt}(z) - \boldsymbol{\theta}_1|\, dt \sum(1 + X_{i-1}^2) \\ =\ &o_P(n^{-1/2})O_P(n^{-1/2})O_P(n) = o_P(1) \qquad (4.9)\end{aligned}$$

For the second term of $\hat{l}_{2n}(z)$, since $\sum \varphi(\epsilon_i)\dot{h}_0(X_{i-1})$ is a zero mean square integrable martingale process,

$$\left|\sum \varphi(\epsilon_i)\dot{h}_0(X_{i-1})\right| = O_P(n^{1/2}). \qquad (4.10)$$

But Theorem 3.2.2 implies that

$$\sup_{|z|\le B}|\boldsymbol{\vartheta}_{1n}(z/n) - \boldsymbol{\vartheta}_{1n}(0)| = o_P(n^{-1/2}).$$

Therefore (4.3) follows from the above two equations and (4.9). This ends the proof of Theorem 4.1.1. □

Thus, the limiting process of the sequence of processes $\{\hat{l}_n(z), z \in \mathcal{R}\}$ is the same as that of the process $\tilde{l}_n(z)$ by Theorem 4.1.1.

4.2 Tightness of $\tilde{l}_n$

We need a preliminary result from our model (2.1). Note that $I(f) < \infty$ implies the boundedness of f (Koul, 1992, p52). From the invariant equation of $g_{\boldsymbol{\theta}}$, $g_{\boldsymbol{\theta}}(y) = \int f(y - h(x, \boldsymbol{\theta}))g_{\boldsymbol{\theta}}(x)dx$, so is $g_{\boldsymbol{\theta}}$.

Proposition 4.2.1 *Suppose that f is continuous, positive everywhere and bounded on $\mathcal{R}$, then for any $0 < c < \infty$, there exists $0 < C < \infty$, such that for any interval $I \subset [-c, c]$ with length $l(I)$, $\forall\ k \geq 1$,*

$$P(X_0 \in I, X_k \in I) \leq C(l(I))^2. \tag{4.11}$$

Proof. We shall use the mathematical induction and the boundedness of f. Let x be any point in $\mathcal{R}$. For $k = 1$,

$$\begin{aligned} P(x, I) &= P(X_1 \in I | X_0 = x) = P(h(X_0, \boldsymbol{\theta}) + \epsilon_1 \in I | X_0 = x) \\ &= \int_I f(y - h(x, \boldsymbol{\theta}))dy \leq Cl(I). \end{aligned}$$

If the assertion holds for $k = m$, then for $k = m + 1$,

$$P^{m+1}(x, I) = \int P(x, dy) P^m(y, I) \leq C\ l(I) \int f(y - h(x, \boldsymbol{\theta}))dy \leq Cl(I).$$

Thus,

$$P(X_0 \in I, X_k \in I) = \int_I P^k(x, I) g_{\boldsymbol{\theta}}(x)dx \leq C(l(I))^2.$$

□

Now for $z \geq 0$, recall (4.1), let $a = a_2 - a_1,\ \beta = \beta_2 - \beta_1$,

$$\tilde{l}_n(z) = -2 \sum \ln \frac{f(\epsilon_i + a + \beta X_{i-1})}{f(\epsilon_i)}\ I(r < X_{i-1} \leq r + z/n).$$

Note that $\tilde{l}_n \in D[0, \infty)$, the following discussion will involve weak convergence on $D[0, \infty)$. To that effect, we need to introduce the modulus of a function in the space $D[0, \infty)$. For any $\phi \in D([0, \infty))$, define $r_{ab}(\phi)(t) = \phi(t),\ a \leq t \leq b,\ a < b$.

Theorem 4.2.1 *(Whitt, 1980) Let P_n, $n \geq 1$, and P be probability measures on $D([0, \infty))$. Then $P_n \Longrightarrow P$ if and only if $P_n r^{-1}_{s_k t_k} \Longrightarrow P r^{-1}_{s_k t_k}$ on $D([s_k, t_k])$ for all k and some sequence $\{[s_k, t_k], k \geq 1\}$ with $\cup_{k=1}^{\infty} [s_k, t_k] = [0, \infty)$.*

Corollary 4.2.1 *P_n is relatively compact if and only if $P_n r^{-1}_{s_k t_k}$ is relatively compact for all k and some sequence $[s_k, t_k]$ such that $\cup_{k=1}^{\infty} [s_k, t_k] = [0, \infty)$.*

Thus, take $[s_k, t_k] = [k, k+1]$, it is enough to work on $D[k, k+1]$, for every $k \geq 0$. That is, for every $\phi \in D[0,\infty)$, define

$$\begin{aligned} \beta_\delta^k(\phi) \;=\; & \sup_{k\le u-\delta\le u'\le u\le u''\le u+\delta\le k+1} [\min\{|\phi(u') - \phi(u)|, |\phi(u'') - \phi(u)|\}] \\ & + \sup_{k\le u\le k+\delta} |\phi(u) - \phi(k)| + \sup_{k+1-\delta\le u\le k+1} |\phi(u) - \phi(k+1)|, \ k \geq 0. \end{aligned}$$

Theorem 4.2.2 *Suppose that f is continuous, positive everywhere and bounded on $\mathcal{R}$, then $(\{\tilde{l}_n(-z), z \geq 0\}, \{\tilde{l}_n(z), z \geq 0\})$ is tight. That is, $\forall\, k \geq 0$, $\forall\, \epsilon > 0$,*

$$\lim_{\delta\to 0} \sup_n P(\beta_\delta^k(\tilde{l}_n) > \epsilon) = 0. \tag{4.12}$$

Proof. We shall only show the tightness of $\{\tilde{l}_n(z), z \geq 0\}$, since the proof is the same for $\{\tilde{l}_n(-z), z \geq 0\}$.

The following argument is similar to the proof of Lemma 3.2 in Ibragimov and Has'minski (1981, pp. 261). Let $A_i = A_i(u, u+\delta]$ be the event that a trajectory of $\tilde{l}_n$ possesses at least i discontinuities on the interval $(u, u+\delta]$. We shall prove the following inequalities:

$$P(A_1) \leq C\delta, \quad P(A_2) \leq C\delta^2. \tag{4.13}$$

A trajectory of $\tilde{l}_n$ has at least one discontinuity on $(u, u+\delta]$ only if at least one $X_{i-1} \in (r + u/n, r + (u+\delta)/n]$. Denote $C_i = \{X_{i-1} \in (r + u/n, r + (u+\delta)/n]\}$, then by the boundedness of f and Remark 2,

$$P(A_1) \leq \sum P(C_i) \leq C\delta. \tag{4.14}$$

A trajectory of $\tilde{l}_n$ has at least two discontinuities on $(u, u+\delta]$ only if at least one pair of $(X_{i-1}, X_{j-1}) \in (r + u/n, r + (u+\delta)/n]^2$, $i \neq j$. Hence, by Proposition 4.2.1 and the stationarity of $\{X_i\}$,

$$P(A_2) \leq \sum_{i\neq j} P(X_{i-1} \in (r+u/n, r+(u+\delta)/n], X_{j-1} \in (r+u/n, r+(u+\delta)/n]) \leq C\delta^2.$$

This and (4.14) prove (4.13).

Now let B be the event that on the interval $[k, k+1]$, there exists at least two points of discontinuities of $\tilde{l}_n$ such that the distance between them is less than 2δ.

Let us divide the interval $[k, k+1]$ into $m = [\delta^{-1}]$ subintervals δ_i of length m^{-1}. Each interval with length less than 2δ is totally contained in $\delta_i \cup (\delta_{i+1} \cup \delta_{i+2})$. Therefore,

$$B \subset \bigcup_{i=1}^{m} A_2(\delta_i) \cup \bigcup_{i=1}^{m-2} A_2(\delta_{i+1} \cup \delta_{i+2}).$$

Hence,

$$P(B) \leq \sum_{i=1}^{m} P(A_2(\delta_i)) + \sum_{i=1}^{m-2} P(A_2(\delta_{i+1} \cup \delta_{i+2})) \leq Cm\delta^2 \leq C\delta. \tag{4.15}$$

Furthermore, as long as the event B does not occur (i.e., the complement of B, say B^c, occurs), any interval of the form $[u-\delta, u+\delta]$ possesses at most one point of discontinuity of $\tilde{l}_n$. So that this function is continuous on either $[u, u+\delta]$ or $[u-\delta, u]$. For example, suppose that $\tilde{l}_n$ is continuous on $[u, u+\delta]$. Then $\tilde{l}_n$ has no jump on $[u, u+\delta]$. Note that $\tilde{l}_n$ is a step function, so $\tilde{l}_n$ is a constant on $[u, u+\delta]$, i.e.

$$\sup_{u \leq u'' \leq u+\delta} |\tilde{l}_n(u) - \tilde{l}_n(u'')| = 0.$$

Finally, on B^c, there is at most one discontinuity point of $\tilde{l}_n$ and $\forall\, \epsilon > 0$,

$$\Big\{ \sup_{k \leq u \leq k+\delta} |\tilde{l}_n(u) - \tilde{l}_n(k)| > \epsilon/2 \Big\} \cap B^c \subset A_1(r + k/n, r + (k+\delta)/n],$$

thus, by (4.13),

$$\begin{aligned} &P(\{ \sup_{k \leq u \leq k+\delta} |\tilde{l}_n(u) - \tilde{l}_n(k)| > \epsilon/2\} \cap B^c) \\ &\leq P(A_1(r + k/n, r + (k+\delta)/n]) \leq C\delta, \end{aligned} \tag{4.16}$$

and

$$\begin{aligned} &P(\{ \sup_{k+1-\delta \leq u \leq k+1} |\tilde{l}_n(u) - \tilde{l}_n(k+1)| > \epsilon/2\} \cap B^c) \\ &\leq P(A_1(r + (k+1-\delta)/n, r + (k+1)/n]) \leq C\delta. \end{aligned} \tag{4.17}$$

Therefore, by (4.15)-(4.17), for every $\epsilon > 0$,

$$\begin{aligned} P(\beta_\delta^k(\tilde{l}_n) > \epsilon) &\leq P(B) + P(B^c \cap \{\beta_\delta^k(\tilde{l}_n) > \epsilon\}) \\ &\leq C\delta + P(\{ \sup_{k \leq u \leq k+\delta} |\tilde{l}_n(u) - \tilde{l}_n(k)| > \epsilon/2\} \cap B^c) \\ &\qquad + P(\{ \sup_{k+1-\delta \leq u \leq k+1} |\tilde{l}_n(u) - \tilde{l}_n(k+1)| > \epsilon/2\} \cap B^c) \\ &\leq 3C\delta. \end{aligned} \tag{4.18}$$

It follows from (4.18) that, $\forall\, k \geq 0$, $\forall\, \epsilon > 0$,

$$\lim_{\delta \to 0} \sup_n P(\beta_\delta^k(\tilde{l}_n) > \epsilon) = 0.$$

Therefore, $\{\tilde{l}_n(z), z \geq 0\}$ is tight. □

4.3 Some problems for future research

Note that for each n, the process $\tilde{l}_n$ is a jump process with finite number of possible jumps at $nX_{i-1}, i = 1, ..., n$. It is thus reasonable to expect that the limiting process of $\tilde{l}_n$ will asymptotically behave like a compound Poisson process with rate $g_{\boldsymbol{\theta}}(r)$ whose left and right jump distributions are given by the conditional distribution of $\zeta_1 = -2\ln[f(\epsilon_1 - a - \beta X_0)/f(\epsilon_1)]$ given $X_0 = r^-$ and the conditional distribution of $\zeta_2 = -2\ln[f(\epsilon_1 + a + \beta X_0)/f(\epsilon_1)]$ given $X_0 = r^+$, respectively. The former conditional distribution is the limiting conditional distribution of ζ_1 given $r - \delta_1 < X_0 \leq r - \delta_2$ and the latter is that of ζ_2 given $r + \delta_1 < X_0 \leq r + \delta_2$ as $\delta_1 \downarrow 0$, $\delta_2 \downarrow 0$ and $\delta_1 \geq 0$, $\delta_2 > 0$.

I am presently working on the above problem.

After obtaining the limiting distribution of $\tilde{l}_n$, it will be easy to obtain some inference on the limiting distribution, which will be related to the compound Poisson process, of the standardized maximum likelihood estimator of threshold parameter r.

Appendix A

Lemma A.0.1 *Suppose that* $p_1(y) = \int |\ln[f(x+y)/f(x)]| dF(x) < \infty, \quad \forall\, y \in \mathbf{R}$. *Then (i) and (ii) below are equivalent.*

(i). $p_1(y)$ *is continuous at* $a + \beta r$,

(ii). $\int |\ln[f(x+a+\beta y)/f(x)] - \ln[f(x+a+\beta r)/f(x)]| dF(x)$ *is continuous at* r.

Proof. Suppose that (i) holds and $y_n \to r$. Let

$$g_n(x) = \ln \frac{f(x+a+\beta y_n)}{f(x)}, \qquad g(x) = \ln \frac{f(x+a+\beta r)}{f(x)}.$$

Then, the continuity of f implies that,

$$g_n \to g, \quad \text{point-wise}$$

and which implies

$$g_n^+ \to g^+, \quad g_n^- \to g^-. \tag{A.1}$$

Thus, by (i),

$$\int |g_n|\, dF \to \int |g|\, dF. \tag{A.2}$$

Combining (A.1) and (A.2) implies that

$$\int g_n^\pm\, dF \to \int g^\pm\, dF. \tag{A.3}$$

Since $0 \le (g^\pm - g_n^\pm)^\pm \le g^\pm$, $(g^\pm - g_n^\pm)^\pm \to 0$ and $\int g^\pm dF \le |g| dF < \infty$, by dominate convergence theorem,

$$\int (g^\pm - g_n^\pm)^+\, dF \to 0. \tag{A.4}$$

The result (A.4) and $\int((g^{\pm} - g_n^{\pm})\, dF \to 0$ imply that

$$\int (g^{\pm} - g_n^{\pm})^{-}\, dF \to 0.$$

Thus,

$$\int |g^{\pm} - g_n^{\pm}|\, dF \to 0 \tag{A.5}$$

which implies

$$\int |g - g_n|\, dF \to 0.$$

The fact that (ii) implies (i) is obvious. □

Lemma A.0.2 *Let $p(y) = \int \ln[f(x+y)/f(x)]dF(x)$, then the continuity of p_2 implies the continuity of p_1 and p.*

Proof. Let $y_n \to r$, denote

$$h_n(x) = \ln \frac{f(x + a + \beta y_n)}{f(x)}, \quad h(x) = \ln \frac{f(x + a + \beta r)}{f(x)}.$$

Then

$$h_n \to h, \qquad \int |h_n|^2\, dF \to \int h^2\, dF. \tag{A.6}$$

A convergence theorem in Hajek-Sidak (1967, pp. 154) and (A.6) imply that

$$\int |h_n - h|^2\, dF \to 0. \tag{A.7}$$

By Lemma A.1, (A.7) is equivalent to the continuity of p_1. The continuity of p follows from (ii) of Lemma A.0.1. □

Lemma A.0.3 *The conditions (C1) and (C2) imply the continuity of*

$$p_2(y) = E\left\{\ln[f(\epsilon_1 + y)/f(\epsilon_1)]\right\}^2 \quad \text{on } \mathcal{R}.$$

Proof. For any x and y in $\mathcal{R}$,

$$\begin{aligned}
&\left|E\left\{\left(\ln \frac{f(\epsilon_1 + x)}{f(\epsilon_1)}\right)^2 - \left(\ln \frac{f(\epsilon_1 + y)}{f(\epsilon_1)}\right)^2\right\}\right| \\
&\le E\left|\ln \frac{f(\epsilon_1 + x)}{f(\epsilon_1 + y)}\right| \left|\ln \frac{f(\epsilon_1 + x)}{f(\epsilon_1)} + \ln \frac{f(\epsilon_1 + y)}{f(\epsilon_1)}\right| \\
&\le 2\, |x - y|\, (E|\varphi(\epsilon_1 + \eta(x - y))|^2)^{1/2} \cdot \\
&\qquad \left\{\left(E\left|\ln \frac{f(\epsilon_1 + x)}{f(\epsilon_1)}\right|^2\right)^{1/2} + \left(E\left|\ln \frac{f(\epsilon_1 + y)}{f(\epsilon_1)}\right|^2\right)^{1/2}\right\},
\end{aligned}$$

where $|\eta| < 1$, which is between x and y. The conditions (C1) and (C2) imply the finiteness of $E|\varphi(\epsilon_1 + \eta(x-y))|^2$ and $E\left|\ln \frac{f(\epsilon_1+y)}{f(\epsilon_1)}\right|^2$ for any x and y. Therefore, $E\left\{\ln[f(\epsilon_1 + y)/f(\epsilon_1)]\right\}^2$ is continuous in $\mathcal{R}$. $\square$

Bibliography

[1] Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle, *2nd Int. Symp. on Inf. Th.* (eds. B. N. Petrov and F. Csaki), 267-281, Budapest: Akademia Kiado.

[2] Billingsley, P. (1961), The Lindeberg-Lèvy theorem for martingales, *Proc. Amer. Math. Soc.,* **12**, 788-792.

[3] Chan, K.S. and Tong, H. (1985), On the use of the deterministic lyapunov function for the ergodicity of stochastic difference equations, *Adv. Appl. Prob,* **17**, 666-678.

[4] Chan, K.S., Petruccelli, J.D., Tong, H. and Woolford, S.W. (1985), A multiple-threshold AR(1) model, *J. Appl. Prob.* **22**, 267-279.

[5] Chan, K.S. (1989), Note on the geometric ergodicity of a Markov chain, *Adv. Appl. Prob.* **21**, 702-704.

[6] Chan, K.S. (1993), Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model, *Ann. Statist.* **21**, 520-533.

[7] Chernoff, H. and Rubin, H. (1956), The estimation of the location of a discontinuity in density, *proc. Third Berkeley Symp. Math. Statist. Probab.* **1**, 19-38 Univ. California Press, Berkeley.

[8] Dhar, S.K. (1990), Functional Least Squares Estimator in an Additive Effects Outliers Model, *Australian Mathematical Society*, Ser. A, **48**, 299-319.

[9] Dhar, S.K. (1991), Minimum Distance Estimation in an Additive Effects Outliers Model, *The Annals of Statistics*, **19**, 205-228.

[10] Dhar, S.K. (1993), Computation of certain minimum distance estimators in AR[k] model, *J. Amer. Statist. Assoc* , **421**, 278-283.

[11] Hajek, J. and Sidak, Z. (1967), *Theory of Rank Tests,* Academic Press.

[12] Hall, P. and Heyde, C. C. (1980), *Martingale limit theory and its applications,* Academic Press, New York.

[13] Huber, P.J. (1967), The behavior of maximum likelihood estimates under non-standard conditions, *proc. Fifth Berkeley Symp. Math. Statist. Probab.*, **1**, 221-234 Univ. California Press, Berkeley.

[14] Ibragimov, I.A. and Has'minskii, R.Z. (1981), *Statistical Estimation: Asymptotic Theory,* **16**, Springer, Berlin.

[15] Koul, H. L. (1986), Minimum Distance Estimation And Goodness-of-fit Tests in First-order Autoregression, *Ann. Statist.* **14**, 1194-1213.

[16] Koul, H. L. (1992), *Weighted empiricals and linear models,* IMS Lecture Notes-Monograph Ser. **21**.

[17] Koul, H.L. and Schick,A. (1995), Efficient Estimation in Nonlinear Time series Models, Manuscript, MSU Tech. Report **545**.

[18] Moran, P.A.P.(1953), The statistical analysis of the Canadian lynx cycle, *Aust. J. Zool.*, **1**, 163-173, 291-298.

[19] Nicholls, D.F. and Quinn, B.G. (1982), *Random coefficient autoregressive models: An introduction,* Lect. Notes in Statistics. **11**, Springer Verlag, New York.

[20] Petruccelli, J. D. and Woolford, S.W. (1984), A threshold AR(1) model, *J. Appl. Prob.*, **21**, 270-286.

[21] Petruccelli, J. D. (1986), On the consistency of least squares estimators for a threshold AR(1) model, *J. Time Series Anal.*, **4**, 269-278.

[22] Priestley, M. B. (1980), State-dependent models: A general approach to non-linear time, *J. of Time Ser. Anal.*, **1**, 47-72.

[23] Pagan, A.R. (1980), Some identification and estimation results for regression models with stochastically varying parameters, *J. of Econometrics,* **13**, 341-363.

[24] Tong, H. (1977), Discussion of a paper by A. J. Lawrance and N. T. Kotlegoda, *J. R. Statist. Soc.* Ser A **140**, 34-35.

[25] Tong, H. (1978a), On a threshold model, *Pattern recognition and signal processing* (ed. C. H. Chen).

[26] Tong, H. (1978b), Threshold autoregrssion, limit cycles and cyclical data, Tech. Rep. No. **101**, Dept. of Maths. (Stats.), UMIST.

[27] Tong, H. (1980), A view on non-linear time series model building, time series (ed. O.D. Anderson), Amsterdam: North-Holland.

[28] Tong, H. (1983), *Threshold Models in Non-linear Time Series Analysis,* Lecture Notes in Statist., **21**, Springer, Berlin.

[29] Tong, H. (1990), *Non-linear Time Series: A Dynamical System Approach,* Oxford Univ. Press, New York.

[30] Tong, H. and Lim, K. S. (1980), Threshold autoregrssion, limit cycles and cyclical data (with discussion), *J. R. Stat. Soc.*, **B42**, 245-92.

[31] Whitt, W. (1980), Some useful functions for functional limit theorems, *Math. Oper. Res.*, **5**, 67-85.