Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2010, Article ID 956907, 30 pages
doi:10.1155/2010/956907

Research Article

QML Estimators in Linear Regression Models with Functional Coefficient Autoregressive Processes

Hongchang Hu

School of Mathematics and Statistics, Hubei Normal University, Huangshi 435002, China

Correspondence should be addressed to Hongchang Hu, retutome@163.com

Received 30 December 2009; Revised 19 March 2010; Accepted 6 April 2010

Academic Editor: Massimo Scalia

This paper studies a linear regression model, whose errors are functional coefficient autoregressive processes. Firstly, the quasi-maximum likelihood (QML) estimators of some unknown parameters are given. Secondly, under general conditions, the asymptotic properties (existence, consistency, and asymptotic distributions) of the QML estimators are investigated. These results extend those of Maller (2003), White (1959), Brockwell and Davis (1987), and so on. Lastly, the validity and feasibility of the method are illuminated by a simulation example and a real example.

1. Introduction

Consider the following linear regression model:

$$y_t = x_t^T \beta + \varepsilon_t, \quad t = 1, 2, \ldots, n, \tag{1.1}$$

where y_t's are scalar response variables, x_t's are explanatory variables, β is a d-dimensional unknown parameter, and the ε_t's are functional coefficient autoregressive processes given as

$$\varepsilon_1 = \eta_1, \qquad \varepsilon_t = f_t(\theta)\varepsilon_{t-1} + \eta_t, \quad t = 2, 3, \ldots, n, \tag{1.2}$$

where η_t's are independent and identically distributed random errors with zero mean and finite variance σ^2, θ is a one-dimensional unknown parameter and $f_t(\theta)$ is a real valued function defined on a compact set Θ which contains the true value θ_0 as an inner point and is a subset of R^1. The values of θ_0 and σ^2 are unknown.

Model (1.1) includes many special cases, such as an ordinary linear regression model when $f_t(\theta) \equiv 0$; see [1–11]. In the sequel, we always assume that $f_t(\theta) \neq 0$, for some $\theta \in \Theta$, is a linear regression model with constant coefficient autoregressive processes (when $f_t(\theta) = \theta$, see Maller [12], Pere [13], and Fuller [14]), time-dependent and functional coefficient autoregressive processes (when $\beta = 0$, see Kwoun and Yajima [15]), constant coefficient autoregressive processes (when $f_t(\theta) = \theta$ and $\beta = 0$, see White [16, 17], Hamilton [18], Brockwell and Davis [19], and Abadir and Lucas [20]), time-dependent or time-varying autoregressive processes (when $f_t(\theta) = a_t$ and $\beta = 0$, see Carsoule and Franses [21], Azrak and Mélard [22], and Dahlhaus [23]), and so forth.

Regression analysis is one of the most mature and widely applied branches of statistics. Linear regression analysis is one of the most widely used statistical techniques. Its applications occur in almost every field, including engineering, economics, the physical sciences, management, life and biological sciences, and the social sciences. Linear regression model is the most important and popular model in the statistical literature, which attracts many statisticians to estimate the coefficients of the regression model. For the ordinary linear regression model (when the errors are independent and identically distributed random variables), Bai and Guo [1], Chen [2], Anderson and Taylor [3], Drygas [4], González-Rodríguez et al. [5], Hampel et al. [6], He [7], Cui [8], Durbin [9], Hoerl and Kennard [10], Li and Yang [11], and Zhang et al. [24] used various estimation methods (Least squares estimate method, robust estimation, biased estimation, and Bayes estimation) to obtain estimators of the unknown parameters in (1.1) and discussed some large or small sample properties of these estimators.

However, the independence assumption for the errors is not always appropriate in applications, especially for sequentially collected economic and physical data, which often exhibit evident dependence on the errors. Recently, linear regression with serially correlated errors has attracted increasing attention from statisticians. One case of considerable interest is that the errors are autoregressive processes and the asymptotic theory of this estimator was developed by Hannan and Kavalieris [25]. Fox and Taqqu [26] established its asymptotic normality in the case of long-memory stationary Gaussian observations errors. Giraitis and Surgailis [27] extended this result to non-Gaussian linear sequences. The asymptotic distribution of the maximum likelihood estimator was studied by Giraitis and Koul in [28] and Koul in [29] when the errors are nonlinear instantaneous functions of a Gaussian long-memory sequence. Koul and Surgailis [30] established the asymptotic normality of the Whittle estimator in linear regression models with non-Gaussian long-memory moving average errors. When the errors are Gaussian, or a function of Gaussian random variables that are strictly stationary and long range dependent, Koul and Mukherjee [31] investigated the linear model. Shiohama and Taniguchi [32] estimated the regression parameters in a linear regression model with autoregressive process.

In addition to (constant or functional or random coefficient) autoregressive model, it has gained much attention and has been applied to many fields, such as economics, physics, geography, geology, biology, and agriculture. Fan and Yao [33], Berk [34], Hannan and Kavalieris [35], Goldenshluger and Zeevi [36], Liebscher [37], An et al. [38], Elsebach [39], Carsoule and Franses [21], Baran et al. [40], Distaso [41], and Harvill and Ray [42] used various estimation methods (the least squares method, the Yule-Walker method, the method of stochastic approximation, and robust estimation method) to obtain some estimators and discussed their asymptotic properties, or investigated hypotheses testing.

This paper discusses the model (1.1)-(1.2) including stationary and explosive processes. The organization of the paper is as follows. In Section 2 some estimators of $\beta, \theta,$

and σ^2 are given by the quasi-maximum likelihood method. Under general conditions, the existence and consistency the quasi-maximum likelihood estimators are investigated, and asymptotic normality as well, in Section 3. Some preliminary lemmas are presented in Section 4. The main proofs are presented in Section 5, with some examples in Section 6.

2. Estimation Method

Write the "true" model as

$$y_t = x_t^T \beta_0 + e_t, \quad t = 1, 2, \ldots, n, \tag{2.1}$$

$$e_1 = \eta_1, \qquad e_t = f_t(\theta_0) e_{t-1} + \eta_t, \quad t = 2, 3, \ldots, n, \tag{2.2}$$

where $f_t'(\theta_0) = (df_t(\theta)/d\theta)|_{\theta=\theta_0} \neq 0$, and η_t's are i.i.d errors with zero mean and finite variance σ_0^2. Define $\prod_{i=0}^{-1} f_{t-i}(\theta_0) = 1$, and by (2.2) we have

$$e_t = \sum_{j=0}^{t-1} \left(\prod_{i=0}^{j-1} f_{t-i}(\theta_0) \right) \eta_{t-j}. \tag{2.3}$$

Thus e_t is measurable with respect to the σ-field H generated by $\eta_1, \eta_2, \ldots, \eta_t$, and

$$Ee_t = 0, \qquad \operatorname{Var}(e_t) = \sigma_0^2 \sum_{j=0}^{t-1} \left(\prod_{i=0}^{j-1} f_{t-i}^2(\theta_0) \right). \tag{2.4}$$

Assume at first that the η_t's are i.i.d. $N(0, \sigma^2)$. Using similar arguments to those of Fuller [14] or Maller [12], we get the log-likelihood of $y_2, y_3, \ldots, y_n$ conditional on y_1:

$$\Psi_n\left(\beta, \theta, \sigma^2\right) = \log L_n = -\frac{1}{2}(n-1)\log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=2}^{n} \left(\varepsilon_t - f_t(\theta)\varepsilon_{t-1}\right)^2 - \frac{1}{2}(n-1)\log(2\pi). \tag{2.5}$$

At this stage we drop the normality assumption, but still maximize (2.5) to obtain QML estimators, denoted by $\widehat{\sigma}_n^2$, $\widehat{\beta}_n$, $\widehat{\theta}_n$ (when they exist):

$$\frac{\partial \Psi_n}{\partial \sigma^2} = -\frac{n-1}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=2}^{n} \left(\varepsilon_t - f_t(\theta)\varepsilon_{t-1}\right)^2, \tag{2.6}$$

$$\frac{\partial \Psi_n}{\partial \theta} = \frac{1}{\sigma^2} \sum_{t=2}^{n} f_t'(\theta) \left(\varepsilon_t - f_t(\theta)\varepsilon_{t-1}\right) \varepsilon_{t-1}, \tag{2.7}$$

$$\frac{\partial \Psi_n}{\partial \beta} = \frac{1}{\sigma^2} \sum_{t=2}^{n} \left(\varepsilon_t - f_t(\theta)\varepsilon_{t-1}\right) \left(x_t - f_t(\theta) x_{t-1}\right). \tag{2.8}$$

Thus $\widehat{\sigma}_n^2,\ \widehat{\beta}_n,\ \widehat{\theta}_n$ satisfy the following estimation equations:

$$\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{t=2}^{n}\left(\widehat{\varepsilon}_t - f_t\left(\widehat{\theta}_n\right)\widehat{\varepsilon}_{t-1}\right)^2, \tag{2.9}$$

$$\sum_{t=2}^{n}\left(\widehat{\varepsilon}_t - f_t\left(\widehat{\theta}_n\right)\widehat{\varepsilon}_{t-1}\right)f_t'\left(\widehat{\theta}_n\right)\widehat{\varepsilon}_{t-1} = 0, \tag{2.10}$$

$$\sum_{t=2}^{n}\left(\widehat{\varepsilon}_t - f_t\left(\widehat{\theta}_n\right)\widehat{\varepsilon}_{t-1}\right)\left(x_t - f_t\left(\widehat{\theta}_n\right)x_{t-1}\right) = 0, \tag{2.11}$$

where

$$\widehat{\varepsilon}_t = y_t - x_t^T\widehat{\beta}_n. \tag{2.12}$$

Remark 2.1. If $f_t(\theta) = \theta$, then the above equations become the same as Maller's [12]. Therefore, we extend the QML estimators of Maller [12].

To calculate the values of the QML estimators, we may use the grid search method, steepest ascent method, Newton-Raphson method, and modified Newton-Raphson method. In order to calculate in Section 6, we introduce the most popular modified Newton-Raphson method proposed by Davidon-Fletcher-Powell (see Hamilton [18]).

Let $(d+2)\times 1$ vector $\overrightarrow{\theta}^{(m)} = (\sigma^{(m)2}, \beta^{(m)}, \theta^{(m)})$ denote an estimator of $\overrightarrow{\theta} = (\sigma^2, \beta, \theta)$ that has been calculated at the mth iteration, and let $A^{(m)}$ denote an estimation of $[\mathbf{H}(\overrightarrow{\theta}^{(m)})]^{-1}$. The new estimator $\overrightarrow{\theta}^{(m+1)}$ is given by

$$\overrightarrow{\theta}^{(m+1)} = \overrightarrow{\theta}^{(m)} + sA^{(m)}g\left(\overrightarrow{\theta}^{(m)}\right) \tag{2.13}$$

for s the positive scalar that maximizes $\Psi_n\{\overrightarrow{\theta}^{(m)} + sA^{(m)}g(\overrightarrow{\theta}^{(m)})\}$, where $(d+2)\times 1$ vector

$$g\left(\overrightarrow{\theta}^{(m)}\right) = \frac{\partial\Psi_n\left(\overrightarrow{\theta}\right)}{\partial\overrightarrow{\theta}}\Big|_{\overrightarrow{\theta}=\overrightarrow{\theta}^{(m)}} = \begin{pmatrix} \dfrac{\partial\Psi_n}{\partial\sigma^2}\Big|_{\sigma^2=\sigma^{(m)2}} \\ \dfrac{\partial\Psi_n}{\partial\beta}\Big|_{\beta=\beta^{(m)}} \\ \dfrac{\partial\Psi_n}{\partial\theta}\Big|_{\theta=\theta^{(m)}} \end{pmatrix} \tag{2.14}$$

and $(d+2)\times(d+2)$ symmetric matrix

$$\mathbf{H}\left(\overrightarrow{\theta}^{(m)}\right) = -\frac{\partial^2\Psi_n\left(\overrightarrow{\theta}\right)}{\partial\overrightarrow{\theta}\partial\overrightarrow{\theta}^T}\Big|_{\overrightarrow{\theta}=\overrightarrow{\theta}^{(m)}} = \begin{pmatrix} \dfrac{\partial^2\Psi_n}{\partial(\sigma^2)^2} & \dfrac{\partial^2\Psi_n}{\partial\sigma^2\partial\beta} & \dfrac{\partial^2\Psi_n}{\partial\sigma^2\partial\theta} \\ * & \dfrac{\partial^2\Psi_n}{\partial\beta\partial\beta^T} & \dfrac{\partial^2\Psi_n}{\partial\beta\partial\theta} \\ * & * & \dfrac{\partial^2\Psi_n}{\partial\theta^2} \end{pmatrix}\Big|_{\overrightarrow{\theta}=\overrightarrow{\theta}^{(m)}}, \tag{2.15}$$

where

$$\frac{\partial^2\Psi_n}{\partial(\sigma^2)^2} = \frac{n-1}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{t=2}^{n}\left(\varepsilon_t - f_t(\theta)\varepsilon_{t-1}\right)^2,$$

$$\frac{\partial^2\Psi_n}{\partial\sigma^2\partial\beta} = -\frac{1}{\sigma^4}\sum_{t=2}^{n}\left(\varepsilon_t - f_t(\theta)\varepsilon_{t-1}\right)\left(x_t - f_t(\theta)x_{t-1}\right)^T, \tag{2.16}$$

$$\frac{\partial^2\Psi_n}{\partial\sigma^2\partial\theta} = -\frac{1}{\sigma^4}\sum_{t=2}^{n}\left(\varepsilon_t - f_t(\theta)\varepsilon_{t-1}\right)f_t'(\theta),$$

$$\frac{\partial^2\Psi_n}{\partial\beta\partial\beta^T} = -\frac{1}{\sigma^2}\sum_{t=2}^{n}\left(x_t - f_t(\theta)x_{t-1}\right)\left(x_t - f_t(\theta)x_{t-1}\right)^T, \tag{2.17}$$

$$\begin{aligned}\frac{\partial^2\Psi_n}{\partial\beta\partial\theta} &= -\frac{1}{\sigma^2}\sum_{t=2}^{n}\left(f_t'(\theta)\varepsilon_{t-1}x_t + f_t'(\theta)\varepsilon_t x_{t-1} - 2f_t(\theta)f_t'(\theta)x_{t-1}\varepsilon_{t-1}\right),\\ \frac{\partial^2\Psi_n}{\partial\theta^2} &= -\frac{1}{\sigma^2}\sum_{t=2}^{n}\left(\left(f_t'^2(\theta) + f_t(\theta)f_t''(\theta)\right)\varepsilon_{t-1}^2 - f_t''(\theta)\varepsilon_t\varepsilon_{t-1}\right).\end{aligned} \tag{2.18}$$

Once $\overrightarrow{\theta}^{(m+1)}$ and the gradient at $\overrightarrow{\theta}^{(m+1)}$ have been calculated, a new estimation $A^{(m+1)}$ is found from

$$A^{(m+1)} = A^{(m)} - \frac{A^{(m)}\left(\Delta g^{(m+1)}\right)\left(\Delta g^{(m+1)}\right)^T A^{(m)}}{\left(\Delta g^{(m+1)}\right)^T A^{(m)}\left(\Delta g^{(m+1)}\right)} - \frac{\left(\Delta\overrightarrow{\theta}^{(m+1)}\right)\left(\Delta\overrightarrow{\theta}^{(m+1)}\right)^T}{\left(\Delta g^{(m+1)}\right)^T\left(\Delta\overrightarrow{\theta}^{(m+1)}\right)}, \tag{2.19}$$

where

$$\Delta\overrightarrow{\theta}^{(m+1)} = \overrightarrow{\theta}^{(m+1)} - \overrightarrow{\theta}^{(m)}, \qquad \Delta g^{(m+1)} = g\left(\overrightarrow{\theta}^{(m+1)}\right) - g\left(\overrightarrow{\theta}^{(m)}\right). \tag{2.20}$$

It is well known that least squares estimators in ordinary linear regression model are very good estimators, so a recursive algorithms procedure is to start the iteration with $\beta^{(0)}$, $\sigma^{(0)2}$ which are least squares estimators of β and σ^2, respectively. Take $\theta^{(0)}$ such that $f_t(\theta^{(0)}) = 0$. Iterations are stopped if some termination criterion is reached, for example, if

$$\frac{\left\|\overrightarrow{\theta}^{(m+1)} - \overrightarrow{\theta}^{(m)}\right\|}{\left\|\overrightarrow{\theta}^{(m)}\right\|} < \delta, \tag{2.21}$$

for some prechosen small number $\delta > 0$.

Up to this point, we obtain the values of QML estimators when the function $f_t(\theta) = f(t,\theta)$ is known. However, the function $f_t(\theta)$ is never the case in practice; we have to estimate it. By (2.12) and (1.2), we obtain

$$\tilde{f}\left(t,\hat{\theta}_n\right) = \frac{\hat{\varepsilon}_t}{\hat{\varepsilon}_{t-1}}, \quad t = 2,3,\ldots,n. \tag{2.22}$$

Based on the dataset $\{\tilde{f}(t,\hat{\theta}_n), t = 2,3,\ldots,n\}$, we may obtain the estimation function $\hat{f}(t,\hat{\theta}_n)$ of $f(t,\theta)$ by some smoothing methods (see Simonff [43], Fan and Yao [33], Green and Silverman [44], Fan and Gijbels [45], etc.)

To obtain our results, the following conditions are sufficient.

(A1) $X_n = \sum_{t=2}^{n} x_t x_t^T$ is positive definite for sufficiently large n and

$$\lim_{n\to\infty} \max_{1\le t\le n} x_t^T X_n^{-1} x_t = 0, \tag{2.23}$$

$$\limsup_{n\to\infty} |\lambda|_{\max}\left(X_n^{-1/2} Z_n X_n^{-T/2}\right) < 1, \tag{2.24}$$

where $Z_n = (1/2)\sum_{t=2}^{n}(x_t x_{t-1}^T + x_{t-1}x_t^T)$ and $|\lambda|_{\max}(\cdot)$ denotes the maximum in absolute value of the eigenvalues of a symmetric matrix.

(A2) There is a constant $\alpha > 0$ such that

$$\sum_{j=1}^{t}\left(\prod_{i=0}^{j-1} f_{t-i}^2(\theta)\right) \le \alpha \tag{2.25}$$

for any $t \in \{1,2,\ldots,n\}$ and $\theta \in \Theta$.

(A3) The derivatives $f_t'(\theta) = df_t(\theta)/d\theta$, $f_t''(\theta) = df_t'(\theta)/d\theta$ exist and are bounded for any t and $\theta \in \Theta$.

Remark 2.2. Maller [12] applied the condition (A1), and Kwoun and Yajima [15] used the conditions (A2) and (A3). Thus our conditions are general. (A1) delineates the class of x_t for which our results hold in the sense required. It is further discussed by Maller in [12]. Kwoun and Yajima [15] call $\{e_t\}$ stable if $\mathrm{Var}(e_t)$ is bounded. Thus (A2) implies that $\{e_t\}$ is stable. However, $\{e_t\}$ is not stationary. In fact, by (2.3), we obtain that

$$\mathrm{Cov}(e_t, e_{t+k}) = \sigma_0^2\left\{\prod_{i=0}^{k-1} f_{t+k-i}(\theta_0) + f_t(\theta_0)\prod_{i=0}^{k} f_{t+k-i}(\theta_0) + \cdots + \prod_{l=0}^{t-2} f_{t-l}(\theta_0)\prod_{i=0}^{t+k-2} f_{t+k-i}(\theta_0)\right\}, \tag{2.26}$$

which is dependent of t.

For ease of exposition, we will introduce the following notations which will be used later in the paper.

Define $(d+1)$-vector $\varphi = (\beta, \theta)$, and

$$S_n(\varphi) = \sigma^2 \frac{\partial \Psi_n}{\partial \varphi} = \sigma^2 \left(\frac{\partial \Psi_n}{\partial \beta}, \frac{\partial \Psi_n}{\partial \theta} \right), \qquad F_n(\varphi) = -\sigma^2 \frac{\partial^2 \Psi_n}{\partial \varphi \partial \varphi^T}. \tag{2.27}$$

By (2.7) and (2.8), we get

$$\mathrm{F}_n(\varphi) = \begin{pmatrix} X_n(\theta) & \sum_{t=2}^{n} \left(f_t'(\theta)\varepsilon_{t-1}x_t + f_t'(\theta)\varepsilon_t x_{t-1} - 2f_t(\theta)f_t'(\theta)x_{t-1}\varepsilon_{t-1} \right) \\ * & \sum_{t=2}^{n} \left(\left(f_t'^2(\theta) + f_t(\theta)f_t''(\theta) \right)\varepsilon_{t-1}^2 - f_t''(\theta)\varepsilon_t\varepsilon_{t-1} \right) \end{pmatrix}, \tag{2.28}$$

where $X_n(\theta) = -\sigma^2(\partial^2\Psi_n/\partial\beta\partial\beta^T)$ and the $*$ indicates that the element is filled in by symmetry. Thus,

$$\begin{aligned} D_n &= E(F_n(\varphi_0)) \\ &= \begin{pmatrix} X_n(\theta_0) & 0 \\ * & \sum_{t=2}^{n} \left(\left(f_t'^2(\theta_0) + f_t(\theta_0)f_t''(\theta_0) \right) Ee_{t-1}^2 - f_t''(\theta_0)E(e_t e_{t-1}) \right) \end{pmatrix} \\ &= \begin{pmatrix} X_n(\theta_0) & 0 \\ * & \sum_{t=2}^{n} f_t'^2(\theta_0)Ee_{t-1}^2 \end{pmatrix} \\ &= \begin{pmatrix} X_n(\theta_0) & 0 \\ * & \Delta_n(\theta_0, \sigma_0) \end{pmatrix}, \end{aligned} \tag{2.29}$$

where

$$\Delta_n(\theta_0, \sigma_0) = \sum_{t=2}^{n} f_t'^2(\theta_0)Ee_{t-1}^2 = \sigma_0^2 \sum_{t=2}^{n} f_t'^2(\theta_0) \sum_{j=0}^{t-2} \left(\prod_{i=0}^{j-1} f_{t-i}^2(\theta) \right) = O(n). \tag{2.30}$$

3. Statement of Main Results

Theorem 3.1. *Suppose that conditions (A1)–(A3) hold. Then there is a sequence $A_n \downarrow 0$ such that, for each $A > 0$, as $n \to \infty$, the probability*

$$P\left\{ \text{there are estimators } \widehat{\varphi}_n, \widehat{\sigma}_n^2 \text{ with } S_n(\widehat{\varphi}_n) = 0, \text{ and } \left(\widehat{\varphi}_n, \widehat{\sigma}_n^2 \right) \in N_n'(A) \right\} \longrightarrow 1. \tag{3.1}$$

Furthermore,

$$\left(\widehat{\varphi}_n, \widehat{\sigma}_n^2\right) \longrightarrow_p \left(\varphi_0, \sigma_0^2\right), \quad n \longrightarrow \infty, \tag{3.2}$$

where, for each $n = 1, 2, \ldots, A > 0$ *and* $A_n \in (0, \sigma_0^2)$*; define neighborhoods*

$$\begin{aligned} N_n(A) &= \left\{\varphi \in R^{d+1} : (\varphi - \varphi_0)^T D_n(\varphi - \varphi_0) \le A^2\right\}, \\ N_n'(A) &= N_n(A) \cap \left\{\sigma^2 \in \left[\sigma_0^2 - A_n, \sigma_0^2 + A_n\right]\right\}. \end{aligned} \tag{3.3}$$

Theorem 3.2. *Suppose that conditions (A1)–(A3) hold. Then*

$$\frac{1}{\widehat{\sigma}_n} F_n^{T/2}(\widehat{\varphi}_n)(\widehat{\varphi}_n - \varphi_0) \longrightarrow_D N(0, I_{d+1}), \quad n \longrightarrow \infty. \tag{3.4}$$

Remark 3.3. For $\theta \in R^m, m \in N$, our results still hold.

In the following, we will investigate some special cases in the model (1.1)-(1.2). Although the following results are directly obtained from Theorems 3.1 and 3.2, we discuss these results in order to compare with the corresponding results.

Corollary 3.4. *Let* $f_t(\theta) = \theta$*. If condition (A1) holds, then, for* $|\theta| \neq 1$, (3.1), (3.2), *and* (3.4) *hold.*

Remark 3.5. These results are the same as the corresponding results of Maller [12].

Corollary 3.6. *If* $\beta = 0$ *and* $f_t(\theta) = \theta$*, then, for* $|\ \theta\ | \neq 1$,

$$\frac{\sqrt{\sum_{t=2}^n \varepsilon_{t-1}^2}}{\widehat{\sigma}_n}\left(\widehat{\theta}_n - \theta_0\right) \longrightarrow_D N(0, 1), \quad n \longrightarrow \infty, \tag{3.5}$$

where

$$\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{t=2}^n \left(\varepsilon_t - \widehat{\theta}_n \varepsilon_{t-1}\right)^2, \quad \widehat{\theta}_n = \frac{\sum_{t=2}^n (\varepsilon_t \varepsilon_{t-1})}{\sum_{t=2}^n \varepsilon_{t-1}^2}. \tag{3.6}$$

Remark 3.7. These estimators are the same as the least squares estimators (see White [16]). For $|\theta| > 1$, $\{\varepsilon_t\}$ are explosive processes. In the case, the corollary is the same as the results of White [17]. While $|\theta| < 1$, notice that $\widehat{\sigma}_n^2 \to_p \sigma_0^2$ and $(1/(n-1)) \sum_{t=2}^n \varepsilon_{t-1}^2 \to_p E\varepsilon_t^2 = \sigma_0^2/(1-\theta_0^2)$, and by Corollary 3.6 we obtain

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \longrightarrow_D N\left(0, 1 - \theta_0^2\right). \tag{3.7}$$

The result was discussed by many authors, such as Fujikoshi and Ochi [46] and Brockwell and Davis [19].

Corollary 3.8. *Let $\beta = 0$. If conditions (A2) and (A3) hold, then*

$$\frac{F_n^{1/2}\left(\widehat{\theta}_n\right)}{\widehat{\sigma}_n}\left(\widehat{\theta}_n - \theta_0\right) \longrightarrow_D N(0,1), \quad n \longrightarrow \infty, \tag{3.8}$$

where

$$\begin{gathered} F_n\left(\widehat{\theta}_n\right) = \sum_{t=2}^{n}\left(\left(f_t'^2\left(\widehat{\theta}_n\right) + f_t\left(\widehat{\theta}_n\right)f_t''\left(\widehat{\theta}_n\right)\right)\varepsilon_{t-1}^2 - f_t''\left(\widehat{\theta}_n\right)\varepsilon_t\varepsilon_{t-1}\right), \\ \widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{t=2}^{n}\left(\varepsilon_t - f_t\left(\widehat{\theta}_n\right)\varepsilon_{t-1}\right)^2. \end{gathered} \tag{3.9}$$

Corollary 3.9. *Let $f_t(\theta) = a_t$. If condition (A1) holds, then*

$$\frac{1}{\widehat{\sigma}_n}\left\{\sum_{t=2}^{n}(x_t - a_t x_{t-1})(x_t - a_t x_{t-1})^T\right\}^{T/2}\left(\widehat{\beta}_n - \beta_0\right) \longrightarrow_D N(0, I_d), \quad n \longrightarrow \infty. \tag{3.10}$$

Remark 3.10. Let $a_t = 0$. Note that $\sum_{t=2}^{n} x_t x_t^T = O(\sqrt{n})$ and $\widehat{\sigma}_n^2 \to_p \sigma_0^2$; we easily obtain asymptotic normality of the (quasi-)maximum likelihood or least squares estimator in ordinary linear regression models from the corollary.

4. Some Lemmas

To prove Theorems 3.1 and 3.2, we first introduce the following lemmas.

Lemma 4.1. *The matrix D_n is positive definite for large enough n with $E(S_n(\varphi_0)) = 0$ and $\mathrm{Var}(S_n(\varphi_0)) = \sigma_0^2 D_n$.*

Proof. It is easy to show that the matrix D_n is positive definite for large enough n. By (2.8), we have

$$\begin{aligned} \sigma_0^2 E\left(\frac{\partial \Psi_n}{\partial \beta}\Big|_{\beta=\beta_0}\right) &= \sum_{t=2}^{n} E(e_t - f_t(\theta_0)e_{t-1})(x_t - f_t(\theta_0)x_{t-1}) \\ &= \sum_{t=2}^{n}(x_t - f_t(\theta_0)x_{t-1})E\eta_t = 0. \end{aligned} \tag{4.1}$$

Note that e_{t-1} and η_t are independent of each other; thus by (2.7) and $E\eta_t = 0$, we have

$$\begin{aligned} \sigma_0^2 E\left(\frac{\partial \Psi_n}{\partial \theta}\Big|_{\theta=\theta_0}\right) &= \sum_{t=2}^{n} E\left((e_t - f_t(\theta_0)e_{t-1})f_t'(\theta_0)e_{t-1}\right) \\ &= \sum_{t=2}^{n} f_t'(\theta_0)E(\eta_t e_{t-1}) = 0. \end{aligned} \tag{4.2}$$

Hence, from (4.1) and (4.2),

$$E(S_n(\varphi_0)) = \sigma_0^2 E\left(\frac{\partial\Psi_n}{\partial\beta}|_{\beta=\beta_0}, \frac{\partial\Psi_n}{\partial\theta}|_{\theta=\theta_0}\right) = 0. \tag{4.3}$$

By (2.8) and (2.17), we have

$$\begin{aligned}\operatorname{Var}\left(\sigma_0^2\frac{\partial\Psi_n}{\partial\beta}|_{\beta=\beta_0}\right) &= \operatorname{Var}\left\{\sum_{t=2}^{n}(e_t - f_t(\theta_0)e_{t-1})(x_t - f_t(\theta_0)x_{t-1})\right\}\\ &= \operatorname{Var}\left\{\sum_{t=2}^{n}\eta_t(x_t - f_t(\theta_0)x_{t-1})\right\} = \sigma_0^2 X_n(\theta_0).\end{aligned} \tag{4.4}$$

Note that $\{f_t'(\theta_0)\eta_t e_{t-1}, H_t\}$ is a martingale difference sequence with

$$\operatorname{Var}(f_t'(\theta_0)\eta_t e_{t-1}) = f_t'^2(\theta_0)E\eta_t^2 Ee_{t-1}^2 = \sigma_0^2 f_t'^2(\theta_0)Ee_{t-1}^2, \tag{4.5}$$

so

$$\begin{aligned}\operatorname{Var}\left(\sigma_0^2\frac{\partial\Psi_n}{\partial\theta}|_{\theta=\theta_0}\right) &= \operatorname{Var}\left\{\sum_{t=2}^{n}\eta_t f_t'(\theta_0)e_{t-1}\right\}\\ &= \sum_{t=2}^{n} f_t'^2(\theta_0)Ee_{t-1}^2 = \sigma_0^2\Delta_n(\theta_0, \sigma_0).\end{aligned} \tag{4.6}$$

By (2.7) and (2.8) and noting that e_{t-1} and η_t are independent of each other, we have

$$\begin{aligned}\operatorname{Cov}\left(\sigma_0^2\frac{\partial\Psi_n}{\partial\beta}|_{\beta=\beta_0}, \sigma_0^2\frac{\partial\Psi_n}{\partial\theta}|_{\theta=\theta_0}\right) &= E\left(\sigma_0^2\frac{\partial\Psi_n}{\partial\beta}|_{\beta=\beta_0}, \sigma_0^2\frac{\partial\Psi_n}{\partial\theta}|_{\theta=\theta_0}\right)\\ &= E\left(\sum_{t=2}^{n}\eta_t^2(x_t - f_t(\theta_0)x_{t-1})f_t'(\theta_0)e_{t-1}\right)\\ &\quad + E\left(\sum_{t=3}^{n}\eta_t(x_t - f_t(\theta_0)x_{t-1})\sum_{s=2}^{t-1}\eta_s f_s'(\theta_0)e_{s-1}\right)\\ &\quad + E\left(\sum_{s=3}^{n}\eta_s f_s'(\theta_0)e_{s-1}\sum_{t=2}^{s-1}\eta_t(x_t - f_t(\theta_0)x_{t-1})\right)\\ &= 0.\end{aligned} \tag{4.7}$$

From (4.4)–(4.7), it follows that $\operatorname{Var}(S_n(\varphi_0)) = \sigma_0^2 D_n$. □

Lemma 4.2. *If condition (A1) holds, then, for any $\theta \in \Theta$, the matrix $X_n(\theta)$ is positive definite for large enough n, and*

$$\lim_{n\to\infty}\max_{1\le t\le n} x_t^T X_n^{-1}(\theta)x_t = 0. \tag{4.8}$$

Proof. Let λ_1 and λ_d be the smallest and largest roots of $|Z_n - \lambda X_n| = 0$. Then from the study of Rao in [47, Ex 22.1],

$$\lambda_1 \le \frac{u^T Z_n u}{u^T X_n u} \le \lambda_d \tag{4.9}$$

for unit vectors u. Thus by (2.24), there are some $\delta \in (\max\{0, 1-(1+\min_{2\le t\le n}| f_t^2(\theta)|)/\max_{2\le t\le n}|f_t(\theta)|\}, 1)$ and $n_0(\delta)$ such that $n \ge N_0$ implies that

$$\left|u^T Z_n u\right| \le (1-\delta)u^T X_n u. \tag{4.10}$$

By (4.10), we have

$$\begin{aligned}
u^T X_n u &= \sum_{t=2}^{n}\left(u^T(x_t - f_t(\theta)x_{t-1})\right)^2 \\
&= \sum_{t=2}^{n}\left(\left(u^T x_t\right)^2 + f_t^2(\theta)\left(u^T x_{t-1}\right)^2 - f_t(\theta)u^T x_{t-1}x_t^T u - f_t(\theta)u^T x_t x_{t-1}^T u\right) \\
&\ge \sum_{t=2}^{n}\left(u^T x_t\right)^2 + \min_{2\le t\le n}\left|f_t^2(\theta)\right|\sum_{t=2}^{n}\left(u^T x_{t-1}\right)^2 - \max_{2\le t\le n}\left|f_t(\theta)\right|u^T Z_n u \\
&\ge u^T X_n u + \min_{2\le t\le n}\left|f_t^2(\theta)\right|u^T X_n u - \max_{2\le t\le n}\left|f_t(\theta)\right|u^T Z_n u \\
&\ge \left(1 + \min_{2\le t\le n}\left|f_t^2(\theta)\right| - \max_{2\le t\le n}\left|f_t(\theta)\right|(1-\delta)\right)u^T X_n u \\
&= C(\theta,\delta)u^T X_n u.
\end{aligned} \tag{4.11}$$

By the study of Rao in [47, page 60] and (2.23), we have

$$\frac{\left(u^T x_t\right)^2}{u^T X_n u} \longrightarrow 0. \tag{4.12}$$

From (4.12) and $C(\theta,\delta) > 0$,

$$x_t^T X_n^{-1}(\theta) = \sup_u \left(\frac{(u^T x_t)^2}{u^T X_n(\theta) u} \right) \le \sup_u \left(\frac{(u^T x_t)^2}{C(\theta,\delta) u^T X_n u} \right) \longrightarrow 0. \tag{4.13}$$

□

Lemma 4.3 (see [48]). *Let W_n be a symmetric random matrix with eigenvalues $\lambda_j(n), 1 \le j \le d$. Then*

$$W_n \longrightarrow_p I \Longleftrightarrow \lambda_j(n) \longrightarrow_p 1, \quad n \longrightarrow \infty. \tag{4.14}$$

Lemma 4.4. *For each $A > 0$,*

$$\sup_{\varphi \in N_n(A)} \left\| D_n^{-1/2} F_n(\varphi) D_n^{-T/2} - \Phi_n \right\| \longrightarrow_p 0, \quad n \longrightarrow \infty, \tag{4.15}$$

and also

$$\Phi_n \longrightarrow_D \Phi, \tag{4.16}$$

$$\lim_{c\to 0} \limsup_{A\to\infty} \limsup_{n\to\infty} P\left\{ \inf_{\varphi\in N_n(A)} \lambda_{\min}\left(D_n^{-1/2} F_n(\varphi) D_n^{-T/2} \right) \le c \right\} = 0, \tag{4.17}$$

where

$$\Phi_n = \begin{pmatrix} I_d & 0 \\ 0 & \dfrac{\sum_{t=2}^n f_t'^2(\theta_0) e_{t-1}^2}{\Delta_n(\theta_0,\sigma_0)} \end{pmatrix}, \quad \Phi = I_{d+1}. \tag{4.18}$$

Proof. Let $X_n(\theta_0) = X_n^{1/2}(\theta_0) X_n^{T/2}(\theta_0)$ be a square root decomposition of $X_n(\theta_0)$. Then

$$D_n = \begin{pmatrix} X_n^{1/2}(\theta_0) & 0 \\ * & \sqrt{\Delta_n(\theta_0,\sigma_0)} \end{pmatrix} \begin{pmatrix} X_n^{T/2}(\theta_0) & 0 \\ * & \sqrt{\Delta_n(\theta_0,\sigma_0)} \end{pmatrix} = D_n^{1/2} D_n^{T/2}. \tag{4.19}$$

Let $\varphi \in N_n(A)$. Then

$$(\varphi - \varphi_0)^T D_n (\varphi - \varphi_0) = (\beta - \beta_0)^T X_n(\theta_0)(\beta - \beta_0) + (\theta - \theta_0)^2 \Delta_n(\theta_0,\sigma_0) \le A^2. \tag{4.20}$$

From (2.28), (2.29), and (4.18),

$$D_n^{-1/2}F_n(\varphi)D_n^{-T/2}-\Phi_n$$
$$=\begin{pmatrix} X_n^{-1/2}(\theta_0)X_n(\theta)X_n^{-T/2}(\theta_0)-I_d & \dfrac{X_n^{-1/2}(\theta_0)\sum_{t=2}^n\left(f_t'(\theta)\varepsilon_{t-1}x_t+f_t'(\theta)\varepsilon_t x_{t-1}-2f_t(\theta)f_t'(\theta)\varepsilon_{t-1}x_{t-1}\right)}{\sqrt{\Delta_n(\theta_0,\sigma_0)}} \\ * & \dfrac{\sum_{t=2}^n\left(\left(f_t'^2(\theta)+f_t(\theta)f_t''(\theta)\right)\varepsilon_{t-1}^2-f_t''(\theta)\varepsilon_t\varepsilon_{t-1}\right)-\sum_{t=2}^n f_t'^2(\theta_0)e_{t-1}^2}{\sqrt{\Delta_n(\theta_0,\sigma_0)}} \end{pmatrix}. \tag{4.21}$$

Let

$$N_n^{\beta}(A)=\left\{\beta:\left|(\beta-\beta_0)^T X_n^{1/2}(\theta_0)\right|^2\le A^2\right\}, \tag{4.22}$$

$$N_n^{\theta}(A)=\left\{\theta:|\theta-\theta_0|\le\frac{A}{\sqrt{\Delta_n(\theta_0,\sigma_0)}}\right\}. \tag{4.23}$$

In the first step, we will show that, for each $A>0$,

$$\sup_{\theta\in N_n^{\theta}(A)}\left\|X_n^{-1/2}(\theta_0)X_n(\theta)X_n^{-T/2}(\theta_0)-I_d\right\|\longrightarrow 0,\quad n\longrightarrow\infty. \tag{4.24}$$

In fact, note that

$$\begin{aligned} X_n^{-1/2}(\theta_0)X_n(\theta)X_n^{-T/2}(\theta_0)-I_d &= X_n^{-1/2}(\theta_0)(X_n(\theta)-X_n(\theta_0))X_n^{-T/2}(\theta_0) \\ &= X_n^{-1/2}(\theta_0)(T_1+T_2-T_3)X_n^{-T/2}(\theta_0), \end{aligned} \tag{4.25}$$

where

$$\begin{aligned} T_1 &= \sum_{t=2}^n\left(f_t(\theta_0)-f_t(\theta)\right)x_{t-1}\left(x_t-f_t(\theta_0)x_{t-1}\right)^T, \\ T_2 &= \sum_{t=2}^n\left(x_t-f_t(\theta_0)x_{t-1}\right)x_{t-1}^T, \\ T_3 &= \sum_{t=2}^n\left(f_t(\theta_0)-f_t(\theta)\right)^2x_{t-1}x_{t-1}^T. \end{aligned} \tag{4.26}$$

Let $u, v \in R^d$, $|u| = |v| = 1$, and let $u_n^T = u^T X_n^{-1/2}(\theta_0)$, $v_n^T = X_n^{-T/2}(\theta_0)v$. By Cauchy-Schwartz inequality, Lemma 4.2, condition (A3), and noting that $\theta \in N_n^\theta(A)$, we have that

$$\begin{aligned}
\left|u_n^T T_1 v_n\right| &= \left|\sum_{t=2}^{n}\left(f_t(\theta_0) - f_t(\theta)\right)u_n^T x_{t-1}\left(x_t - f_t(\theta_0)x_{t-1}\right)^T v_n\right| \\
&\leq \max_{2\leq t\leq n}\left|f_t(\theta_0) - f_t(\theta)\right|\left|\sum_{t=2}^{n}u_n^T x_{t-1}\left(x_t - f_t(\theta_0)x_{t-1}\right)^T v_n\right| \\
&\leq \max_{2\leq t\leq n}\left|f_t(\theta_0) - f_t(\theta)\right|\left(\sum_{t=2}^{n}u_n^T x_{t-1}x_{t-1}^T u_n\right)^{1/2} \\
&\quad\cdot\left(\sum_{t=2}^{n}v_n^T\left(x_t - f_t(\theta_0)x_{t-1}\right)\left(x_t - f_t(\theta_0)x_{t-1}\right)^T v_n\right)^{1/2} \\
&\leq \max_{2\leq t\leq n}\left|f_t(\theta_0) - f_t(\theta)\right|\left(\sum_{t=2}^{n}u_n^T x_t x_t^T u_n\right)^{1/2} \\
&\leq \max_{2\leq t\leq n}\left|f_t'\left(\widetilde{\theta}\right)\right||\theta_0 - \theta| \cdot \sqrt{n}\max_{1\leq t\leq n}\left(x_t^T X_n^{-1}(\theta_0)x_t\right) \\
&\leq C\sqrt{\frac{n}{\Delta_n(\theta_0, \sigma_0)}}o(1) \longrightarrow 0.
\end{aligned} \tag{4.27}$$

Here $\widetilde{\theta} = a\theta + (1-a)\theta_0$ for some $0 \leq a \leq 1$. Similar to the proof of T_1, we easily obtain that

$$\left|u_n^T T_2 v_n\right| \longrightarrow 0. \tag{4.28}$$

By Cauchy-Schwartz inequality, Lemma 4.2, condition (A3), and noting that $N_n^\theta(A)$, we have that

$$\begin{aligned}
\left|u_n^T T_3 v_n\right| &= \left|u_n^T \sum_{t=2}^{n}\left(f_t(\theta_0) - f_t(\theta)\right)^2 x_{t-1}x_{t-1}^T v_n\right| \\
&\leq \max_{2\leq t\leq n}\left|f_t(\theta_0) - f_t(\theta)\right|^2\left(\sum_{t=2}^{n}u_n^T x_t x_t^T u_n \sum_{t=2}^{n}v_n^T x_t x_t^T v_n\right)^{1/2} \\
&\leq n\max_{2\leq t\leq n}\left|f_t'\left(\widetilde{\theta}\right)\right|^2|\theta_0 - \theta|^2\max_{1\leq t\leq n}\left(x_t^T X_n^{-1}(\theta_0)x_t\right) \\
&\leq \frac{nA^2}{\Delta_n(\theta_0, \sigma_0)}o(1) \longrightarrow 0.
\end{aligned} \tag{4.29}$$

Hence, (4.24) follows from (4.25)–(4.29).

In the second step, we will show that

$$\frac{X_n^{-1/2}(\theta_0)\sum_{t=2}^{n}\left(f_t'(\theta)\varepsilon_{t-1}x_t + f_t'(\theta)\varepsilon_t x_{t-1} - 2f_t(\theta)f_t'(\theta)\varepsilon_{t-1}x_{t-1}\right)}{\sqrt{\Delta_n(\theta_0,\sigma_0)}}\longrightarrow_p 0. \tag{4.30}$$

Note that

$$\begin{aligned}\varepsilon_t &= y_t - x_t^T\beta = x_t^T(\beta_0-\beta)+e_t,\\ \varepsilon_t - f_t(\theta_0)\varepsilon_{t-1} &= \left(x_t - f_t(\theta_0)x_{t-1}\right)^T(\beta_0-\beta)+\eta_t.\end{aligned} \tag{4.31}$$

Consider

$$\begin{aligned}J &= \sum_{t=2}^{n}\left(f_t'(\theta)\varepsilon_{t-1}x_t + f_t'(\theta)\varepsilon_t x_{t-1} - 2f_t(\theta)f_t'(\theta)\varepsilon_{t-1}x_{t-1}\right)\\ &= \sum_{t=2}^{n}\left(\varepsilon_{t-1}f_t'(\theta)\left(x_t - f_t(\theta_0)x_{t-1}\right) + f_t'(\theta)\left(\varepsilon_t - f_t(\theta_0)\varepsilon_{t-1}\right)x_{t-1}\right)\\ &\quad + 2f_t(\theta)\left(f_t(\theta_0)-f_t(\theta)\right)\varepsilon_{t-1}x_{t-1}\\ &= T_1+T_2+T_3+T_4+2T_5+2T_6,\end{aligned} \tag{4.32}$$

where

$$\begin{gathered}T_1 = \sum_{t=2}^{n}x_{t-1}^T f_t'(\theta)(\beta_0-\beta)\left(x_t - f_t(\theta_0)x_{t-1}\right),\qquad T_2 = \sum_{t=2}^{n}f_t'(\theta)e_{t-1}\left(x_t - f_t(\theta_0)x_{t-1}\right),\\ T_3 = \sum_{t=2}^{n}f_t'(\theta)\left(x_t - f_t(\theta_0)x_{t-1}\right)^T(\beta_0-\beta)x_{t-1},\qquad T_4 = \sum_{t=2}^{n}f_t'(\theta)\eta_t x_{t-1},\\ T_5 = \sum_{t=2}^{n}f_t'(\theta)\left(f_t(\theta_0)-f_t(\theta)\right)x_{t-1}^T(\beta_0-\beta)x_{t-1},\qquad T_6 = \sum_{t=2}^{n}f_t'(\theta)\left(f_t(\theta_0)-f_t(\theta)\right)e_{t-1}x_{t-1}.\end{gathered} \tag{4.33}$$

For $\beta\in N_n^{\beta}(A)$ and each $A>0$, we have

$$\begin{aligned}\left|(\beta_0-\beta)^T x_t\right|^2 &= (\beta_0-\beta)^T X_n^{1/2}(\theta_0)X_n^{-1/2}(\theta_0)x_t x_t^T X_n^{-T/2}(\theta_0)X_n^{T/2}(\theta_0)(\beta_0-\beta)\\ &\le \max_{1\le t\le n}\left(x_t^T X_n^{-1}(\theta_0)x_t\right)(\beta_0-\beta)^T X_n(\theta_0)(\beta_0-\beta)\\ &\le A^2\max_{1\le t\le n}\left(x_t^T X_n^{-1}(\theta_0)x_t\right).\end{aligned} \tag{4.34}$$

By (4.34) and Lemma 4.2, we have

$$\sup_{\beta\in N_n^{\beta}(A)} \max_{1\le t\le n}\left|(\beta_0-\beta)^T x_t\right| \longrightarrow 0, \quad n\longrightarrow\infty,\ A>0. \tag{4.35}$$

Using Cauchy-Schwartz inequality, condition (A3), and (4.35), we obtain

$$\begin{aligned} u_n^T T_1 &= \sum_{t=2}^{n} u_n^T x_{t-1}^T (\beta_0-\beta) f_t'(\theta)(x_t - f_t(\theta_0)x_{t-1}) \\ &\le \left\{\sum_{t=2}^{n}\left(x_{t-1}^T(\beta_0-\beta)\right)^2\right\}^{1/2} \\ &\quad \times \left\{\sum_{t=2}^{n}\left(f_t'^2(\theta) u_n^T (x_t - f_t(\theta_0)x_{t-1})(x_t - f_t(\theta_0)x_{t-1})^T u_n\right)^2\right\}^{1/2} \\ &\le \sqrt{n}\max_{1\le t\le n}\left|(\beta_0-\beta)^T x_t\right| \max_{1\le t\le n}\left|f_t'(\theta)\right| \\ &= o(\sqrt{n}). \end{aligned} \tag{4.36}$$

Let

$$a_{tn} = u_n^T f_t'(\theta)(x_t - f_t(\theta_0)x_{t-1}). \tag{4.37}$$

Then from Lemma 4.2,

$$\begin{aligned} \max_{2\le t\le n} a_{tn}^2 &= \max_{2\le t\le n}\left|f_t'^2(\theta)\right| \\ &\quad \times \max_{2\le t\le n}\left\{u^T (x_t - f_t(\theta_0)x_{t-1}) X_n^{-1}(\theta_0)(x_t - f_t(\theta_0)x_{t-1})^T u\right\} \\ &= o(1). \end{aligned} \tag{4.38}$$

By condition (A2) and (4.38), we have

$$\begin{aligned}
\operatorname{Var}\left(u_n^T T_2\right) &= \operatorname{Var}\left(\sum_{t=2}^{n} a_{tn} e_{t-1}\right) = \operatorname{Var}\left(\sum_{t=2}^{n} a_{tn} e_{t-1}\right) \\
&= \operatorname{Var}\left\{\sum_{j=1}^{n-1} \eta_j \left(\sum_{t=j+1}^{n} a_{tn} \prod_{i=0}^{t-j-1} f_{t-i}(\theta_0)\right)\right\} \\
&= \sigma_0^2 \sum_{j=1}^{n-1} \left(\sum_{t=j+1}^{n} a_{tn} \prod_{i=0}^{t-j-1} f_{t-i}(\theta_0)\right) \\
&\leq \sigma_0^2 \max_{2\leq t\leq n} |a_{tn}| \sum_{j=1}^{n-1} \left(\prod_{i=0}^{t-j-1} f_{t-i}(\theta_0)\right) \\
&\leq \alpha \sigma_0^2 \max_{2\leq t\leq n} |a_{tn}| n = o(n).
\end{aligned} \tag{4.39}$$

Thus by Chebychev inequality and (4.39),

$$u_n^T T_2 = o_p\left(\sqrt{n}\right). \tag{4.40}$$

Using the similar argument as T_1, we obtain that

$$u_n^T T_3 = o_p\left(\sqrt{n}\right). \tag{4.41}$$

Using the similar argument as T_2, we obtain that

$$u_n^T T_4 = o_p\left(\sqrt{n}\right), \qquad u_n^T T_6 = o_p\left(\sqrt{n}\right). \tag{4.42}$$

By Cauchy-Schwartz inequality, (4.35), and (4.27), we get

$$\begin{aligned}
u_n^T T_5 &= \sum_{t=2}^{n} f_t'(\theta)\left(f_t(\theta_0) - f_t(\theta)\right) x_{t-1}^T (\beta_0 - \beta) u_t^T x_{t-1} \\
&\leq \left\{\sum_{t=2}^{n} f_t'^2(\theta)\left(f_t(\theta_0) - f_t(\theta)\right)^2 \left(x_{t-1}^T (\beta_0 - \beta)\right)^2 \sum_{t=2}^{n} \left(u_t^T x_{t-1}\right)^2\right\}^{1/2} \\
&\leq \max_{2\leq t\leq n} \left|f_t'(\theta)\right| \max_{2\leq t\leq n} \left|f_t'\left(\tilde{\theta}\right)\right| |\theta_0 - \theta| \cdot \left\{\sum_{t=2}^{n} \left(x_{t-1}^T (\beta_0 - \beta)\right)^2 \sum_{t=2}^{n} \left(u_t^T x_{t-1}\right)^2\right\}^{1/2} \\
&\leq C \frac{A}{\sqrt{\Delta_n(\theta_0, \sigma_0)}} \sqrt{n} o(1) = o\left(\sqrt{n}\right).
\end{aligned} \tag{4.43}$$

Thus (4.30) follows immediately from (4.32), (4.36), and (4.40)–(4.43).

In the third step, we will show that

$$\frac{\sum_{t=2}^{n}\left(\left(f_t'^2(\theta)+f_t(\theta)f_t''(\theta)\right)\varepsilon_{t-1}^2-f_t''(\theta)\varepsilon_t\varepsilon_{t-1}\right)-\sum_{t=2}^{n}f_t'^2(\theta_0)e_{t-1}^2}{\sqrt{\Delta_n(\theta_0,\sigma_0)}}\longrightarrow_p 0. \tag{4.44}$$

Let

$$J=\sum_{t=2}^{n}\left(\left(f_t'^2(\theta)+f_t(\theta)f_t''(\theta)\right)\varepsilon_{t-1}^2-f_t''(\theta)\varepsilon_t\varepsilon_{t-1}\right)-\sum_{t=2}^{n}f_t'^2(\theta_0)e_{t-1}^2. \tag{4.45}$$

Then

$$\begin{aligned}
J&=\sum_{t=2}^{n}\left(\left(f_t'^2(\theta)+f_t(\theta)f_t''(\theta)\right)\varepsilon_{t-1}^2-f_t''(\theta)\left(f_t(\theta)\varepsilon_{t-1}+\eta_t\right)\varepsilon_{t-1}\right)-\sum_{t=2}^{n}f_t'^2(\theta_0)e_{t-1}^2\\
&=\sum_{t=2}^{n}\left\{f_t'^2(\theta)\varepsilon_{t-1}^2-f_t''(\theta)\eta_t\varepsilon_{t-1}-f_t'^2(\theta_0)e_{t-1}^2\right\}\\
&=\sum_{t=2}^{n}\left\{f_t'^2\left(x_{t-1}^T(\beta_0-\beta)\right)^2+\left(f_t'^2(\theta)-f_t'^2(\theta_0)\right)^2e_{t-1}^2\right\}\\
&\quad+\sum_{t=2}^{n}\left\{2f_t'^2(\theta_0)x_{t-1}^T(\beta_0-\beta)e_{t-1}-f_t''(\theta)x_{t-1}^T(\beta_0-\beta)\eta_{t-1}-f_t''(\theta_0)\eta_te_{t-1}\right\}\\
&=T_1+T_2+2T_3-T_4-T_5.
\end{aligned} \tag{4.46}$$

By (4.34), it is easy to show that

$$T_1=o(n). \tag{4.47}$$

From condition (A3), (2.30), and (4.23), we obtain that

$$\begin{aligned}
|ET_2|&=\left|\left(f_t'^2(\theta)-f_t'^2(\theta_0)\right)^2Ee_{t-1}^2\right|\\
&=\left|f_t''\left(\tilde{\theta}\right)(\theta-\theta_0)\left(\frac{f_t'(\theta)+f_t'(\theta_0)}{f_t'^2(\theta_0)}\right)f_t'^2(\theta_0)Ee_{t-1}^2\right|\\
&\le\max_{2\le t\le n}\left|f_t''\left(\tilde{\theta}\right)\left(\frac{f_t'(\theta)+f_t'(\theta_0)}{f_t'^2(\theta_0)}\right)\right||\theta-\theta_0|\Delta_n(\theta_0,\sigma_0)\\
&\le\max_{2\le t\le n}\left|f_t''\left(\tilde{\theta}\right)\left(\frac{f_t'(\theta)+f_t'(\theta_0)}{f_t'^2(\theta_0)}\right)\right|\frac{A}{\sqrt{\Delta_n(\theta_0,\sigma_0)}}\Delta_n(\theta_0,\sigma_0)\\
&=o\left(\sqrt{\Delta_n(\theta_0,\sigma_0)}\right).
\end{aligned} \tag{4.48}$$

Hence, by Markov inequality,

$$T_2 = O_p\left(\sqrt{\Delta_n(\theta_0, \sigma_0)}\right). \tag{4.49}$$

Using the similar argument as (4.40), we easily obtain that

$$T_3 = o_p(\sqrt{n}). \tag{4.50}$$

By Markov inequality and noting that

$$\operatorname{Var}(T_4) \le \sigma_0^2 n \max_{2\le t\le n} \left| f_t''(\theta) \right|^2 \max_{2\le t\le n} \left(x_t^T (\beta_0 - \beta) \right) = o(n), \tag{4.51}$$

we have that

$$T_4 = o_p(\sqrt{n}). \tag{4.52}$$

Using the similar argument as (4.6), we easily obtain that

$$T_5 = O_p\left(\sqrt{\Delta_n(\theta_0, \sigma_0)}\right). \tag{4.53}$$

Hence, (4.44) follows immediately from (4.46), (4.47), and (4.49)–(4.53).

This completes the proof of (4.15) from (4.21), (4.24), (4.30), and (4.44). To prove (4.16), we need to show that

$$\frac{\sum_{t=2}^{n} f_t'^2(\theta_0) e_{t-1}^2}{\Delta_n(\theta_0, \sigma_0)} \longrightarrow_p 1, \quad n \longrightarrow \infty. \tag{4.54}$$

This follows immediately from (2.27) and Markov inequality.

Finally, we will prove (4.17). By (4.15) and (4.16), we have

$$D_n^{-1/2} F(\varphi) D_n^{-T/2} \longrightarrow_p I, \quad n \longrightarrow \infty, \tag{4.55}$$

uniformly in $\varphi \in N_n(A)$ for each $A > 0$. Thus, by Lemma 4.3,

$$\lambda_{\min}\left(D_n^{-1/2} F(\varphi) D_n^{-T/2}\right) \longrightarrow_p 1, \quad n \longrightarrow \infty. \tag{4.56}$$

This implies (4.17). □

Lemma 4.5 (see [49]). *Let $\{S_{ni}, F_{ni}, 1 \le i \le k_n, n \ge 1\}$ be a zero-mean, square-integrable martingale array with differences X_{ni}, and let η^2 be an a.s. finite random variable. Suppose that $\sum_i E\{X_{ni}^2 I(|X_{ni}| > \varepsilon) \mid F_{n,i-1}\} \to_p 0$, for all $\varepsilon \to 0$, and $\sum_i E\{X_{ni}^2 \mid F_{n,i-1}\} \to_p \eta^2$. Then*

$$S_{nk_n} = \sum_i X_{ni} \longrightarrow_D Z, \tag{4.57}$$

where the r.v. Z has characteristic function $E\{\exp(-(1/2)\eta^2 t^2)\}$.

5. Proof of Theorems

5.1. Proof of Theorem 3.1

Take $A > 0$, let

$$M_n(A) = \left\{\varphi \in R^{d+1} : (\varphi - \varphi_0)^T D_n(\varphi - \varphi_0) = A^2\right\} \tag{5.1}$$

be the boundary of $N_n(A)$, and let $\varphi \in M_n(A)$. Using (2.27) and Taylor expansion, for each $\sigma^2 > 0$, we have

$$\begin{aligned}\Psi_n\left(\varphi, \sigma^2\right) &= \Psi_n\left(\varphi_0, \sigma^2\right) + (\varphi - \varphi_0)^T \frac{\partial \Psi_n(\varphi_0, \sigma^2)}{\partial \varphi} + \frac{1}{2}(\varphi - \varphi_0)^T \frac{\partial^2 \Psi_n(\varphi_0, \sigma^2)}{\partial \varphi \partial \varphi^T}(\varphi - \varphi_0) \\ &= \frac{1}{\sigma^2}\Psi_n\left(\varphi_0, \sigma^2\right) + (\varphi - \varphi_0)^T S_n(\varphi_0) - \frac{1}{2\sigma^2}(\varphi - \varphi_0)^T F_n(\tilde{\varphi})(\varphi - \varphi_0),\end{aligned} \tag{5.2}$$

where $\tilde{\varphi} = a\varphi + (1-a)\varphi_0$ for some $0 \le a \le 1$.

Let $Q_n(\varphi) = (1/2)(\varphi - \varphi_0)^T F_n(\tilde{\varphi})(\varphi - \varphi_0)$ and $v_n(\varphi) = (1/A) D_n^{T/2}(\varphi - \varphi_0)$. Take $c > 0$ and $\varphi \in M_n(A)$, and by (5.2) we obtain that

$$\begin{aligned}&P\left\{\Psi_n\left(\varphi, \sigma^2\right) \ge \Psi_n\left(\varphi_0, \sigma^2\right) \text{ for some } \varphi \in M_n(A)\right\} \\ &\quad\le P\left\{(\varphi - \varphi_0)^T S_n(\varphi_0) \ge Q_n(\varphi), Q_n(\varphi) > cA^2 \text{ for some } \varphi \in M_n(A)\right\} \\ &\qquad + P\left\{Q_n(\varphi) \le cA^2 \text{ for some } \varphi \in M_n(A)\right\} \\ &\quad\le P\left\{v_n^T(\varphi) D_n^{-1/2} S_n(\varphi_0) > cA \text{ for some } \varphi \in M_n(A)\right\} \\ &\qquad + P\left\{v_n^T(\varphi) D_n^{-1/2} F_n(\tilde{\varphi}) D_n^{-T/2} v_n(\varphi) \le c \text{ for some } \varphi \in M_n(A)\right\} \\ &\quad\le P\left\{\left|D_n^{-1/2} S_n(\varphi_0)\right| > cA\right\} + P\left\{\inf_{\varphi \in N_n(A)} \lambda_{\min}\left(D_n^{-1/2} F_n(\tilde{\varphi}) D_n^{-T/2}\right) \le c\right\}.\end{aligned} \tag{5.3}$$

By Lemma 4.1 and Chebychev inequality, we obtain

$$P\left\{\left|D_n^{-1/2}S_n(\varphi_0)\right| > cA\right\} \leq \frac{\operatorname{Var}\left(D_n^{-1/2}S_n(\varphi_0)\right)}{c^2A^2} = \frac{\sigma_0^2}{c^2A^2}. \tag{5.4}$$

Let $A \to \infty$, then $c \downarrow 0$, and using (4.17), we have

$$P\left\{\inf_{\varphi\in N_n(A)} \lambda_{\min}\left(D_n^{-1/2}F_n(\tilde{\varphi})D_n^{-T/2}\right) \leq c\right\} \longrightarrow 0. \tag{5.5}$$

By (5.3)–(5.5), we have

$$\lim_{A\to\infty}\liminf_{n\to\infty} P\left\{\Psi_n\left(\varphi,\sigma^2\right) < \Psi_n\left(\varphi_0,\sigma^2\right)\ \forall\varphi \in M_n(A)\right\} = 1. \tag{5.6}$$

By Lemma 4.3, $\lambda_{\min}(X_n(\theta_0)) \to \infty$ as $n \to \infty$. Hence $\lambda_{\min}(D_n) \to \infty$. Moreover, from (4.17), we have

$$\inf_{\varphi\in N_n(A)} \lambda_{\min}(F_n(\varphi)) \longrightarrow_p \infty. \tag{5.7}$$

This implies that $\Psi_n(\varphi,\sigma^2)$ is concave on $N_n(A)$. Noting this fact and (5.6), we get

$$\lim_{A\to\infty}\liminf_{n\to\infty} P\left\{\sup_{\varphi\in M_n(A)} \Psi_n\left(\varphi,\sigma^2\right) < \Psi_n\left(\varphi_0,\sigma^2\right), \Psi_n\left(\varphi,\sigma^2\right) \text{ is concave on } N_n(A)\right\} = 1. \tag{5.8}$$

On the event in the brackets, the continuous function $\Psi_n(\varphi,\sigma^2)$ has a unique maximum in φ over the compact neighborhood $N_n(A)$. Hence

$$\lim_{A\to\infty}\liminf_{n\to\infty} P\{S_n(\hat{\varphi}_n(A)) = 0 \text{ for a unique } \hat{\varphi}_n(A) \in N_n(A)\} = 1. \tag{5.9}$$

Moreover, there is a sequence $A_n \to \infty$ such that $\hat{\varphi}_n = \hat{\varphi}(A_n)$ satisfies

$$\liminf_{n\to\infty} P\left\{S_n(\hat{\varphi}_n) = 0 \text{ and } \hat{\varphi}_n \text{ maximizes } \Psi_n\left(\varphi,\sigma^2\right) \text{ uniquely in } N_n(A)\right\} = 1. \tag{5.10}$$

Thus the $\hat{\varphi}_n = (\hat{\beta}_n,\hat{\theta}_n)$ is a QML estimator for φ_0. It is clearly consistent, and

$$\lim_{A\to\infty}\liminf_{n\to\infty} P\{\hat{\varphi}_n \in N_n(A)\} = 1. \tag{5.11}$$

Since $\hat{\varphi}_n = (\hat{\beta}_n,\hat{\theta}_n)$ is a QML estimator for φ_0, $\hat{\sigma}_n^2$ is a QML estimator for σ_0^2 from (2.9).

To complete the proof, we will show that $\widehat{\sigma}_n^2 \to \sigma_0^2$ as $n \to \infty$. If $\widehat{\varphi}_n \in N_n(A)$, then $\widehat{\beta}_n \in N_n^{\beta}(A)$ and $\widehat{\theta}_n \in N_n^{\theta}(A)$. By (2.12) and (2.1), we have

$$\widehat{\varepsilon}_t - f_t\left(\widehat{\theta}_n\right)\widehat{\varepsilon}_{t-1} = \left(x_t - f_t\left(\widehat{\theta}_n\right)x_{t-1}\right)^T\left(\beta_0 - \widehat{\beta}_n\right) + \left(e_t - f_t\left(\widehat{\theta}_n e_{t-1}\right)\right). \tag{5.12}$$

By (2.9), (2.11), and (5.12), we have

$$\begin{aligned}
(n-1)\widehat{\sigma}_n^2 &= \sum_{t=2}^{n}\left(\widehat{\varepsilon}_t - f_t\left(\widehat{\theta}_n\right)\widehat{\varepsilon}_{t-1}\right)^2 \\
&= \sum_{t=2}^{n}\left(\widehat{\varepsilon}_t - f_t\left(\widehat{\theta}_n\right)\widehat{\varepsilon}_{t-1}\right)\left\{\left(x_t - f_t\left(\widehat{\theta}_n\right)x_{t-1}\right)^T\left(\beta_0 - \widehat{\beta}_n\right) + \left(e_t - f_t\left(\widehat{\theta}_n\right)e_{t-1}\right)\right\} \\
&= \sum_{t=2}^{n}\left(\widehat{\varepsilon}_t - f_t\left(\widehat{\theta}_n\right)\widehat{\varepsilon}_{t-1}\right)\left(x_t - f_t\left(\widehat{\theta}_n\right)x_{t-1}\right)^T\left(\beta_0 - \widehat{\beta}_n\right) \\
&\quad + \sum_{t=2}^{n}\left(\widehat{\varepsilon}_t - f_t\left(\widehat{\theta}_n\right)\widehat{\varepsilon}_{t-1}\right)\left(e_t - f_t\left(\widehat{\theta}_n\right)e_{t-1}\right) \\
&= \sum_{t=2}^{n}\left(\widehat{\varepsilon}_t - f_t\left(\widehat{\theta}_n\right)\widehat{\varepsilon}_{t-1}\right)\left(e_t - f_t\left(\widehat{\theta}_n\right)e_{t-1}\right).
\end{aligned} \tag{5.13}$$

From (5.12), it follows that

$$\begin{aligned}
\sum_{t=2}^{n}\left\{\left(x_t - f_t\left(\widehat{\theta}_n\right)x_{t-1}\right)^T\left(\beta_0 - \widehat{\beta}_n\right)\right\}^2 &= \sum_{t=2}^{n}\left(\widehat{\varepsilon}_t - f_t\left(\widehat{\theta}_n\right)\widehat{\varepsilon}_{t-1}\right)^2 \\
&\quad - 2\sum_{t=2}^{n} f_t(\widehat{\theta}_n)x_{t-1})^{\ T}\left(\beta_0 - \widehat{\beta}_n\right)\left(e_t - f_t\left(\widehat{\theta}_n\right)e_{t-1}\right) \\
&\quad + \sum_{t=2}^{n}\left(e_t - f_t\left(\widehat{\theta}_n\right)e_{t-1}\right)^2.
\end{aligned} \tag{5.14}$$

From (2.2),

$$\begin{aligned}
\sum_{t=2}^{n}\left(e_t - f_t\left(\widehat{\theta}_n\right)e_{t-1}\right)^2 &= \sum_{t=2}^{n}\left(f_t(\theta_0)e_{t-1} - f_t\left(\widehat{\theta}_n\right)e_{t-1}\right)^2 \\
&= \sum_{t=2}^{n}\eta_t^2 + 2\sum_{t=2}^{n}\left(f_t(\theta_0) - f_t\left(\widehat{\theta}_n\right)\right)\eta_t e_{t-1} \\
&\quad + \sum_{t=2}^{n}\left(f_t(\theta_0) - f_t\left(\widehat{\theta}_n\right)\right)^2 e_{t-1}^2.
\end{aligned} \tag{5.15}$$

By (5.13)–(5.15), we have

$$\begin{aligned}(n-1)\widehat{\sigma}_n^2 &= \sum_{t=2}^{n}\left(e_t - f_t\left(\widehat{\theta}_n\right)e_{t-1}\right)^2 - \sum_{t=2}^{n}\left(\left(x_t - f_t\left(\widehat{\theta}_n\right)x_{t-1}\right)^T\left(\beta_0-\widehat{\beta}_n\right)\right)^2 \\ &= \sum_{t=2}^{n}\eta_t^2 + 2\sum_{t=2}^{n}\left(f_t(\theta_0) - f_t\left(\widehat{\theta}_n\right)\right)\eta_t e_{t-1} + \sum_{t=2}^{n}\left(f_t(\theta_0) - f_t\left(\widehat{\theta}_n\right)\right)^2 e_{t-1}^2 \\ &\quad - \sum_{t=2}^{n}\left(\left(x_t - f_t\left(\widehat{\theta}_n\right)x_{t-1}\right)^T\left(\beta_0-\widehat{\beta}_n\right)\right)^2 \\ &= T_1 + 2T_2 + T_3 - T_4.\end{aligned} \tag{5.16}$$

By the law of large numbers,

$$\frac{1}{n-1}T_1 = \frac{1}{n-1}\sum_{t=2}^{n}\eta_t^2 \longrightarrow \sigma_0^2, \quad \text{a.s. } (n\longrightarrow\infty). \tag{5.17}$$

Since $\{(f_t(\theta_0) - f_t(\widehat{\theta}_n))\eta_t e_{t-1}, H_{t-1}\}$ is a martingale difference sequence with

$$\begin{aligned}\operatorname{Var}\left\{\left(f_t(\theta_0) - f_t\left(\widehat{\theta}_n\right)\right)\eta_t e_{t-1}\right\} &= \left(f_t(\theta_0) - f_t\left(\widehat{\theta}_n\right)\right)^2\sigma_0^2 E e_{t-1}^2, \\ \operatorname{Var}(T_2) &= \sum_{t=2}^{n}E\left(\left(f_t(\theta_0) - f_t\left(\widehat{\theta}_n\right)\right)\eta_t e_{t-1}\right)^2 \\ &= \sigma_0^2\sum_{t=2}^{n}\left(f_t(\theta_0) - f_t\left(\widehat{\theta}_n\right)\right)^2 E e_{t-1}^2 \\ &= \sigma_0^2\sum_{t=2}^{n}\frac{f_t'^2\left(\widetilde{\theta}\right)}{f_t'^2(\theta_0)}\left(\theta_0-\widehat{\theta}_n\right)^2 f_t'^2(\theta_0) E e_{t-1}^2 \\ &\le \sigma_0^2 \max_{2\le t\le n}\left|\frac{f_t'(\widetilde{\theta})}{f_t'(\theta_0)}\right|^2 \left|\left(\theta_0-\widehat{\theta}_n\right)\right|^2 \Delta_n\left(\theta_0,\sigma_0^2\right) \\ &\le CA^2.\end{aligned} \tag{5.18}$$

By Chebychev inequality, we have

$$\frac{1}{n-1}T_2 \longrightarrow_p 0, \quad (n\longrightarrow\infty). \tag{5.19}$$

By Markov inequality and noting that $ET_3 \le CA^2$, we obtain that

$$\frac{1}{n-1}T_3 \longrightarrow_p 0, \quad (n\longrightarrow\infty). \tag{5.20}$$

Write

$$\begin{aligned} T_4 &= \sum_{t=2}^{n}\left(\left(x_t - f_t(\theta_0)x_{t-1}\right)^T\left(\beta_0 - \widehat{\beta}_n\right) + \left(f_t(\theta_0) - f_t\left(\widehat{\theta}_n\right)\right)x_{t-1}^T\left(\beta_0 - \widehat{\beta}_n\right)\right)^2 \\ &= \left(\beta_0 - \widehat{\beta}_n\right)^T X_n(\theta_0)\left(\beta_0 - \widehat{\beta}_n\right) + \sum_{t=2}^{n}\left(f_t(\theta_0) - f_t\left(\widehat{\theta}_n\right)\right)^2\left(x_{t-1}^T\left(\beta_0 - \widehat{\beta}_n\right)\right)^2 \\ &\quad + 2\sum_{t=2}^{n}\left(x_t - f_t(\theta_0)x_{t-1}\right)^T\left(\beta_0 - \widehat{\beta}_n\right)\left(f_t(\theta_0) - f_t\left(\widehat{\theta}_n\right)\right)x_{t-1}^T\left(\beta_0 - \widehat{\beta}_n\right) \\ &= I_1 + I_2 + 2I_3. \end{aligned} \tag{5.21}$$

Noting that $\beta \in N_n^{\beta}(A)$, we have

$$I_1 = O_p(1), \quad (n \longrightarrow \infty). \tag{5.22}$$

By (4.34) and condition (A3), we have

$$|I_2| \leq \max_{2\leq t\leq n}\left|x_{t-1}^T(\beta_0 - \widehat{\beta}_n)\right|^2 n \max_{2\leq t\leq n}\left|f_t'(\widetilde{\theta})\right|^2\left|\theta_0 - \widehat{\theta}_n\right|^2 = o(1), \quad (n \longrightarrow \infty). \tag{5.23}$$

By (4.34), condition (A3), and Cauchy-Schwartz inequality, we have

$$\begin{aligned} |I_3|^2 &\leq \sum_{t=2}^{n}\left(\beta_0 - \widehat{\beta}_n\right)^T\left(x_t - f_t(\theta_0)x_{t-1}\right)^T\left(x_t - f_t(\theta_0)x_{t-1}\right)\left(\beta_0 - \widehat{\beta}_n\right) \\ &\quad \cdot \sum_{t=2}^{n}\left(f_t(\theta_0) - f_t\left(\widehat{\theta}_n\right)\right)^2\left(x_{t-1}^T\left(\beta_0 - \widehat{\beta}_n\right)\right)^2 \\ &\leq \left(\beta_0 - \widehat{\beta}_n\right)^T X_n(\theta_0)\left(\beta_0 - \widehat{\beta}_n\right) n \max_{2\leq t\leq n}\left|f_t'(\widetilde{\theta})\right|^2\left|\theta_0 - \widehat{\theta}_n\right|^2 \\ &\quad \cdot \max_{2\leq t\leq n}\left|x_{t-1}^T(\beta_0 - \widehat{\beta}_n)\right|^2 \\ &\leq O_p(1) n \frac{A^2}{\Delta_n(\theta_0, \sigma_0^2)} o(1) = o_p(1). \end{aligned} \tag{5.24}$$

By (5.21)–(5.24), we obtain

$$\frac{1}{n-1} T_4 \longrightarrow_p 0, \quad (n \longrightarrow \infty). \tag{5.25}$$

From (5.16), (5.17), (5.19), (5.20), and (5.25), we have $\widehat{\sigma}_n^2 \rightarrow \sigma_0^2$. We therefore complete the proof of Theorem 3.1.

5.2. Proof of Theorem 3.2

It is easy to know that $S_n(\widehat{\varphi}_n) = 0$ and $F_n(\widehat{\varphi}_n)$ is nonsingular from Theorem 3.1. By Taylor's expansion, we have

$$0 = S_n(\widehat{\varphi}_n) = S_n(\varphi_0) - F_n(\widetilde{\varphi}_n)(\widehat{\varphi}_n - \varphi_0). \tag{5.26}$$

Since $\widehat{\varphi}_n \in N_n(A)$, also $\widetilde{\varphi}_n \in N_n(A)$. By (4.15), we have

$$F_n(\widetilde{\varphi}_n) = D_n^{1/2}\left(\Phi_n + \widetilde{A}_n\right)D_n^{T/2}, \tag{5.27}$$

where $\widetilde{A}_n$ is a symmetric matrix with $\widetilde{A}_n \to_p 0$. By (5.26) and (5.27), we have

$$D_n^{T/2}(\widehat{\varphi}_n - \varphi_0) = D_n^{T/2}F_n^{-1}(\widetilde{\varphi}_n)S_n(\varphi_0) = \left(\Phi_n + \widetilde{A}_n\right)^{-1}D_n^{-1/2}S_n(\varphi_0). \tag{5.28}$$

Similar to (5.27), we have

$$\begin{aligned} F_n(\widehat{\varphi}_n) &= D_n^{1/2}\left(\Phi_n + \widehat{A}_n\right)D_n^{T/2} = \left(D_n^{1/2}\left(\Phi_n + \widehat{A}_n\right)^{1/2}\right)\left(\left(\Phi_n + \widehat{A}_n\right)^{T/2}D_n^{T/2}\right) \\ &= F_n^{1/2}(\widehat{\varphi}_n)F_n^{T/2}(\widehat{\varphi}_n). \end{aligned} \tag{5.29}$$

Here $\widehat{A}_n \to_p 0$. By (5.28), (5.29), and noting that $\widehat{\sigma}_n^2 \to_p \sigma_0^2$ and $D_n^{-1/2}S_n(\varphi_0) = O_p(1)$, we obtain that

$$\begin{aligned} \frac{F_n^{T/2}(\widehat{\varphi}_n)(\widehat{\varphi}_n - \varphi_0)}{\widehat{\sigma}_n} &= \frac{\left(\Phi_n + \widehat{A}_n\right)^{1/2}\left(\Phi_n + \widetilde{A}_n\right)^{-1}D_n^{-1/2}S_n(\varphi_0)}{\widehat{\sigma}_n} \\ &= \frac{\Phi_n^{-1/2}D_n^{-1/2}S_n(\varphi_0)}{\sigma_0} + o_p(1). \end{aligned} \tag{5.30}$$

From (2.7) and (2.8), we have

$$\frac{S_n(\varphi_0)}{\sigma_0} = \frac{1}{\sigma_0}\left(\sum_{t=2}^{n}\eta_t(x_t - f_t(\theta_0)x_{t-1}), \sum_{t=2}^{n}f_t'(\theta_0)\eta_t e_{t-1}\right). \tag{5.31}$$

From (2.29) and (4.18), we have

$$\Phi_n^{-1/2} D_n^{-1/2} = \begin{pmatrix} I_d & 0 \\ 0 & \sqrt{\dfrac{\Delta_n(\theta_0,\sigma_0)}{\sum_{t=2}^n f_t'^2(\theta_0) e_{t-1}^2}} \end{pmatrix} \begin{pmatrix} X_n^{-1/2}(\theta_0) & 0 \\ 0 & \dfrac{1}{\sqrt{\Delta_n(\theta_0,\sigma_0)}} \end{pmatrix} = \begin{pmatrix} X_n^{-1/2}(\theta_0) & 0 \\ 0 & \dfrac{1}{\sqrt{\sum_{t=2}^n f_t'^2(\theta_0) e_{t-1}^2}} \end{pmatrix}. \tag{5.32}$$

By (5.30)–(5.32), we have

$$\frac{\Phi_n^{-1/2} D_n^{-1/2} S_n(\varphi_0)}{\sigma_0} = \frac{1}{\sigma_0}\left(\sum_{t=2}^n \eta_t X_n^{-1/2}(\theta_0)(x_t - f_t(\theta_0)x_{t-1}), \frac{\sum_{t=2}^n f_t'(\theta_0)\eta_t e_{t-1}}{\sqrt{\sum_{t=2}^n f_t'^2(\theta_0) e_{t-1}^2}} \right). \tag{5.33}$$

Let $u \in R^d$ with $|u| = 1$, and $a_{tn} = u X_n^{-1/2}(\theta_0)(x_t - f_t(\theta_0)x_{t-1})$. Then $\max_{2\le t\le n} a_{tn} = o(1)$, and we will consider the limiting distribution of the following 2-vector:

$$\frac{1}{\sigma_0}\left(\sum_{t=2}^n a_{tn}\eta_t, \frac{\sum_{t=2}^n f_t'(\theta_0)\eta_t e_{t-1}}{\sqrt{\sum_{t=2}^n f_t'^2(\theta_0) e_{t-1}^2}} \right). \tag{5.34}$$

By Cramer-Wold device, it will suffice to find the asymptotic distribution of the following random:

$$\sum_{t=2}^n \eta_t \left(\frac{u_1 a_{tn}}{\sigma_0} + \frac{u_2 f_t'(\theta_0) e_{t-1}}{\sigma_0 \sqrt{\Delta_n(\theta_0,\sigma_0)}} \right) = \sum_{t=2}^n \eta_t m_t(n), \tag{5.35}$$

where $(u_1, u_2) \in R^2$ with $u_1^2 + u_2^2 = 1$. Note that $E\{\eta_t m_t(n) \mid H_{t-1}\} = 0$, so the sums in (5.35) are partial sums of a martingale triangular array with respect to $\{H_t\}$, and we will verify the Lindeberg conditions for their convergence to normality as follows:

$$\begin{aligned} \sum_{t=2}^n E\left(\eta_t^2 m_t^2(n) \mid H_{t-1}\right) &= \sigma_0^2 \sum_{t=2}^n m_t^2(n) \\ &= u_1^2 \sum_{t=2}^n a_{tn}^2 + \frac{2u_1u_2}{\sqrt{\Delta_n(\theta_0,\sigma_0)}} \sum_{t=2}^n f_t'(\theta_0) a_{tn} e_{t-1} \\ &\quad + \frac{u_2^2}{\Delta_n(\theta_0,\sigma_0)} \sum_{t=2}^n f_t'^2(\theta_0) e_{t-1}^2 \\ &= u_1^2 + o_p(1) + u_2^2 = 1 + o_p(1). \end{aligned} \tag{5.36}$$

Noting that $\max_{1\le t\le n}(e_t^2/n) = o_p(1)$ and $\max_{2\le t\le n} a_{tn} = o(1)$, we obtain that

$$\max_{2\le t\le n}|m_t(n)| \le \max_{2\le t\le n}\left|\frac{u_1 a_{tn}}{\sigma_0}\right| + \max_{2\le t\le n}\left|\frac{u_2 f_t'(\theta_0)e_{t-1}}{\sigma_0\sqrt{\Delta_n(\theta_0,\sigma_0)}}\right| = o_p(1). \tag{5.37}$$

Hence, for given $\delta > 0$, there is a set whose probability approaches 1 as $n \to \infty$ on which $\max_{2\le t\le n}|m_t(n)| \le \delta$. On this event, for any $c > 0$,

$$\begin{aligned}\sum_{t=2}^{n} E\left\{\eta_t^2 m_t^2(n) I(|\eta_t m_t(n)| > c) \mid H_{t-1}\right\} &= \sum_{t=2}^{\infty}\int_c^{\infty} y^2 dp\{|\eta_t m_t(n)| \le y \mid H_{t-1}\}\\ &= \sum_{t=2}^{n} m_t^2(n)\int_{c/m_t(n)}^{\infty} y^2 dp\{|\eta_1| \le y \mid H_{t-1}\}\\ &\le \sum_{t=2}^{n} m_t^2(n)\int_{c/\delta}^{\infty} y^2 dp\{|\eta_1| \le y \mid H_{t-1}\}\\ &= o_\delta \sum_{t=2}^{n} m_t^2(n) = o_\delta O_p(1) \longrightarrow 0, \quad n \longrightarrow \infty.\end{aligned} \tag{5.38}$$

Here $o_\delta \to 0$ as $\delta \to 0$. This verifies the Lindeberg conditions, and by Lemma 4.5

$$\sum_{t=2}^{n} \eta_t m_t(n) \longrightarrow_D N(0,1). \tag{5.39}$$

Thus we complete the proof of Theorem 3.2.

6. Numerical Examples

6.1. Simulation Example

We will simulate a regression model (1.1), where $x_t = t/20$, $\beta = 3.5$, $t = 1, 2, \ldots, 100$, and the random errors

$$\varepsilon_t = \frac{1}{2}\sin\left(\frac{\pi}{2}t + \theta\right)\varepsilon_{t-1} + \eta_t, \tag{6.1}$$

where $\theta = 2, \eta_t \sim N(0,1)$.

By the ordinary least squares method, we obtain the least squares estimators $\hat{\beta}_{LS} = 3.5136$, and $\hat{\sigma}_{LS}^2 = 1.0347$. So we take $\beta^{(0)} = 3.5136, \sigma^{(0)2} = 1.0347$, $\theta^{(0)} = \pi/2$, and $\delta = 0.01$. Therefore, using the iterative computing method, we obtain

$$\vec{\theta}^{(1)} = (1.0159, 3.5076, 1.6573)^T, \qquad \vec{\theta}^{(2)} = (1.0283, 3.5076, 1.7599)^T. \tag{6.2}$$

Since $\|\overrightarrow{\theta}^{(2)} - \overrightarrow{\theta}^{(1)}\| / \|\overrightarrow{\theta}^{(1)}\| < 0.01$, the QML estimators of β, θ, and σ^2 are given by

$$\widehat{\sigma}_n^2 = 1.0283, \qquad \widehat{\beta}_n = 3.5076, \qquad \widehat{\theta}_n = 1.7599. \tag{6.3}$$

These values closely approximate their true values, so our method is successful, especially in estimating the parameters β and σ^2.

6.2. Empirical Example

We will use the data studied by Fuller in [14]. The data pertain to the consumption of spirits in the United Kingdom from 1870 to 1983. The dependent variable y_t is the annual per capita consumption of spirits in the United Kingdom. The explanatory variables x_{t1} and x_{t2} are per capita income and price of spirits, respectively, both deflated by a general price index. All data are in logarithms. The model suggested by Prest can be written as

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \varepsilon_t, \tag{6.4}$$

where 1869 is the origin for t, $x_{t3} = t/100$, and $x_{t4} = (t-35)^2/10^4$, and assuming that ε_t is a stationary time series.

Fuller [14] obtained the estimated generalized least squares equation

$$\begin{gathered} \widehat{y}_t = 2.36 + 0.72 x_{t1} - 0.80 x_{t2} - 0.81 x_{t3} - 0.92 x_{t4}, \\ \varepsilon_t = 0.7633 \varepsilon_{t-1} + \eta_t, \end{gathered} \tag{6.5}$$

where η_t is a sequence of uncorrelated $(0, 0.000417)$ random variables.

Using our method, we obtain the following models:

$$\begin{gathered} \widehat{y}_t = 2.3607 + 0.7437 x_{t1} - 0.8210 x_{t2} - 0.7857 x_{t3} - 0.9178 x_{t4}, \\ \varepsilon_t = (0.70054 + 0.00424t) \varepsilon_{t-1} + \eta_t, \end{gathered} \tag{6.6}$$

where η_t is a sequence of uncorrelated $(0, 0.000413)$ random variables.

By the models (6.6), the residual mean square is 0.000413, which is smaller than 0.000417 calculated by the models (6.5).

From the above examples, it can be seen that our method is successful and valid. However, a further discussion of fitting the function $f_t(\theta)$ is needed so that we can find a good method to use in practical applications.

Acknowledgments

The author would like to thank an anonymous referee for useful comments and suggestions. The paper was supported by the Key Project of Chinese Ministry of Education (no. 209078) and Scientific Research Item of Department of Education, Hubei (no. D20092207).

References

[1] Z. D. Bai and M. Guo, "A paradox in least-squares estimation of linear regression models," *Statistics & Probability Letters*, vol. 42, no. 2, pp. 167–174, 1999.

[2] X. Chen, "Consistency of LS estimates of multiple regression under a lower order moment condition," *Science in China A*, vol. 38, no. 12, pp. 1420–1431, 1995.

[3] T. W. Anderson and J. B. Taylor, "Strong consistency of least squares estimates in normal linear regression," *The Annals of Statistics*, vol. 4, no. 4, pp. 788–790, 1976.

[4] H. Drygas, "Weak and strong consistency of the least squares estimators in regression models," *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 34, no. 2, pp. 119–127, 1976.

[5] G. González-Rodríguez, A. Blanco, N. Corral, and A. Colubi, "Least squares estimation of linear regression models for convex compact random sets," *Advances in Data Analysis and Classification*, vol. 1, no. 1, pp. 67–81, 2007.

[6] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, New York, NY, USA, 1986.

[7] X. He, "A local breakdown property of robust tests in linear regression," *Journal of Multivariate Analysis*, vol. 38, no. 2, pp. 294–305, 1991.

[8] H. Cui, "On asymptotics of t-type regression estimation in multiple linear model," *Science in China A*, vol. 47, no. 4, pp. 628–639, 2004.

[9] J. Durbin, "A note on regression when there is extraneous information about one of the coefficients," *Journal of the American Statistical Association*, vol. 48, pp. 799–808, 1953.

[10] A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[11] Y. Li and H. Yang, "A new stochastic mixed ridge estimator in linear regression model," *Statistical Papers*. In press.

[12] R. A. Maller, "Asymptotics of regressions with stationary and nonstationary residuals," *Stochastic Processes and Their Applications*, vol. 105, no. 1, pp. 33–67, 2003.

[13] P. Pere, "Adjusted estimates and Wald statistics for the AR(1) model with constant," *Journal of Econometrics*, vol. 98, no. 2, pp. 335–363, 2000.

[14] W. A. Fuller, *Introduction to Statistical Time Series*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, New York, NY, USA, 2nd edition, 1996.

[15] G. H. Kwoun and Y. Yajima, "On an autoregressive model with time-dependent coefficients," *Annals of the Institute of Statistical Mathematics*, vol. 38, no. 2, pp. 297–309, 1986.

[16] J. S. White, "The limiting distribution of the serial correlation coefficient in the explosive case," *Annals of Mathematical Statistics*, vol. 29, pp. 1188–1197, 1958.

[17] J. S. White, "The limiting distribution of the serial correlation coefficient in the explosive case. II," *Annals of Mathematical Statistics*, vol. 30, pp. 831–834, 1959.

[18] J. D. Hamilton, *Time Series Analysis*, Princeton University Press, Princeton, NJ, USA, 1994.

[19] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer Series in Statistics, Springer, New York, NY, USA, 1987.

[20] K. M. Abadir and A. Lucas, "A comparison of minimum MSE and maximum power for the nearly integrated non-Gaussian model," *Journal of Econometrics*, vol. 119, no. 1, pp. 45–71, 2004.

[21] F. Carsoule and P. H. Franses, "A note on monitoring time-varying parameters in an autoregression," *Metrika*, vol. 57, no. 1, pp. 51–62, 2003.

[22] R. Azrak and G. Mélard, "Asymptotic properties of quasi-maximum likelihood estimators for ARMA models with time-dependent coefficients," *Statistical Inference for Stochastic Processes*, vol. 9, no. 3, pp. 279–330, 2006.

[23] R. Dahlhaus, "Fitting time series models to nonstationary processes," *The Annals of Statistics*, vol. 25, no. 1, pp. 1–37, 1997.

[24] W. Zhang, L. Wei, and Y. Yang, "The superiority of empirical Bayes estimator of parameters in linear model," *Statistics and Probability Letters*, vol. 72, no. 1, pp. 43–50, 2005.

[25] E. J. Hannan, "The asymptotic theory of linear time-series models," *Journal of Applied Probability*, vol. 10, pp. 130–145, 1973.

[26] R. Fox and M. S. Taqqu, "Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series," *The Annals of Statistics*, vol. 14, no. 2, pp. 517–532, 1986.

[27] L. Giraitis and D. Surgailis, "A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate," *Probability Theory and Related Fields*, vol. 86, no. 1, pp. 87–104, 1990.
[28] L. Giraitis and H. Koul, "Estimation of the dependence parameter in linear regression with long-range-dependent errors," *Stochastic Processes and Their Applications*, vol. 71, no. 2, pp. 207–224, 1997.
[29] H. L. Koul, "Estimation of the dependence parameter in non-linear regression with random designs and long memory errors," in *Proceedings of the Triennial Conference on Nonparametric Inference and Related Fields*, Calcutta University, 1997.
[30] H. L. Koul and D. Surgailis, "Asymptotic normality of the Whittle estimator in linear regression models with long memory errors," *Statistical Inference for Stochastic Processes*, vol. 3, no. 1-2, pp. 129–147, 2000.
[31] H. L. Koul and K. Mukherjee, "Asymptotics of R-, MD- and LAD-estimators in linear regression models with long range dependent errors," *Probability Theory and Related Fields*, vol. 95, no. 4, pp. 535–553, 1993.
[32] T. Shiohama and M. Taniguchi, "Sequential estimation for time series regression models," *Journal of Statistical Planning and Inference*, vol. 123, no. 2, pp. 295–312, 2004.
[33] J. Fan and Q. Yao, *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer, New York, NY, USA, 2005.
[34] K. N. Berk, "Consistent autoregressive spectral estimates," *The Annals of Statistics*, vol. 2, pp. 489–502, 1974.
[35] E. J. Hannan and L. Kavalieris, "Regression, autoregression models," *Journal of Time Series Analysis*, vol. 7, no. 1, pp. 27–49, 1986.
[36] A. Goldenshluger and A. Zeevi, "Nonasymptotic bounds for autoregressive time series modeling," *The Annals of Statistics*, vol. 29, no. 2, pp. 417–444, 2001.
[37] E. Liebscher, "Strong convergence of estimators in nonlinear autoregressive models," *Journal of Multivariate Analysis*, vol. 84, no. 2, pp. 247–261, 2003.
[38] H.-Z. An, L.-X. Zhu, and R.-Z. Li, "A mixed-type test for linearity in time series," *Journal of Statistical Planning and Inference*, vol. 88, no. 2, pp. 339–353, 2000.
[39] R. Elsebach, "Evaluation of forecasts in AR models with outliers," *OR Spectrum*, vol. 16, no. 1, pp. 41–45, 1994.
[40] S. Baran, G. Pap, and M. C. A. van Zuijlen, "Asymptotic inference for unit roots in spatial triangular autoregression," *Acta Applicandae Mathematicae*, vol. 96, no. 1–3, pp. 17–42, 2007.
[41] W. Distaso, "Testing for unit root processes in random coefficient autoregressive models," *Journal of Econometrics*, vol. 142, no. 1, pp. 581–609, 2008.
[42] J. L. Harvill and B. K. Ray, "Functional coefficient autoregressive models for vector time series," *Computational Statistics and Data Analysis*, vol. 50, no. 12, pp. 3547–3566, 2006.
[43] J. S. Simonoff, *Smoothing Methods in Statistics*, Springer Series in Statistics, Springer, New York, NY, USA, 1996.
[44] P. J. Green and B. W. Silverman, *Nonparametric Regression and Generalized Linear Models*, vol. 58 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London, UK, 1994.
[45] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*, vol. 66 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London, UK, 1996.
[46] Y. Fujikoshi and Y. Ochi, "Asymptotic properties of the maximum likelihood estimate in the first order autoregressive process," *Annals of the Institute of Statistical Mathematics*, vol. 36, no. 1, pp. 119–128, 1984.
[47] C. R. Rao, *Linear Statistical Inference and Its Applications*, Wiley Series in Probability and Mathematical Statistic, John Wiley & Sons, New York, NY, USA, 2nd edition, 1973.
[48] R. A. Maller, "Quadratic negligibility and the asymptotic normality of operator normed sums," *Journal of Multivariate Analysis*, vol. 44, no. 2, pp. 191–219, 1993.
[49] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application. Probability and Mathematical Statistic*, Academic Press, New York, NY, USA, 1980.