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

Copyright © 2010 Hongchang Hu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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:

\[ y_t = x_t^T \hat{\beta} + \varepsilon_t, \quad t = 1, 2, \ldots, n, \quad (1.1) \]

where \( y_t \)'s are scalar response variables, \( x_t \)'s are explanatory variables, \( \hat{\beta} \) is a \( d \)-dimensional unknown parameter, and the \( \varepsilon_t \)'s are functional coefficient autoregressive processes given as

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

where \( \eta_t \)'s are independent and identically distributed random errors with zero mean and finite variance \( \sigma^2 \), \( \theta \) is a one-dimensional unknown parameter and \( f_t(\theta) \) is a real valued function defined on a compact set \( \Theta \) which contains the true value \( \theta_0 \) as an inner point and is a subset of \( \mathbb{R}^1 \). The values of \( \theta_0 \) and \( \sigma^2 \) are unknown.
Model (1.1) includes many special cases, such as an ordinary linear regression model when $f_1(\theta) \equiv 0$; see [1–11]. In the sequel, we always assume that $f_i(\theta) \neq 0$, for some $\theta \in \Theta$, is a linear regression model with constant coefficient autoregressive processes (when $f_i(\theta) = \theta$, 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 $f_i(\theta) = \theta$ 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 $f_i(\theta) = a_i$ and $\beta = 0$, see Carsoule and Franses [21], Azrak and Melard [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-Rodriguez 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 Kou in [28] and Kou in [29] when the errors are nonlinear instantaneous functions of a Gaussian long-memory sequence. Kou 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, Kou 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 \( \sigma^2 \) 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

\[
y_t = x_t^T \beta_0 + e_t, \quad t = 1, 2, \ldots, n, \\
e_1 = \eta_1, \quad e_t = f_t(\theta_0)e_{t-1} + \eta_t, \quad t = 2, 3, \ldots, n,
\]

(2.1)

(2.2)

where \( f_t'(\theta_0) = (df_t(\theta)/d\theta)|_{\theta=\theta_0} \neq 0 \), and \( \eta_t \)'s are i.i.d errors with zero mean and finite variance \( \sigma_0^2 \). Define \( \prod_{i=0}^{-1} f_{t-i}(\theta_0) = 1 \), and by (2.2) we have

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

(2.3)

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

\[
Ee_t = 0, \quad \text{Var}(e_t) = \sigma^2 \sum_{j=0}^{t-1} \left( \prod_{i=0}^{j-1} f_{t-i}(\theta_0) \right).
\]

(2.4)

Assume at first that the \( \eta_t \)'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 \( y_1 \):

\[
\Psi_n(\beta, \theta, \sigma^2) = \log L_n = -\frac{1}{2} (n - 1) \log \sigma^2 - \frac{1}{2 \sigma^2} \sum_{t=2}^{n} (e_t - f_t(\theta)e_{t-1})^2 - \frac{1}{2} (n - 1) \log(2\pi).
\]

(2.5)

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

\[
\frac{\partial \Psi_n}{\partial \sigma^2} = \frac{n-1}{2\sigma^2} + \frac{1}{2\sigma^2} \sum_{t=2}^{n} (e_t - f_t(\theta)e_{t-1})^2,
\]

(2.6)

\[
\frac{\partial \Psi_n}{\partial \theta} = \frac{1}{\sigma^2} \sum_{t=2}^{n} f_t'(\theta) (e_t - f_t(\theta)e_{t-1}) e_{t-1},
\]

(2.7)

\[
\frac{\partial \Psi_n}{\partial \beta} = \frac{1}{\sigma^2} \sum_{t=2}^{n} (e_t - f_t(\theta)e_{t-1}) (x_t - f_t(\theta)x_{t-1}).
\]

(2.8)
Thus $\hat{\sigma}_n^2, \hat{\beta}_n, \hat{\theta}_n$ satisfy the following estimation equations:

\[
\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{t=2}^{n} (\hat{\epsilon}_t - f_t(\hat{\theta}_n)\hat{\epsilon}_{t-1})^2, \tag{2.9}
\]

\[
\sum_{t=2}^{n} (\hat{\epsilon}_t - f_t(\hat{\theta}_n)\hat{\epsilon}_{t-1}) f_t'(\hat{\theta}_n)\hat{\epsilon}_{t-1} = 0, \tag{2.10}
\]

\[
\sum_{t=2}^{n} (\hat{\epsilon}_t - f_t(\hat{\theta}_n)\hat{\epsilon}_{t-1})(x_t - f_t(\hat{\theta}_n)x_{t-1}) = 0, \tag{2.11}
\]

where

\[
\hat{\epsilon}_t = y_t - x_t^T \hat{\beta}_n. \tag{2.12}
\]

**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 $\bar{\theta}^{(m)} = (\sigma^{(m)}, \beta^{(m)}, \theta^{(m)})$ denote an estimator of $\bar{\theta} = (\sigma, \beta, \theta)$ that has been calculated at the $m$th iteration, and let $A^{(m)}$ denote an estimation of $[H(\bar{\theta}^{(m)})]^{-1}$. The new estimator $\bar{\theta}^{(m+1)}$ is given by

\[
\bar{\theta}^{(m+1)} = \bar{\theta}^{(m)} + sA^{(m)}g(\bar{\theta}^{(m)}) \tag{2.13}
\]

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

\[
g(\bar{\theta}^{(m)}) = \left. \frac{\partial \Psi_n(\bar{\theta})}{\partial \bar{\theta}} \right|_{\bar{\theta}=\bar{\theta}^{(m)}} = \begin{pmatrix}
\frac{\partial \Psi_n}{\partial \sigma^2} |_{\sigma=\sigma^{(m)}} \\
\frac{\partial \Psi_n}{\partial \beta} |_{\beta=\beta^{(m)}} \\
\frac{\partial \Psi_n}{\partial \theta} |_{\theta=\theta^{(m)}}
\end{pmatrix} \tag{2.14}
\]

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

\[
H(\bar{\theta}^{(m)}) = \left. \frac{\partial^2 \Psi_n(\bar{\theta})}{\partial \bar{\theta} \partial \bar{\theta}^T} \right|_{\bar{\theta}=\bar{\theta}^{(m)}} = \begin{pmatrix}
\frac{\partial^2 \Psi_n}{\partial \sigma^2 \partial \sigma} & \frac{\partial^2 \Psi_n}{\partial \sigma^2 \partial \beta} & \frac{\partial^2 \Psi_n}{\partial \sigma^2 \partial \theta} \\
\frac{\partial^2 \Psi_n}{\partial \beta \partial \sigma} & \frac{\partial^2 \Psi_n}{\partial \beta \partial \beta} & \frac{\partial^2 \Psi_n}{\partial \beta \partial \theta} \\
\frac{\partial^2 \Psi_n}{\partial \theta \partial \sigma} & \frac{\partial^2 \Psi_n}{\partial \theta \partial \beta} & \frac{\partial^2 \Psi_n}{\partial \theta^2} & 0
\end{pmatrix} |_{\bar{\theta}=\bar{\theta}^{(m)}}, \tag{2.15}
\]

**Remark 2.2.** The above equations can be simplified for the case of the QML estimators of Maller [12].
where

\[
\frac{\partial^2 \Psi_n}{\partial (\sigma^2)^2} = \frac{n - 1}{2\sigma^4} - \frac{1}{\sigma^2} \sum_{i=2}^{n} (\varepsilon_i - f_i(\theta)\varepsilon_{i-1})^2,
\]

\[
\frac{\partial^2 \Psi_n}{\partial \sigma^2 \partial \beta} = -\frac{1}{\sigma^4} \sum_{i=2}^{n} (\varepsilon_i - f_i(\theta)\varepsilon_{i-1}) (x_i - f_i(\theta)x_{i-1})^T,
\] (2.16)

\[
\frac{\partial^2 \Psi_n}{\partial \sigma^2 \partial \theta} = -\frac{1}{\sigma^2} \sum_{i=2}^{n} (\varepsilon_i - f_i(\theta)\varepsilon_{i-1}) f_i'(\theta),
\]

\[
\frac{\partial^2 \Psi_n}{\partial \beta \partial \beta^T} = -\frac{1}{\sigma^2} \sum_{i=2}^{n} (x_i - f_i(\theta)x_{i-1}) (x_i - f_i(\theta)x_{i-1})^T,
\] (2.17)

\[
\frac{\partial^2 \Psi_n}{\partial \beta \partial \theta} = -\frac{1}{\sigma^2} \sum_{i=2}^{n} (f_i'(\theta)\varepsilon_{i-1}x_i + f_i(\theta)\varepsilon_i x_{i-1} - 2f_i(\theta)f_i'(\theta)x_{i-1}\varepsilon_{i-1}),
\]

\[
\frac{\partial^2 \Psi_n}{\partial \theta^2} = -\frac{1}{\sigma^2} \sum_{i=2}^{n} \left( (f_i''(\theta) + f_i(\theta)f_i''(\theta))\varepsilon_{i-1}^2 - f_i''(\theta)\varepsilon_i \varepsilon_{i-1} \right).
\] (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

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

where

\[
\Delta \overrightarrow{\theta}^{(m+1)} = \overrightarrow{\theta}^{(m+1)} - \overrightarrow{\theta}^{(m)}, \quad \Delta g^{(m+1)} = g(\overrightarrow{\theta}^{(m+1)}) - g(\overrightarrow{\theta}^{(m)}).
\] (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 \( \beta^{(0)}, \sigma^{(0)} \) which are least squares estimators of \( \beta \) and \( \sigma^2 \), respectively. Take \( \theta^{(0)} \) such that \( f_i(\theta^{(0)}) = 0 \). Iterations are stopped if some termination criterion is reached, for example, if

\[
\frac{\| \overrightarrow{\theta}^{(m+1)} - \overrightarrow{\theta}^{(m)} \|}{\| \overrightarrow{\theta}^{(m)} \|} < \delta,
\] (2.21)

for some prechosen small number \( \delta > 0 \).
Up to this point, we obtain the values of QML estimators when the function \( f_i(\theta) = f(t, \theta) \) is known. However, the function \( f_i(\theta) \) is never the case in practice; we have to estimate it. By (2.12) and (1.2), we obtain

\[
\tilde{f}(t, \hat{\theta}_n) = \frac{\hat{\epsilon}_t}{\hat{\epsilon}_{t-1}}, \quad t = 2, 3, \ldots, n. \tag{2.22}
\]

Based on the dataset \( \{\tilde{f}(t, \hat{\theta}_n), t = 2, 3, \ldots, n\} \), we may obtain the estimation function \( \tilde{f}(t, \hat{\theta}_n) \) 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) \( X_n = \sum_{i=2}^{n} x_i x_i^T \) is positive definite for sufficiently large \( n \) and

\[
\lim_{n \to \infty} \max_{1 \leq j \leq n} x_j^T X_n^{-1} x_i = 0, \quad (2.23)
\]

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

where \( Z_n = (1/2) \sum_{i=2}^{n} (x_i x_i^T + x_{i-1} x_{i-1}^T) \) 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

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

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

(A3) The derivatives \( f'_i(\theta) = d f_i(\theta) / d \theta, \ f''_i(\theta) = d f'_i(\theta) / d \theta \) 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 \( \{\varepsilon_t\} \) stable if \( \text{Var} (\varepsilon_t) \) is bounded. Thus (A2) implies that \( \{\varepsilon_t\} \) is stable. However, \( \{\varepsilon_t\} \) is not stationary. In fact, by (2.3), we obtain that

\[
\text{Cov}(\varepsilon_t, \varepsilon_{t+k}) = \sigma_0^2 \left\{ \prod_{i=0}^{k-1} f_{t+i}(\theta_0) + f_{t}(\theta_0) \prod_{i=0}^{k} f_{t+k-i}(\theta_0) + \cdots + \prod_{i=0}^{t-2} f_{t-i}(\theta_0) \prod_{i=0}^{t+k-2} f_{t+k-i}(\theta_0) \right\}, \tag{2.26}
\]

which is dependent of \( t \).
Thus, Theorem 3.1. Suppose that conditions (A1)–(A3) hold. Then there is a sequence later in the paper.

2. Statement of Main Results

For ease of exposition, we will introduce the following notations which will be used

\[
S_n(\varphi) = \sigma^2 \frac{\partial^2 \Psi_n}{\partial \varphi} = \sigma^2 \left( \frac{\partial^2 \Psi_n}{\partial \beta} + \frac{\partial^2 \Psi_n}{\partial \theta} \right), \quad F_n(\varphi) = -\sigma^2 \frac{\partial^2 \Psi_n}{\partial \theta \partial \varphi^T}.
\] (2.27)

By (2.7) and (2.8), we get

\[
F_n(\varphi) = \left( X_n(\theta) \sum_{i=2}^n (f_i'(\theta)\epsilon_{i-1} x_i + f_i'(\theta)\epsilon_i x_i - 2 f_i'(\theta) f_i'(\theta) x_i \epsilon_{i-1}) \right) \times \left( \sum_{i=2}^n \left( f_i''(\theta) + f_i'(\theta) f_i'''(\theta) \right) \epsilon_{i-1} - f_i''(\theta) \epsilon_i \right),
\] (2.28)

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

\[
D_n = E(F_n(\varphi_0)) = \begin{pmatrix} X_n(\theta_0) & 0 \\ \ast & \sum_{i=2}^n \left( f_i''(\theta_0) + f_i'(\theta_0) f_i'''(\theta_0) \right) E\epsilon_{i-1}^2 - f_i''(\theta_0) E(\epsilon_i \epsilon_{i-1}) \end{pmatrix}
\] (2.29)

\[
= \begin{pmatrix} X_n(\theta_0) & 0 \\ \ast & \sum_{i=2}^n f_i''(\theta_0) E\epsilon_{i-1}^2 \end{pmatrix}
\]

\[
= \begin{pmatrix} \Delta_n(\theta_0, \sigma_0) \\ \ast \end{pmatrix},
\]

where

\[
\Delta_n(\theta_0, \sigma_0) = \sum_{i=2}^n f_i''(\theta_0) E\epsilon_{i-1}^2 = \sigma_0^2 \sum_{i=2}^n f_i''(\theta_0) \sum_{j=0}^{i-2} \left( \prod_{k=0}^{j-1} f_k''(\theta) \right) = O(n).
\] (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

\[
P\left\{ \text{there are estimators } \hat{\varphi}_n, \hat{\sigma}_n^2 \text{ with } S_n(\hat{\varphi}_n) = 0, \text{ and } \left( \hat{\varphi}_n, \hat{\sigma}_n^2 \right) \in N_n'(A) \right\} \to 1.
\] (3.1)
Furthermore,

$$\left( \hat{\sigma}_n, \sigma_n^2 \right) \xrightarrow{p} \left( \sigma_0, \sigma_0^2 \right), \quad n \to \infty, \quad (3.2)$$

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

$$N_n(A) = \left\{ \varphi \in \mathbb{R}^{d+1} : (\varphi - \varphi_0)^T D_n(\varphi - \varphi_0) \leq A^2 \right\},$$

$$N'_n(A) = N_n(A) \cap \left\{ \sigma^2 \in [\sigma_0^2 - A_n, \sigma_0^2 + A_n] \right\}. \quad (3.3)$$

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

$$\frac{1}{\hat{\sigma}_n} F_n^{1/2}(\hat{\varphi}_n)(\hat{\varphi}_n - \varphi_0) \xrightarrow{D} N(0, I_{d+1}), \quad n \to \infty. \quad (3.4)$$

**Remark 3.3.** For $\theta \in \mathbb{R}^m, m \in \mathbb{N}$, 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 $f_i(\theta) = \theta$. 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 $f_i(\theta) = \theta$, then, for $|\theta| \neq 1$,

$$\sqrt{\frac{1}{n-1} \sum_{t=2}^{n} \varepsilon_t^2} \left( \theta_n - \theta_0 \right) \xrightarrow{D} N(0, 1), \quad n \to \infty. \quad (3.5)$$

where

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

**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 $\hat{\sigma}_n^2 \xrightarrow{p} \sigma_0^2$ and $(1/(n-1)) \sum_{t=2}^{n} \varepsilon_t^2 \xrightarrow{p} \sigma_0^2 / (1 - \theta_0^2)$, and by Corollary 3.6 we obtain

$$\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \xrightarrow{D} N \left( 0, 1 - \theta_0^2 \right). \quad (3.7)$$

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

$$
\frac{F_n^{1/2}(\hat{\beta})}{\hat{\sigma}_n} (\hat{\theta}_n - \theta_0) \rightarrow_D N(0,1), \quad n \to \infty,
$$

where

$$
F_n(\hat{\theta}_n) = \sum_{t=2}^{n} \left( \left( f_1''(\hat{\theta}_n) + f_1(\hat{\theta}_n) f_1'(\hat{\theta}_n) \right) \epsilon_{t-1}^2 - f_1'(\hat{\theta}_n) \epsilon_t \epsilon_{t-1} \right),
$$

$$
\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{t=2}^{n} \left( \epsilon_t - f_1(\hat{\theta}_n) \epsilon_{t-1} \right)^2.
$$

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

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

Remark 3.10. Let $a_t = 0$. Note that $\sum_{t=2}^{n} x_t^T = O(\sqrt{n})$ and $\hat{\sigma}_n^2 \to \sigma_0^2$; 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 $\text{Var}(S_n(\varphi_0)) = \sigma_0^2 D_n$.

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

$$
\sigma_0^2 E \left( \frac{\partial \Psi_n}{\partial \beta} \right)_{\beta=\beta_0} = \sum_{t=2}^{n} E(e_t - f_1(\theta_0) \epsilon_{t-1})(x_t - f_1(\theta_0) x_{t-1})
$$

$$
= \sum_{t=2}^{n} (x_t - f_1(\theta_0) x_{t-1}) E \eta_t = 0.
$$

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

$$
\sigma_0^2 E \left( \frac{\partial \Psi_n}{\partial \theta} \right)_{\theta=\theta_0} = \sum_{t=2}^{n} E((e_t - f_1(\theta_0) \epsilon_{t-1}) f_1'(\theta_0) \epsilon_{t-1})
$$

$$
= \sum_{t=2}^{n} f_1'(\theta_0) E(\eta_t \epsilon_{t-1}) = 0.
$$
Hence, from (4.1) and (4.2),

\[ E(S_n(\theta_0)) = \sigma_0^2 E \left( \frac{\partial \Psi_n}{\partial \beta} \Big|_{\beta = \beta_0}, \frac{\partial \Psi_n}{\partial \theta} \Big|_{\theta = \theta_0} \right) = 0. \quad (4.3) \]

By (2.8) and (2.17), we have

\[ \text{Var} \left( \sigma_0^2 \frac{\partial \Psi_n}{\partial \beta} \Big|_{\beta = \beta_0} \right) = \text{Var} \left\{ \sum_{t=2}^{n} (e_t - f_t(\theta_0)e_{t-1}) (x_t - f_t(\theta_0)x_{t-1}) \right\} \]

\[ = \text{Var} \left\{ \sum_{t=2}^{n} \eta_t (x_t - f_t(\theta_0)x_{t-1}) \right\} = \sigma_0^2 X_n(\theta_0). \quad (4.4) \]

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

\[ \text{Var} \left( f'_t(\theta_0)\eta_t e_{t-1} \right) = f'_t^2(\theta_0) \text{Var} \left( \eta_t^2 E e_{t-1}^2 = \sigma_0^2 f'_t^2(\theta_0) E e_{t-1}^2. \quad (4.5) \right. \]

so

\[ \text{Var} \left( \sigma_0^2 \frac{\partial \Psi_n}{\partial \theta} \right) = \text{Var} \left\{ \sum_{t=2}^{n} \eta_t f'_t(\theta_0)e_{t-1} \right\} \]

\[ = \sum_{t=2}^{n} f'_{t}^2(\theta_0) E e_{t-1}^2 = \sigma_0^2 \Delta_n(\theta_0, \sigma_0). \quad (4.6) \]

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

\[ \text{Cov} \left( \sigma_0^2 \frac{\partial \Psi_n}{\partial \beta} \Big|_{\beta = \beta_0}, \sigma_0^2 \frac{\partial \Psi_n}{\partial \theta} \Big|_{\theta = \theta_0} \right) = E \left( \sigma_0^2 \frac{\partial \Psi_n}{\partial \beta} \Big|_{\beta = \beta_0}, \sigma_0^2 \frac{\partial \Psi_n}{\partial \theta} \Big|_{\theta = \theta_0} \right) \]

\[ = E \left( \sum_{t=2}^{n} \eta_t^2 (x_t - f_t(\theta_0)x_{t-1}) f'_t(\theta_0)e_{t-1} \right) \]

\[ + E \left( \sum_{t=3}^{n} \eta_t (x_t - f_t(\theta_0)x_{t-1}) \sum_{s=2}^{t-1} \eta_s f'_s(\theta_0)e_{s-1} \right) \]

\[ + E \left( \sum_{s=3}^{n} \eta_s f'_s(\theta_0)e_{s-1} \sum_{t=2}^{s-1} \eta_t (x_t - f_t(\theta_0)x_{t-1}) \right) \]

\[ = 0. \quad (4.7) \]

From (4.4)–(4.7), it follows that \( \text{Var}(S_n(\theta_0)) = \sigma_0^2 D_n. \)
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

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

Proof. Let $\lambda_1$ and $\lambda_d$ be the smallest and largest roots of $|Z_n - \lambda X_n| = 0$. Then from the study of Rao in [47, Ex 22.1],

$$\lambda_1 \leq \frac{u^T Z_n u}{u^T X_n u} \leq \lambda_d \quad (4.9)$$

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

$$|u^T Z_n u| \leq (1 - \delta)u^T X_n u. \quad (4.10)$$

By (4.10), we have

$$u^T X_n u = \sum_{i=2}^{n} \left( u^T (x_i - f_i(\theta)x_{i-1}) \right)^2$$

$$= \sum_{i=2}^{n} \left( \left( u^T x_i \right)^2 + f_i^2(\theta) \left( u^T x_{i-1} \right)^2 - f_i(\theta) u^T x_{i-1} x_i^T u - f_i(\theta) u^T x_i x_{i-1}^T u \right)$$

$$\geq \sum_{i=2}^{n} \left( u^T x_i \right)^2 + \min_{2 \leq t \leq n} f_t^2(\theta) \sum_{i=2}^{n} \left( u^T x_{i-1} \right)^2 - \max_{2 \leq t \leq n} f_t(\theta) |u^T Z_n u| \quad (4.11)$$

$$\geq u^T X_n u + \min_{2 \leq t \leq n} f_t^2(\theta) \left| u^T X_n u - \max_{2 \leq t \leq n} f_t(\theta) |u^T Z_n u| \right|$$

$$\geq \left( 1 + \min_{2 \leq t \leq n} f_t^2(\theta) \right) \left| u^T X_n u - \max_{2 \leq t \leq n} f_t(\theta) |u^T Z_n u| \right|$$

$$= C(\theta, \delta) u^T X_n u.$$ 

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

$$\frac{(u^T x_i)^2}{u^T X_n u} \to 0. \quad (4.12)$$
From (4.12) and $C(\theta, \delta) > 0$,

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

\[\square\]

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

$$W_n \overset{p}{\to} I \iff \lambda_j(n) \overset{p}{\to} 1, \quad n \to \infty. \tag{4.14}$$

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

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

and also

$$\Phi_n \overset{d}{\to} \Phi, \tag{4.16}$$

where

$$\Phi_n = \begin{pmatrix} I_d & 0 \\ 0 & \frac{\sum_{i=2}^{n} f_i^2(\theta_0) e_i^2_{i-1}}{\Delta_n(\theta_0, \sigma_0)} \end{pmatrix}, \quad \Phi = I_{d+1}. \tag{4.18}$$

**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

$$(\varphi - \varphi_0)^T D_n (\varphi - \varphi_0) = (\beta - \beta_0)^T X_n(\theta_0) (\beta - \beta_0) + (\theta - \theta_0)^2 \Delta_n(\theta_0, \sigma_0) \leq A^2. \tag{4.20}$$
From (2.28), (2.29), and (4.18),

\[
D_n^{-1/2} F_n(\varphi) D_n^{-T/2} - \Phi_n = \left( X_n^{-1/2}(\theta_0) X_n(\theta) X_n^{-T/2}(\theta_0) - I_d \right) \times \frac{X_n^{-1/2}(\theta_0) \sum_{i=2}^n (f_i(\theta) \varepsilon_{i-1} x_i + f_i'(\theta) \varepsilon_{i-1} x_{i+1}) - 2 f_i(\theta) x_i \varepsilon_{i-1} x_{i-1})}{\sqrt{\Delta_n(\theta_0, \sigma_0)}}
\]

\[
\sum_{i=2}^n \left( f_i'(\theta) + f_i'(\theta) \right) \varepsilon_{i-1}^2 - f_i''(\theta) \varepsilon_{i-1}^2 - \sum_{i=2}^n f_i''(\theta) \varepsilon_{i-1}^2 \right) \frac{\sum_{i=2}^n \left( f_i'(\theta) \right) \varepsilon_{i-1}^2}{\sqrt{\Delta_n(\theta_0, \sigma_0)}}.
\]

Let

\[
N^\beta_n(A) = \left\{ \beta : \left( \beta - \beta_0 \right)^T X_n^{1/2}(\theta_0) \right\} \leq A^2 \}
\]

(4.22)

\[
N^\theta_n(A) = \left\{ \theta : |\theta - \theta_0| \leq \frac{A}{\sqrt{\Delta_n(\theta_0, \sigma_0)}} \}
\]

(4.23)

In the first step, we will show that, for each \( A > 0 \),

\[
\sup_{\theta \in N^\beta_n(A)} \left\| X_n^{-1/2}(\theta_0) X_n(\theta) X_n^{-T/2}(\theta_0) - I_d \right\| \to 0, \quad n \to \infty.
\]

(4.24)

In fact, note that

\[
X_n^{-1/2}(\theta_0) X_n(\theta) X_n^{-T/2}(\theta_0) - I_d = X_n^{-1/2}(\theta_0) (X_n(\theta) - X_n(\theta_0)) X_n^{-T/2}(\theta_0)
\]

(4.25)

\[
= X_n^{-1/2}(\theta_0) (T_1 + T_2 - T_3) X_n^{-T/2}(\theta_0),
\]

where

\[
T_1 = \sum_{i=2}^n \left( f_i(\theta_0) - f_i(\theta) \right) x_{i-1} \left( x_i - f_i(\theta_0) x_{i-1} \right)^T,
\]

\[
T_2 = \sum_{i=2}^n \left( x_i - f_i(\theta_0) x_{i-1} \right) x_{i-1}^T,
\]

(4.26)

\[
T_3 = \sum_{i=2}^n (f_i(\theta_0) - f_i(\theta))^2 x_{i-1} x_{i-1}^T.
\]
Let $u, v \in \mathbb{R}^d$, $|u| = |v| = 1$, and let $u_i^T = u^T X_n^{-1/2}(\theta_0)$, $v_i^T = X_n^{-T/2}(\theta_0)v$. By Cauchy-Schwartz inequality, Lemma 4.2, condition (A3), and noting that $\theta \in N_n^\theta(A)$, we have that

$$
\left| u_i^T T_1 v_n \right| = \left| \sum_{i=1}^{n} (f_i(\theta_0) - f_i(\bar{\theta})) u_i^T x_{i-1} (x_i - f_i(\theta_0)x_{i-1})^T v_n \right|
$$

$$
\leq \max_{2 \leq i \leq n} \left| f_i(\theta_0) - f_i(\bar{\theta}) \right| \left| \sum_{i=1}^{n} u_i^T x_{i-1} (x_i - f_i(\theta_0)x_{i-1})^T v_n \right|
$$

$$
\leq \max_{2 \leq i \leq n} \left| f_i(\theta_0) - f_i(\bar{\theta}) \right| \left( \sum_{i=1}^{n} u_i^T x_{i-1} x_i^T u_n \right)^{1/2}
$$

$$
\cdot \left( \sum_{i=2}^{n} v_i^T (x_i - f_i(\theta_0)x_{i-1}) (x_i - f_i(\theta_0)x_{i-1})^T v_n \right)^{1/2} \quad (4.27)
$$

$$
\leq \max_{2 \leq i \leq n} \left| f_i(\theta_0) - f_i(\bar{\theta}) \right| \left( \sum_{i=1}^{n} u_i^T x_i^T u_n \right)^{1/2}
$$

$$
\leq \max_{2 \leq i \leq n} \left| f_i(\theta_0) - f_i(\bar{\theta}) \right| \sqrt{n} \max_{1 \leq i \leq n} \left( x_i^T X_n^{-1}(\theta_0)x_i \right)
$$

$$
\leq C \sqrt{\frac{n}{\Delta_n(\theta_0, \sigma_0)}} o(1) \rightarrow 0.
$$

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

$$
\left| u_i^T T_2 v_n \right| \rightarrow 0. \quad (4.28)
$$

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

$$
\left| u_i^T T_3 v_n \right| = \left| u_i^T \sum_{i=2}^{n} (f_i(\theta_0) - f_i(\bar{\theta})) x_{i-1}^T x_i^T v_n \right|
$$

$$
\leq \max_{2 \leq i \leq n} \left| f_i(\theta_0) - f_i(\bar{\theta}) \right| \left( \sum_{i=2}^{n} u_i^T x_i^T u_n \sum_{i=2}^{n} v_i^T x_i^T v_n \right)^{1/2} \quad (4.29)
$$

$$
\leq n \max_{2 \leq i \leq n} \left| f_i(\theta_0) - f_i(\bar{\theta}) \right| \sqrt{\frac{n}{\Delta_n(\theta_0, \sigma_0)}} o(1) \rightarrow 0.
$$

Hence, (4.24) follows from (4.25)–(4.29).
In the second step, we will show that

\[
X_n^{-1/2}(\theta_0) \sum_{i=2}^{n} (f_i'(\theta) \epsilon_{i-1} x_i + f_i'(\theta) \epsilon_i x_{i-1} - 2 f_i(\theta) f_i'(\theta) \epsilon_{i-1} x_{i-1}) \to_p 0. 
\]  

Note that

\[
\epsilon_t = y_t - x_t^T \beta = x_t^T (\beta_0 - \beta) + \epsilon_t, 
\]

\[
\epsilon_t - f_i(\theta_0) \epsilon_{i-1} = (x_t - f_i(\theta_0) x_{i-1})^T (\beta_0 - \beta) + \eta_t. 
\]

Consider

\[
J = \sum_{i=2}^{n} (f_i'(\theta) \epsilon_{i-1} x_i + f_i'(\theta) \epsilon_i x_{i-1} - 2 f_i(\theta) f_i'(\theta) \epsilon_{i-1} x_{i-1}) 
\]

\[
= \sum_{i=2}^{n} (\epsilon_{i-1} f_i'(\theta) (x_t - f_i(\theta_0) x_{i-1}) + f_i'(\theta) (\epsilon_t - f_i(\theta_0) \epsilon_{i-1}) x_{i-1}) + 2 f_i(\theta) (f_i(\theta_0) - f_i(\theta)) \epsilon_{i-1} x_{i-1} 
\]

\[
= T_1 + T_2 + T_3 + T_4 + 2T_5 + 2T_6, 
\]

where

\[
T_1 = \sum_{i=2}^{n} x_{i-1}^T f_i'(\theta) (\beta_0 - \beta) (x_t - f_i(\theta_0) x_{i-1}), \quad T_2 = \sum_{i=2}^{n} f_i'(\theta) \epsilon_{i-1} (x_t - f_i(\theta_0) x_{i-1}), 
\]

\[
T_3 = \sum_{i=2}^{n} f_i'(\theta) (x_t - f_i(\theta_0) x_{i-1})^T (\beta_0 - \beta) x_{i-1}, \quad T_4 = \sum_{i=2}^{n} f_i'(\theta) \eta_i x_{i-1}, 
\]

\[
T_5 = \sum_{i=2}^{n} f_i'(\theta) (f_i(\theta_0) - f_i(\theta)) x_{i-1}^T (\beta_0 - \beta) x_{i-1}, \quad T_6 = \sum_{i=2}^{n} f_i'(\theta) (f_i(\theta_0) - f_i(\theta)) \epsilon_{i-1} x_{i-1}. 
\]

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

\[
\left| (\beta_0 - \beta)^T x_i \right|^2 = (\beta_0 - \beta)^T X_n^{-1/2}(\theta_0) X_n^{-1/2}(\theta_0) x_i x_i^T X_n^{-T/2}(\theta_0) X_n^{T/2}(\theta_0) (\beta_0 - \beta) 
\]

\[
\leq \max_{1 \leq s \leq n} \left( x_i^T X_n^{-1}(\theta_0) x_i \right) (\beta_0 - \beta)^T X_n(\theta_0) (\beta_0 - \beta) 
\]

\[
\leq A^2 \max_{1 \leq s \leq n} \left( x_i^T X_n^{-1}(\theta_0) x_i \right). 
\]
By (4.34) and Lemma 4.2, we have

$$\sup_{\beta \in N^0(A)} \left\{ (\beta_0 - \beta)^T x_t \right\} \rightarrow 0, \quad n \rightarrow \infty, \quad A > 0. \quad (4.35)$$

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

$$u_n^T T_3 = \sum_{t=2}^n u_n^T x_{t-1} (\beta_0 - \beta) f_i'(\theta) (x_t - f_i(\theta_0) x_{t-1})$$

$$\leq \left\{ \sum_{t=2}^n (x_{t-1}^T (\beta_0 - \beta))^2 \right\}^{1/2}$$

$$\times \left\{ \sum_{t=2}^n \left( f_i^2(\theta) u_n^T (x_t - f_i(\theta_0) x_{t-1}) (x_t - f_i(\theta_0) x_{t-1})^T u_n \right)^2 \right\}^{1/2} \quad (4.36)$$

$$\leq \sqrt{n} \max_{1 \leq t \leq n} \left| (\beta_0 - \beta)^T x_t \right| \max_{1 \leq t \leq n} \left| f_i'(\theta) \right|$$

$$= o(\sqrt{n}).$$

Let

$$a_{nT} = u_n^T f_i'(\theta) (x_t - f_i(\theta_0) x_{t-1}). \quad (4.37)$$

Then from Lemma 4.2,

$$\max a_{nT}^2 = \max_{2 \leq t \leq n} \left| f_i^2(\theta) \right|$$

$$\times \max_{2 \leq t \leq n} \left\{ u_n^T (x_t - f_i(\theta_0) x_{t-1}) X_n^{-1}(\theta_0) (x_t - f_i(\theta_0) x_{t-1})^T u \right\} \quad (4.38)$$

$$= o(1).$$
By condition (A2) and (4.38), we have

\[
\text{Var}(u_n^T T_2) = \text{Var} \left( \sum_{i=2}^{n} a_{in} e_{i-1} \right) = \text{Var} \left( \sum_{i=2}^{n} a_{in} e_{i-1} \right) \\
= \text{Var} \left\{ \sum_{j=1}^{n-1} \eta_j \left( \sum_{l=j+1}^{n} a_{in} \prod_{i=0}^{l-j-1} f_{i-1}(\theta) \right) \right\} \\
= \sigma_0^2 \sum_{j=1}^{n-1} \left( \sum_{l=j+1}^{n} a_{in} \prod_{i=0}^{l-j-1} f_{i-1}(\theta) \right) \\
\leq \sigma_0^2 \max_{2 \leq l \leq n} |a_{in}| \sum_{j=1}^{n-1} \left( \prod_{i=0}^{l-j-1} f_{i-1}(\theta) \right) \\
\leq a \sigma_0^2 \max_{2 \leq l \leq n} |a_{in}| n = o(n). \tag{4.39}
\]

Thus by Chebychev inequality and (4.39),

\[
u_n^T T_2 = o_p(\sqrt{n}). \tag{4.40}
\]

Using the similar argument as \( T_1 \), we obtain that

\[
u_n^T T_3 = o_p(\sqrt{n}). \tag{4.41}
\]

Using the similar argument as \( T_2 \), we obtain that

\[
u_n^T T_4 = o_p(\sqrt{n}), \quad u_n^T T_6 = o_p(\sqrt{n}). \tag{4.42}
\]

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

\[
u_n^T T_5 = \sum_{l=2}^{n} f_i^2(\theta) (f_i(\theta_0) - f_i(\theta)) x_{l-1}^T (\beta_0 - \beta) u_l^T x_{l-1} \leq \left\{ \sum_{i=2}^{n} f_i^2(\theta) (f_i(\theta_0) - f_i(\theta))^2 (x_{l-1}^T (\beta_0 - \beta))^2 \sum_{l=2}^{n} (u_l^T x_{l-1})^2 \right\}^{1/2} \leq \max_{2 \leq l \leq n} |f_i^2(\theta)| \max_{2 \leq l \leq n} |f_i(\theta)| |\theta_0 - \theta| \cdot \left\{ \sum_{l=2}^{n} (x_{l-1}^T (\beta_0 - \beta))^2 \sum_{l=2}^{n} (u_l^T x_{l-1})^2 \right\}^{1/2} \leq C \frac{A}{\sqrt{\Delta_n(\theta_0, \theta_0)}} \sqrt{n} o(1) = o(\sqrt{n}). \tag{4.43}
\]

Thus (4.30) follows immediately from (4.32), (4.36), and (4.40)–(4.43).
In the third step, we will show that

\[
\sum_{t=2}^{n} \left( \left( f_i^2(\theta) + f_i(\theta) f_i''(\theta) \right) \epsilon_{t-1}^2 - f_i''(\theta) \epsilon_{t-1} \right) - \sum_{t=2}^{n} f_i^2(\theta_0) \epsilon_{t-1}^2 \right) \frac{e_n(\theta_0, \sigma_0)}{\sqrt{\Delta_n(\theta_0, \sigma_0)}} \rightarrow 0. \tag{4.44}
\]

Let

\[
J = \sum_{t=2}^{n} \left( \left( f_i^2(\theta) + f_i(\theta) f_i''(\theta) \right) \epsilon_{t-1}^2 - f_i''(\theta) \epsilon_{t-1} \right) - \sum_{t=2}^{n} f_i^2(\theta_0) \epsilon_{t-1}^2. \tag{4.45}
\]

Then

\[
J = \sum_{t=2}^{n} \left( \left( f_i^2(\theta) + f_i(\theta) f_i''(\theta) \right) \epsilon_{t-1}^2 - f_i''(\theta) \epsilon_{t-1} \right) - \sum_{t=2}^{n} f_i^2(\theta_0) \epsilon_{t-1}^2 \\
= \sum_{t=2}^{n} \left\{ f_i^2(\theta) \epsilon_{t-1}^2 - f_i''(\theta) \epsilon_{t-1} - f_i^2(\theta_0) \epsilon_{t-1}^2 \right\} \\
= \sum_{t=2}^{n} \left\{ f_i^2(\theta) \epsilon_{t-1}^2 - f_i''(\theta) \epsilon_{t-1} - f_i^2(\theta_0) \epsilon_{t-1}^2 \right\} \\
+ \sum_{t=2}^{n} \left\{ 2 f_i^2(\theta_0) \epsilon_{t-1}^2 - f_i''(\theta) \epsilon_{t-1} - f_i^2(\theta_0) \epsilon_{t-1}^2 \right\} \\
= T_1 + T_2 + 2T_3 - T_4 - T_5.
\]

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

\[
|ET_2| = \left| \left( f_i^2(\theta) - f_i^2(\theta_0) \right) E \epsilon_{t-1}^2 \right| \\
= \left| f_i^2(\theta) \left( \theta - \theta_0 \right) \left( \frac{f_i^2(\theta) + f_i^2(\theta_0)}{f_i^2(\theta_0)} \right) \right| \left| \epsilon_{t-1}^2 \right| \\
\leq \max_{2 \leq t \leq n} \left| f_i^2(\theta) \left( \frac{f_i^2(\theta) + f_i^2(\theta_0)}{f_i^2(\theta_0)} \right) \right| \left| \theta - \theta_0 \right| \Delta_n(\theta_0, \sigma_0) \tag{4.48}
\]

\[
\leq \max_{2 \leq t \leq n} \left| f_i^2(\theta) \left( \frac{f_i^2(\theta) + f_i^2(\theta_0)}{f_i^2(\theta_0)} \right) \right| \left| \epsilon_{t-1}^2 \right| \Delta_n(\theta_0, \sigma_0) \\
= o\left( \sqrt{\Delta_n(\theta_0, \sigma_0)} \right).
\]
Hence, by Markov inequality,

\[ T_2 = O_p \left( \sqrt{\Delta_n(\theta_0, \sigma_0)} \right). \]  

(4.49)

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

\[ T_3 = o_p(\sqrt{n}). \]  

(4.50)

By Markov inequality and noting that

\[ \text{Var}(T_4) \leq \sigma_0^2 n \max_{2 \leq i \leq n} \left| f_i' \right|^2 \max_{2 \leq i \leq n} \left( x_i^T \left( \beta - \beta_0 \right) \right) = o(n), \]  

(4.51)

we have that

\[ T_4 = o_p(\sqrt{n}). \]  

(4.52)

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

\[ T_5 = O_p \left( \sqrt{\Delta_n(\theta_0, \sigma_0)} \right). \]  

(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

\[ \frac{\sum_{i=2}^{n} f_i^2(\theta_0) e_i^2}{\Delta_n(\theta_0, \sigma_0)} \rightarrow_p 1, \quad n \rightarrow \infty. \]  

(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

\[ D_n^{-1/2} F(\varphi) D_n^{-T/2} \rightarrow_p I, \quad n \rightarrow \infty, \]  

(4.55)

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

\[ \lambda_{\min} \left( D_n^{-1/2} F(\varphi) D_n^{-T/2} \right) \rightarrow_p 1, \quad n \rightarrow \infty. \]  

(4.56)

This implies (4.17). \( \square \)
Lemma 4.5 (see [49]). Let \( \{S_{ni}, F_{ni}, 1 \leq i \leq k_n, n \geq 1\} \) be a zero-mean, square-integrable martingale array with differences \( X_{ni} \), and let \( \eta^2 \) be an a.s. finite random variable. Suppose that \( \sum_i E \{X_{ni}^2, I(|X_{ni}| > \varepsilon) \mid F_{ni-1}\} \rightarrow_p 0 \), for all \( \varepsilon \rightarrow 0 \), and \( \sum_i E \{X_{ni}^2 \mid F_{ni-1}\} \rightarrow_p \eta^2 \). Then

\[
S_{nk_n} = \sum_i X_{ni} \longrightarrow_{D} Z,
\]

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

5. Proof of Theorems

5.1. Proof of Theorem 3.1

Take \( A > 0 \), let

\[
M_n(A) = \{ \varphi \in \mathbb{R}^{d+1} : (\varphi - \varphi_0)^T D_n(\varphi - \varphi_0) = A^2 \}\]

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

\[
\Psi_n(\varphi, \sigma^2) = \Psi_n(\varphi_0, \sigma^2) + (\varphi - \varphi_0)^T \frac{\partial \Psi_n(\varphi_0, \sigma^2)}{\partial \varphi} + \frac{1}{2} (\varphi - \varphi_0)^T \frac{\partial^2 \Psi_n(\varphi_0, \sigma^2)}{\partial \varphi \partial \varphi^T} (\varphi - \varphi_0)
\]

\[
= \frac{1}{\sigma^2} \Psi_n(\varphi_0, \sigma^2) + (\varphi - \varphi_0)^T S_n(\varphi_0) - \frac{1}{2\sigma^2} (\varphi - \varphi_0)^T F_n(\varphi) (\varphi - \varphi_0),
\]

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

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

\[
P\{\Psi_n(\varphi, \sigma^2) \geq \Psi_n(\varphi_0, \sigma^2) \text{ for some } \varphi \in M_n(A)\}
\]

\[
\leq P\{(\varphi - \varphi_0)^T S_n(\varphi_0) \geq Q_n(\varphi), Q_n(\varphi) > cA^2 \text{ for some } \varphi \in M_n(A)\}
\]

\[
+ P\{Q_n(\varphi) \leq cA^2 \text{ for some } \varphi \in M_n(A)\}
\]

\[
\leq P\{v_n^T(\varphi)D_n^{-1/2}S_n(\varphi_0) > cA \text{ for some } \varphi \in M_n(A)\}
\]

\[
+ P\{v_n^T(\varphi)D_n^{-1/2}F_n(\tilde{\varphi})D_n^{-1/2}v_n(\varphi) \leq c \text{ for some } \varphi \in M_n(A)\}
\]

\[
\leq P\{|D_n^{-1/2}S_n(\varphi_0)| > cA\} + P\left\{\inf_{\varphi \in N_n(A)} \lambda_{\min}\left(D_n^{-1/2}F_n(\tilde{\varphi})D_n^{-1/2}\right) \leq c\right\}.
\]
By Lemma 4.1 and Chebychev inequality, we obtain

\[ P\left\{ \left| D_n^{-1/2} S_n(\varphi_0) \right| > cA \right\} \leq \frac{\text{Var}\left(D_n^{-1/2} S_n(\varphi_0)\right)}{c^2 A^2} = \frac{\sigma_0^2}{c^2 A^2}. \]  

(5.4)

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

\[ P\left\{ \inf_{\varphi \in N_n(A)} \lambda_{\min}\left(D_n^{-1/2} F_n(\bar{\varphi}) D_n^{-T/2}\right) \leq c \right\} \to 0. \]  

(5.5)

By (5.3)–(5.5), we have

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

(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

\[ \inf_{\varphi \in N_n(A)} \lambda_{\min}(F_n(\varphi)) \to \infty. \]  

(5.7)

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

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

(5.8)

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

\[ \lim_{A \to \infty} \liminf_{n \to \infty} P\left\{ S_n(\bar{\varphi}_n(A)) = 0 \text{ for a unique } \bar{\varphi}_n(A) \in N_n(A) \right\} = 1. \]  

(5.9)

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

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

(5.10)

Thus the \( \bar{\varphi}_n = (\bar{\beta}_n, \bar{\theta}_n) \) is a QML estimator for \( \varphi_0 \). It is clearly consistent, and

\[ \lim_{A \to \infty} \liminf_{n \to \infty} P\left\{ \bar{\varphi}_n \in N_n(A) \right\} = 1. \]  

(5.11)

Since \( \bar{\varphi}_n = (\bar{\beta}_n, \bar{\theta}_n) \) is a QML estimator for \( \varphi_0 \), \( \bar{\sigma}_n^2 \) is a QML estimator for \( \sigma_0^2 \) from (2.9).
To complete the proof, we will show that $\hat{\sigma}_n^2 \to \sigma_0^2$ as $n \to \infty$. If $\tilde{\varphi}_n \in N_n(A)$ and $\tilde{\vartheta}_n \in N_n^0(A)$. By (2.12) and (2.1), we have

$$
\hat{\varepsilon}_t - f_i(\tilde{\vartheta}_n)\hat{\varepsilon}_{t-1} = (x_t - f_i(\tilde{\vartheta}_n)x_{t-1})^T(\hat{\beta}_0 - \hat{\beta}_n) + (e_t - f_i(\tilde{\vartheta}_n)e_{t-1}).
$$

(5.12)

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

$$(n - 1)\hat{\sigma}_n^2 = \sum_{t=2}^{n} (\hat{\varepsilon}_t - f_i(\tilde{\vartheta}_n)\hat{\varepsilon}_{t-1})^2
$$

$$
= \sum_{t=2}^{n} \left\{ (x_t - f_i(\tilde{\vartheta}_n)x_{t-1})^T(\hat{\beta}_0 - \hat{\beta}_n) + (e_t - f_i(\tilde{\vartheta}_n)e_{t-1}) \right\}^2
$$

$$
= \sum_{t=2}^{n} \left\{ \hat{\varepsilon}_t - f_i(\tilde{\vartheta}_n)\hat{\varepsilon}_{t-1} \right\} (x_t - f_i(\tilde{\vartheta}_n)x_{t-1})^T(\hat{\beta}_0 - \hat{\beta}_n)
$$

$$
+ \sum_{t=2}^{n} \left\{ \hat{\varepsilon}_t - f_i(\tilde{\vartheta}_n)\hat{\varepsilon}_{t-1} \right\} (e_t - f_i(\tilde{\vartheta}_n)e_{t-1})
$$

$$
= \sum_{t=2}^{n} \left\{ \hat{\varepsilon}_t - f_i(\tilde{\vartheta}_n)\hat{\varepsilon}_{t-1} \right\} (e_t - f_i(\tilde{\vartheta}_n)e_{t-1}).
$$

(5.13)

From (5.12), it follows that

$$
\sum_{t=2}^{n} \left\{ (x_t - f_i(\tilde{\vartheta}_n)x_{t-1})^T(\hat{\beta}_0 - \hat{\beta}_n) \right\}^2 = \sum_{t=2}^{n} (\hat{\varepsilon}_t - f_i(\tilde{\vartheta}_n)\hat{\varepsilon}_{t-1})^2
$$

$$
- 2\sum_{t=2}^{n} f_i(\tilde{\vartheta}_n)x_{t-1})^T(\hat{\beta}_0 - \hat{\beta}_n)(e_t - f_i(\tilde{\vartheta}_n)e_{t-1})
$$

$$
+ \sum_{t=2}^{n} (e_t - f_i(\tilde{\vartheta}_n)e_{t-1})^2.
$$

(5.14)

From (2.2),

$$
\sum_{t=2}^{n} (e_t - f_i(\tilde{\vartheta}_n)e_{t-1})^2 = \sum_{t=2}^{n} (f_i(\theta_0)e_{t-1} - f_i(\tilde{\vartheta}_n)e_{t-1})^2
$$

$$
= \sum_{t=2}^{n} \eta_t^2 + 2\sum_{t=2}^{n} (f_i(\theta_0) - f_i(\tilde{\vartheta}_n))\eta_t e_{t-1}
$$

$$
+ \sum_{t=2}^{n} (f_i(\theta_0) - f_i(\tilde{\vartheta}_n))^2 e_{t-1}^2.
$$

(5.15)
By (5.13)–(5.15), we have

\[(n - 1)\hat{\sigma}_n^2 = \sum_{t=2}^{n} (e_t - f_t(\hat{\theta}_n)e_{t-1})^2 - \sum_{t=2}^{n} \left( x_t - f_t(\hat{\theta}_n)x_{t-1} \right)^T \left( \beta_0 - \hat{\beta}_n \right)^2 \]

\[= \sum_{t=2}^{n} \eta_t^2 + 2 \sum_{t=2}^{n} (f_t(\theta_0) - f_t(\hat{\theta}_n)) \eta_t e_{t-1} + \sum_{t=2}^{n} (f_t(\theta_0) - f_t(\hat{\theta}_n))^2 e_{t-1}^2 \]

\[- \sum_{t=2}^{n} \left( x_t - f_t(\hat{\theta}_n)x_{t-1} \right)^T \left( \beta_0 - \hat{\beta}_n \right)^2 \]

\[= T_1 + 2T_2 + T_3 - T_4. \tag{5.16} \]

By the law of large numbers,

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

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

\[\text{Var}\left\{ (f_t(\theta_0) - f_t(\hat{\theta}_n))\eta_t e_{t-1} \right\} = (f_t(\theta_0) - f_t(\hat{\theta}_n))^2 \sigma_0^2 E e_{t-1}^2, \]

\[\text{Var}(T_2) = \sum_{t=2}^{n} E\left( (f_t(\theta_0) - f_t(\hat{\theta}_n))\eta_t e_{t-1} \right)^2 \]

\[= \sigma_0^2 \sum_{t=2}^{n} (f_t(\theta_0) - f_t(\hat{\theta}_n))^2 E e_{t-1}^2 \]

\[= \sigma_0^2 \sum_{t=2}^{n} f_t^2(\hat{\theta}_n) (\theta_0 - \hat{\theta}_n)^2 f_t^2(\theta_0) E e_{t-1}^2 \]

\[\leq \sigma_0^2 \max_{2 \leq t \leq n} \left| \frac{f_t(\hat{\theta})}{f_t(\theta_0)} \right|^2 \left| (\theta_0 - \hat{\theta}_n) \right|^2 \Delta_n(\theta_0, \sigma_0^2) \]

\[\leq CA^2. \tag{5.18} \]

By Chebychev inequality, we have

\[\frac{1}{n-1}T_2 \rightarrow_p 0, \quad (n \rightarrow \infty). \tag{5.19} \]

By Markov inequality and noting that \(ET_3 \leq CA^2\), we obtain that

\[\frac{1}{n-1}T_3 \rightarrow_p 0, \quad (n \rightarrow \infty). \tag{5.20} \]
Write
\[
T_4 = \sum_{i=2}^{n} \left( (x_i - f_i(\theta_0)x_{i-1})^T (\beta_0 - \bar{\beta}_n) + (f_i(\theta_0) - f_i(\bar{\theta}_n)) x_{i-1}^T (\beta_0 - \bar{\beta}_n) \right)^2 \\
= \left( \beta_0 - \bar{\beta}_n \right)^T X_n(\theta_0) \left( \beta_0 - \bar{\beta}_n \right) + \sum_{i=2}^{n} \left( f_i(\theta_0) - f_i(\bar{\theta}_n) \right)^2 \left( x_{i-1}^T (\beta_0 - \bar{\beta}_n) \right)^2 \\
+ 2 \sum_{i=2}^{n} (x_i - f_i(\theta_0)x_{i-1})^T (\beta_0 - \bar{\beta}_n) \left( f_i(\theta_0) - f_i(\bar{\theta}_n) \right) x_{i-1}^T (\beta_0 - \bar{\beta}_n) \\
= I_1 + I_2 + 2I_3. 
\] (5.21)

Noting that \( \beta \in N_\alpha^\beta(A) \), we have
\[
I_1 = O_p(1), \quad (n \to \infty). \tag{5.22}
\]

By (4.34) and condition (A3), we have
\[
|I_2| \leq \max_{2 \leq i \leq n} \left| x_{i-1}^T (\beta_0 - \bar{\beta}_n) \right|^2 \left( n \max_{2 \leq i \leq n} |f_i'(\bar{\theta})| \right) \left| x_{i-1} - \bar{\theta}_n \right|^2 = o(1), \quad (n \to \infty). \tag{5.23}
\]

By (4.34), condition (A3), and Cauchy-Schwartz inequality, we have
\[
|I_3|^2 \leq \sum_{i=2}^{n} \left( \beta_0 - \bar{\beta}_n \right)^T (x_i - f_i(\theta_0)x_{i-1})^T (x_i - f_i(\theta_0)x_{i-1}) (\beta_0 - \bar{\beta}_n) \\
\cdot \sum_{i=2}^{n} \left( f_i(\theta_0) - f_i(\bar{\theta}_n) \right)^2 \left( x_{i-1}^T (\beta_0 - \bar{\beta}_n) \right)^2 \\
\leq \left( \beta_0 - \bar{\beta}_n \right)^T X_n(\theta_0) \left( \beta_0 - \bar{\beta}_n \right) n \max_{2 \leq i \leq n} \left| f_i'(\bar{\theta}) \right|^2 \left| \theta_0 - \bar{\theta}_n \right|^2 \\
\cdot \max_{2 \leq i \leq n} \left| x_{i-1}^T (\beta_0 - \bar{\beta}_n) \right|^2 \\
\leq O_p(1) n \frac{A^2}{\Delta_n(\theta_0, \sigma_0^2)} o(1) = o_p(1). \tag{5.24}
\]

By (5.21)–(5.24), we obtain
\[
\frac{1}{n - 1} T_4 \xrightarrow{p} 0, \quad (n \to \infty). \tag{5.25}
\]

From (5.16), (5.17), (5.19), (5.20), and (5.25), we have \( \hat{\sigma}_n^2 \to \sigma_0^2 \). We therefore complete the proof of Theorem 3.1.
5.2. Proof of Theorem 3.2

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

\[
0 = S_n(\bar{\varphi}_n) = S_n(\varphi_0) - F_n(\bar{\varphi}_n)(\bar{\varphi}_n - \varphi_0). \tag{5.26}
\]

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

\[
F_n(\bar{\varphi}_n) = D_n^{1/2}(\Phi_n + \bar{A}_n)D_n^{T/2}, \tag{5.27}
\]

where \(\bar{A}_n\) is a symmetric matrix with \(\bar{A}_n \to p0\). By (5.26) and (5.27), we have

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

Similar to (5.27), we have

\[
F_n(\bar{\varphi}_n) = D_n^{1/2}(\Phi_n + \hat{A}_n)D_n^{T/2} = \left(D_n^{1/2}(\Phi_n + \bar{A}_n)^{1/2}\right)^T \left(D_n^{1/2}(\Phi_n + \bar{A}_n)^{T/2}\right) = F_n^{1/2}(\bar{\varphi}_n)F_n^{T/2}(\bar{\varphi}_n). \tag{5.29}
\]

Here \(\hat{A}_n \to p0\). By (5.28), (5.29), and noting that \(\hat{\sigma}_n^2 \to p\sigma_0^2\) and \(D_n^{-1/2}S_n(\varphi_0) = O_p(1)\), we obtain that

\[
\frac{F_n^{T/2}(\bar{\varphi}_n)(\bar{\varphi}_n - \varphi_0)}{\hat{\sigma}_n} = \frac{\left(\Phi_n + \bar{A}_n\right)^{1/2}}{\hat{\sigma}_n} \frac{\left(\Phi_n + \bar{A}_n\right)^{-1}D_n^{-1/2}S_n(\varphi_0)}{\hat{\sigma}_n} = \frac{\Phi_n^{1/2}D_n^{1/2}S_n(\varphi_0)}{\sigma_0} + O_p(1). \tag{5.30}
\]

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

\[
\Phi_{n}^{-1/2} D_{n}^{-1/2} = \begin{pmatrix}
I_d & 0 \\
0 & \sqrt{\frac{\Delta_n(\theta_0, \sigma_0)}{\sum_{t=2}^{n} f_t^2(\theta_0)e_{t-1}^2}}
\end{pmatrix} \begin{pmatrix}
X_n^{-1/2}(\theta_0) & 0 \\
0 & \frac{1}{\sqrt{\Delta_n(\theta_0, \sigma_0)}}
\end{pmatrix}
\]

\[
= \begin{pmatrix}
X_n^{-1/2}(\theta_0) & 0 \\
0 & \frac{1}{\sqrt{\sum_{t=2}^{n} f_t^2(\theta_0)e_{t-1}^2}}
\end{pmatrix}.
\]

(5.32)

By (5.30)–(5.32), we have

\[
\frac{\Phi_{n}^{-1/2} D_{n}^{-1/2} S_n(\psi_0)}{\sigma_0} = \frac{1}{\sigma_0} \left( \frac{n}{\dot{\sum}_{t=2} f_t(\theta_0)\eta_t e_{t-1}} + \frac{\sum_{t=2} f_t^2(\theta_0)\eta_t e_{t-1}}{\sqrt{\sum_{t=2} f_t^2(\theta_0)e_{t-1}^2}} \right).
\]

(5.33)

Let \( u \in \mathbb{R}^d \) with \(|u| = 1\), and \( a_{in} = uX_n^{-1/2}(\theta_0)(x_i - f_i(\theta_0)x_{i-1}) \). Then \( \max_{2 \leq t \leq n} a_{in} = o(1) \), and we will consider the limiting distribution of the following 2-vector:

\[
\frac{1}{\sigma_0} \left( \sum_{t=2} a_{in} \eta_t + \frac{\sum_{t=2} f_t^2(\theta_0)\eta_t e_{t-1}}{\sqrt{\sum_{t=2} f_t^2(\theta_0)e_{t-1}^2}} \right).
\]

(5.34)

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

\[
\sum_{t=2} \eta_t \left( \frac{u_1 a_{in}}{\sigma_0} + \frac{u_2 f_t^2(\theta_0)e_{t-1}}{\sigma_0 \sqrt{\Delta_n(\theta_0, \sigma_0)}} \right) = \sum_{t=2} \eta_t m_t(n),
\]

(5.35)

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

\[
\sum_{t=2} E\left( \eta_t^2 m_t^2(n) | H_{i-1} \right) = \sigma_0^2 \sum_{t=2} m_t^2(n)
\]

\[
= u_1^2 \sum_{t=2} a_{in}^2 + \frac{2u_1u_2}{\Delta_n(\theta_0, \sigma_0)} \sum_{t=2} f_t^2(\theta_0)a_{in}e_{t-1}
\]

\[
+ \frac{u_2^2}{\Delta_n(\theta_0, \sigma_0)} \sum_{t=2} f_t^2(\theta_0)e_{t-1}^2
\]

(5.36)

\[= u_1^2 + o_p(1) + u_2^2 = 1 + o_p(1).\]
Mathematical Problems in Engineering

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

\[
\max_{1 \leq t \leq n}|m_t(n)| \leq \max_{1 \leq t \leq n}\left| \frac{u_{1a_{in}}}{\sigma_0} \right| + \max_{2 \leq t \leq n}\left| \frac{u_2f'_{(\theta_0)}e_{t-1}}{\sigma_0\sqrt{\Delta_n(\theta_0, \theta_0)}} \right| = o_p(1). \tag{5.37}
\]

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

\[
\sum_{t=1}^{n} E\left\{ \eta_t^2 m_t^2(n) I\left( |\eta_t m_t(n)| > c \right) | H_{t-1} \right\} = \sum_{t=2}^{\infty} \int_{c/m_t(n)}^{\infty} y^2 dp\left\{ |\eta_t m_t(n)| \leq y | H_{t-1} \right\}
\]
\[
= \sum_{t=2}^{n} m_t^2(n) \int_{c/m_t(n)}^{\infty} y^2 dp\left\{ |\eta_t| \leq y | H_{t-1} \right\}
\]
\[
\leq \sum_{t=2}^{n} m_t^2(n) \int_{c/\delta}^{\infty} y^2 dp\left\{ |\eta_t| \leq y | H_{t-1} \right\}
\]
\[
= o_\delta \sum_{t=2}^{n} m_t^2(n) = o_\delta O_p(1) \to 0, \quad n \to \infty. \tag{5.38}
\]

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

\[
\sum_{t=1}^{n} \eta_t m_t(n) \to_D N(0, 1). \tag{5.39}
\]

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

\[
\epsilon_t = \frac{1}{2} \sin\left( \frac{\pi}{2} + \theta \right) \epsilon_{t-1} + \eta_t, \tag{6.1}
\]

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 \( \hat{\beta}^{(0)} = 3.5136, \sigma^{(0)} = 1.0347, \theta^{(0)} = \pi/2 \), and \( \delta = 0.01 \). Therefore, using the iterative computing method, we obtain

\[
\hat{\theta}^{(1)} = (1.0159, 3.5076, 1.6573)^T, \quad \hat{\theta}^{(2)} = (1.0283, 3.5076, 1.7599)^T. \tag{6.2}
\]
Since $\|\hat{\theta}^{(2)} - \hat{\theta}^{(1)}\|/\|\hat{\theta}^{(1)}\| < 0.01$, the QML estimators of $\beta, \theta,$ and $\sigma^2$ are given by

$$\hat{\sigma}^2_n = 1.0283, \quad \hat{\beta}_n = 3.5076, \quad \hat{\theta}_n = 1.7599.$$ (6.3)

These values closely approximate their true values, so our method is successful, especially in estimating the parameters $\beta$ and $\sigma^2$.

### 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

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

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

Fuller [14] obtained the estimated generalized least squares equation

$$\hat{y}_t = 2.36 + 0.72 x_{t1} - 0.80 x_{t2} - 0.81 x_{t3} - 0.92 x_{t4},$$
$$\varepsilon_t = 0.7633 \varepsilon_{t-1} + \eta_t,$$ (6.5)

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

Using our method, we obtain the following models:

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

where $\eta_t$ 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


[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

[28] L. Giraitis and H. Koul, “Estimation of the dependence parameter in linear regression with long-

[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

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.


NY, USA, 2005.

1974.


USA, 1996.


