DISSERTATION

ESTIMATION FOR SOME LINEAR AND NONLINEAR TIME SERIES MODELS

Submitted by

Rongning Wu

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2007

COLORADO STATE UNIVERSITY

June 18, 2007

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY RONGNING WU ENTITLED " ESTIMATION FOR SOME LINEAR AND NONLINEAR TIME SERIES MODELS " BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

P. J. Brockwell

Wang Haonan

Michael Kirby

Richard A. Davis

Adviser

Department Head

ABSTRACT OF DISSERTATION

ESTIMATION FOR SOME LINEAR AND NONLINEAR TIME SERIES MODELS

This dissertation concerns parameter estimation for two different classes of models. One class is parameter-driven generalized linear models (GLMs) for time series, which is an important tool in modeling non-Gaussian time series. We first consider a negative binomial logit regression model for studying time series of count data. Serial dependence among observed data is introduced by incorporating a latent process in the link function of the model. We apply a standard GLM estimation by ignoring the latent process and maximizing the resulting pseudo-likelihood. We show the consistency and asymptotic normality of the GLM estimator under two cases: where the latent process is a stationary Gaussian process, and where it is a stationary strongly mixing process. We also study parameter-driven GLMs for general time series, where the observation variable, conditional on covariates and a latent process, is assumed to have a distribution from the one-parameter exponential family. We generalize the asymptotic results of the GLM estimator under suitable conditions, and thus unify in a common framework the results for Poisson log-linear regression models (Davis, Dunsmuir, and Wang, 2000), negative binomial logit regression models, and other models.

Another class of models that we study consists of noncausal and/or noninvertible autoregressive-moving average (ARMA) models. The ARMA models are a class of linear time series models, which provides a general framework for studying stationary processes. In the classical Gaussian framework, causality and invertibility are assumed in order to eliminate the nonidentifiability of the parameterization. In a non-Gaussian setup, however, the assumptions are artificial because causal and noncausal (or invertible and noninvertible) models are identifiable. In this dissertation, we remove the assumption of causality and invertibility under non-Gaussian setups, and investigate exclusively least absolute deviation

(LAD) estimation, which is widely used in the non-Gaussian setting, especially when observations are heavy-tailed. We first consider MA(1) models. Consistency and asymptotic normality are established for the local LAD estimator, the global LAD estimator, and the linearized LAD estimator in both the invertible and noninvertible cases. Then, we investigate LAD estimation for noncausal and/or noninvertible ARMA(p, q) models. We establish a functional limit theorem for random processes, from which the asymptotic results of the LAD estimator follow.

Rongning Wu
Department of Statistics
Colorado State University
Fort Collins, CO 80523
Summer 2007

ACKNOWLEDGMENTS

I have been fortunate to benefit from many people during the course of my study at Colorado State University. I am really delighted to be able to acknowledge them upon the completion of my Ph.D. program.

First of all, I express my sincere gratitude to my adviser, Professor Richard Davis, for his constant guidance and inspiration throughout my dissertation research. I also want to express my appreciation of help and assistance of the advisory committee, Professors Peter Brockwell, Michael Kirby, and Haonan Wang. In addition, I would like to thank Professors Jay Breidt and Fang Yao for their previous service on the advisory committee. I particularly want to thank Professors Peter Brockwell and Hari Iyer for their encouragement and advice. I am also grateful to the people at the Department of Statistics for all helps I have been offered. Moreover, I acknowledge the support of NSF Grant DMS-0338109.

Finally, I am greatly indebted to my parents and my family; special thanks go to my wife, Fang Sheng.

To the memory of my mother

CONTENTS

1 INTRODUCTION

This dissertation concerns parameter estimation for two different classes of models. The first is a class of nonlinear time series models, which is an important tool in modeling non-Gaussian time series. For example, if the time series consists of counts, especially when the counts may be small, then Gaussian models are inappropriate. Examples of modeling time series of counts can be found in many applications, including time series of monthly polio counts in the USA (Zeger, 1988; Davis, Dunsmuir, and Wang, 2000), daily asthma presentation at a hospital in Sydney (Davis, Dunsmuir, and Wang, 2000), and traffic accidents in the county of Västerbotten, Sweden (Brännäs and Johansson, 1994). In the dissertation we apply the generalized linear model (GLM) estimation for this class of models. The second class of models consists of noncausal and/or noninvertible autoregressive-moving average (ARMA) models, which arise frequently in time series modeling and real applications. For example, noncausal models have been used for deconvolution of seismic signals (Wiggins, 1978) and modeling the trading volume data of Microsoft stock (Breidt, Davis, and Trindade, 2001) (see Breidt, Davis, and Trindade (2001) for a list of applications for noncausal models); and noninvertible models have appeared in vocal tract filters (Chien, Yang, and Chi, 1997), analysis of monthly time series of unemployment in the USA (Huang and Pawitan, 2000), and seismogram deconvolution (Andrews, Davis, and Breidt, 2006). We study the least absolute deviation (LAD) estimation for noncausal and/or noninvertible ARMA models.

1.1 A Class of Nonlinear Time Series Models

In the first part of this dissertation, we study parameter-driven GLMs for time series. Our research is motivated by Davis, Dunsmuir, and Wang (2000), who studied Poisson log-linear regression modeling for a time series of observed counts. In their model, the mean function was specified by a linear predictor modified by a latent process. The asymptotic properties of the standard GLM estimator were derived and applied to real-life data sets. The goal of our research is to extend their asymptotic results for Poisson log-linear regression model, first to the negative binomial logit regression and then to a general setup, assuming that the observation at a specific time, conditional on covariates and a latent process, has a distribution from a one-parameter exponential family.

We consider first GLMs for time series of count data and then GLMs for general time series. The count data cannot be satisfactorily handled within a classical Gaussian framework because they are nonnegative and integer-valued. A common phenomenon in count data is overdispersion (the variance is larger than the mean). Although the overdispersion has little effect on estimates of the regression coefficients, standard errors may be seriously in error without an appropriate accommodation of it (Xue and Deddens, 1992). Many authors have considered overdispersed Poisson and overdispersed binomial regression models for studying the overdispersion in count data. In this dissertation we consider overdispersed negative binomial regression models. The negative binomial, as well as Poisson and binomial, regression models can actually be treated as special cases of a model class: generalized linear models. As an extension of classical linear models, GLMs are a widely used family of models for analyzing count data (see McCullagh and Nelder, 1989).

On the other hand, a typical characteristic with time series data is serial dependence. Modeling time series of counts requires one to accommodate serial dependence among data as well as discreteness of data. When incorporated in serial dependence, GLMs provide a class of models for studying time series of counts. In the literature, many different approaches have been proposed to add serial dependence to the standard GLM setting. Cox (1981) classified the time series models for serially dependent data into two classes: observation-driven models and parameter-driven models. The observation-driven models insert the dependence into a model by making the current observation depend explicitly on past outcomes. In contrast, the parameter-driven models specify an underlying autocorrelated latent process to introduce dependence into observations. Both model specifications are in fact able to handle the overdispersion in count data as well. GLMs for time series that we study in this dissertation are a type of parameter-driven models. An unobserved autocorrelated process is incorporated into the link function of a GLM, which introduces serial dependence among observed data.

A parameter-driven GLM for time series consists of two components: random component and systematic component. The random component states the conditional distribution $p(Y_t|\mathbf{x}_t, \alpha_t)$ of observation variable Y_t given covariates $\mathbf{x}_t$ and a latent process α_t. Observation variables are assumed to be conditionally independent. On the other hand, the systematic component specifies a link function $f(\cdot)$ such that $f(u_t) = \mathbf{x}_t^T\boldsymbol{\beta} + \alpha_t$, where u_t is the conditional mean of Y_t. The unobserved latent process $\{\alpha_t\}$ introduces serial dependence into the model; it is assumed to evolve independently of the observed data.

Due to the existence of the latent process, the likelihood cannot be expressed in a closed form of observations and the limit theory for maximum likelihood estimator (MLE) is difficult. One way to get around the difficulty is to ignore the latent process in the model and consider the GLM estimator obtained by maximizing the resulting pseudo-likelihood. The idea of GLM estimation is well-illustrated by Wang (2002). Consider a classical linear model with time series noise, namely, $Y_t = \mathbf{x}_t^T\boldsymbol{\beta} + Z_t$, where $\{Z_t\}$ may be a stationary Gaussian process. To fit the model, we can first ignore serial dependence of $\{Z_t\}$ and estimate the parameter vector $\boldsymbol{\beta}$ using ordinary least squares (OLS) method by regressing the data vector $(Y_1, \ldots, Y_n)^T$ onto the regressor vector $\mathbf{x}_t$. Although the resulting model is misspecified, the OLS estimator of $\boldsymbol{\beta}$ has the same asymptotic efficiency as its MLE under a wide class of models for $\{Z_t\}$ (Hannan, 1970). Once a consistent estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained, the autocovariance function (ACVF) of $\{Z_t\}$ can be consistently estimated from the sample ACVF of residuals. Then, a model can be selected for $\{Z_t\}$, and the parameter vector $\boldsymbol{\beta}$ and also parameters of the model for $\{Z_t\}$ can be re-estimated using MLE.

1.2 Noncausal and/or Noninvertible ARMA Models

The second part of the dissertation considers inference for noncausal and/or noninvertible ARMA models. We investigate LAD estimation for possibly noninvertible MA(1) models (moving average models of order 1) and noncausal and/or noninvertible ARMA(p,q) models, and explore the asymptotic behavior of the LAD estimator.

The following definitions and notation are based on Brockwell and Davis (1996). A time series $\{X_t\}$ is called an ARMA process of order p and q, denoted as ARMA(p,q), if it is stationary and for every t satisfies the recursion,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \tag{1.1}$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ (white noise with mean zero and variance σ^2), $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ are constants, and the polynomials $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ have no common factors. A stationary solution $\{X_t\}$ of the equation (1.1) exists (and is also unique) if and only if $\phi(z) \neq 0$ for all $|z| = 1$. If $\theta(z) = 1$, then the equation (1.1) reduces to

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t \tag{1.2}$$

and $\{X_t\}$ is said to be an autoregressive process of order p, or AR(p). Likewise, if $\phi(z) = 1$, then the equation (1.1) reduces to

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \tag{1.3}$$

and $\{X_t\}$ is said to be an MA process of order q, or MA(q).

Causality means that X_t can be expressed in terms of current and past Z_t's. That is, there exist constants $\{\psi_j\}$ such that $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ for all t, where $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Or we can say that, the observation at a given time is caused by current and past values of the noise. Causality is equivalent to the condition that all the roots of the polynomial $\phi(z)$ are outside the unit circle in the complex plane. Then ψ_j, $j = 0, 1, \ldots,$ is the coefficient of

z^j in the Taylor series expansion of $\theta(z)/\phi(z)$, $|z| < 1$. On the other hand, if there exists any root inside the unit circle, then the model (1.1) is said to be noncausal. Furthermore, if $\phi(z)$ has all its roots inside the unit circle, i.e., $\phi(z) \neq 0$ for $|z| \geq 1$, then the model (1.1) is purely noncausal.

Likewise, an ARMA(p, q) process is said to be invertible if there exist constants $\{\pi_j\}$ such that $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$ for all t, where $\sum_{j=0}^{\infty} |\pi_j| < \infty$. Invertibility means that Z_t is expressible in terms of $\{X_s; s \leq t\}$, and is equivalent to the condition that all the roots of the polynomial $\theta(z)$ are outside the unit circle. An ARMA(p, q) process is said to be noninvertible if it is not invertible; moreover, it is purely noninvertible if Z_t is expressible in terms of $\{X_s; s \geq t\}$.

The ARMA models are a class of linear time series models, which provides a general framework for studying stationary processes (Brockwell and Davis, 1996). However, in the classical Gaussian framework, one usually assumes causality and invertibility of an ARMA model for identifiability reason. In our research, we remove the assumption of causality and invertibility under the non-Gaussian setup; the motivation of studying noncausal and/or noninvertible ARMA models relates to different model representations of $\{X_t\}$. Suppose for a causal-invertible ARMA(p, q) process, the AR polynomial $\phi(z)$ and MA polynomial $\theta(z)$ have no roots on the unit circle, and all roots of $\phi(z)$ and $\theta(z)$ are real and distinct. By flipping roots of $\phi(z)$ and $\theta(z)$ from outside the unit circle to inside the unit circle, one can obtain 2^{p+q} model representations of $\{X_t\}$ that correspond to the same autocorrelation function (ACF). For example, the following two MA(1) processes

$$\begin{aligned} X_t &= Z_t + 2Z_{t-1}, \qquad & Z_t &\sim \mathrm{WN}\left(0, \sigma^2\right), \\ Y_t &= W_t + .5W_{t-1}, \qquad & W_t &\sim \mathrm{WN}\left(0, \tau^2\right), \end{aligned} \tag{1.4}$$

have the same autocorrelation function

$$\rho(h) = \begin{cases} 1, & \text{if } h = 0, \\ 2/5, & \text{if } |h| = 1, \\ 0, & \text{if } |h| > 1. \end{cases}$$

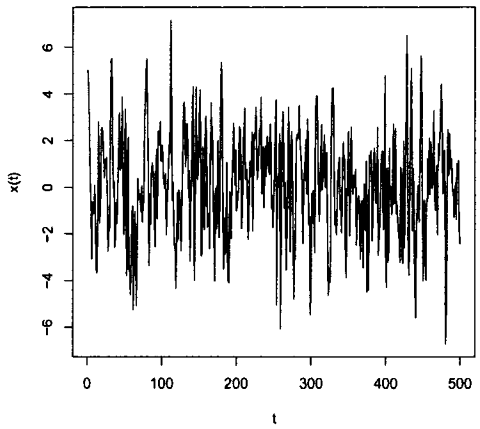

Figure 1.1: Sample path of noninvertible MA(1) process $X_t = Z_t + 2Z_{t-1}$, $Z_t \sim$ IID N(0, 1).

The first process in (1.4) is noninvertible, while the second process is invertible. Notice that the two coefficients are the reciprocal of one another.

If the underlying noise is Gaussian, then the probability structure of $\{X_t\}$ is completely determined by its mean and autocorrelation functions. It follows that the probability structure of $\{X_t\}$ is the same for all 2^{p+q} representations up to a scale factor. This implies that causal and noncausal models are not identifiable; neither are invertible and noninvertible models. In other words, there is no point to discriminate between causal and noncausal models or to discriminate between invertible and noninvertible models in the classical Gaussian setup. The causality and invertibility are assumed only to remove the nonidentifiability and configure the model uniquely. To see this, let us look at an example. In Figure 1.1, five hundred observations were simulated from the noninvertible MA(1) process $X_t = Z_t + 2Z_{t-1}$, where $Z_t \sim$ IID N(0, 1) (IID stands for independent and identically distributed). We fit noninvertible and invertible models to the data. For each fitted model, we compute the residuals and plot the ACF of residuals. From Figure 1.2 (a) and (b), we see that for each fitted model the residuals are white noise since the autocorrelations at most of the time lags ≥ 1 are within the 95% confidence bands. We further check the independence of residuals for each fitted model by plotting the ACF of their absolute values (Figure 1.2 (c) and (d)). We see that residuals from both fitted models also appear independent. This indicates that

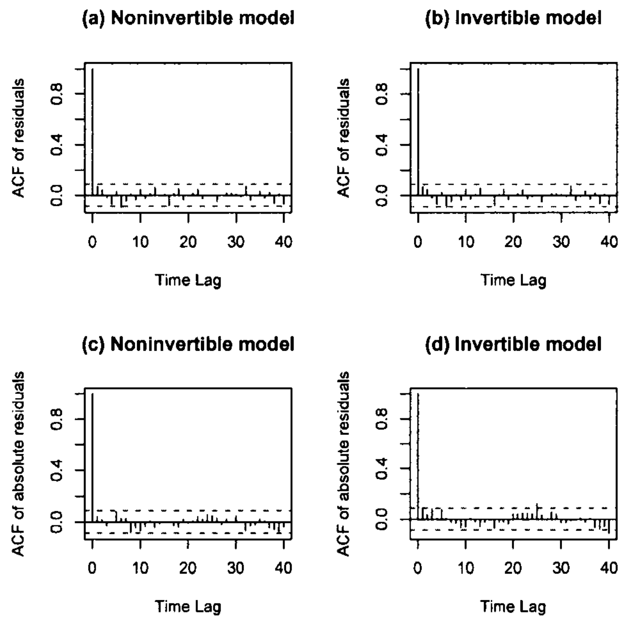

Figure 1.2: (a) ACF of residuals from noninvertible model. (b) ACF of residuals from invertible model. (c) ACF of absolute values of residuals from noninvertible model. (d) ACF of absolute values of residuals from invertible model.

both fitted models perform equally well.

On the other hand, if the underlying noise is non-Gaussian, then different configurations lead to different probability structures of $\{X_t\}$. That is, causal and noncausal models are identifiable; so are invertible and noninvertible models. For illustration, we simulated another five hundred observations from the same model $X_t = Z_t + 2Z_{t-1}$ but where $\{Z_t\}$ is IID having a Student's t distribution with 3 degrees of freedom (see Figure 1.3). As before, we fit both noninvertible and invertible models to the data, compute the residuals and plot ACFs of residuals and absolute residuals for each fitted model, which are displayed in Figure 1.4. From Figure 1.4 (a) and (b), we see that the residuals from both fitted models are white noise, same as in the previous example. However, Figure 1.4 (d) suggests that the residuals from the invertible fitting are not independent, since the absolute residuals have significant autocorrelations at time lags 1 and 2. Therefore, the noninvertible fitting is more appropriate in this example.

Now it should be clear that, for non-Gaussian ARMA processes, distinguishing between

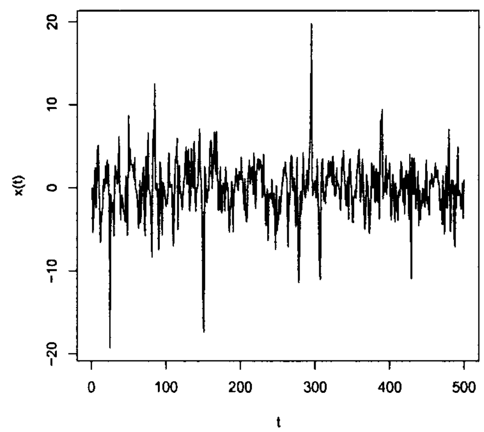

Figure 1.3: Sample path of noninvertible MA(1) process $X_t = Z_t + 2Z_{t-1}$, $Z_t \sim$ IID $t\,(3)$.

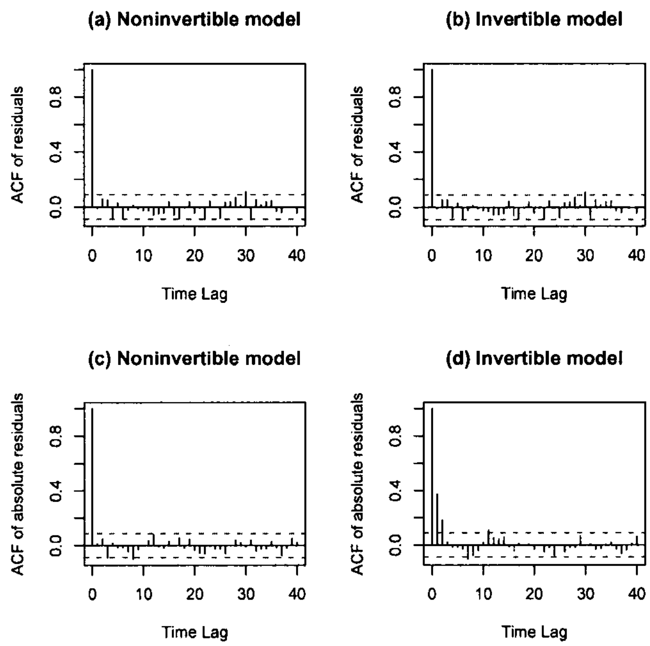

Figure 1.4: (a) ACF of residuals from noninvertible model. (b) ACF of residuals from invertible model. (c) ACF of absolute values of residuals from noninvertible model. (d) ACF of absolute values of residuals from invertible model.

invertible and noninvertible models and distinguishing between causal and noncausal models are meaningful.

In a non-Gaussian setup, if the distribution of underlying noise is known, then we can apply the likelihood approach to parameter estimation for a noncausal and/or noninvertible ARMA model. For example, Lii and Rosenblatt (1992) proposed a procedure to estimate the coefficients of a noninvertible MA process and showed the consistency and asymptotic normality of the estimator under some regularity conditions. Their method is analogous to that proposed by Breidt *et al.* (1991) for the noncausal AR process driven by IID non-Gaussian noise, who established the asymptotic results for the MLE of model parameters.

But, in practice we usually do not know the actual distribution of underlying noise. In this situation, several approaches have been suggested. For example, we may use a pseudo-likelihood method, namely, we approximate the noise distribution with some known distribution and take the corresponding likelihood as an objective function. As an alternative, we may use the LAD method, especially when observations display a heavy-tailed characteristic. Actually, an approximation to the likelihood, assuming that the underlying noise is Laplacian, yields an absolute deviation criterion.

The LAD estimation, which minimizes the absolute norm of residuals, is widely used in non-Gaussian settings. It is an analogue to the least squares method, which minimizes the Euclidean norm of residuals. The least squares method has been widely employed and performs well when the noise is Gaussian or near-Gaussian. However, in real life, time series driven by non-Gaussian noise exists abundantly; for example, it is common in econometrics to observe heavy-tailed time series, such as stock price and returns on financial assets. In such cases the LAD estimation is more appealing than the least squares method; LAD estimator is more robust since it is less sensitive to outlying observations.

1.3 Outline of Dissertation

In Chapter 2, we study parameter-driven GLMs for time series. We give a brief overview of negative binomial distributions and GLMs. A literature review on parameter estimation for parameter-driven time series models is also given. We first consider regression analysis of time series of count data in a negative binomial logit setup, and show the consistency and asymptotic normality of the GLM estimator. The proof is given under two cases: where the latent process is assumed to be a stationary Gaussian process, and where it is assumed to be a stationary strongly mixing process. Then, we extend the study to regression analysis of general time series, where the conditional distribution of the observation variable Y_t, given covariates $\mathbf{x}_t$ and a latent process α_t, is assumed to come from the one-parameter exponential family. The asymptotic results of the GLM estimator are proved in the appendix of the chapter. Also included in the appendix are technical details. A simulation study is undertaken, and an application to real data is given.

In Chapter 3, we investigate LAD estimation for MA(1) models in both the invertible and noninvertible cases. The LAD criterion is derived based on Breidt, Davis, Hsu, and Rosenblatt (2006). We show the existence of a sequence of local LAD estimators that is consistent and asymptotic normal. Consistency and asymptotic normality are established for the global LAD estimator with additional assumptions on the distribution of underlying noise. We also propose a linearized LAD estimator and derive its asymptotic properties. A numerical study is conducted to evaluate the asymptotic theory about LAD estimators. Technical details are given in the appendix of the chapter.

In Chapter 4, LAD estimation for noncausal and/or noninvertible ARMA(p, q) models are investigated. We deconstruct an ARMA(p, q) model into its causal, purely noncausal, invertible, and purely noninvertible components, and focus our study on the deconstructed model. We derive the LAD criterion by a likelihood approximation assuming the Laplacian underlying noise. With a local approximation technique, we establish a functional limit theorem for random processes, from which the consistency and asymptotic normality of the LAD estimator are established. A simulation study is conducted for ARMA$(1, 1)$ models, and applications to real data are given.

In Chapter 5, we summarize the methods and results in previous chapters, and discuss future work and open questions.

2 GLM ESTIMATION FOR A CLASS OF NONLINEAR TIME SERIES MODELS

2.1 Preliminaries

In this chapter we first study parameter-driven generalized linear models (GLMs) for time series of count data in a negative binomial logit setup. We begin by giving a brief overview of negative binomial distributions and GLMs. We also give a literature review on parameter estimation for parameter-driven models in this section.

2.1.1 Negative Binomial Distributions

The negative binomial distribution, also known as the Pascal distribution or Polya distribution, gives the probability of y failures in the first $y+r-1$ trials with a success on the $(y+r)^{th}$ trial for some fixed number $r \geq 1$. Its probability density function (pdf) is given by

$$f_Y(y) = \binom{y+r-1}{r-1} p^r (1-p)^y, \qquad y = 0, 1, 2, \ldots, \tag{2.1}$$

where p is the probability of success in each trial; and its characteristic function is given by

$$\phi_Y(s) \equiv \mathrm{E}\left(e^{isY}\right) = \left(\frac{1}{p} - \frac{1-p}{p} e^{is}\right)^{-r}. \tag{2.2}$$

We denote the negative binomial distribution with density (2.1) by NegBin(r, p). Actually, the negative binomial distribution can be derived from several models, e.g., as a generalization of the geometric distribution, or as a mixture of Poisson distributions. Accordingly, there are a variety of definitions in the literature. The negative binomial distribution has been found to provide useful representations in many fields. For wide-ranging applications of the negative binomial distribution, see Johnson, Kemp, and Kotz (2005, pp 232-233).

When modeling random counts, one usually considers using the Poisson distribution. However, many processes are not adequately modeled by Poisson distribution, e.g., when counts tend to occur in clusters. As a natural extension of the Poisson distribution, the negative binomial distribution is more flexible and allows for overdispersion. Negative binomial regression models are widely employed for analyzing count data. For example, Lawless (1987) studied negative binomial regression models and examined efficiency and robustness properties of inference procedures based on them; Xue and Deddens (1992) developed tests for extra variation relative to negative binomial regression models; Booth *et al.* (2003) extended the negative binomial log-linear model to dependent counts, where dependence among the counts is handled by including linear combinations of random effects in the linear predictor; Byers *et al.* (2003) presented a case study using the negative binomial regression model for discrete outcome data arising from a clinical trial. However, due to the difficulty of using negative binomial distributions to analyze time series of counts, there is little in the literature to fully illustrate such applications. It is worth mentioning that Benjamin *et al.* (2003) used a negative binomial conditional distribution to model the monthly cases of poliomyelitis in the USA to demonstrate the application of their GARMA model, which is a type of observation-driven models.

2.1.2 Generalized Linear Models

An important milestone in the development of regression models for count data was the emergence of GLMs. The GLMs were introduced by Nelder and Wedderburn (1972), and a large literature on the subject has since developed.

In a classical Gaussian linear model, we assume that an $n \times 1$ response vector $\mathbf{Y}$ has multivariate normal distribution with mean vector $\boldsymbol{\mu} \equiv \mathrm{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ and covariance matrix $\sigma^2\mathbf{I}$, where $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ is $n \times l$ design matrix and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_l)^T$ is l-dimensional vector of regression parameters. The linear predictor $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^T$, defined by $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, specifies the systematic component of the model. So, $\boldsymbol{\eta} = \boldsymbol{\mu}$ in the classical Gaussian framework.

There are two extensions in GLMs from classical Gaussian linear models. Firstly, the assumption of a normal distribution for $\mathbf{Y}$, which describes the random component of the model, is extended to any distribution from the exponential family. This generalization allows errors to come from Poisson, binomial, multinomial, negative binomial, gamma and inverse gamma distributions, among other possibilities. Secondly, a link function $f(\cdot)$ specifies the relation between linear predictor $\boldsymbol{\eta}$ and mean $\boldsymbol{\mu}$, namely $\boldsymbol{\eta} = f(\boldsymbol{\mu})$. The link function may be any monotonic differentiable function with inverse function $h(\cdot)$, so the mean $\boldsymbol{\mu}$ can be written in the form $\boldsymbol{\mu} = h(\mathbf{X}\boldsymbol{\beta})$.

A GLM is defined by the distribution assumed for $\mathbf{Y}$, the form of link function, and the terms in the linear predictor. Within this framework, a classical Gaussian linear model has normal observation variables and identity link; a log-linear model has Poisson observation variables and log link; a logistic regression model has binomial observation variables and logit link; and so on.

Suppose that the component Y_t of $\mathbf{Y}$ has a distribution from the one-parameter exponential family, then its density can be written as

$$f_{Y_t}(y_t; \theta_t) = \exp\{y_t\theta_t - b(\theta_t) + c(y_t)\}, \tag{2.3}$$

where $b(\cdot)$ and $c(\cdot)$ are given measurable functions. It is well-known that

$$\mu_t \equiv \mathrm{E}(Y_t) = b'(\theta_t) \qquad \text{and} \qquad \mathrm{Var}(Y_t) = b''(\theta_t), \tag{2.4}$$

where $b'(\cdot)$ and $b''(\cdot)$ denote the first and second derivatives of function $b(\cdot)$, respectively. The link function $f(\cdot)$ is called a canonical link if $\eta_t = f(\mu_t) = \theta_t$. Suppose $Y_t \overset{\text{indep}}{\sim} \mathrm{NegBin}(r, p_t)$, for $t = 1, \ldots, n$. That is, the observation variables $Y_1, \cdots, Y_n$ are independent, each having a negative binomial distribution. If we express the density of Y_t in form

(2.3), then we have

$$\begin{aligned}
\theta_t &= \log(1-p_t), \\
b(\theta_t) &= -r\log p_t = -r\log\left(1-e^{\theta_t}\right), \\
c(y_t) &= \log\binom{y_t+r-1}{r-1},
\end{aligned}$$

and

$$\mu_t = \frac{r(1-p_t)}{p_t} = \frac{re^{\theta_t}}{1-e^{\theta_t}}.$$

The canonical link function for negative binomial data is given by

$$\theta_t = \log\left(\frac{\mu_t}{\mu_t+r}\right) = \eta_t.$$

Nelder and Wedderburn (1972) gave a unified procedure for model fitting based on likelihood. For GLMs with canonical link function, Haberman (1977) provided conditions under which the MLE is consistent and asymptotically normal. For GLMs with a non-canonical link function, consistency and asymptotic normality of the MLE were established by Fahrmeir and Kaufmann (1985). Wedderburn (1974) discussed estimation for GLMs based on quasi-likelihood function. The asymptotic results of the maximum quasi-likelihood estimator of the regression coefficient vector were given by McCullagh (1983). For details of inference based on the likelihood function, see McCullagh and Nelder (1989).

2.1.3 Literature Review

The analysis of time series of count data is one of the rapidly developing areas in time series modeling, and has been investigated by researchers in a great variety of contexts. In addition to applications mentioned in Chapter 1, Campbell (1994) investigated the relationship between sudden infant death syndrome and environmental temperature; Johansson (1996) used time series of counts to assess the effect of lowered speed limits on the number of road casualties; Jørgensen *et al.* (1996) studied the relationship between respiratory morbidity and air pollution; Cameron and Trivedi (1996) discussed count data models for financial data. Good reviews can be found in MacDonald and Zucchini (1997) and Cameron and Trivedi (1998).

An overview of parameter driven models for time series of counts can be found in Zeger (1988), Davis, Dunsmuir, and Wang (1999), and Kelsall, Zeger, and Samet (1999). Zeger (1988) studied Poisson log-linear regression models for regression analysis with a time series of counts, where serial dependence is introduced by adding a latent process to the linear predictor. An estimating equation approach was used for parameter estimation, and asymptotic results of quasi-likelihood estimator were established. Blais *et al.* (2000) extended Zeger's asymptotic results for Poisson log-linear regression to the GLM setup where each observation, conditional on a latent process, is assumed to have an exponential family distribution, and the latent process is assumed to be stationary and strongly mixing. Gourieroux, Monfort, and Trognon (1984) applied a pseudo maximum likelihood method to Poisson log-linear regression models, where the latent process is IID, and showed the consistency and asymptotic normality of the pseudo MLE. Davis, Dunsmuir, and Wang (2000) developed a practical approach to diagnosing the existence of latent process in Poisson log-linear regression models, and derived the asymptotic properties of GLM estimator for the case where an autocorrelated latent process is present. Harvey and Fernandes (1989) studied a structural model for time series of counts and qualitative data using natural-conjugate distributions. Jørgensen *et al.* (1999) proposed a nonstationary state space model for multivariate longitudinal count data driven by a latent gamma Markov process; and estimation was based on the Kalman smoother.

Since the likelihood of parameter-driven models for time series of counts is very complex, Chan and Ledolter (1995) proposed a Monte Carlo EM (MCEM) algorithm and showed that the resulting estimator of an MCEM algorithm gets close to MLE with a high probability. Kuk and Cheng (1997) proposed a Monte Carlo Newton-Raphson procedure as a viable alternative to the MCEM algorithm; it is computationally more efficient than the MCEM algorithm. Durbin and Koopman (1997) considered state space models for observations having a non-Gaussian distribution, and developed a likelihood estimation procedure using importance sampling. Kuk (1999) showed that Durbin and Koopman's method is closely related to a method proposed by Geyer (1994) for simulating the likelihood of a random-effect model and to a method proposed by Schall (1991) for approximating the maximum likelihood estimate of a generalized linear mixed model. Kuk (1999) proposed a hybrid method for approximating the likelihood function as opposed to Durbin and Koopman's pointwise approximation. Durbin and Koopman (2000) discussed the analysis of non-Gaussian time series using state space models from both classical and Bayesian perspectives.

2.2 GLM Estimator for Regression of Time Series of Count Data

In this section we consider regression analysis of time series of count data, where the observed process of counts is modeled by a negative binomial distribution based GLM and an unobservable latent process is incorporated in the mean function of the model to introduce serial dependence among observed data. Like Poisson regression, the negative binomial regression is one of the basic models for studying count data. Actually, Poisson(λ) is a limiting case of NegBin(r, p) as $\lambda = r(1-p)$ and $r \to \infty$.

2.2.1 Setup

Let $\{Y_t : t = 1, \dots, n\}$ denote a time series of counts and suppose that for each t, $\mathbf{x}_t$ is an l-dimensional vector of observed regressors whose first component is 1. In some cases $\mathbf{x}_t$ may depend on the sample size n and form a triangular array $\mathbf{x}_{nt}$. We assume that, conditional on the regressor $\{\mathbf{x}_{nt}\}$ and some latent process $\{\alpha_t\}$, the random variables $Y_1, \dots, Y_n$ are independent, where the conditional distribution of Y_t depends only on α_t and $\mathbf{x}_{nt}$ and is specified by a negative binomial distribution. To be specific, we consider the following parameter-driven model:

$$Y_t|\alpha_t, \mathbf{x}_{nt} \overset{\text{indep}}{\sim} \text{NegBin}\left(r, p_t\right), \tag{2.5}$$

where p_t satisfies the logit model

$$\log \frac{p_t}{1-p_t} = \mathbf{x}_{nt}^T \boldsymbol{\beta} + \alpha_t. \tag{2.6}$$

So, the conditional density function is given by

$$\mathrm{P}\left(Y_t = y_t | \alpha_t, \mathbf{x}_{nt}\right) = \binom{y_t + r - 1}{r-1} p_t^r \left(1 - p_t\right)^{y_t}$$

for $y_t = 0, 1, \ldots$. In (2.6), $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_l)^T$ is the vector of regression coefficients of interest, and we assume that $\mathbf{x}_{nt}$ is fixed and that $\{\alpha_t\}$ is a stationary Gaussian linear process. Instead of dealing with $\{\alpha_t\}$ in the derivation, it is more convenient to use process $\{\epsilon_t\}$, where $\epsilon_t \equiv e^{-\alpha_t}$. Note that $\{\epsilon_t\}$ is a strictly stationary nonnegative time series. Furthermore, we center $\{\alpha_t\}$ such that $\{\epsilon_t\}$ has mean 1 and autocovariance function (ACVF)

$$\gamma_\epsilon(h) = \mathrm{E}\left[\left(\epsilon_{t+h} - 1\right)\left(\epsilon_t - 1\right)\right].$$

The condition of $\mathrm{E}(\epsilon_t) = 1$ is imposed for identification purposes; without this assumption, one cannot specify the mean of ϵ_t (see Wang, 2002). The conditional mean of Y_t given α_t can be written in terms of ϵ_t:

$$\mathrm{E}\left(Y_t | \alpha_t\right) = \frac{r\left(1 - p_t\right)}{p_t} = r \exp\left(-\mathbf{x}_{nt}^T \boldsymbol{\beta} - \alpha_t\right) = r \exp\left(-\mathbf{x}_{nt}^T \boldsymbol{\beta}\right) \epsilon_t. \tag{2.7}$$

Since $\{\alpha_t\}$ is a stationary Gaussian linear process, one can derive an explicit relationship between ACVF's of $\{\epsilon_t\}$ and $\{\alpha_t\}$. To be specific, in order to satisfy the identifiability condition of $\mathrm{E}(e^{-\alpha_t}) = 1$, it is required that $\alpha_t \sim \mathrm{N}\left(\sigma_\alpha^2/2, \sigma_\alpha^2\right)$, where σ_α^2 (or $\gamma_\alpha(0)$) is the variance of process $\{\alpha_t\}$. The connection between two ACVF's $\gamma_\epsilon(h)$ and $\gamma_\alpha(h)$ is given by $\gamma_\epsilon(h) = \exp\left(\gamma_\alpha(h)\right) - 1$ for all h.

2.2.2 Asymptotic Properties of GLM Estimator

The GLM estimator $\widehat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$ is obtained by ignoring $\{\alpha_t\}$ in the model and maximizing log-likelihood function of the misspecified GLM. We investigate the large sample behavior of $\widehat{\boldsymbol{\beta}}_n$, and show that $\widehat{\boldsymbol{\beta}}_n$ is consistent and asymptotically normal under suitable conditions.

Before deriving the asymptotic properties of $\widehat{\boldsymbol{\beta}}_n$, we evaluate the first and second moments of the observed process $\{Y_t\}$. The mean, variance, and covariance are conditional upon the linear regressors $\mathbf{x}_{nt}$. By (2.7), we obtain the mean μ_t of Y_t as follows:

$$\mu_t = \mathrm{E}\left(Y_t\right) = \mathrm{E}\left[\mathrm{E}\left(Y_t|\alpha_t\right)\right] = \mathrm{E}\left[r \exp\left(-\mathbf{x}_{nt}^T\boldsymbol{\beta}\right)\epsilon_t\right] = re^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}},$$

where the last equality follows from the assumption of $\mathrm{E}(\epsilon_t) = 1$. From the logit model (2.6), we obtain

$$\begin{aligned}
p_t &= \frac{e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}+\alpha_t}}{1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}+\alpha_t}} = \frac{1}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}-\alpha_t}} = \frac{1}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}}\epsilon_t}, \\
q_t &\equiv 1-p_t = \frac{1}{1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}+\alpha_t}} = \frac{e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}-\alpha_t}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}-\alpha_t}} = \frac{e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}}\epsilon_t}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}}\epsilon_t}.
\end{aligned}$$

Then, the variance of Y_t is

$$\begin{aligned}
\mathrm{Var}\left(Y_t\right) &= \mathrm{E}\left[\mathrm{Var}\left(Y_t|\alpha_t\right)\right] + \mathrm{Var}\left[\mathrm{E}\left(Y_t|\alpha_t\right)\right] \\
&= \mathrm{E}\left[\frac{r\left(1-p_t\right)}{p_t^2}\right] + \mathrm{Var}\left[r\exp\left(-\mathbf{x}_{nt}^T\boldsymbol{\beta}\right)\epsilon_t\right] \\
&= \mathrm{E}\left[re^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}}\epsilon_t\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}}\epsilon_t\right)\right] + r^2\exp\left(-2\mathbf{x}_{nt}^T\boldsymbol{\beta}\right)\mathrm{Var}\left(\epsilon_t\right) \\
&= re^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}} + re^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}}\left(\gamma_\epsilon\left(0\right)+1\right) + r^2e^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}}\gamma_\epsilon\left(0\right) \\
&= \mu_t + \mu_t^2\frac{\gamma_\epsilon\left(0\right)+1}{r} + \mu_t^2\gamma_\epsilon\left(0\right),
\end{aligned}$$

and the ACVF of Y_t is

$$\begin{aligned}
\mathrm{Cov}(Y_{t+h}, Y_t) &= \mathrm{Cov}\left[\mathrm{E}\left(Y_{t+h}|\alpha_{t+h}\right), \mathrm{E}\left(Y_t|\alpha_t\right)\right] + 0 \\
&= \mathrm{Cov}\left(re^{-\mathbf{x}_{n,t+h}^T\boldsymbol{\beta}}\epsilon_{t+h}, re^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}}\epsilon_t\right)
\end{aligned}$$

$$= r^2 e^{-(\mathbf{x}_{n,t+h}^T + \mathbf{x}_{nt}^T)\boldsymbol{\beta}} \gamma_\epsilon(h)$$
$$= \mu_{t+h}\mu_t \gamma_\epsilon(h),$$

for $h \neq 0$.

Let $Y_1, \ldots, Y_n$ be observations from model (2.5) and (2.6) with the true parameter vector $\boldsymbol{\beta}_0$. The GLM estimator $\widehat{\boldsymbol{\beta}}_n$ is obtained by maximizing the following pseudo log-likelihood function:

$$\begin{aligned} l(\boldsymbol{\beta}) &= \log\left\{\prod_{t=1}^{n} \binom{Y_t+r-1}{r-1} p_t^r (1-p_t)^{Y_t}\right\} \\ &= r\sum_{t=1}^{n} \log p_t + \sum_{t=1}^{n} Y_t \log(1-p_t) + \log \prod_{t=1}^{n} \binom{Y_t+r-1}{r-1} \\ &= -r\sum_{t=1}^{n} \log\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}}\right) - \sum_{t=1}^{n} Y_t \log\left(1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}}\right) + \log \prod_{t=1}^{n} \binom{Y_t+r-1}{r-1}. \end{aligned} \tag{2.8}$$

Assumptions on the regressors $\mathbf{x}_{nt}$ are needed in order to establish the asymptotic results of $\widehat{\boldsymbol{\beta}}_n$. We assume that there exists a sequence of nonsingular matrices $\mathbf{M}_n$ such that the regressors $\mathbf{x}_{nt}$ satisfy the following conditions:

$$\mathbf{M}_n^T \left(\sum_{t=1}^{n} \mathbf{x}_{nt}\mathbf{x}_{nt}^T e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\mathbf{M}_n \rightarrow \frac{\Omega_1(\boldsymbol{\beta}_0)}{r}, \tag{2.9}$$

$$\mathbf{M}_n^T \left(\sum_{t=1}^{n} \mathbf{x}_{nt}\mathbf{x}_{nt}^T e^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\mathbf{M}_n \rightarrow \frac{\Omega_2(\boldsymbol{\beta}_0)}{r(\gamma_\epsilon(0)+1)}, \tag{2.10}$$

and

$$\mathbf{M}_n^T \left(\sum_{t=1}^{n} \mathbf{x}_{nt}\mathbf{x}_{n,t+h}^T e^{-(\mathbf{x}_{nt}^T+\mathbf{x}_{n,t+h}^T)\boldsymbol{\beta}_0}\right)\mathbf{M}_n \rightarrow \frac{\Omega_h(\boldsymbol{\beta}_0)}{r^2} \tag{2.11}$$

uniformly in h as $n \to \infty$. Further, for $h \leq 0$,

$$\mathbf{M}_n^T \left(\sum_{t=1}^{1-h} \mathbf{x}_{nt}\mathbf{x}_{n,t+h}^T e^{-(\mathbf{x}_{nt}^T+\mathbf{x}_{n,t+h}^T)\boldsymbol{\beta}_0}\right)\mathbf{M}_n \rightarrow 0 \tag{2.12}$$

with the left-hand side being uniformly bounded in h as $n \to \infty$, and for $h > 0$,

$$\mathbf{M}_n^T \left(\sum_{t=n-h}^{n} \mathbf{x}_{nt}\mathbf{x}_{n,t+h}^T e^{-(\mathbf{x}_{nt}^T+\mathbf{x}_{n,t+h}^T)\boldsymbol{\beta}_0}\right)\mathbf{M}_n \rightarrow 0 \tag{2.13}$$

with the left-hand side again being uniformly bounded in h as $n \to \infty$. (See Davis, Dunsmuir, and Wang (2000) for a wide range of regressors that satisfy the conditions.)

We define

$$\begin{aligned}
\Sigma_n(\boldsymbol{\beta}_0) &\equiv \operatorname{Var}\left[\mathbf{M}_n^T \sum_{t=1}^n \mathbf{x}_{nt}(Y_t - \mu_t)\right] \\
&= \mathbf{M}_n^T \sum_{t=1}^n \mathbf{x}_{nt}\mathbf{x}_{nt}^T \operatorname{Var}(Y_t)\,\mathbf{M}_n + \mathbf{M}_n^T \sum_{s,t=1,s\neq t}^n \mathbf{x}_{ns}\mathbf{x}_{nt}^T \operatorname{Cov}(Y_s, Y_t)\mathbf{M}_n \\
&= \mathbf{M}_n^T \sum_{t=1}^n \mathbf{x}_{nt}\mathbf{x}_{nt}^T \left[\mu_t + \mu_t^2 \frac{\gamma_\epsilon(0)+1}{r} + \mu_t^2\gamma_\epsilon(0)\right]\mathbf{M}_n \\
&\quad + \mathbf{M}_n^T \sum_{s,t=1,s\neq t}^n \mathbf{x}_{ns}\mathbf{x}_{nt}^T \mu_s\mu_t\gamma_\epsilon(s-t)\,\mathbf{M}_n \\
&= \mathbf{M}_n^T \sum_{t=1}^n \mathbf{x}_{nt}\mathbf{x}_{nt}^T \left[\mu_t + \mu_t^2 \frac{\gamma_\epsilon(0)+1}{r}\right]\mathbf{M}_n + \mathbf{M}_n^T \sum_{s,t=1}^n \mathbf{x}_{ns}\mathbf{x}_{nt}^T \mu_s\mu_t\gamma_\epsilon(s-t)\,\mathbf{M}_n.
\end{aligned}$$

Then the asymptotic results of $\widehat{\boldsymbol{\beta}}_n$ when the latent process $\{\alpha_t\}$ is a stationary Gaussian linear process are summarized in the following theorem.

Theorem 2.1 *Let $\widehat{\boldsymbol{\beta}}_n$ be the GLM estimator of parameter vector $\boldsymbol{\beta}$ in model (2.5) and (2.6). Assume that the regressors $\mathbf{x}_{nt}$ satisfy (2.9)-(2.13), $\sup_{1\le t\le n}\left|\mathbf{M}_n^T\mathbf{x}_{nt}\right| = O(1/\sqrt{n})$, and $\sum_{h=0}^{\infty}|\gamma_\epsilon(h)| < \infty$. Then*

$$\Sigma_n(\boldsymbol{\beta}_0) \to \Omega_1(\boldsymbol{\beta}_0) + \Omega_2(\boldsymbol{\beta}_0) + \Omega_3(\boldsymbol{\beta}_0),$$

where $\Omega_3(\boldsymbol{\beta}_0) = \sum_{h=-\infty}^{\infty}\Omega_h(\boldsymbol{\beta}_0)\gamma_\epsilon(h)$. Moreover, $\widehat{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}_0$ in probability and

$$\mathbf{M}_n^{-1}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right) \xrightarrow{d} N(\mathbf{0}, \left[\Omega_1^\dagger(\boldsymbol{\beta}_0)\right]^{-1}\left[\Omega_1^\ddagger(\boldsymbol{\beta}_0) + \Omega_2^\ddagger(\boldsymbol{\beta}_0) + \Omega_3^\ddagger(\boldsymbol{\beta}_0)\right]\left[\Omega_1^\dagger(\boldsymbol{\beta}_0)\right]^{-1})$$

as $n \to \infty$, where

$$\Omega_1^\dagger(\boldsymbol{\beta}_0) = \lim_{n\to\infty} \mathbf{M}_n^T \sum_{t=1}^n \frac{\mathbf{x}_{nt}\mathbf{x}_{nt}^T\mu_t}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\mathbf{M}_n,$$

$$\Omega_1^{\ddagger}(\boldsymbol{\beta}_0) = \lim_{n\to\infty} \mathbf{M}_n^T \sum_{t=1}^{n} \frac{\mathbf{x}_{nt}\mathbf{x}_{nt}^T\mu_t}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2}\mathbf{M}_n,$$

$$\Omega_2^{\ddagger}(\boldsymbol{\beta}_0) = \frac{\gamma_\epsilon(0)+1}{r} \lim_{n\to\infty} \mathbf{M}_n^T \sum_{t=1}^{n} \frac{\mathbf{x}_{nt}\mathbf{x}_{nt}^T\mu_t^2}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2}\mathbf{M}_n,$$

and

$$\Omega_3^{\ddagger}(\boldsymbol{\beta}_0) = \sum_{h=-\infty}^{\infty} \Omega_h^{\ddagger}(\boldsymbol{\beta}_0)\gamma_\epsilon(h),$$

with

$$\Omega_h^{\ddagger}(\boldsymbol{\beta}_0) = \lim_{n\to\infty} \mathbf{M}_n^T \sum_{t=1}^{n} \frac{\mathbf{x}_{nt}\mathbf{x}_{n,t+h}^T\mu_t\mu_{t+h}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\left(1+e^{-\mathbf{x}_{n,t+h}^T\boldsymbol{\beta}_0}\right)}\mathbf{M}_n.$$

Proof. Maximizing $l(\boldsymbol{\beta})$ in (2.8) is equivalent to minimizing

$$\begin{aligned} -[l(\boldsymbol{\beta}) - l(\boldsymbol{\beta}_0)] &= r\sum_{t=1}^{n}\left[\log\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}}\right) - \log\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\right] \\ &\quad + \sum_{t=1}^{n} Y_t\left[\log\left(1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}}\right) - \log\left(1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\right] \\ &= r\sum_{t=1}^{n}\left[\log\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}}\right) - \log\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\right] \\ &\quad + r\sum_{t=1}^{n} e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\left[\log\left(1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}}\right) - \log\left(1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\right] \\ &\quad + \sum_{t=1}^{n}\left(Y_t - re^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\left[\log\left(1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}}\right) - \log\left(1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\right]. \end{aligned} \tag{2.14}$$

Note that

$$\log\left(1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}}\right) = \log\left[\left(e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}}+1\right)e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}}\right] = \log\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}}\right) + \mathbf{x}_{nt}^T\boldsymbol{\beta},$$

and similarly $\log\left(1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right) = \log\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right) + \mathbf{x}_{nt}^T\boldsymbol{\beta}_0$. We can rewrite equation (2.14) as

$$-[l(\boldsymbol{\beta}) - l(\boldsymbol{\beta}_0)] = r\sum_{t=1}^{n}\left[\log\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}}\right) - \log\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\right]$$

$$
\begin{aligned}
& +r\sum_{t=1}^{n} e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\left[\log\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}}\right)+\mathbf{x}_{nt}^T\boldsymbol{\beta}-\log\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right] \\
& +\sum_{t=1}^{n}\left(Y_t-re^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\left[\log\left(1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}}\right)-\log\left(1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\right] \\
= \; & r\sum_{t=1}^{n} e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\mathbf{x}_{nt}^T\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right) \\
& +r\sum_{t=1}^{n}\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\left[\log\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}}\right)-\log\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\right] \\
& +\sum_{t=1}^{n}\left(Y_t-re^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\left[\log\left(1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}}\right)-\log\left(1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\right]. \qquad (2.15)
\end{aligned}
$$

We define vector $\mathbf{u}$ by

$$\mathbf{u} \equiv \mathbf{M}_n^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right),$$

and function $g_n(\mathbf{u})$ by

$$g_n(\mathbf{u}) \equiv -\left[l(\boldsymbol{\beta})-l(\boldsymbol{\beta}_0)\right] = -l\left(\boldsymbol{\beta}_0+\mathbf{M}_n\mathbf{u}\right)+l(\boldsymbol{\beta}_0).$$

Then

$$g_n(\mathbf{u}) = g_{n,1}(\mathbf{u}) - g_{n,2}(\mathbf{u}),$$

where

$$
\begin{aligned}
g_{n,1}(\mathbf{u}) \equiv \; & r\sum_{t=1}^{n} e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u} \\
& +r\sum_{t=1}^{n}\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\left[\log\left(1+e^{-\mathbf{x}_{nt}^T(\boldsymbol{\beta}_0+\mathbf{M}_n\mathbf{u})}\right)-\log\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\right],
\end{aligned}
$$

and

$$g_{n,2}(\mathbf{u}) \equiv -\sum_{t=1}^{n}\left(Y_t-re^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\left[\log\left(1+e^{\mathbf{x}_{nt}^T(\boldsymbol{\beta}_0+\mathbf{M}_n\mathbf{u})}\right)-\log\left(1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\right].$$

For any fixed $\mathbf{u}$, we will show that

$$g_{n,1}(\mathbf{u}) \to \frac{1}{2}\mathbf{u}^T \Omega_1^\dagger(\boldsymbol{\beta}_0)\mathbf{u} \tag{2.16}$$

and

$$g_{n,2}(\mathbf{u}) \xrightarrow{d} \mathbf{u}^T \cdot \mathrm{N}(\mathbf{0}, \Omega_1^\ddagger(\boldsymbol{\beta}_0) + \Omega_2^\ddagger(\boldsymbol{\beta}_0) + \Omega_3^\ddagger(\boldsymbol{\beta}_0)). \tag{2.17}$$

Consider (2.16) first. Let function $\vartheta(x) \equiv \log(1+e^x)$, then its first three derivatives are

$$\vartheta'(x) = \frac{e^x}{1+e^x}, \qquad \vartheta''(x) = \frac{e^x}{(1+e^x)^2}, \qquad \text{and} \qquad \vartheta^{(3)}(x) = \frac{e^x(1-e^x)}{(1+e^x)^3}.$$

Since $\sup_{1\le t\le n}\left|\mathbf{M}_n^T\mathbf{x}_{nt}\right| \to 0$, expanding $\log\left(1+e^{-\mathbf{x}_{nt}^T(\boldsymbol{\beta}_0+\mathbf{M}_n\mathbf{u})}\right)$ in a third degree Taylor series about $-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0$ yields

$$\begin{aligned}\log\left(1+e^{-\mathbf{x}_{nt}^T(\boldsymbol{\beta}_0+\mathbf{M}_n\mathbf{u})}\right) &= \log\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right) - \frac{e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u} \\ &\quad + \frac{e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{2\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2}\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^2 - C_{nt}^1\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^3,\end{aligned}$$

where $C_{nt}^1 = \vartheta^{(3)}\left(-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \delta_{nt}^1\right)/3!$ for some δ_{nt}^1 between 0 and $-\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}$. It follows that

$$\begin{aligned} g_{n,1}(\mathbf{u}) &= r\sum_{t=1}^n e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u} \\ &\quad + r\sum_{t=1}^n\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\left[-\frac{e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u} + \frac{e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{2\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2}\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^2\right] \\ &\quad - r\sum_{t=1}^n C_{nt}^1\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^3 \\ &= \frac{r}{2}\sum_{t=1}^n\frac{e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^2 + E_n^1(\mathbf{u}),\end{aligned}$$

where $E_n^1(\mathbf{u}) = -r\sum_{t=1}^n C_{nt}^1\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^3$. We will see that $E_n^1(\mathbf{u}) \to 0$, and hence (2.16) follows.

Turning to (2.17), we similarly expand $\log\left(1+e^{\mathbf{x}_{nt}^T(\boldsymbol{\beta}_0+\mathbf{M}_n\mathbf{u})}\right)$ in a second degree Taylor series about $\mathbf{x}_{nt}^T\boldsymbol{\beta}_0$ and obtain

$$\log\left(1+e^{\mathbf{x}_{nt}^T(\boldsymbol{\beta}_0+\mathbf{M}_n\mathbf{u})}\right)=\log\left(1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)+\frac{e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}+C_{nt}^2\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^2,$$

where $C_{nt}^2=\vartheta''\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\delta_{nt}^2\right)/2!$ and δ_{nt}^2 lies between 0 and $\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}$. Then,

$$\begin{aligned}
g_{n,2}(\mathbf{u}) &= -\sum_{t=1}^{n}\left(Y_t-re^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\left[\frac{e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{1+e^{\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}+C_{nt}^2\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^2\right]\\
&= -\sum_{t=1}^{n}\left(Y_t-re^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\frac{\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}-\sum_{t=1}^{n}C_{nt}^2\left(Y_t-re^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^2\\
&\equiv -\mathbf{u}^T\mathbf{U}_n-E_n^2(\mathbf{u}),
\end{aligned}$$

where

$$\mathbf{U}_n=\sum_{t=1}^{n}\left(Y_t-re^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\frac{\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}} \tag{2.18}$$

and $E_n^2(\mathbf{u})=\sum_{t=1}^{n}C_{nt}^2\left(Y_t-re^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^2$. We will see that $E_n^2(\mathbf{u})\to 0$ in probability. So, in order to show (2.17), it suffices to show

$$\mathbf{U}_n\xrightarrow{d}\mathrm{N}(\mathbf{0},\Omega_1^\ddagger(\boldsymbol{\beta}_0)+\Omega_2^\ddagger(\boldsymbol{\beta}_0)+\Omega_3^\ddagger(\boldsymbol{\beta}_0)).$$

Recall that $Y_t|\alpha_t\overset{\text{indep}}{\sim}\mathrm{NegBin}(r,p_t)$, and then the characteristic function of Y_t given α_t is

$$\phi_{Y_t|\alpha_t}(s)=\mathrm{E}\left(e^{isY_t}|\alpha_t\right)=\left(\frac{1}{p_t}-\frac{1-p_t}{p_t}e^{is}\right)^{-r}.$$

We rewrite $\mathbf{U}_n$ as

$$\mathbf{U}_n=-r\sum_{t=1}^{n}\frac{\mathbf{M}_n^T\mathbf{x}_{nt}e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}+\sum_{t=1}^{n}\frac{\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}Y_t.$$

For every real vector $\mathbf{s}$,

$$\begin{aligned}
\mathrm{E}\left(e^{i\mathbf{s}^T\mathbf{U}_n}|\alpha_t\right) &= \exp\left(-r\sum_{t=1}^{n} i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\frac{e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right)\times\prod_{t=1}^{n}\mathrm{E}\left[\exp\left(\frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}Y_t\right)|\alpha_t\right] \\
&= \exp\left(-r\sum_{t=1}^{n} i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\frac{e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right) \\
&\quad\times\prod_{t=1}^{n}\left[\left(1+\epsilon_t e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)-\epsilon_t e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\exp\left(\frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right)\right]^{-r},
\end{aligned}$$

and

$$\begin{aligned}
\log\mathrm{E}\left(e^{i\mathbf{s}^T\mathbf{U}_n}|\alpha_t\right) &= -r\sum_{t=1}^{n} i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\frac{e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}} \\
&\quad -r\sum_{t=1}^{n}\log\left\{1-\epsilon_t e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\left[\exp\left(\frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right)-1\right]\right\}.
\end{aligned}$$

Let $Q_{nt}\equiv\epsilon_t e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\left[\exp\left(\frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right)-1\right]$. It is easy to show that $\forall\ \varepsilon>0$

$$\mathrm{P}\left(\max_{1\le t\le n}|Q_{nt}|\ge\varepsilon\right)\to 0 \qquad \text{as } n\to\infty.$$

Applying a Taylor expansion to $\log(1-Q_{nt})$, we get

$$\begin{aligned}
&\sum_{t=1}^{n}\log(1-Q_{nt})\,1_{\left\{\max_{1\le t\le n}|Q_{nt}|<\varepsilon\right\}} \\
&\quad= -\sum_{t=1}^{n}\left[Q_{nt}+\tfrac{(Q_{nt})^2}{2}\right]1_{\left\{\max_{1\le t\le n}|Q_{nt}|<\varepsilon\right\}}-E_n^3 1_{\left\{\max_{1\le t\le n}|Q_{nt}|<\varepsilon\right\}},
\end{aligned}$$

where $1_{\{\cdot\}}$ is indicator function and $E_n^3=\sum_{t=1}^{n}\sum_{j=3}^{\infty}(Q_{nt})^j/j$. For n large enough, therefore,

$$\begin{aligned}
\log\mathrm{E}\left(e^{i\mathbf{s}^T\mathbf{U}_n}|\alpha_t\right) &= -r\sum_{t=1}^{n} i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\frac{e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}+r\sum_{t=1}^{n}\left[Q_{nt}+\frac{(Q_{nt})^2}{2}\right]+rE_n^3 \\
&= \{r\sum_{t=1}^{n}e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\left[\exp\left(\frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right)-1-\frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right]
\end{aligned}$$

$$
\begin{aligned}
& -\frac{(\gamma_\epsilon(0)+1)r}{2}\sum_{t=1}^{n}\frac{\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2 e^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2}\} \\
& +\{r\sum_{t=1}^{n}(\epsilon_t-1)e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\left[\exp\left(\frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right)-1\right] \\
& +\frac{(\gamma_\epsilon(0)+1)r}{2}\sum_{t=1}^{n}\frac{\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2 e^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2} \\
& +\frac{r}{2}\sum_{t=1}^{n}\epsilon_t^2 e^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\left[\exp\left(\frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right)-1\right]^2\}+rE_n^3 \\
\equiv\; & D_n+F_n+rE_n^3.
\end{aligned}
$$

That is,

$$
\mathrm{E}\left(e^{i\mathbf{s}^T\mathbf{U}_n}|\alpha_t\right)=\exp\left(D_n+F_n+rE_n^3\right).
$$

We will see that $E_n^3\to 0$ in probability. As to D_n, we have

$$
\begin{aligned}
D_n &= r\sum_{t=1}^{n}e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\left[\exp\left(\frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right)-1-\frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right] \\
&\quad -\frac{(\gamma_\epsilon(0)+1)r}{2}\sum_{t=1}^{n}\frac{\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2 e^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2} \\
&= -\frac{r}{2}\sum_{t=1}^{n}\frac{\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2 e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2}+E_n^4-\frac{(\gamma_\epsilon(0)+1)r}{2}\sum_{t=1}^{n}\frac{\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2 e^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2},
\end{aligned}
$$

where $E_n^4=r\sum_{t=1}^{n}e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\sum_{j=3}^{\infty}\frac{1}{j!}\frac{\left(i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^j}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^j}$. We will see that $E_n^4\to 0$. So,

$$
D_n\to-\frac{1}{2}\mathbf{s}^T\left[\Omega_1^\ddagger(\boldsymbol{\beta}_0)+\Omega_2^\ddagger(\boldsymbol{\beta}_0)\right]\mathbf{s}. \tag{2.19}
$$

Moving on to F_n, we will show

$$
F_n\xrightarrow{d}V, \tag{2.20}
$$

where V is a Gaussian random variable with mean zero and variance $\mathbf{s}^T\Omega_3^{\ddagger}(\boldsymbol{\beta}_0)\,\mathbf{s}$, by showing

$$|F_n - iC_n(\mathbf{s})| \xrightarrow{P} 0 \tag{2.21}$$

and

$$C_n(\mathbf{s}) \xrightarrow{d} V, \tag{2.22}$$

where $C_n(\mathbf{s})$ is defined by

$$C_n(\mathbf{s}) \equiv r\sum_{t=1}^{n} \frac{\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}(\epsilon_t - 1).$$

To be more specific on (2.21), we have

$$\begin{aligned}
F_n - iC_n(\mathbf{s}) &= F_n - r\sum_{t=1}^{n}\frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}(\epsilon_t - 1)\\
&= r\sum_{t=1}^{n}(\epsilon_t - 1)e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\left[\exp\left(\frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right) - 1 - \frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right]\\
&\quad + \frac{r}{2}\sum_{t=1}^{n}(\gamma_\epsilon(0)+1)\frac{\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2 e^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2}\\
&\quad + \frac{r}{2}\sum_{t=1}^{n}\epsilon_t^2 e^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\left[\exp\left(\frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right) - 1\right]^2\\
&= r\sum_{t=1}^{n}(\epsilon_t - 1)e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\left[\exp\left(\frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right) - 1 - \frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right]\\
&\quad - \frac{r}{2}\sum_{t=1}^{n}\left[\epsilon_t^2 - (\gamma_\epsilon(0)+1)\right]\frac{\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2 e^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2}\\
&\quad + \frac{r}{2}\sum_{t=1}^{n}\epsilon_t^2 e^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\left\{\left[\exp\left(\frac{i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}\right) - 1\right]^2 + \frac{\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2}\right\}.
\end{aligned}$$

So

$$|F_n - iC_n(\mathbf{s})| \leq \frac{r}{2}\left|\sum_{t=1}^{n}\frac{\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2 e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2}(\epsilon_t - 1)\right|$$

$$
\begin{aligned}
&+\text{Const}\sum_{t=1}^{n}\frac{\left|\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right|^3 e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^3}\left|\epsilon_t-1\right| \\
&+\frac{r}{2}\left|\sum_{t=1}^{n}\frac{\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2 e^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2}\left[\epsilon_t^2-\left(\gamma_\epsilon(0)+1\right)\right]\right| \\
&+\text{Const}\sum_{t=1}^{n}\frac{\left|\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right|^3 e^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^3}\epsilon_t^2 \\
= &\ \frac{r}{2}\left|T_1\right|+\text{Const}\cdot T_2+\frac{r}{2}\left|T_3\right|+\text{Const}\cdot T_4, \qquad (2.23)
\end{aligned}
$$

where the two Const's are not necessarily identical. We show that the terms $T_i, i=1,2,3,4$, go to zero in probability, respectively, and hence establish (2.21). The variance of T_1 is

$$
\begin{aligned}
\operatorname{Var}(T_1) &= \operatorname{Var}\left[\sum_{t=1}^{n}\frac{\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2 e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2}\left(\epsilon_t-1\right)\right] \\
&= \sum_{j=1}^{n}\sum_{t=1}^{n}\frac{\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nj}\right)^2 e^{-\mathbf{x}_{nj}^T\boldsymbol{\beta}_0}\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2 e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nj}^T\boldsymbol{\beta}_0}\right)^2\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2}\operatorname{Cov}\left(\epsilon_j,\epsilon_t\right) \\
&\le \left(\sup_{1\le t\le n}\left|\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right|\right)^4\cdot n\sum_{|h|<n}\left|\gamma_\epsilon(h)\right| \\
&\to 0,
\end{aligned}
$$

because $\sup_{1\le t\le n}\left|\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right| = O\left(1/\sqrt{n}\right)$ and $\sum_h\left|\gamma_\epsilon(h)\right|<\infty$. It follows that T_1 converges to its mean zero in probability. Likewise, the term T_3 also converges to its mean zero in probability because its variance goes to zero. That is,

$$
\begin{aligned}
\operatorname{Var}(T_3) &= \operatorname{Var}\left\{\sum_{t=1}^{n}\frac{\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2 e^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2}\left[\epsilon_t^2-\left(\gamma_\epsilon(0)+1\right)\right]\right\} \\
&= \sum_{j=1}^{n}\sum_{t=1}^{n}\frac{\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nj}\right)^2 e^{-2\mathbf{x}_{nj}^T\boldsymbol{\beta}_0}\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2 e^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nj}^T\boldsymbol{\beta}_0}\right)^2\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^2}\operatorname{Cov}\left(\epsilon_j^2,\epsilon_t^2\right) \\
&\le \left(\sup_{1\le t\le n}\left|\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right|\right)^4\cdot n\sum_{|h|<n}\left|\gamma_{\epsilon^2}(h)\right| \\
&\to 0.
\end{aligned}
$$

The nonnegative term T_2 converges to zero in probability since its expectation converges to zero:

$$\mathrm{E}\left[\sum_{t=1}^{n}\frac{\left|\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right|^3 e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^3}\left|\epsilon_t-1\right|\right]=\left[\sum_{t=1}^{n}\frac{\left|\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right|^3 e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)^3}\right]\mathrm{E}\left|\epsilon_t-1\right|\to 0$$

because $\mathrm{E}|\epsilon_t-1|$ is bounded and $\sup_{1\le t\le n}\left|\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right|^3=O\left(n^{-\frac{3}{2}}\right)$. Likewise, the term T_4 converges to zero in probability. On the other hand, the convergence (2.22) can be shown by adapting the technique for proving the central limit theorem of linear processes (see Section 2.5 Appendix). The variance $\mathbf{s}^T\Omega_3^{\ddagger}(\boldsymbol{\beta}_0)\,\mathbf{s}$ of random variable V is the limit of $\mathrm{Var}[C_n(\mathbf{s})]$:

$$\begin{aligned}
\mathrm{Var}\left[C_n(\mathbf{s})\right] &= \mathrm{E}\left[r\sum_{t=1}^{n}\frac{\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}(\epsilon_t-1)\right]^2\\
&= r^2\mathbf{s}^T\mathbf{M}_n^T\sum_{j=1}^{n}\sum_{t=1}^{n}\frac{\mathbf{x}_{nj}e^{-\mathbf{x}_{nj}^T\boldsymbol{\beta}_0}\mathbf{x}_{nt}^T e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nj}^T\boldsymbol{\beta}_0}\right)\left(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\right)}\mathrm{Cov}\left(\epsilon_j,\epsilon_t\right)\mathbf{M}_n\mathbf{s}\\
&\to \mathbf{s}^T\Omega_3^{\ddagger}(\boldsymbol{\beta}_0)\,\mathbf{s}.
\end{aligned}$$

Moreover, following the same lines in deriving the convergence of the term T_1, we can show that $E_n^2(\mathbf{u})\to 0$ in probability; following the same lines in deriving the convergence of the term T_2, we can show that $E_n^3\to 0$ in probability, $E_n^1(\mathbf{u})\to 0$, and $E_n^4\to 0$.

Putting all the pieces together we obtain

$$D_n+F_n+rE_n^3\xrightarrow{d}-\frac{1}{2}\mathbf{s}^T\left[\Omega_1^{\ddagger}(\boldsymbol{\beta}_0)+\Omega_2^{\ddagger}(\boldsymbol{\beta}_0)\right]\mathbf{s}+iV,$$

and hence

$$\begin{aligned}
\mathrm{E}\left(e^{i\mathbf{s}^T\mathbf{U}_n}\right) &= \mathrm{E}\left[\mathrm{E}\left(e^{i\mathbf{s}^T\mathbf{U}_n}|\alpha_t\right)\right]\\
&= \mathrm{E}\left[\exp\left(D_n+F_n+rE_n^3\right)\right]\\
&\to \mathrm{E}\left\{\exp\left\{-\frac{1}{2}\mathbf{s}^T\left[\Omega_1^{\ddagger}(\boldsymbol{\beta}_0)+\Omega_2^{\ddagger}(\boldsymbol{\beta}_0)\right]\mathbf{s}+iV\right\}\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \exp\left\{-\frac{1}{2}\mathbf{s}^T\left[\Omega_1^\ddagger(\boldsymbol{\beta}_0)+\Omega_2^\ddagger(\boldsymbol{\beta}_0)\right]\mathbf{s}\right\}\cdot \mathrm{E}\left(e^{iV}\right)\\
&= \exp\left\{-\frac{1}{2}\mathbf{s}^T\left[\Omega_1^\ddagger(\boldsymbol{\beta}_0)+\Omega_2^\ddagger(\boldsymbol{\beta}_0)\right]\mathbf{s}\right\}\cdot \exp\left\{-\frac{1}{2}\mathbf{s}^T\Omega_3^\ddagger(\boldsymbol{\beta}_0)\,\mathbf{s}\right\}\\
&= \exp\left\{-\frac{1}{2}\mathbf{s}^T\left[\Omega_1^\ddagger(\boldsymbol{\beta}_0)+\Omega_2^\ddagger(\boldsymbol{\beta}_0)+\Omega_3^\ddagger(\boldsymbol{\beta}_0)\right]\mathbf{s}\right\}.
\end{aligned}
$$

That is, $\mathbf{U}_n \xrightarrow{d} \mathrm{N}(\mathbf{0},\Omega_1^\ddagger(\boldsymbol{\beta}_0)+\Omega_2^\ddagger(\boldsymbol{\beta}_0)+\Omega_3^\ddagger(\boldsymbol{\beta}_0))$. Hence (2.17) follows.

Since g_n has convex sample paths,

$$
g_n(\mathbf{u}) \to g(\mathbf{u}) \equiv \frac{1}{2}\mathbf{u}^T\Omega_1^\dagger(\boldsymbol{\beta}_0)\,\mathbf{u} - \mathbf{u}^T\cdot \mathrm{N}(\mathbf{0},\Omega_1^\ddagger(\boldsymbol{\beta}_0)+\Omega_2^\ddagger(\boldsymbol{\beta}_0)+\Omega_3^\ddagger(\boldsymbol{\beta}_0))
$$

in distribution on the space $C\left(\mathbb{R}^l\right)$ (see Rockafellar (1970) and Pollard (1991)). It follows that, the estimator $\widehat{\mathbf{u}}_n = \mathbf{M}_n^{-1}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right)$, which minimizes $g_n(\mathbf{u})$, converges in distribution to the minimizer $\widehat{\mathbf{u}}$ of $g(\mathbf{u})$. But

$$
\widehat{\mathbf{u}} = \left[\Omega_1^\dagger(\boldsymbol{\beta}_0)\right]^{-1}\cdot \mathrm{N}(\mathbf{0},\Omega_1^\ddagger(\boldsymbol{\beta}_0)+\Omega_2^\ddagger(\boldsymbol{\beta}_0)+\Omega_3^\ddagger(\boldsymbol{\beta}_0)),
$$

or equivalently

$$
\widehat{\mathbf{u}} \sim \mathrm{N}(\mathbf{0},\left[\Omega_1^\dagger(\boldsymbol{\beta}_0)\right]^{-1}\left[\Omega_1^\ddagger(\boldsymbol{\beta}_0)+\Omega_2^\ddagger(\boldsymbol{\beta}_0)+\Omega_3^\ddagger(\boldsymbol{\beta}_0)\right]\left[\Omega_1^\dagger(\boldsymbol{\beta}_0)\right]^{-1}).
$$

Therefore,

$$
\mathbf{M}_n^{-1}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right) \to \mathrm{N}(\mathbf{0},\left[\Omega_1^\dagger(\boldsymbol{\beta}_0)\right]^{-1}\left[\Omega_1^\ddagger(\boldsymbol{\beta}_0)+\Omega_2^\ddagger(\boldsymbol{\beta}_0)+\Omega_3^\ddagger(\boldsymbol{\beta}_0)\right]\left[\Omega_1^\dagger(\boldsymbol{\beta}_0)\right]^{-1})
$$

in distribution. This completes the proof. ■

REMARK 2.1. For function $\vartheta(x) = \log(1+e^x)$, we have $\vartheta(-x) = \vartheta(x) - x$. Thus $(-1)^j\vartheta^{(j)}(-x) = \vartheta^{(j)}(x)$ for $j \geq 2$, where $\vartheta^{(j)}(x)$ denotes the j^{th} derivative of $\vartheta(x)$. It follows that

$$
\vartheta^{(j)}\left(-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\left(-\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^j = \vartheta^{(j)}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^j, \qquad \text{for } j \geq 2.
$$

2.2.3 Strongly Mixing Latent Process Case

The theorem also holds for the case where the latent process $\{\epsilon_t\}$ is assumed to be a stationary strongly mixing (α-mixing) process, provided that $\{\epsilon_t\}$ satisfies the following two assumptions:

(A1) There exists a positive constant λ such that $\mathrm{E}\left[(\epsilon_t)^{\lambda+4}\right] < \infty$;

(A2) The process $\{\epsilon_t\}$ is strongly mixing with mixing coefficient $\alpha(m)$ such that

$$\sum\nolimits_m \alpha(m)^{\frac{\lambda+2}{\lambda}} < \infty.$$

We give the definition of strongly mixing process and state a central limit theorem of Davidson (1992) for strongly mixing process.

Definition 2.2 *A process $\{X_t\}$ is strongly mixing (α-mixing) if the mixing coefficient*

$$\alpha(m) \equiv \sup_{A\in\mathcal{F}_{-\infty}^{0}, B\in\mathcal{F}_{m}^{\infty}} |P(AB) - P(A)P(B)| \to 0$$

as $m \to \infty$, where $\mathcal{F}_{-\infty}^{0}$ and $\mathcal{F}_{m}^{\infty}$ are σ-fields generated by $\{X_t, t \leq 0\}$ and $\{X_t, t \geq m\}$, respectively.

Proposition 2.3 *(Davidson, 1992) Let $\{X_{nt}, t = 1, \ldots, n, n \geq 1\}$ denote a triangular array of random variables defined on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ such that*

1. *$E(X_{nt}) = 0$ and $E(\sum_{t=1}^{n} X_{nt})^2 = 1$;*

2. *There exists a positive constant array $\{c_{nt}\}$ and a constant $\gamma > 2$ such that $\left[E\left(\frac{X_{nt}}{c_{nt}}\right)^{\gamma}\right]^{\frac{1}{\gamma}}$ is uniformly bounded in t and n;*

3. *For each n, the sequence $\{X_{nt}\}$ is strongly mixing with mixing coefficient $\alpha(m)$ such that $\sum_m \alpha(m)^{\frac{\gamma}{\gamma-2}} < \infty$; and*

4. $\sup_n \left\{n\left(\max_{1\leq t\leq n} c_{nt}\right)^2\right\} < \infty$.

Then the central limit theorem holds:

$$\sum_{t=1}^{n} X_{nt} \xrightarrow{d} N(0,1).$$

The asymptotic results of the GLM estimator are stated in the following theorem for the case where the latent process is a stationary strongly mixing process.

Theorem 2.4 *Suppose $\{\epsilon_t\}$ is strongly mixing and satisfies assumptions (A1) and (A2). Also suppose that the conditions of Theorem 2.1 are satisfied. Then the same asymptotic results hold for the GLM estimator $\widehat{\boldsymbol{\beta}}_n$.*

Proof. It suffices to show that the convergence (2.22) holds true for $\{\epsilon_t\}$. Let

$$\begin{aligned}
\tau_n^2(\mathbf{s}) &\equiv \operatorname{Var}\left[C_n(\mathbf{s})\right] \\
&= r^2 \mathbf{s}^T \mathbf{M}_n^T \sum_{j=1}^{n}\sum_{t=1}^{n} \frac{\mathbf{x}_{nj} e^{-\mathbf{x}_{nj}^T \boldsymbol{\beta}_0} \mathbf{x}_{nt}^T e^{-\mathbf{x}_{nt}^T \boldsymbol{\beta}_0}}{\left(1+e^{-\mathbf{x}_{nj}^T \boldsymbol{\beta}_0}\right)\left(1+e^{-\mathbf{x}_{nt}^T \boldsymbol{\beta}_0}\right)} \operatorname{Cov}\left(\epsilon_j, \epsilon_t\right) \mathbf{M}_n \mathbf{s}.
\end{aligned}$$

Then $\tau_n^2(\mathbf{s}) \to \mathbf{s}^T \Omega_3^{\ddagger}(\boldsymbol{\beta}_0)\, \mathbf{s}$. Define

$$Z_{nt} \equiv \frac{r}{\tau_n(\mathbf{s})} \cdot \frac{\mathbf{s}^T \mathbf{M}_n^T \mathbf{x}_{nt} e^{-\mathbf{x}_{nt}^T \boldsymbol{\beta}_0}}{1+e^{-\mathbf{x}_{nt}^T \boldsymbol{\beta}_0}} \left(\epsilon_t - 1\right),$$

then we show $\sum_{t=1}^{n} Z_{nt} \xrightarrow{d} \mathrm{N}(0,1)$ by verifying that $\{Z_{nt}\}$ satisfies conditions in Proposition 2.3. First, it is clear that

$$\mathrm{E}\left(Z_{nt}\right) = 0 \quad \text{and} \quad \mathrm{E}\left(\sum_{t=1}^{n} Z_{nt}\right)^2 = \mathrm{E}\left(\frac{C_n(\mathbf{s})}{\tau_n(\mathbf{s})}\right)^2 = 1.$$

Next, take $c_{nt} = \frac{1}{\sqrt{n}}$, then $\sup_n \left\{ n \left(\max_{1\le t\le n} c_{nt}\right)^2 \right\} = 1 < \infty$. For $\lambda > 0$,

$$\left[\mathrm{E}\left(\frac{Z_{nt}}{c_{nt}}\right)^{\lambda+2}\right]^{\frac{1}{\lambda+2}} = \left[\mathrm{E}\left(\frac{r}{\tau_n(\mathbf{s})} \cdot \frac{\mathbf{s}^T \mathbf{M}_n^T \mathbf{x}_{nt} e^{-\mathbf{x}_{nt}^T \boldsymbol{\beta}_0}}{c_{nt}\left(1+e^{-\mathbf{x}_{nt}^T \boldsymbol{\beta}_0}\right)} \left(\epsilon_t - 1\right)\right)^{\lambda+2}\right]^{\frac{1}{\lambda+2}}$$

$$= \frac{r}{\tau_n(\mathbf{s})} \cdot \frac{\mathbf{s}^T \mathbf{M}_n^T \mathbf{x}_{nt} e^{-\mathbf{x}_{nt}^T \boldsymbol{\beta}_0}}{c_{nt}\left(1+e^{-\mathbf{x}_{nt}^T \boldsymbol{\beta}_0}\right)} \left[\mathrm{E}\left(\epsilon_t - 1\right)^{\lambda+2}\right]^{\frac{1}{\lambda+2}},$$

which is bounded uniformly in t and n by assumption that $\sup_{1\le t\le n}\left|\mathbf{M}_n^T \mathbf{x}_{nt}\right| = O\left(1/\sqrt{n}\right)$ and the assumption (A1). The third condition of Proposition 2.3 is assured by the assumption (A2). Applying Proposition 2.3 to $\{Z_{nt}\}$ yields $\sum_{t=1}^n Z_{nt} \xrightarrow{d} \mathrm{N}(0,1)$. Therefore the convergence (2.22) follows, namely,

$$C_n(\mathbf{s}) = r\sum_{t=1}^n \frac{\mathbf{s}^T \mathbf{M}_n^T \mathbf{x}_{nt} e^{-\mathbf{x}_{nt}^T \boldsymbol{\beta}_0}}{1+e^{-\mathbf{x}_{nt}^T \boldsymbol{\beta}_0}} \left(\epsilon_t - 1\right) \xrightarrow{d} \mathrm{N}\left(0, \mathbf{s}^T \Omega_3^\ddagger\left(\boldsymbol{\beta}_0\right)\mathbf{s}\right).$$

This completes the proof. ■

2.3 Generalization: One-Parameter Exponential Family Setup

In this section, we study time series regression models in a more general setup, where the random component of a GLM is specified by a one-parameter exponential family distribution. We generalize the asymptotic results of GLM estimator.

Let $Y_1, \ldots, Y_n$ denote observed variables of a time series that are independent conditional upon a latent process $\{\alpha_t\}$ and the regressor $\{\mathbf{x}_{nt}\}$, where the conditional distribution depends only on α_t and $\mathbf{x}_{nt}$ and is specified by

$$p\left(Y_t|\alpha_t, \mathbf{x}_{nt}\right) = \exp\left[\theta_t Y_t - b\left(\theta_t\right) + c\left(Y_t\right)\right] \tag{2.24}$$

where $\theta_t = \mathcal{G}\left(\mathbf{x}_{nt}^T \boldsymbol{\beta} + \alpha_t\right)$ for a real function $\mathcal{G}(\cdot)$. That is, the conditional distribution $p\left(Y_t|\alpha_t, \mathbf{x}_{nt}\right)$, $t = 1, \ldots, n$, belongs to one-parameter exponential family. Let $\eta_t \equiv \mathbf{x}_{nt}^T \boldsymbol{\beta} + \alpha_t$ represent the linear predictor $\mathbf{x}_{nt}^T \boldsymbol{\beta}$ modified by the latent process $\{\alpha_t\}$. The link function $\mathcal{F}$, which is assumed to be monotone and differentiable, is chosen such that

$$\mathrm{E}\left(Y_t|\alpha_t\right) = \mathcal{H}\left(\eta_t\right) = \mathcal{H}\left(\mathbf{x}_{nt}^T \boldsymbol{\beta} + \alpha_t\right), \tag{2.25}$$

where $\mathcal{H}$ is the inverse function of $\mathcal{F}$. On the other hand, for the one-parameter exponential family (2.24), it is well known that

$$\mathrm{E}\left(Y_t|\alpha_t\right) = b'\left(\theta_t\right) = \left(b' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta} + \alpha_t\right).$$

So, $\left(b' \circ \mathcal{G}\right)(\cdot) = \mathcal{H}(\cdot)$.

Now suppose $Y_1, \ldots, Y_n$ are observations from the model with the true parameter vector $\boldsymbol{\beta}_0$. The GLM estimator $\widehat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$ is defined as the maximizer of the pseudo log-likelihood function

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{t=1}^{n}\left[\mathcal{G}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}\right)Y_t - \left(b \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}\right) + c\left(Y_t\right)\right] \\ &= -\sum_{t=1}^{n}\left(b \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}\right) + \sum_{t=1}^{n} Y_t\mathcal{G}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}\right) + \sum_{t=1}^{n} c\left(Y_t\right), \end{aligned} \tag{2.26}$$

which ignores the latent process $\{\alpha_t\}$ in the model.

2.3.1 Assumptions

Assumptions are needed on link function, regressors, etc. in order to establish the asymptotic properties of GLM estimator in this general one-parameter exponential family setup. As one can see, these assumptions reduce once a concrete form of the conditional distribution of Y_t given α_t and $\mathbf{x}_{nt}$ and that of link function are given, e.g., in negative binomial logit or Poisson log-linear setup.

- Assumptions on link function

We assume that with the link function $\mathcal{F}$ the log-likelihood function (2.26) is concave and $\mathrm{E}\left[\mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta} + \alpha_t\right)\right] = \mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}\right)$. Such a link function exists for Poisson, negative binomial, Gaussian, among other cases. Under this assumption, the mean of Y_t is

$$\mu_t = \mathrm{E}\left(Y_t\right) = \mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right).$$

Since $\mathrm{Var}(Y_t|\alpha_t) = b''(\theta_t) = (b'' \circ \mathcal{G})\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)$, we obtain the variance of Y_t by

$$\mathrm{Var}(Y_t) = \mathrm{E}\left[(b'' \circ \mathcal{G})\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)\right] + \mathrm{Var}\left[\mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)\right].$$

Moreover, the ACVF of Y_t is given by

$$\mathrm{Cov}(Y_{t+h}, Y_t) = \mathrm{Cov}\left[\mathcal{H}\left(\mathbf{x}_{n,t+h}^T\boldsymbol{\beta}_0 + \alpha_{t+h}\right), \mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)\right]$$

for $h \neq 0$.

- Assumptions on regressors

We assume that there exists a sequence of nonsingular matrices $\mathbf{M}_n$ such that regressors satisfy the following conditions:

$$\mathbf{M}_n^T \sum_{t=1}^{n} \mathbf{x}_{nt}\mathbf{x}_{nt}^T \mathrm{E}\left[(b'' \circ \mathcal{G})\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)\right] \mathbf{M}_n \to \Omega_1(\boldsymbol{\beta}_0) \tag{2.27}$$

and

$$\mathbf{M}_n^T \sum_{j,t=1}^{n} \mathbf{x}_{nj}\mathbf{x}_{nt}^T \mathrm{Cov}\left[\mathcal{H}\left(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0 + \alpha_j\right), \mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)\right] \mathbf{M}_n \to \Omega_3(\boldsymbol{\beta}_0) \tag{2.28}$$

as $n \to \infty$. Also

$$\mathbf{M}_n^T \sum_{t=1}^{n} \mathbf{x}_{nt}\mathbf{x}_{nt}^T \mathrm{E}\left[(b'' \circ \mathcal{G})\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)\right]\left[\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\right]^2 \mathbf{M}_n \to \Omega_1^{\ddagger}(\boldsymbol{\beta}_0), \tag{2.29}$$

$$\mathbf{M}_n^T \sum_{t=1}^{n} \mathbf{x}_{nt}\mathbf{x}_{nt}^T (b'' \circ \mathcal{G})\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\left[\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\right]^2 \mathbf{M}_n \to \Omega_1^{\dagger}(\boldsymbol{\beta}_0), \tag{2.30}$$

$$\begin{aligned}\mathbf{M}_n^T \sum_{j,t=1}^{n} \mathbf{x}_{nj}\mathbf{x}_{nt}^T &\mathrm{Cov}\left[\mathcal{H}\left(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0 + \alpha_j\right), \mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)\right] \\ &\times \mathcal{G}'\left(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0\right)\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\mathbf{M}_n \to \Omega_3^{\ddagger}(\boldsymbol{\beta}_0),\end{aligned} \tag{2.31}$$

and

$$\sup_{1\le t\le n}\left|\mathbf{M}_n^T\mathbf{x}_{nt}\right| = O\left(1/\sqrt{n}\right). \tag{2.32}$$

- Other assumptions

We assume that each of $\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)$, $\mathcal{G}''\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)$, $\mathcal{G}^{(3)}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)$, and $(b\circ\mathcal{G})^{(3)}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)$ is uniformly bounded in t. Actually, if $\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)$ is uniformly bounded in t, then (2.29) is implied by the convergence (2.27), and (2.31) is satisfied given the convergence (2.28). We also assume that

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{t=1}^{n}\left|\mathrm{Cov}\left[\left(b''\circ\mathcal{G}\right)\left(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0+\alpha_j\right),\left(b''\circ\mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\alpha_t\right)\right]\right| < \infty. \tag{2.33}$$

2.3.2 Asymptotic Properties of GLM Estimator

We define

$$\begin{aligned}
\boldsymbol{\Sigma}_n\left(\boldsymbol{\beta}_0\right) &\equiv \mathrm{Cov}\left[\mathbf{M}_n^T\sum_{t=1}^{n}\mathbf{x}_{nt}\left(Y_t-\mu_t\right)\right]\\
&= \mathbf{M}_n^T\sum_{t=1}^{n}\mathbf{x}_{nt}\mathbf{x}_{nt}^T\mathrm{Var}\left(Y_t\right)\mathbf{M}_n+\mathbf{M}_n^T\sum_{j,t=1,j\neq t}^{n}\mathbf{x}_{nj}\mathbf{x}_{nt}^T\mathrm{Cov}(Y_j,Y_t)\mathbf{M}_n\\
&= \mathbf{M}_n^T\sum_{t=1}^{n}\mathbf{x}_{nt}\mathbf{x}_{nt}^T\left\{\mathrm{E}\left[\left(b''\circ\mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\alpha_t\right)\right]+\mathrm{Var}\left[\mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\alpha_t\right)\right]\right\}\mathbf{M}_n\\
&\quad+\mathbf{M}_n^T\sum_{j,t=1,j\neq t}^{n}\mathbf{x}_{nj}\mathbf{x}_{nt}^T\mathrm{Cov}\left[\mathcal{H}\left(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0+\alpha_j\right),\mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\alpha_t\right)\right]\mathbf{M}_n\\
&= \mathbf{M}_n^T\sum_{t=1}^{n}\mathbf{x}_{nt}\mathbf{x}_{nt}^T\mathrm{E}\left[\left(b''\circ\mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\alpha_t\right)\right]\mathbf{M}_n\\
&\quad+\mathbf{M}_n^T\sum_{j,t=1}^{n}\mathbf{x}_{nj}\mathbf{x}_{nt}^T\mathrm{Cov}\left[\mathcal{H}\left(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0+\alpha_j\right),\mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\alpha_t\right)\right]\mathbf{M}_n.
\end{aligned}$$

The asymptotic results of GLM estimator are stated in the following theorem.

Theorem 2.5 *Let $\widehat{\boldsymbol{\beta}}_n$ be the GLM estimator of $\boldsymbol{\beta}$ for the parameter driven model (2.24)* *and (2.25), where all forementioned assumptions in Section 2.3.1 are satisfied. Then*

$$\Sigma_n(\boldsymbol{\beta}_0) \to \Omega_1(\boldsymbol{\beta}_0) + \Omega_3(\boldsymbol{\beta}_0)$$

as $n \to \infty$. Moreover, if

$$C_n(\mathbf{s}) \equiv \sum_{t=1}^{n} \mathbf{s}^T \mathbf{M}_n^T \mathbf{x}_{nt} \mathcal{G}'\left(\mathbf{x}_{nt}^T \boldsymbol{\beta}_0\right)\left[\mathcal{H}\left(\mathbf{x}_{nt}^T \boldsymbol{\beta}_0 + \alpha_t\right) - \mathcal{H}\left(\mathbf{x}_{nt}^T \boldsymbol{\beta}_0\right)\right] \xrightarrow{d} V$$

where $V \sim N\Big(0, \mathbf{s}^T \Omega_3^{\ddagger}(\boldsymbol{\beta}_0)\,\mathbf{s}\Big)$, then $\widehat{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}_0$ in probability, and

$$\mathbf{M}_n^{-1}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \left[\Omega_1^{\dagger}(\boldsymbol{\beta}_0)\right]^{-1}\left[\Omega_1^{\ddagger}(\boldsymbol{\beta}_0) + \Omega_3^{\ddagger}(\boldsymbol{\beta}_0)\right]\left[\Omega_1^{\dagger}(\boldsymbol{\beta}_0)\right]^{-1}\right).$$

Proof. See Section 2.5 Appendix. ■

2.3.3 Examples

1. Poisson case

Suppose the random variables $Y_1, \ldots, Y_n$, conditional on a stationary latent process $\{\alpha_t\}$ and the regressor $\{\mathbf{x}_{nt}\}$, are independent, where the conditional distribution depends only on α_t and $\mathbf{x}_{nt}$ and is specified by

$$Y_t | \alpha_t, \mathbf{x}_{nt} \overset{\text{indep}}{\sim} \text{Po}(\lambda_t),$$

with $\log \lambda_t = \mathbf{x}_{nt}^T \boldsymbol{\beta} + \alpha_t$. Rewriting the conditional density function in the canonical form (2.24) of one-parameter exponential family, we obtain

$$p(Y_t | \alpha_t, \mathbf{x}_{nt}) = \exp\left[Y_t \log \lambda_t - \lambda_t - \log(Y_t!)\right].$$

That is, $\theta_t = \log \lambda_t = \mathbf{x}_{nt}^T \boldsymbol{\beta} + \alpha_t$ and $b(\theta_t) = \lambda_t = e^{\theta_t}$. It follows that $b'(\theta_t) = b''(\theta_t) = e^{\theta_t}$. In this setup we have a canonical link function, namely $\mathcal{F}(z) = \log(z)$, and inverse function of the link is $\mathcal{H}(z) = \exp(z)$. The function $\mathcal{G}(z) = z$.

In the Poisson setup,

$$\lambda_t = \mathrm{E}\left(Y_t|\alpha_t\right) = \exp\left(\mathbf{x}_{nt}^T\boldsymbol{\beta} + \alpha_t\right) = \exp\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}\right)\epsilon_t,$$

where $\epsilon_t \equiv \exp\left(\alpha_t\right)$. Since $b''\left(\theta_t\right) = e^{\theta_t}$ and $\mathcal{G}\left(z\right) = z$, we have

$$\left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right) = \exp\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\epsilon_t.$$

In addition, with the assumption that $\mathrm{E}(\epsilon_t) = 1$, we have

$$\mathrm{E}\left[\left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)\right] = \exp\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right) = b''\left(\mathcal{G}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\right).$$

That is, $\Omega_1^{\ddagger}\left(\boldsymbol{\beta}_0\right)$ and $\Omega_1^{\dagger}\left(\boldsymbol{\beta}_0\right)$ are identical. Furthermore, since $\mathcal{G}'\left(\cdot\right) = 1$, then

$$\Omega_1\left(\boldsymbol{\beta}_0\right) = \Omega_1^{\dagger}\left(\boldsymbol{\beta}_0\right) = \Omega_1^{\ddagger}\left(\boldsymbol{\beta}_0\right) \quad \text{and} \quad \Omega_3\left(\boldsymbol{\beta}_0\right) = \Omega_3^{\ddagger}\left(\boldsymbol{\beta}_0\right).$$

Also, for any j and t,

$$\mathrm{Cov}\left[\mathcal{H}\left(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0 + \alpha_j\right), \mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)\right] = r^2 e^{\left(\mathbf{x}_{nj}^T + \mathbf{x}_{nt}^T\right)\boldsymbol{\beta}_0}\gamma_\epsilon\left(j-t\right),$$

and

$$\mathrm{Cov}\left[\left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0 + \alpha_j\right), \left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)\right] = r^2 e^{\left(\mathbf{x}_{nj}^T + \mathbf{x}_{nt}^T\right)\boldsymbol{\beta}_0}\gamma_\epsilon\left(j-t\right).$$

Putting all these together, we can see that Theorem 2.5 reduces to Theorem 1 of Davis, Dunsmuir, and Wang (2000).

2. Negative binomial case

Suppose the random variables $Y_1, \ldots, Y_n$, conditional on a stationary latent process $\{\alpha_t\}$ and the regressor $\{\mathbf{x}_{nt}\}$, are independent, where the conditional distribution depends only on α_t and $\mathbf{x}_{nt}$ and is specified by

$$Y_t|\alpha_t, \mathbf{x}_{nt} \stackrel{\text{indep}}{\sim} \text{NegBin}\,(r, p_t)\,,$$

with a known positive integer r and $\log(p_t/(1-p_t)) = \mathbf{x}_{nt}^T\boldsymbol{\beta} + \alpha_t$. Rewriting the conditional density function in the form (2.24), we obtain

$$p\,(Y_t|\alpha_t, \mathbf{x}_{nt}) = \exp\left[Y_t \log\left(1-p_t\right) - \left(-r\log p_t\right) + \log\binom{Y_t + r - 1}{r-1}\right].$$

That is, $\theta_t = \log\left(1-p_t\right)$ and $b\left(\theta_t\right) = -r\log p_t = -r\log\left(1-e^{\theta_t}\right)$. It follows that $b'\left(\theta_t\right) = \frac{re^{\theta_t}}{1-e^{\theta_t}}$ and $b''\left(\theta_t\right) = \frac{re^{\theta_t}}{\left(1-e^{\theta_t}\right)^2}$. In this setup we have a link function with inverse $\mathcal{H}\left(z\right) = r\exp\left(-z\right)$. The function $\mathcal{G}\left(z\right)$ becomes $-\log\left(1+e^z\right)$, from which it follows that $\mathcal{G}'\left(z\right) = -\frac{1}{1+e^{-z}}$ and $\mathcal{G}''\left(z\right) = -\frac{e^z}{(1+e^z)^2}$.

In the negative binomial case,

$$\text{E}\left(Y_t|\alpha_t\right) = r\exp\left(-\mathbf{x}_{nt}^T\boldsymbol{\beta} - \alpha_t\right) = r\exp\left(-\mathbf{x}_{nt}^T\boldsymbol{\beta}\right)\epsilon_t,$$

where $\epsilon_t \equiv \exp\left(-\alpha_t\right)$. Further derivations show that

$$\left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right) = re^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\epsilon_t + re^{-2\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}\epsilon_t^2,$$

which, combined with the assumption $\text{E}(\epsilon_t) = 1$, implies

$$\text{E}\left[\left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)\right] = \mu_t + \mu_t^2\frac{\gamma_\epsilon\left(0\right)+1}{r},$$

where $\mu_t = re^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0}$. So, $\Omega_1\left(\boldsymbol{\beta}_0\right)$ in Theorem 2.5 is actually $\Omega_1\left(\boldsymbol{\beta}_0\right) + \Omega_2\left(\boldsymbol{\beta}_0\right)$ in Theorem 2.1 for the negative binomial setup. Likewise, since $\mathcal{G}'(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0) = 1/(1+e^{-\mathbf{x}_{nt}^T\boldsymbol{\beta}_0})$, $\Omega_1^\ddagger\left(\boldsymbol{\beta}_0\right)$ in Theorem 2.5 is actually $\Omega_1^\ddagger\left(\boldsymbol{\beta}_0\right) + \Omega_2^\ddagger\left(\boldsymbol{\beta}_0\right)$ in Theorem 2.1.

It is easy to verify that Theorem 2.5 reduces to Theorem 2.1.

3. Gaussian case

Suppose the random variables $Y_1, \ldots, Y_n$, conditional on a stationary latent process $\{\alpha_t\}$ and the regressor $\{\mathbf{x}_{nt}\}$, are independent, where the conditional distribution depends only on α_t and $\mathbf{x}_{nt}$ and is specified by

$$Y_t|\alpha_t, \mathbf{x}_{nt} \overset{\text{indep}}{\sim} \mathrm{N}\left(\nu_t, \tau^2\right),$$

with a known constant τ and $\nu_t = \mathbf{x}_{nt}^T\boldsymbol{\beta} + \alpha_t$. Rewriting the conditional density function in the form (2.24), we obtain

$$p\left(Y_t|\alpha_t, \mathbf{x}_{nt}\right) = \exp\left\{Y_t\left(\frac{\nu_t}{\tau^2}\right) - \frac{\nu_t^2}{2\tau^2} - \frac{1}{2}\left[\frac{Y_t^2}{\tau^2} + \log\left(2\pi\tau^2\right)\right]\right\}.$$

That is, $\theta_t = \nu_t/\tau^2$ and $b\left(\theta_t\right) = \nu_t^2/\left(2\tau^2\right) = \theta_t^2\tau^2/2$. It follows that $b'\left(\theta_t\right) = \tau^2\theta_t$ and $b''\left(\theta_t\right) = \tau^2$. In this setup we have a link function with inverse $\mathcal{H}(z) = z$. The function $\mathcal{G}(z)$ becomes z/τ^2, from which it follows that $\mathcal{G}'(z) = 1/\tau^2$ and $\mathcal{G}''(z) = 0$.

In the Gaussian case,

$$\mathrm{E}\left(Y_t|\alpha_t\right) = \mathbf{x}_{nt}^T\boldsymbol{\beta} + \alpha_t,$$

which yields an additive model, instead of a multiplicative model as in the Poisson or negative binomial setup. Since $\left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right) = \mathrm{E}\left[\left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)\right] = \tau^2$, $\Omega_1^{\ddagger}\left(\boldsymbol{\beta}_0\right)$ and $\Omega_1^{\dagger}\left(\boldsymbol{\beta}_0\right)$ are identical; moreover, since $\mathcal{G}'(\cdot) = 1/\tau^2$,

$$\Omega_1^{\dagger}\left(\boldsymbol{\beta}_0\right) = \Omega_1^{\ddagger}\left(\boldsymbol{\beta}_0\right) = \Omega_1\left(\boldsymbol{\beta}_0\right)/\tau^4 \quad \text{and} \quad \Omega_3^{\ddagger}\left(\boldsymbol{\beta}_0\right) = \Omega_3\left(\boldsymbol{\beta}_0\right)/\tau^4.$$

Also, for any j and t,

$$\mathrm{Cov}\left[\mathcal{H}\left(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0 + \alpha_j\right), \mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)\right] = \gamma_\alpha\left(j - t\right),$$

and

$$\mathrm{Cov}\left[\left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0 + \alpha_j\right), \left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)\right] = 0.$$

The asymptotic results of $\widehat{\boldsymbol{\beta}}_n$ in the Gaussian setup reduce to

$$\mathbf{M}_n^{-1}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \tau^4 \left[\Omega_1(\boldsymbol{\beta}_0)\right]^{-1}\left[\Omega_1(\boldsymbol{\beta}_0) + \Omega_3(\boldsymbol{\beta}_0)\right]\left[\Omega_1(\boldsymbol{\beta}_0)\right]^{-1}\right).$$

2.4 Numerical Study

2.4.1 Simulation Study

We undertake a simulation study to evaluate the finite sample performance of GLM estimator. We consider three models; two of them have a negative binomial setup and the other has a Bernoulli setup, which are specified as follows.

Model 1. $Y_t|\alpha_t, \mathbf{x}_t \stackrel{\text{indep}}{\sim} \text{NegBin}(4, p_t)$, where $\log \frac{p_t}{1-p_t} = \mathbf{x}_t^T\boldsymbol{\beta} + \alpha_t$, the regressor $\mathbf{x}_t = 1$, and $\alpha_t = \phi\alpha_{t-1} + Z_t$, $Z_t \sim$ IID $\mathrm{N}(0, \sigma^2)$.

Model 2. $Y_t|\alpha_t, \mathbf{x}_t \stackrel{\text{indep}}{\sim} \text{NegBin}(4, p_t)$, where $\log \frac{p_t}{1-p_t} = \mathbf{x}_t^T\boldsymbol{\beta} + \alpha_t$, the regressor $\mathbf{x}_t = (1, t/n, \cos(2\pi t/6), \sin(2\pi t/6))^T$, and $\alpha_t = \phi\alpha_{t-1} + Z_t$, $Z_t \sim$ IID $\mathrm{N}(0, \sigma^2)$.

Model 3. $Y_t|\alpha_t, \mathbf{x}_t \stackrel{\text{indep}}{\sim} \text{Bern}(p_t)$, where $-\log p_t = \mathbf{x}_t^T\boldsymbol{\beta} + \alpha_t$, the regressor $\mathbf{x}_t = 1$, and $\alpha_t = \phi\alpha_{t-1} + Z_t$, $Z_t \sim$ IID Exp(1).

For each model, we simulate 1000 replications under each of different sets of given parameter values. Then, we estimate the parameters of interest for each case, report the empirical mean and standard deviation of estimates, and give normal probability plots and boxplots of estimates.

Model 1

Model 1 has a constant regressor and an AR(1) latent process. We consider sample sizes of 100 and 200. The value of the true parameter β_0 is taken to be 0.7 and -0.7. The value of ϕ in AR(1) latent process is taken to be 0.1, 0.5, and 0.9, and σ is chosen such that $\text{Var}(\alpha_t)$ is one.

The method of a combination of golden section search and successive parabolic interpolation is used to search for the maximizer of log-likelihood function, which is implemented with the function "optimize" in R package.

Table 2.1: $\beta_0 = 0.7$

ϕ	0.1		0.5		0.9	
n	100	200	100	200	100	200
$\widehat{\beta}$	0.7196	0.7005	0.7399	0.7067	0.7720	0.7522
SD	0.1805	0.1253	0.2270	0.1618	0.4567	0.3276
ASD	0.1770	0.1252	0.2300	0.1626	0.5114	0.3616

Table 2.2: $\beta_0 = -0.7$

ϕ	0.1		0.5		0.9	
n	100	200	100	200	100	200
$\widehat{\beta}$	-0.6938	-0.6899	-0.6718	-0.6859	-0.5879	-0.6484
SD	0.1628	0.1177	0.2162	0.1501	0.4289	0.3277
ASD	0.1660	0.1174	0.2216	0.1567	0.5077	0.3590

Simulation results are reported in Table 2.1 for $\beta_0 = 0.7$ and Table 2.2 for $\beta_0 = -0.7$. In all cases, the empirical standard deviation (SD) is pretty close to the asymptotic standard deviation (ASD). When $\phi = 0.1$ and 0.5, the GLM estimates are approximately unbiased. But there is some bias of the GLM estimates when $\phi = 0.9$. This is not surprising because when there is a strong autocorrelation among data, estimation by ignoring the serial dependence makes difference. In this case, we would expect a slow convergence to the limit distribution. We also conduct simulation with sample sizes $n = 500$ and 1000 for the case of $\phi = 0.9$ (the results are not shown here), showing that the accuracy increases as sample size increases.

Figures 2.1 and 2.2 are the boxplots and normal probability plots of β estimates for the case where $\beta_0 = 0.7$, when sample size is 100 and 200.

Model 2

Model 2 includes a standardized trend and two harmonic function components. The latent process is specified by an AR(1) as in Model 1. The value of the true parameter vector $\boldsymbol{\beta}_0$ is taken to be $(0.1, 0.3, 0.5, 0.7)^T$. The value of ϕ in AR(1) latent process is taken to be 0.5 and 0.8, and σ is chosen such that $\mathrm{Var}(\alpha_t)$ is one. Searching of the maximizer of

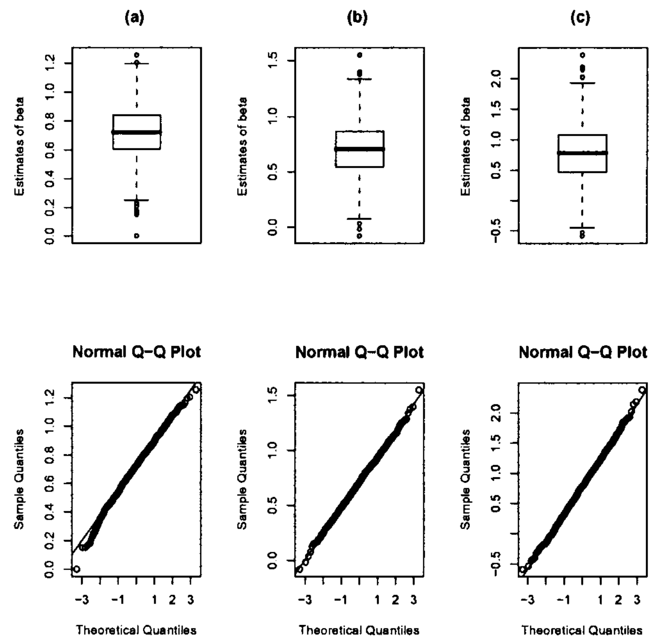

Figure 2.1: Boxplots and Normal probability plots of β estimates for Model 1 when $\beta_0 = 0.7$ and $n = 100$. (a) $\phi = 0.1$. (b) $\phi = 0.5$. (c) $\phi = 0.9$.

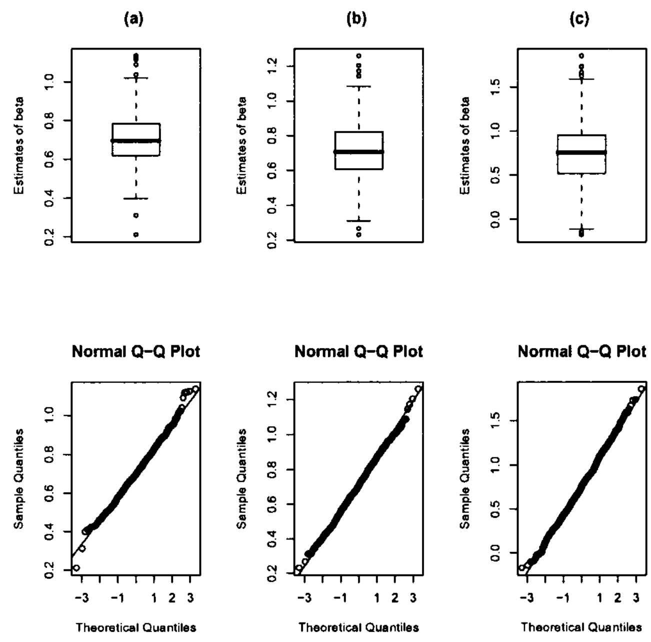

Figure 2.2: Boxplots and Normal probability plots of β estimates for Model 1 when $\beta_0 = 0.7$ and $n = 200$. (a) $\phi = 0.1$. (b) $\phi = 0.5$. (c) $\phi = 0.9$.

Table 2.3: $\boldsymbol{\beta}_0 = (0.1, 0.3, 0.5, 0.7)^T, \phi = 0.5$

n	500				1000			
$\widehat{\boldsymbol{\beta}}$	0.1133	0.3055	0.4976	0.7013	0.1086	0.2948	0.4988	0.7013
SD	0.1973	0.3419	0.1067	0.1092	0.1447	0.2532	0.0736	0.0781
ASD	0.2052	0.3578	0.1082	0.1083	0.1450	0.2529	0.0764	0.0766

Table 2.4: $\boldsymbol{\beta}_0 = (0.1, 0.3, 0.5, 0.7)^T, \phi = 0.8$

n	500				1000			
$\widehat{\boldsymbol{\beta}}$	0.1466	0.2563	0.4931	0.6986	0.1090	0.2916	0.4989	0.6982
SD	0.3084	0.5419	0.0906	0.0866	0.2221	0.3808	0.0630	0.0637
ASD	0.3207	0.5555	0.0918	0.0921	0.2274	0.3941	0.0649	0.0651

log-likelihood function is implemented with the function "optim" in R package; the method used is "BFGS", which is a quasi-Newton method and uses function values and gradients to build up a picture of the surface to be optimized (see R Reference Manual).

It shows that, the convergence of estimates of the intercept term is slow for this model. So we only report the results for larger sample sizes of 500 and 1000, which are displayed in Table 2.3 and Table 2.4. In all cases, the empirical standard deviation (SD) is close to the asymptotic standard deviation (ASD). The GLM estimates are approximately unbiased except for the case where $\phi = 0.8$ with the sample size $n = 500$, which is due to slow convergence to the limit distribution when ignoring a more strongly autocorrelated latent process.

Figures 2.3 and 2.4 are the boxplots and normal probability plots of $\boldsymbol{\beta}$ estimates for the case where $\boldsymbol{\beta}_0 = (0.1, 0.3, 0.5, 0.7)^T$ and $\phi = 0.5$, when sample size is 500 and 1000.

Model 3

Model 3 has a constant regressor. However, due to the model structure, it is not appropriate for its coefficient to be negative. Hence we set the true parameter β_0 to be 0.7. Likewise, the AR(1) latent process is desired to take nonnegative values, so we use a sequence of IID Exp(1) random variables as the innovation process $\{Z_t\}$, and restrict ϕ to

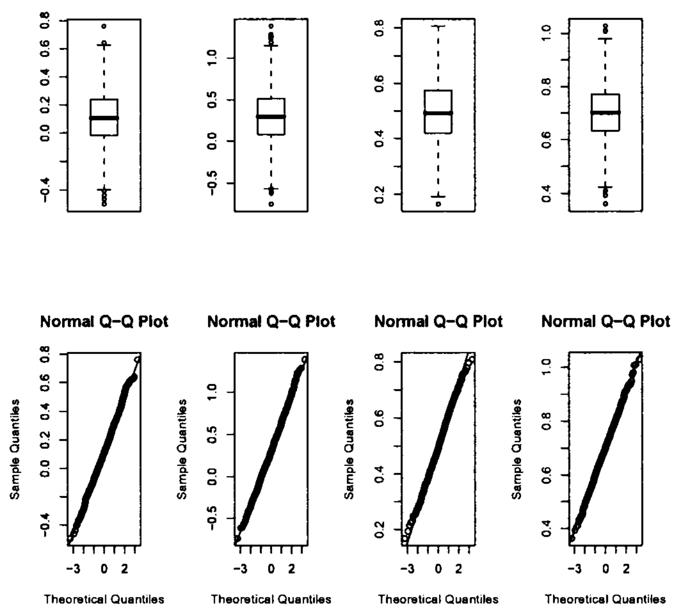

Figure 2.3: Boxplots and Normal probability plots of $\boldsymbol{\beta}$ estimates for Model 2 when $\boldsymbol{\beta}_0 = (0.1, 0.3, 0.5, 0.7)^T$, $\phi = 0.5$, and $n = 500$.

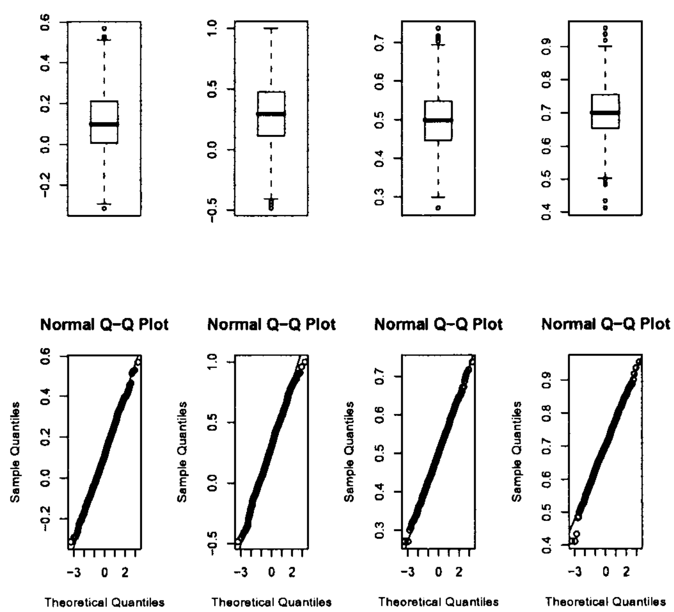

Figure 2.4: Boxplots and Normal probability plots of $\boldsymbol{\beta}$ estimates for Model 2 when $\boldsymbol{\beta}_0 = (0.1, 0.3, 0.5, 0.7)^T$, $\phi = 0.5$, and $n = 1000$.

Table 2.5: $\beta_0 = 0.7$

ϕ	0.1			0.5		
n	100	200	500	100	200	500
$\widehat{\beta}$	0.7175	0.7092	0.7046	0.7436	0.7234	0.7080
SD	0.2045	0.1375	0.0819	0.3443	0.2341	0.1413
ASD	0.1894	0.1339	0.0847	0.3156	0.2232	0.1412

be within $[0, 1)$, which in the simulation is taken to be 0.1 and 0.5. We consider sample sizes of 100, 200, and 500.

Actually, in this case, the condition of $\mathrm{E}(\epsilon_t) = 1$ for Theorem 2.5 is not satisfied, where $\epsilon_t \equiv e^{-\alpha_t}$. However, we still can show the asymptotic normality of the GLM estimator with a constant bias. That is,

$$\sqrt{n}\left(\widehat{\beta}_n - (\beta_0 - d)\right) \xrightarrow{d} \mathrm{N}\left(0, \left[\Omega_1^{\ddagger}(\beta_0) + \Omega_3^{\ddagger}(\beta_0)\right]\left[\Omega_1^{\dagger}(\beta_0)\right]^{-2}\right),$$

where $d = \log(\mathrm{E}\epsilon_t)$ and

$$\begin{aligned}
\Omega_1^{\dagger}(\beta_0) &= \frac{1}{(e^{\beta_0 - d} - 1)}, \\
\Omega_1^{\ddagger}(\beta_0) &= \frac{e^{-2d}}{(e^{\beta_0 - d} - 1)^2}\left(e^{\beta_0}\mathrm{E}\epsilon_t - \mathrm{E}\epsilon_t^2\right), \\
\Omega_3^{\ddagger}(\beta_0) &= \frac{e^{-2d}}{(e^{\beta_0 - d} - 1)^2}\sum_{h=-\infty}^{\infty}\gamma_\epsilon(h).
\end{aligned}$$

Notice that, if $\mathrm{E}(\epsilon_t) = 1$, then $d = 0$ and $\widehat{\beta}_n$ is consistent.

Simulation results are reported in Table 2.5. In all cases, the empirical standard deviation (SD) is pretty close to the asymptotic standard deviation (ASD). But there is some bias in the GLM estimates when $\phi = 0.5$. The boxplots and normal probability plots of estimates are given in Figures 2.5 and 2.6. Further studies show that, the GLM estimation does not perform well when $\phi > 0.6$. This is because that the latent process $\{\alpha_t\}$ becomes large and dominates β_0; it follows that $\{p_t = e^{-(\beta_0 + \alpha_t)}\}$ tend to be very small and most of the simulated values of $\{Y_t\}$ are zeros, which makes the estimation much more difficult.

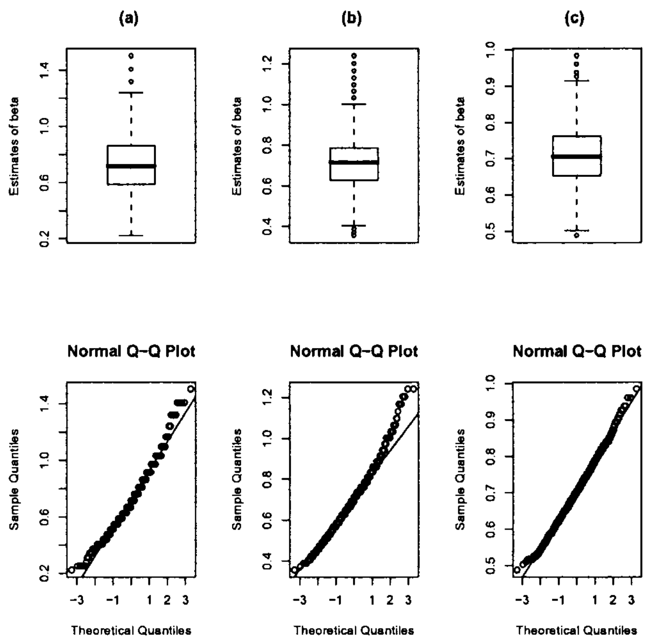

Figure 2.5: Boxplots and Normal probability plots of β estimates for Model 3 when $\beta_0 = 0.7$ and $\phi = 0.1$. (a) $n = 100$. (b) $n = 200$. (c) $n = 500$.

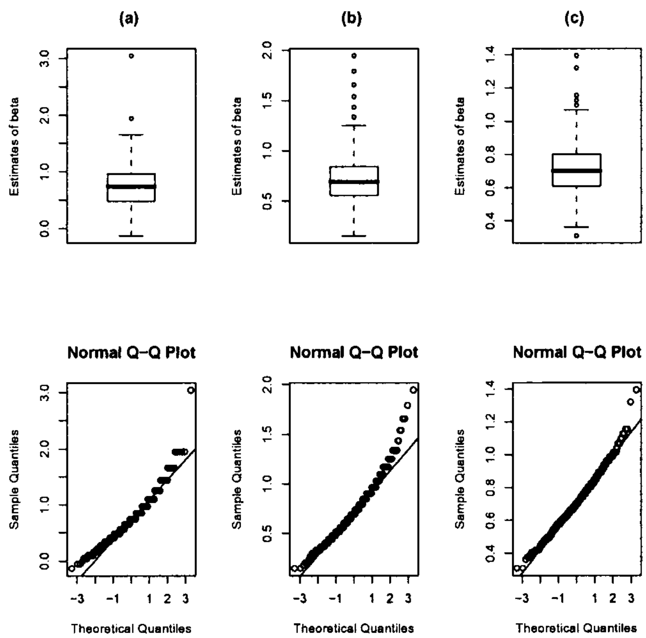

Figure 2.6: Boxplots and Normal probability plots of β estimates for Model 3 when $\beta_0 = 0.7$ and $\phi = 0.5$. (a) $n = 100$. (b) $n = 200$. (c) $n = 500$.

2.4.2 Application to Polio Data

We apply the results of Theorem 2.1 to the polio data shown in Figure 2.7. The 168 counts in the data set are monthly numbers of poliomyelitis cases in the U.S.A. from the year 1970 to 1983 as reported by the Centers for Disease Control. The data set was first considered by Zeger (1988) who proposed a parameter-driven model and used the theory of estimating equations for parameter estimation. Davis, Dunsmuir, and Wang (2000) also studied the polio data in the context of classical GLM estimation in a parameter-driven model. Both Zeger (1988) and Davis, Dunsmuir, and Wang (2000) modeled the counts as Poisson distributed given a latent process. Now we model the counts by a parameter-driven GLM with a negative binomial conditional distribution as described in Section . That is, the count variables Y_t are modeled by

$$Y_t|\alpha_t, \mathbf{x}_t \overset{\text{indep}}{\sim} \text{NegBin}\,(r, p_t)\,.$$

We use the same regressors as in the analysis of Zeger (1988) and Davis, Dunsmuir, and Wang (2000), namely

$$\mathbf{x}_t = \left(1, t'/1000, \cos(2\pi t'/12), \sin(2\pi t'/12), \cos(2\pi t'/6), \cos(2\pi t'/6)\right)^T,$$

where $t' = t - 73$ is used to locate the intercept term at January 1976. In order to compare our results to those obtained by Davis, Dunsmuir, and Wang (2000), we use the link function

$$\log \frac{r(1-p_t)}{p_t} = \mathbf{x}_t^T\boldsymbol{\beta} + \alpha_t,$$

where $\{\alpha_t\}$ is assumed to be an AR(1) latent process with Gaussian innovations. Note that this specification is slightly different than $\log \frac{p_t}{1-p_t} = \mathbf{x}_t^T\boldsymbol{\beta} + \alpha_t$ in Section 2.2. With this specification,

$$\mu_t \equiv \text{E}\,(Y_t) = \text{E}\,[\text{E}\,(Y_t|\alpha_t)] = \text{E}\left[\frac{r\,(1-p_t)}{p_t}\right] = \text{E}\left(e^{\mathbf{x}_t^T\boldsymbol{\beta}+\alpha_t}\right) = e^{\mathbf{x}_t^T\boldsymbol{\beta}}$$

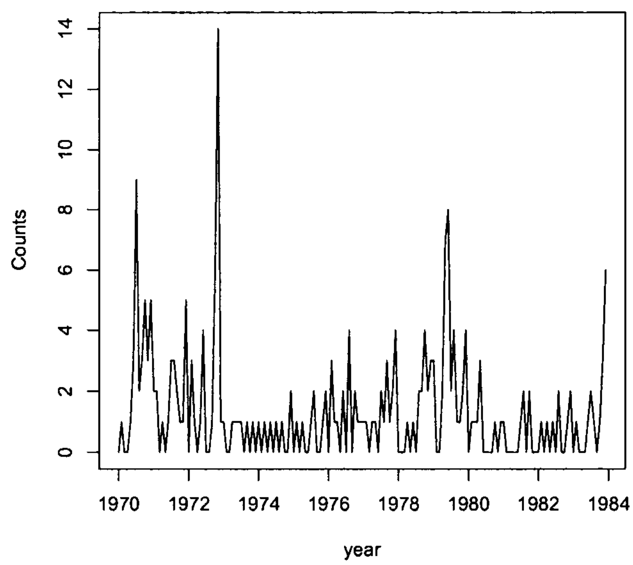

Figure 2.7: Monthly number of poliomyelitis cases in the U.S.A from 1970 to 1983

Table 2.6: Parameter estimates for polio data based on standard negative binomial and Poisson GLM fits

	Poisson		Negative binomial			Simulations	
Covariate	$\hat{\beta}_{Po}$	s.e.	$\hat{\beta}_{NB}$	s.e.	Asy. s.e.	mean	std. dev.
Intercept	0.207	0.075	0.209	0.096	0.167	0.162	0.173
Trend	-4.799	1.403	-4.332	1.847	3.311	-4.381	3.190
$\cos(2\pi t'/12)$	-0.149	0.097	-0.143	0.129	0.156	-0.153	0.158
$\sin(2\pi t'/12)$	-0.532	0.109	-0.503	0.137	0.165	-0.512	0.163
$\cos(2\pi t'/6)$	0.169	0.099	0.168	0.132	0.144	0.179	0.149
$\sin(2\pi t'/6)$	-0.432	0.101	-0.421	0.131	0.146	-0.424	0.145

under the assumption of $E(e^{\alpha_t}) = 1$. Hence the parameters in $\boldsymbol{\beta}$ have the same interpretation as those in Davis, Dunsmuir, and Wang (2000).

The data are fitted by a standard negative binomial GLM using the function "glm.nb" in the R statistical software package. The estimate of r is $\hat{r} = 1.763$. The estimate $\hat{\boldsymbol{\beta}}_{NB}$ of $\boldsymbol{\beta}$ and its standard error are reported in Table 2.6, columns 4-5. The AIC for this model fit is 521.656. For comparison, the results from a standard Poisson GLM fit of Davis, Dunsmuir, and Wang (2000) are also included in Table 2.6 (columns 2-3); the AIC value of the standard Poisson GLM model fit is 557.898. Based on the AIC values, the standard

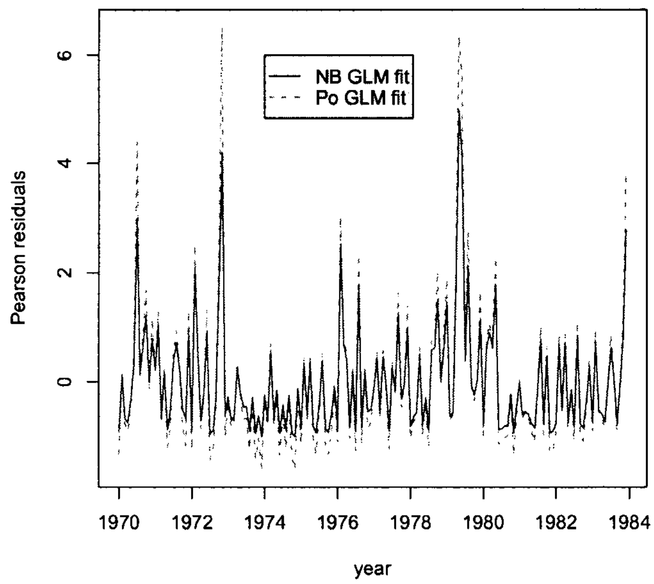

Figure 2.8: Pearson residuals from standard negative binomial and Poisson GLM fits

negative binomial GLM fit is better than the standard Poisson GLM fit.

The estimates and standard errors in the two cases are similar. These standard error calculations ignore the possibility of the presence of a latent process. We examine the Pearson residuals to check for the existence of a latent process and a suitable model if such a process exists. The Pearson residuals from both model-fittings are plotted in Figure 2.8. Since the serial dependence among observations is ignored when fitting a standard GLM, we would expect that Pearson residuals from both model-fittings display the same autocorrelation structure as that of the observations. In both cases, the ACF and PACF plots of Pearson residuals (see Figure 2.9) support the assumption of AR(1) latent process.

The asymptotic standard errors in the sixth column of Table 2.6 are obtained based on Theorem 2.1 assuming an AR(1) latent process $\alpha_t = \phi\alpha_{t-1} + Z_t$, $Z_t \sim$ IID N$(0, \sigma^2)$. We use method of moments to provide estimates of ϕ and σ^2. Recall that,

$$\mathrm{Var}\,(Y_t) = \mu_t + \mu_t^2 \frac{\sigma_\epsilon^2 + 1}{r} + \mu_t^2 \sigma_\epsilon^2 = \mu_t + \mu_t^2 \frac{(r+1)\,\sigma_\epsilon^2 + 1}{r},$$

where $\epsilon_t \equiv e^{\alpha_t}$. So, the values of $\hat{\sigma}_\epsilon^2 = 0.3586$ and $\hat{\rho}_\epsilon(1) = 0.7719$ can be obtained

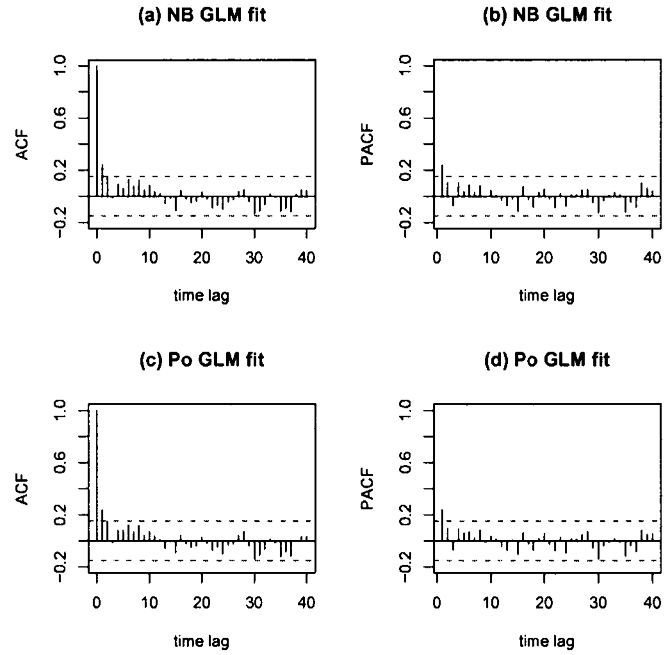

Figure 2.9: ACF and PACF plots of Pearson residuals from standard negative binomial and Poisson GLM fits

using ordinary least squares (OLS) type of estimators suggested by Brännäs and Johansson (1994):

$$\hat{\sigma}_\epsilon^2 = \frac{1}{r+1}\left\{\frac{r\sum_{t=1}^n \hat{\mu}_t^2\left[(Y_t-\hat{\mu}_t)^2-\hat{\mu}_t\right]}{\sum_{t=1}^n \hat{\mu}_t^4} - 1\right\},$$

$$\hat{\rho}_\epsilon(1) = \frac{\hat{\sigma}_\epsilon^{-2}\sum_{t=2}^n \hat{\mu}_t\hat{\mu}_{t-1}(Y_t-\hat{\mu}_t)(Y_{t-1}-\hat{\mu}_{t-1})}{\sum_{t=2}^n \hat{\mu}_t^2\hat{\mu}_{t-1}^2}.$$

Once these two estimates are available, the values of $\hat{\sigma}_\alpha^2 = 0.3065$ and $\hat{\rho}_\alpha(1) = 0.7973$ can be obtained using the relationship between ACVF's of $\{\epsilon_t\}$ and $\{\alpha_t\}$, namely $\gamma_\epsilon(h) = e^{\gamma_\alpha(h)} - 1$ for all h. Then $\hat{\phi} = 0.7973$ and $\hat{\sigma}^2 = 0.1117$ can be calculated; and $\hat{\gamma}_\alpha(h)$ and $\hat{\gamma}_\epsilon(h)$ are readily to be computed.

Now, with these estimates, we can give approximations to the asymptotic standard errors of $\hat{\boldsymbol{\beta}}_{NB}$ using the formula

$$\text{Var}\left(\hat{\boldsymbol{\beta}}_{NB}\right) = \left(\hat{\Omega}_{1,n}^\dagger\right)^{-1}\left(\hat{\Omega}_{1,n}^\ddagger + \hat{\Omega}_{2,n}^\ddagger + \hat{\Omega}_{3,n}^\ddagger\right)\left(\hat{\Omega}_{1,n}^\dagger\right)^{-1},$$

where

$$
\begin{aligned}
\hat{\Omega}^{\dagger}_{1,n} &= \sum_{t=1}^{n} \frac{\mathbf{x}_t \mathbf{x}_t^T \hat{\mu}_t}{1+\hat{\mu}_t/\hat{r}}, \\
\hat{\Omega}^{\ddagger}_{1,n} &= \sum_{t=1}^{n} \frac{\mathbf{x}_t \mathbf{x}_t^T \hat{\mu}_t}{(1+\hat{\mu}_t/\hat{r})^2}, \\
\hat{\Omega}^{\ddagger}_{2,n} &= \frac{\hat{\sigma}_\epsilon^2+1}{\hat{r}} \sum_{t=1}^{n} \frac{\mathbf{x}_t \mathbf{x}_t^T \hat{\mu}_t^2}{(1+\hat{\mu}_t/\hat{r})^2}, \\
\hat{\Omega}^{\ddagger}_{3,n} &= \sum_{t=1}^{n}\sum_{s=1}^{n} \frac{\mathbf{x}_t \mathbf{x}_s^T \hat{\mu}_t \hat{\mu}_s}{(1+\hat{\mu}_t/\hat{r})(1+\hat{\mu}_s/\hat{r})} \hat{\gamma}_\epsilon(t-s),
\end{aligned}
$$

with $\hat{\mu}_t = e^{\mathbf{x}_t^T \hat{\boldsymbol{\beta}}_{NB}}$.

While for the seasonal components (harmonics at periods of 6 and 12 months) the standard errors are close to the asymptotic standard errors, the standard errors for the intercept and linear trend are much smaller than their asymptotic counterparts. One of the main objectives in modeling the polio data is to investigate whether the incidence of polio has been decreasing since 1970. The results show that, the negative trend is not significant using the standard error that includes a latent process; a false significance would be obtained if using the standard error of 1.847 produced by the standard negative binomial GLM estimation. This is in agreement with the conclusion of Davis, Dunsmuir, and Wang (2000) with standard Poisson GLM estimation.

To further check the parameter estimates and asymptotic standard errors based on the negative binomial GLM fit, we simulate 1000 replications of a time series of length $n = 168$ using $\hat{r}$ and $\hat{\boldsymbol{\beta}}_{NB}$ as true parameter values. The latent process $\{\alpha_t\}$ is assumed to be a Gaussian AR(1) with $\phi = 0.7973$ and marginal distribution $\mathrm{N}(-\sigma_\alpha^2/2, \sigma_\alpha^2)$, where $\sigma_\alpha^2 = 0.3065$. The empirical means and standard deviations of parameter estimates from simulations are reported in the last two columns of Table 2.6. The estimates are approximately unbiased, except for the intercept term where the empirical mean of 0.162 is significantly different from the true value of 0.209 used for simulating data. The empirical standard deviations agree with the asymptotic standard errors pretty well.

Table 2.7: Parameter estimates for polio data based on parameter-driven negative binomial and Poisson GLM fits

	Poisson		Negative binomial	
Covariate	$\hat{\beta}_{Po}$	s.e.	$\hat{\beta}_{NB}$	s.e.
Intercept	0.090	0.141	0.106	0.177
Trend	-3.600	2.751	-3.467	3.375
$\cos(2\pi t'/12)$	-0.098	0.143	-0.109	0.129
$\sin(2\pi t'/12)$	-0.478	0.154	-0.488	0.140
$\cos(2\pi t'/6)$	0.190	0.121	0.182	0.122
$\sin(2\pi t'/6)$	-0.355	0.122	-0.365	0.123

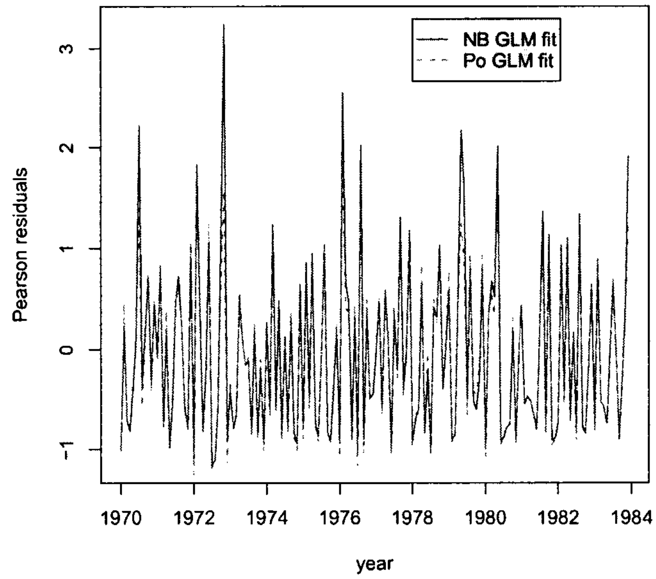

Figure 2.10: Pearson residuals from parameter-driven negative binomial and Poisson GLM fits

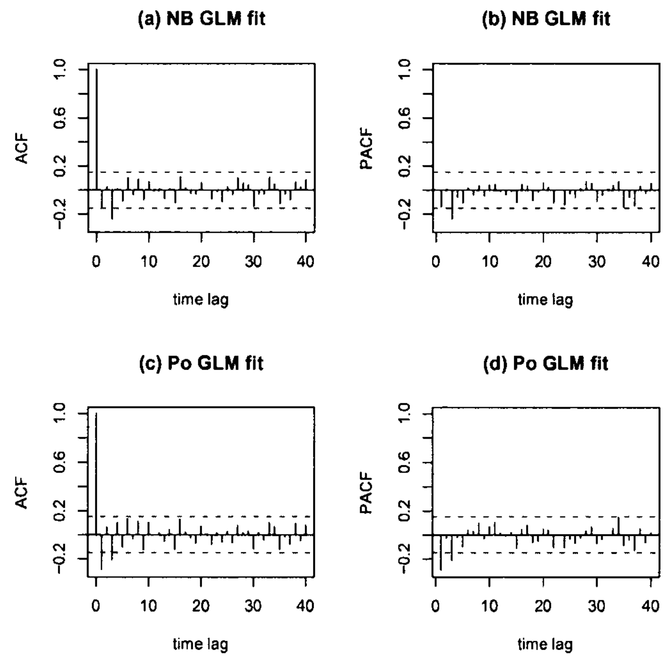

Figure 2.11: ACF and PACF plots of Pearson residuals from parameter-driven negative binomial and Poisson GLM fits

We also fit the polio data by parameter-driven negative binomial and Poisson GLMs using the procedure "glimmix" in the SAS statistical software package, assuming that the latent process follows a Gaussian AR(1) model in both cases. The estimates of the coefficients and their standard errors are reported in Table 2.7; the estimate of r in the parameter-driven negative binomial GLM is obtained to be $\hat{r} = 4.146$. Because AIC values generated by "glimmix" are based on pseudo-likelihood estimation, they are not useful for comparing models. Hence, we perform residual diagnostics of the models by studying the Pearson residuals. The Pearson residuals from both model-fittings are plotted in Figure 2.10; ACF and PACF plots of the Pearson residuals are given in Figure 2.11. Since the serial dependence among the observations has been taken into account in these two model-fittings, we should expect Pearson residuals to be independent in both cases. Figure 2.11 (a) shows that the Pearson residuals from the parameter-driven negative binomial GLM fit appear independent because only the lag 2 autocorrelation is outside the 95% confidence bands among autocorrelations of time lags up to 40. A Ljung-Box test of randomness yields a p-value of 0.14550, which implies that there is no evidence against independence. However, the Pearson residuals from the parameter-driven Poisson GLM fit are not likely to be independent because the autocorrelations at both time lags 1 and 2 seem significant. The p-value 0.00097 of the Ljung-Box test of randomness indicates strong evidence to reject independence. Therefore, the parameter-driven negative binomial GLM is more appropriate for fitting the polio data.

2.5 Appendix

2.5.1 Proof of Convergence (2.22)

We provide a proof of (2.22) in the one-dimensional case where regressors form a triangular array with $\mathbf{x}_{nt} = g(t/n)$ and $g(\cdot)$ is a continuous function; the general proof can be established similarly.

In this case, the normalizing matrix $\mathbf{M}_n$ reduces to $1/\sqrt{n}$. So, it suffices to study the asymptotic behavior of

$$\widetilde{C}_n \equiv \frac{1}{\sqrt{n}} \sum\nolimits_{t=1}^{n} g(t/n)(\epsilon_t - 1), \tag{2.34}$$

where $\epsilon_t = e^{-\alpha_t}$. Recall that, the latent process $\{\alpha_t\}$ is assumed to be a stationary Gaussian linear process, which can be expressed by

$$\alpha_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \qquad \{Z_t\} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2).$$

Lemma 2.6 *Suppose that* $X_n \sim N(0, \nu_n)$*, where* $\nu_n \to \nu$ *and* $0 < \nu < \infty$*. Then* $X_n \xrightarrow{d} X \sim N(0, \nu)$*.*

Proof. See Brockwell and Davis (1991), page 216. ■

Lemma 2.7 *Let* $X_n, n = 1, 2, \ldots,$ *and* $Y_{nj}, j = 1, 2, \ldots; n = 1, 2, \ldots,$ *be random* k*-vectors such that*

(i) $Y_{nj} \xrightarrow{d} Y_j$ *as* $n \to \infty$ *for each* $j = 1, 2, \ldots,$

(ii) $Y_j \xrightarrow{d} Y$ *as* $j \to \infty$, *and*

(iii) $\lim_{j\to\infty} \limsup_{n\to\infty} \Pr(|X_n - Y_{nj}| > \varepsilon) = 0$ *for every* $\varepsilon > 0$.

Then

$$X_n \xrightarrow{d} Y \text{ as } n \to \infty.$$

Proof. See Brockwell and Davis (1991), Proposition 6.3.9. ■

Theorem 2.8 $\widetilde{C}_n \xrightarrow{d} \xi \sim N(0, \tau)$, *where* $\tau = \int_0^1 g^2(x)\,dx \sum_h \gamma_\epsilon(h)$.

Proof. The proof is broken into several steps. We define

$$\epsilon_{t,m} = \exp\left[-\left(\mu + \sum\nolimits_{|j|\le m} \psi_j Z_{t-j}\right)\right].$$

Then $\{\epsilon_{t,m}\}$ is a strictly stationary m'-dependent sequence, where $m' = 2m+1$, with mean

$$\mu_m = \mathrm{E}(\epsilon_{t,m}) = e^{-\mu} \cdot \exp\left(\frac{\sigma^2}{2} \sum\nolimits_{|j|\le m} \psi_j^2\right),$$

and autocovariance $\gamma_{\epsilon\cdot m}(h)$, which takes value zero when $h > m'$ and is evaluated as follows when $0 \le h \le m'$:

$$\begin{aligned}
\gamma_{\epsilon\cdot m}(h) &= \mathrm{E}(\epsilon_{h,m}\epsilon_{0,m}) - \mathrm{E}(\epsilon_{h,m})\,\mathrm{E}(\epsilon_{0,m}) \\
&= \mathrm{E}\left[e^{-2\mu} \cdot \exp\left(-\sum\nolimits_{|j|\le m} \psi_j Z_{h-j}\right) \exp\left(-\sum\nolimits_{|k|\le m} \psi_k Z_{-k}\right)\right] - \mu_m^2 \\
&= e^{-2\mu}\mathrm{E}\{\exp[-(\sum\nolimits_{-m\le j<h-m} \psi_{-j} Z_j + \sum\nolimits_{m<j\le h+m} \psi_{-j+h} Z_j \\
&\qquad + \sum\nolimits_{h-m\le j\le m} (\psi_{-j+h} + \psi_{-j}) Z_j)]\} - \mu_m^2 \\
&= e^{-2\mu} \exp\{\frac{\sigma^2}{2}[\sum\nolimits_{-m\le j<h-m} \psi_{-j}^2 + \sum\nolimits_{m<j\le h+m} \psi_{-j+h}^2 \\
&\qquad + \sum\nolimits_{h-m\le j\le m} (\psi_{-j+h} + \psi_{-j})^2]\} - \mu_m^2.
\end{aligned}$$

The first step is to show that, for any given m,

$$\widetilde{C}_{n,m} \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g\left(\frac{t}{n}\right)(\epsilon_{t,m} - \mu_m) \xrightarrow{d} \xi_m \sim \mathrm{N}(0, \tau_m), \tag{2.35}$$

where $\tau_m = \int_0^1 g^2(x)\,dx \sum_h \gamma_{\epsilon \cdot m}(h)$. For each $k > 2m'$, let

$$\chi_{nj,m} \equiv \frac{1}{\sqrt{n}} \sum_{i=(j-1)k+1}^{jk-m'} g\left(\frac{i}{n}\right)(\epsilon_{i,m} - \mu_m)$$

for $j = 1, \ldots, r$, and $r = \left[\frac{n}{k}\right]$. Then $\chi_{n1,m}, \chi_{n2,m}, \ldots, \chi_{nr,m}$ are independent with mean zero. Moreover, we define $S_{nk,m} \equiv \sum_{j=1}^{r} \chi_{nj,m}$, of which the variance is

$$\begin{aligned}
s_{nk,m}^2 &= \sum_{j=1}^{r} \mathrm{Var}(\chi_{nj,m}) \\
&= \frac{1}{n}\sum_{j=1}^{r} \mathrm{Var}\left[\sum_{i=(j-1)k+1}^{jk-m'} g\left(\frac{i}{n}\right)(\epsilon_{i,m} - \mu_m)\right] \\
&= \frac{1}{n}\sum_{j=1}^{r} \sum_{i=(j-1)k+1}^{jk-m'} \sum_{i'=(j-1)k+1}^{jk-m'} g\left(\frac{i}{n}\right) g\left(\frac{i'}{n}\right) \mathrm{Cov}(\epsilon_{i,m}, \epsilon_{i',m}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{i'=1}^{n} g\left(\frac{i}{n}\right) g\left(\frac{i'}{n}\right) \gamma_{\epsilon \cdot m}(i - i') \\
&\quad -\frac{1}{n}\sum_{j=1}^{r}\left(\sum_{|i'-i|\le m'} \sum_{i=jk-m'+1}^{jk} + \sum_{|i-i'|\le m'} \sum_{i'=jk-m'+1}^{jk}\right) g\left(\frac{i}{n}\right) g\left(\frac{i'}{n}\right) \gamma_{\epsilon \cdot m}(i - i') \\
&\quad +\frac{1}{n}\sum_{j=1}^{r} \sum_{i=jk-m'+1}^{jk} \sum_{i'=jk-m'+1}^{jk} g\left(\frac{i}{n}\right) g\left(\frac{i'}{n}\right) \gamma_{\epsilon \cdot m}(i - i') \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{i'=1}^{n} g\left(\frac{i}{n}\right) g\left(\frac{i'}{n}\right) \gamma_{\epsilon \cdot m}(i - i') \\
&\quad -\frac{2}{n}\sum_{j=1}^{r} \sum_{|i'-i|\le m'} \sum_{i=jk-m'+1}^{jk} g\left(\frac{i}{n}\right) g\left(\frac{i'}{n}\right) \gamma_{\epsilon \cdot m}(i - i') \\
&\quad +\frac{1}{n}\sum_{j=1}^{r} \sum_{i=jk-m'+1}^{jk} \sum_{i'=jk-m'+1}^{jk} g\left(\frac{i}{n}\right) g\left(\frac{i'}{n}\right) \gamma_{\epsilon \cdot m}(i - i').
\end{aligned} \tag{2.36}$$

Since $g(x)$ is continuous, $|g(x)|$ is bounded on $[0,1]$, i.e., $|g(x)| \le D$ for some constant D. Hence, the last two pieces in (2.36) converge to two finite terms, say $2R_{k,m}^1$ and $R_{k,m}^2$, respectively as $n \to \infty$. It follows that

$$s_{nk,m}^2 \to s_{k,m}^2 = \int_0^1 g^2(x)\,dx \sum_h \gamma_{\epsilon.m}(h) + 2R_{k,m}^1 + R_{k,m}^2 < \infty$$

as $n \to \infty$. It can further be shown that both $R_{k,m}^1$ and $R_{k,m}^2$ go to zero as $k \to \infty$. Now check Lyapounov's condition. Since

$$\begin{aligned} |\chi_{nj,m}|^3 &= \frac{1}{n^{3/2}} \left| \sum\nolimits_{i=(j-1)k+1}^{jk-m'} g\left(\frac{i}{n}\right)(\epsilon_{i,m} - \mu_m) \right|^3 \\ &\le \frac{D^3}{n^{3/2}} \left[\sum\nolimits_{i=(j-1)k+1}^{jk-m'} |\epsilon_{i,m} - \mu_m| \right]^3, \end{aligned}$$

we obtain

$$\begin{aligned} \sum_{j=1}^{r} \mathrm{E}\left(|\chi_{nj,m}|^3\right) &\le \sum_{j=1}^{r} \frac{D^3}{n^{3/2}} \mathrm{E}\left[\sum\nolimits_{i=(j-1)k+1}^{jk-m'} |\epsilon_{i,m} - \mu_m| \right]^3 \\ &= \left[\frac{n}{k}\right] \frac{D^3}{n^{3/2}} \mathrm{E}\left[\sum\nolimits_{i=1}^{k-m'} |\epsilon_{i,m} - \mu_m| \right]^3 \\ &\to 0 \quad \text{as } n \to \infty. \end{aligned}$$

It follows that

$$\lim_{n\to\infty} \sum_{j=1}^{r} \frac{1}{s_{nk,m}^3} \mathrm{E}\left(|\chi_{nj,m}|^3\right) = 0.$$

Applying the Lindeberg's central limit theorem (Billingsley 1995, page 359), we establish that $S_{nk,m}/s_{nk,m} \xrightarrow{d} \mathrm{N}(0,1)$, and hence

$$S_{nk,m} \xrightarrow{d} S_{k,m} \sim \mathrm{N}\left(0, s_{k,m}^2\right), \tag{2.37}$$

as $n \to \infty$ for each $k = 1, 2, \ldots$. On the other hand, since

$$s_{k,m}^2 \to \tau_m \qquad \text{as } k \to \infty,$$

we obtain, by applying Lemma 2.6,

$$S_{k,m} \xrightarrow{d} \xi_m \sim \mathrm{N}(0, \tau_m), \tag{2.38}$$

as $k \to \infty$. Next, we show that, $\forall\ \varepsilon > 0$,

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathrm{P}\left(\left|\widetilde{C}_{n,m} - S_{nk,m}\right| > \varepsilon\right) = 0. \tag{2.39}$$

Note that

$$\begin{aligned} \widetilde{C}_{n,m} - S_{nk,m} &= \frac{1}{\sqrt{n}} \sum_{j=1}^{r-1} \sum_{i=jk-m'+1}^{jk} g\left(\frac{i}{n}\right)(\epsilon_{i,m} - \mu_m) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=rk-m'+1}^{rk} g\left(\frac{i}{n}\right)(\epsilon_{i,m} - \mu_m) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=rk+1}^{n} g\left(\frac{i}{n}\right)(\epsilon_{i,m} - \mu_m), \end{aligned}$$

and

$$\mathrm{Var}\left(\widetilde{C}_{n,m} - S_{nk,m}\right) \leq \frac{1}{n}\left[\left(\left[\frac{n}{k}\right] - 1\right) DR_m\right] + \frac{1}{n} R_{h(n),m},$$

where $R_m = \mathrm{Var}\left(\sum_{i=1}^{m'} \epsilon_{i,m}\right)$, $R_{h(n),m} = \mathrm{Var}\left(\sum_{i=1}^{h(n)} \epsilon_{i,m}\right)$ and $h(n) = n - k\left[\frac{n}{k}\right] + m'$. Note that, the term R_m is independent of n, and the term $R_{h(n),m}$ is a function bounded in n since $0 \leq h(n) \leq k + m'$. Hence,

$$\limsup_{n\to\infty} \mathrm{Var}\left(\widetilde{C}_{n,m} - S_{nk,m}\right) \leq \frac{D}{k} R_m,$$

and it follows that

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathrm{Var}\left(\widetilde{C}_{n,m} - S_{nk,m}\right) = 0. \tag{2.40}$$

Now (2.39) follows from (2.40) and the Chebychev's inequality. Therefore, applying Lemma 2.7 we obtain (2.35).

In the second step we show

$$\xi_m \xrightarrow{d} \xi \sim \mathrm{N}(0,\tau) \tag{2.41}$$

Note that, for given h,

$$\begin{aligned}
\lim_{m\to\infty} \gamma_{\epsilon.m}(h) &= e^{-2\mu}\exp\left\{\frac{\sigma^2}{2}\left[\sum_{j=-\infty}^{\infty}(\psi_{-j+h}+\psi_{-j})^2\right]\right\}-1 \\
&= e^{-2\mu}\exp\left(\sigma^2\sum_{j=-\infty}^{\infty}\psi_j^2\right)\exp\left(\sigma^2\sum_{j=-\infty}^{\infty}\psi_{j+h}\psi_j\right)-1 \\
&= \exp(\gamma_\alpha(h))-1 \\
&= \gamma_\epsilon(h).
\end{aligned}$$

It follows that $\tau_m \to \tau$ as $m\to\infty$. Applying Lemma 2.6 we obtain (2.41).

The final step is to show that Condition (iii) of Lemma 2.7 is satisfied.

$$\begin{aligned}
&\mathrm{Var}\left(\widetilde{C}_n - \widetilde{C}_{n,m}\right) \\
&= \mathrm{Var}\left[\frac{1}{\sqrt{n}}\left(\sum_{t=1}^{n} g\left(\frac{t}{n}\right)(\epsilon_t-1)-\sum_{t=1}^{n} g\left(\frac{t}{n}\right)(\epsilon_{t,m}-\mu_m)\right)\right] \\
&= \frac{1}{n}\mathrm{E}\left\{\sum_{t=1}^{n} g\left(\frac{t}{n}\right)\left[(\epsilon_t-1)-(\epsilon_{t,m}-\mu_m)\right]\right\}^2 \\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{t'=1}^{n} g\left(\frac{t}{n}\right)g\left(\frac{t'}{n}\right)\mathrm{E}\left\{\left[(\epsilon_t-1)-(\epsilon_{t,m}-\mu_m)\right]\left[(\epsilon_{t'}-1)-(\epsilon_{t',m}-\mu_m)\right]\right\} \\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{t'=1}^{n} g\left(\frac{t}{n}\right)g\left(\frac{t'}{n}\right) \\
&\qquad \times\{\gamma_\epsilon(t-t') - [e^{-\mu}\exp(\frac{\sigma^2}{2}\sum_{|j|\le m}\psi_j^2)\exp(\sigma^2\sum_{|j|\le m}\psi_{t'-t+j}\psi_j)-\mu_m] \\
&\qquad\quad - [e^{-\mu}\exp(\frac{\sigma^2}{2}\sum_{|j|\le m}\psi_j^2)\exp(\sigma^2\sum_{|j|\le m}\psi_{t-t'+j}\psi_j)-\mu_m] + \gamma_{\epsilon.m}(t-t')\} \\
&\to \int_0^1 g^2(x)\,dx\sum_h \gamma_\epsilon(h) \\
&\quad -2\int_0^1 g^2(x)\,dx\sum_h[e^{-\mu}\exp(\frac{\sigma^2}{2}\sum_{|j|\le m}\psi_j^2)\exp(\sigma^2\sum_{|j|\le m}\psi_{h+j}\psi_j)-\mu_m] \\
&\quad +\int_0^1 g^2(x)\,dx\sum_h \gamma_{\epsilon.m}(h) \qquad (\text{as } n\to\infty) \\
&\to \int_0^1 g^2(x)\,dx\sum_h\gamma_\epsilon(h) - 2\int_0^1 g^2(x)\,dx\sum_h\gamma_\epsilon(h) + \int_0^1 g^2(x)\,dx\sum_h\gamma_\epsilon(h)
\end{aligned}$$

$$= 0 \qquad (\text{as } m \to \infty),$$

which, in conjunction with the Chebychev's inequality, implies that Condition (iii) of Lemma 2.7 is satisfied.

In view of the results in previous steps, we apply Lemma 2.7 and conclude that

$$\widetilde{C}_n \xrightarrow{d} \xi \sim \mathrm{N}(0, \tau).$$

This completes the proof. ■

2.5.2 Proof of Theorem 2.5

The GLM estimator $\widehat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$ is obtained by minimizing

$$\begin{aligned}
&-\left[l(\boldsymbol{\beta}) - l(\boldsymbol{\beta}_0)\right] \\
&= \sum_{t=1}^{n} \left[(b \circ \mathcal{G})(\mathbf{x}_{nt}^T \boldsymbol{\beta}) - (b \circ \mathcal{G})(\mathbf{x}_{nt}^T \boldsymbol{\beta}_0)\right] - \sum_{t=1}^{n} Y_t \left[\mathcal{G}(\mathbf{x}_{nt}^T \boldsymbol{\beta}) - \mathcal{G}(\mathbf{x}_{nt}^T \boldsymbol{\beta}_0)\right] \\
&= \sum_{t=1}^{n} \left\{\left[(b \circ \mathcal{G})(\mathbf{x}_{nt}^T \boldsymbol{\beta}) - (b \circ \mathcal{G})(\mathbf{x}_{nt}^T \boldsymbol{\beta}_0)\right] - \mathcal{H}(\mathbf{x}_{nt}^T \boldsymbol{\beta}_0)\left[\mathcal{G}(\mathbf{x}_{nt}^T \boldsymbol{\beta}) - \mathcal{G}(\mathbf{x}_{nt}^T \boldsymbol{\beta}_0)\right]\right\} \\
&\quad - \sum_{t=1}^{n} \left[Y_t - \mathcal{H}(\mathbf{x}_{nt}^T \boldsymbol{\beta}_0)\right]\left[\mathcal{G}(\mathbf{x}_{nt}^T \boldsymbol{\beta}) - \mathcal{G}(\mathbf{x}_{nt}^T \boldsymbol{\beta}_0)\right]. \qquad (2.42)
\end{aligned}$$

Defining the vector $\mathbf{u}$ by

$$\mathbf{u} \equiv \mathbf{M}_n^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0),$$

we can rewrite (2.42) by

$$g_n(\mathbf{u}) \equiv -\left[l(\boldsymbol{\beta}) - l(\boldsymbol{\beta}_0)\right] = -\left[l(\boldsymbol{\beta}_0 + \mathbf{M}_n \mathbf{u}) - l(\boldsymbol{\beta}_0)\right] = g_{n,1}(\mathbf{u}) - g_{n,2}(\mathbf{u}),$$

where

$$g_{n,1}(\mathbf{u}) \quad = \quad \sum_{t=1}^{n} \left[(b \circ \mathcal{G})(\mathbf{x}_{nt}^T \boldsymbol{\beta}_0 + \mathbf{x}_{nt}^T \mathbf{M}_n \mathbf{u}) - (b \circ \mathcal{G})(\mathbf{x}_{nt}^T \boldsymbol{\beta}_0)\right]$$

$$-\sum_{t=1}^{n}\mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\left[\mathcal{G}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)-\mathcal{G}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\right],$$

and

$$g_{n,2}(\mathbf{u})=\sum_{t=1}^{n}\left[Y_t-\mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\right]\left[\mathcal{G}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)-\mathcal{G}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\right].$$

For any fixed $\mathbf{u}$, we show

$$g_{n,1}(\mathbf{u})\to\frac{1}{2}\mathbf{u}^T\Omega_1^\dagger(\boldsymbol{\beta}_0)\mathbf{u} \tag{2.43}$$

and

$$g_{n,2}(\mathbf{u})\xrightarrow{d}\mathbf{u}^T\cdot\mathrm{N}\left(\mathbf{0},\Omega_1^\ddagger(\boldsymbol{\beta}_0)+\Omega_3^\ddagger(\boldsymbol{\beta}_0)\right). \tag{2.44}$$

Since $\sup_{1\le t\le n}\left|\mathbf{M}_n^T\mathbf{x}_{nt}\right|\to 0$, we apply a third degree Taylor expansion of $b\circ\mathcal{G}(\cdot)$ about $\mathbf{x}_{nt}^T\boldsymbol{\beta}_0$ when n is large enough and obtain

$$\begin{aligned}
&(b\circ\mathcal{G})\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)\\
&=(b\circ\mathcal{G})(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)+(b'\circ\mathcal{G})(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)\cdot\mathcal{G}'(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)\cdot\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\\
&\quad+\frac{1}{2}[(b'\circ\mathcal{G})'(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)\cdot\mathcal{G}'(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)+(b'\circ\mathcal{G})(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)\mathcal{G}''(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)](\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u})^2\\
&\quad+C_{nt}^1(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u})^3\\
&=(b\circ\mathcal{G})(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)+\mathcal{H}(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)\mathcal{G}'(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\\
&\quad+\frac{1}{2}[\mathcal{H}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\mathcal{G}'(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)+\mathcal{H}(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)\mathcal{G}''(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)](\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u})^2\\
&\quad+C_{nt}^1(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u})^3,
\end{aligned}$$

where $C_{nt}^1=(b\circ\mathcal{G})^{(3)}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\delta_{nt}^1\right)/3!$ with δ_{nt}^1 lying between 0 and $\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}$. Likewise, expanding $\mathcal{G}(\cdot)$ in a third degree Taylor series about $\mathbf{x}_{nt}^T\boldsymbol{\beta}_0$ yields

$$\begin{aligned}
\mathcal{G}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right) &= \mathcal{G}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)+\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\\
&\quad+\frac{1}{2}\mathcal{G}''\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^2+C_{nt}^2\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^3,
\end{aligned}$$

where $C_{nt}^2=\mathcal{G}^{(3)}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\delta_{nt}^2\right)/3!$ for some δ_{nt}^2 lying between 0 and $\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}$. It follows that

$$g_{n,1}(\mathbf{u}) = \frac{1}{2}\sum_{t=1}^{n} \mathcal{H}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right) \mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right) \left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^2 + \sum_{t=1}^{n}\left(C_{nt}^1 - C_{nt}^2\right)\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^3. \quad (2.45)$$

The term $E_n^1(\mathbf{u}) \equiv \sum_{t=1}^n \left(C_{nt}^1 - C_{nt}^2\right)\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^3$ converges to zero because

$$\left|E_n^1(\mathbf{u})\right| \leq \sum_{t=1}^{n}\left|C_{nt}^1 - C_{nt}^2\right|\left|\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right|^3 \to 0,$$

where $\left|C_{nt}^1 - C_{nt}^2\right|$ is bounded uniformly in t and $\sup_{1\leq t\leq n}\left|\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right|^3 = O(n^{-\frac{3}{2}})$. Therefore, (2.43) follows immediately from (2.45).

To show (2.44), we first expand $\mathcal{G}(\cdot)$ in a second degree Taylor series about $\mathbf{x}_{nt}^T\boldsymbol{\beta}_0$ to obtain

$$\mathcal{G}\left(\mathbf{x}_{nt}^T\left(\boldsymbol{\beta}_0 + \mathbf{M}_n\mathbf{u}\right)\right) = \mathcal{G}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right) + \mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u} + C_{nt}^3\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^2,$$

where $C_{nt}^3 = \mathcal{G}''\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \delta_{nt}^3\right)/2$ for some δ_{nt}^3 lying between 0 and $\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}$. It follows that

$$\begin{aligned} g_{n,2}(\mathbf{u}) &= \sum_{t=1}^{n}\left(Y_t - \mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\right)\left[\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u} + C_{nt}^3\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^2\right] \\ &\equiv \mathbf{u}^T\mathbf{U}_n + E_n^2(\mathbf{u}), \end{aligned} \quad (2.46)$$

where

$$\mathbf{U}_n = \sum_{t=1}^{n}\left(Y_t - \mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\right)\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\mathbf{M}_n^T\mathbf{x}_{nt},$$

and $E_n^2(\mathbf{u}) = \sum_{t=1}^n C_{nt}^3\left(Y_t - \mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\right)\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^2$. The variance of $E_n^2(\mathbf{u})$ goes to zero because

$$\begin{aligned} \operatorname{Var}\left(E_n^2(\mathbf{u})\right) &= \operatorname{Var}\left[\sum_{t=1}^{n} C_{nt}^3\left(Y_t - \mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\right)\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^2\right] \\ &= \sum_{j=1}^{n}\sum_{t=1}^{n} C_{nj}^3 C_{nt}^3\left(\mathbf{x}_{nj}^T\mathbf{M}_n\mathbf{u}\right)^2\left(\mathbf{x}_{nt}^T\mathbf{M}_n\mathbf{u}\right)^2\operatorname{Cov}\left(Y_j, Y_t\right) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{t=1}^{n} \left(C_{nt}^{3}\right)^{2} \left(\mathbf{x}_{nt}^{T}\mathbf{M}_{n}\mathbf{u}\right)^{4} \mathrm{E}\left[\left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^{T}\boldsymbol{\beta}_{0}+\alpha_{t}\right)\right] \\
&\quad + \sum_{j=1}^{n}\sum_{t=1}^{n} C_{nj}^{3}C_{nt}^{3} \left(\mathbf{x}_{nj}^{T}\mathbf{M}_{n}\mathbf{u}\right)^{2}\left(\mathbf{x}_{nt}^{T}\mathbf{M}_{n}\mathbf{u}\right)^{2} \\
&\qquad\qquad \times \mathrm{Cov}\left[\mathcal{H}\left(\mathbf{x}_{nj}^{T}\boldsymbol{\beta}_{0}+\alpha_{j}\right), \mathcal{H}\left(\mathbf{x}_{nt}^{T}\boldsymbol{\beta}_{0}+\alpha_{t}\right)\right] \\
&\leq \mathrm{Const} \cdot \left(\sup_{1\leq t\leq n} \left|\mathbf{x}_{nt}^{T}\mathbf{M}_{n}\mathbf{u}\right|\right)^{2} \\
&\quad \times \{\mathbf{u}^{T}\{\mathbf{M}_{n}^{T}\sum_{t=1}^{n}\mathbf{x}_{nt}\mathbf{x}_{nt}^{T}\mathrm{E}[\left(b''\circ\mathcal{G}\right)\left(\mathbf{x}_{nt}^{T}\boldsymbol{\beta}_{0}+\alpha_{t}\right)]\mathbf{M}_{n}\}\mathbf{u} \\
&\qquad + \mathbf{u}^{T}\{\mathbf{M}_{n}^{T}\sum_{j,t=1}^{n}\mathbf{x}_{nj}\mathbf{x}_{nt}^{T}\mathrm{Cov}[\mathcal{H}(\mathbf{x}_{nj}^{T}\boldsymbol{\beta}_{0}+\alpha_{j}), \mathcal{H}(\mathbf{x}_{nt}^{T}\boldsymbol{\beta}_{0}+\alpha_{t})]\mathbf{M}_{n}\}\mathbf{u}\}.
\end{aligned}
$$

It follows that, $E_n^2(\mathbf{u})$ converges to its mean zero in probability. Then, in order to show (2.44), it suffices to show

$$
\mathbf{U}_n \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \Omega_1^{\ddagger}(\boldsymbol{\beta}_0) + \Omega_3^{\ddagger}(\boldsymbol{\beta}_0)\right). \tag{2.47}
$$

The characteristic function of a one-parameter exponential family distribution is given by

$$
\phi(s) = \exp\left[b(\theta_t + is) - b(\theta_t)\right].
$$

Recall that $\theta_t = \mathcal{G}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)$. Then, for every real vector $\mathbf{s}$,

$$
\begin{aligned}
&\mathrm{E}\left(e^{i\mathbf{s}^{T}\mathbf{U}_{n}}|\alpha_{t}\right) \\
&\quad = \exp\left[-\sum_{t=1}^{n} i\mathbf{s}^{T}\mathbf{M}_{n}^{T}\mathbf{x}_{nt}\mathcal{H}\left(\mathbf{x}_{nt}^{T}\boldsymbol{\beta}_{0}\right)\mathcal{G}'\left(\mathbf{x}_{nt}^{T}\boldsymbol{\beta}_{0}\right)\right] \\
&\qquad \times \prod_{t=1}^{n}\mathrm{E}\left\{\exp\left[i\mathbf{s}^{T}\mathbf{M}_{n}^{T}\mathbf{x}_{nt}\mathcal{G}'\left(\mathbf{x}_{nt}^{T}\boldsymbol{\beta}_{0}\right)Y_{t}\right]|\alpha_{t}\right\} \\
&\quad = \exp\left[-\sum_{t=1}^{n} i\mathbf{s}^{T}\mathbf{M}_{n}^{T}\mathbf{x}_{nt}\mathcal{G}'\left(\mathbf{x}_{nt}^{T}\boldsymbol{\beta}_{0}\right)\mathcal{H}\left(\mathbf{x}_{nt}^{T}\boldsymbol{\beta}_{0}\right)\right] \\
&\qquad \times \exp\left\{\sum_{t=1}^{n}\left[\left(b\circ\mathcal{G}\right)\left(\mathbf{x}_{nt}^{T}\boldsymbol{\beta}_{0}+\alpha_{t}+i\mathbf{s}^{T}\mathbf{M}_{n}^{T}\mathbf{x}_{nt}\mathcal{G}'\left(\mathbf{x}_{nt}^{T}\boldsymbol{\beta}_{0}\right)\right)-\left(b\circ\mathcal{G}\right)\left(\mathbf{x}_{nt}^{T}\boldsymbol{\beta}_{0}+\alpha_{t}\right)\right]\right\}.
\end{aligned}
$$

When n is large enough, by applying a Taylor expansion we get

$$\begin{aligned}
&(b \circ \mathcal{G})\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t + i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\right) \\
&\quad = (b \circ \mathcal{G})\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right) + i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right) \\
&\qquad - \frac{1}{2}\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2\left[\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\right]^2 \cdot \left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right) \\
&\qquad + \sum_{j=3}^{\infty}\left(i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^j\left[\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\right]^j\left(b^{(j)} \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)/j!\,.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathrm{E}\left(e^{i\mathbf{s}^TU_n}|\alpha_t\right) \\
&\quad = \exp[-\sum_{t=1}^{n} i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\mathcal{G}'(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)\mathcal{H}(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)] \\
&\qquad \times \exp[\sum_{t=1}^{n} i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\mathcal{G}'(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)\mathcal{H}(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t)] \\
&\qquad \times \exp[-\frac{1}{2}\sum_{t=1}^{n}\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2[\mathcal{G}'(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)]^2\left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)] \times \exp\left(E_n^3\right) \\
&\quad = \exp\{\sum_{t=1}^{n} i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\mathcal{G}'(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)[\mathcal{H}(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t) - \mathcal{H}(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)]\} \\
&\qquad \times \exp\{-\frac{1}{2}\sum_{t=1}^{n}\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2[\mathcal{G}'(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)]^2\mathrm{E}[\left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)]\} \\
&\qquad \times \exp\{-\frac{1}{2}\sum_{t=1}^{n}\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2[\mathcal{G}'(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)]^2\cdot \\
&\qquad\qquad \{\left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right) - \mathrm{E}[\left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)]\}\} \\
&\qquad \times \exp\left(E_n^3\right) \\
&\quad \equiv \exp\left(iC_n\left(\mathbf{s}\right) + D_n + F_n + E_n^3\right),
\end{aligned}$$

where $E_n^3 = \sum_{t=1}^{n}\sum_{j=3}^{\infty}\left(i\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^j\left[\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\right]^j\left(b^{(j)} \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)/j!$. It is easy to check that $E_n^3 \to 0$ in probability. As for D_n, we have

$$D_n \equiv -\frac{1}{2}\sum_{t=1}^{n}\left(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right)^2\left[\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\right]^2\mathrm{E}\left[\left(b'' \circ \mathcal{G}\right)\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0 + \alpha_t\right)\right]$$

$$\rightarrow \quad -\frac{1}{2}\mathbf{s}^T\Omega_1^{\ddagger}(\boldsymbol{\beta}_0)\,\mathbf{s}. \tag{2.48}$$

Moving on to F_n, its mean is zero and its variance goes to zero:

$$\begin{aligned}
\mathrm{Var}(F_n) &= \mathrm{Var}\{-\frac{1}{2}\sum_{t=1}^{n}(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt})^2[\mathcal{G}'(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)]^2\cdot(b''\circ\mathcal{G})(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\alpha_t)\} \\
&= \frac{1}{4}\sum_{j=1}^{n}\sum_{t=1}^{n}\{(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nj})^2(\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt})^2[\mathcal{G}'(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0)]^2[\mathcal{G}'(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0)]^2 \\
&\qquad\qquad \times \mathrm{Cov}[(b''\circ\mathcal{G})(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0+\alpha_j),(b''\circ\mathcal{G})(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\alpha_t)]\} \\
&\le \mathrm{Const}\cdot(\sup_{1\le t\le n}\left|\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right|)^4 \\
&\quad \times\sum_{j=1}^{n}\sum_{t=1}^{n}\left|\mathrm{Cov}[(b''\circ\mathcal{G})(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0+\alpha_j),(b''\circ\mathcal{G})(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\alpha_t)]\right| \\
&\rightarrow 0,
\end{aligned}$$

under the assumptions that $\sup_{1\le t\le n}\left|\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\right| = O(1/\sqrt{n})$ and

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{t=1}^{n}\left|\mathrm{Cov}\left[(b''\circ\mathcal{G})\left(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0+\alpha_j\right),(b''\circ\mathcal{G})\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\alpha_t\right)\right]\right| < \infty.$$

It follows that

$$F_n \xrightarrow{P} 0. \tag{2.49}$$

Moreover, if

$$C_n(\mathbf{s}) \xrightarrow{d} V \sim \mathrm{N}\left(0, \mathbf{s}^T\Omega_3^{\ddagger}(\boldsymbol{\beta}_0)\,\mathbf{s}\right),$$

then

$$iC_n(\mathbf{s}) + D_n + F_n + E_n^3 \xrightarrow{d} iV - \frac{1}{2}\mathbf{s}^T\Omega_1^{\ddagger}(\boldsymbol{\beta}_0)\,\mathbf{s},$$

and

$$\begin{aligned}
\mathrm{E}\left(e^{i\mathbf{s}^T\mathbf{U}_n}\right) &= \mathrm{E}\left(\mathrm{E}\left(e^{i\mathbf{s}^T\mathbf{U}_n}|\alpha_t\right)\right) \\
&= \mathrm{E}\left[\exp\left(iC_n(\mathbf{s})+D_n+F_n+E_n^3\right)\right] \\
&\rightarrow \exp\left\{-\frac{1}{2}\mathbf{s}^T\left[\Omega_1^{\ddagger}(\boldsymbol{\beta}_0)+\Omega_3^{\ddagger}(\boldsymbol{\beta}_0)\right]\mathbf{s}\right\}
\end{aligned}$$

which implies (2.47). The variance $\mathbf{s}^T\Omega_3^{\ddagger}(\boldsymbol{\beta}_0)\,\mathbf{s}$ of V is actually the limit of $\text{Var}[C_n(\mathbf{s})]$, that is,

$$\begin{aligned}
\text{Var}\left[C_n(\mathbf{s})\right] &= \text{Var}[\sum_{t=1}^{n}\mathbf{s}^T\mathbf{M}_n^T\mathbf{x}_{nt}\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right)\mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\alpha_t\right)]^2 \\
&= \mathbf{s}^T\mathbf{M}_n^T\sum_{j=1}^{n}\sum_{t=1}^{n}\mathbf{x}_{nj}\mathbf{x}_{nt}^T\mathcal{G}'\left(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0\right)\mathcal{G}'\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0\right) \\
&\qquad\qquad \times \text{Cov}\left[\mathcal{H}\left(\mathbf{x}_{nj}^T\boldsymbol{\beta}_0+\alpha_j\right),\mathcal{H}\left(\mathbf{x}_{nt}^T\boldsymbol{\beta}_0+\alpha_t\right)\right]\mathbf{M}_n\mathbf{s} \\
&\to \mathbf{s}^T\Omega_3^{\ddagger}(\boldsymbol{\beta}_0)\,\mathbf{s}.
\end{aligned}$$

Since g_n has convex sample paths,

$$g_n(\mathbf{u}) \to g(\mathbf{u}) \equiv \frac{1}{2}\mathbf{u}^T\Omega_1^{\dagger}(\boldsymbol{\beta}_0)\,\mathbf{u} - \mathbf{u}^T\cdot \text{N}\left(\mathbf{0},\Omega_1^{\ddagger}(\boldsymbol{\beta}_0)+\Omega_3^{\ddagger}(\boldsymbol{\beta}_0)\right)$$

in distribution on the space $C\left(\mathbb{R}^l\right)$ (see Rockafellar (1970) and Pollard (1991)). Therefore, the estimator $\widehat{\mathbf{u}}_n = \mathbf{M}_n^{-1}\left(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)$, which minimizes $g_n(\mathbf{u})$, converges in distribution to the minimizer $\widehat{\mathbf{u}}$ of $g(\mathbf{u})$. But

$$\widehat{\mathbf{u}} = \left[\Omega_1^{\dagger}(\boldsymbol{\beta}_0)\right]^{-1}\cdot \text{N}\left(\mathbf{0},\Omega_1^{\ddagger}(\boldsymbol{\beta}_0)+\Omega_3^{\ddagger}(\boldsymbol{\beta}_0)\right).$$

So

$$\mathbf{M}_n^{-1}\left(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right) \xrightarrow{d} \text{N}\left(\mathbf{0},\left[\Omega_1^{\dagger}(\boldsymbol{\beta}_0)\right]^{-1}\left[\Omega_1^{\ddagger}(\boldsymbol{\beta}_0)+\Omega_3^{\ddagger}(\boldsymbol{\beta}_0)\right]\left[\Omega_1^{\dagger}(\boldsymbol{\beta}_0)\right]^{-1}\right).$$

This finishes the proof. ■

3 LAD ESTIMATION FOR MA(1) MODELS

In this chapter we study LAD parameter estimation for MA(1) models $X_t = Z_t - \theta Z_{t-1}$, where $|\theta| \neq 1$. The unit root case ($|\theta| = 1$), which is not the concern of this dissertation, was first studied by Kang (1975); and it has been investigated by many researchers since then. See Davis and Dunsmuir (1996) for a review of the Gaussian likelihood estimation for MA(1) processes with unit root. Recently, Breidt, Davis, Hsu, and Rosenblatt (2006) studied the pile-up phenomenon of MA(1) with unit root for the Laplacian likelihood estimator $\widehat{\theta}$, and developed the limiting pile-up probabilities $\lim \mathrm{P}(\widehat{\theta} = 1)$. For general noninvertible MA(q) models, Lii and Rosenblatt (1992) studied noninvertible MA processes driven by IID non-Gaussian noise. They proposed an approximate maximum likelihood procedure for the parameter estimation, and established the consistency and asymptotic normality of the estimator. Huang and Pawitan (2000) studied the pseudo-likelihood associated with the Laplacian model for MA(q) processes. Safi and Zeroual (2002) proposed a high-order (≥ 2) cumulant-based procedure to estimate the parameters of a noninvertible MA process and performed comparison with existing methods regarding the algorithm efficiency. Breidt and Hsu (2005) studied best mean square prediction for noninvertible MA processes and presented stable numerical recursions for computation of residuals and evaluation of unnormalized conditional distributions.

We begin with the derivation of the LAD criterion. Then we show the existence of a sequence of local LAD estimators that is consistent and asymptotic normal. Next we study the global LAD estimator; with additional assumptions on the distribution of the underlying noise, the global LAD estimator is strongly consistent and asymptotically normal having the same asymptotic normal law as for the local LAD estimator. We also propose a linearized LAD estimator, which is motivated by the simplification of searching for a minimum. Simulation study is undertaken to evaluate the asymptotic theory about the LAD estimators.

The following part is the development of the LAD criterion based on Breidt, Davis, Hsu, and Rosenblatt (2006). Let $\{X_t\}$ be an MA(1) process satisfying

$$X_t = Z_t - \theta Z_{t-1}, \tag{3.1}$$

where $\theta \in \mathbb{R}$ and $\{Z_t\}$ is a sequence of IID random variables with mean zero. Given the observed data $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)^T$, the LAD criterion is derived by likelihood approximation assuming that the underlying noise $\{Z_t\}$ is Laplacian (or, double exponential). In order to derive the criterion, it is convenient to have an initial variable from $\{Z_t\}$; moreover, this initial variable is augmented to account for both invertible and noninvertible cases. The augmented initial variable Z_{aug} is defined by

$$Z_{aug} = \begin{cases} Z_0, & \text{if } |\theta| \leq 1, \\ Z_n, & \text{otherwise.} \end{cases}$$

Simple derivations yield the following likelihood of θ given $\mathbf{X}_n = \mathbf{x}_n$ and $Z_{aug} = z_{aug}$:

$$\begin{aligned} \mathcal{L}(\theta|\mathbf{x}_n, z_{aug}) &= f(z_0, z_1, \cdots, z_n)\left[1_{\{|\theta|\leq 1\}} + |\theta|^{-n} 1_{\{|\theta|>1\}}\right] \\ &= \prod_{t=0}^{n} f_Z(z_t)\left[1_{\{|\theta|\leq 1\}} + |\theta|^{-n} 1_{\{|\theta|>1\}}\right], \end{aligned}$$

where $f_Z(\cdot)$ is the density of the underlying noise $\{Z_t\}$, and the residual z_t, $t = 0, \ldots, n$, is a function of θ, $\mathbf{X}_n = \mathbf{x}_n$, and $Z_{aug} = z_{aug}$, which will be denoted by $z_t(\theta)$ from now on, and can be solved forward by $z_t(\theta) = X_t + \theta z_{t-1}(\theta)$ for $t = 1, 2, \ldots, n$ with $z_0(\theta) = z_{aug}$ if $|\theta| \leq 1$, and backward by $z_{t-1}(\theta) = (z_t(\theta) - X_t)/\theta$ for $t = n, n-1, \ldots, 1$ with $z_n(\theta) = z_{aug}$ if $|\theta| > 1$. Suppose the density of underlying noise is Laplacian; that is, suppose $f_Z(z) = \frac{1}{2\sigma}\exp\left(-\frac{|z|}{\sigma}\right)$. If we view z_{aug} as a parameter, then the joint Laplacian log-likelihood is given by

$$-(n+1)\log(2\sigma)-\frac{1}{\sigma}\sum_{t=0}^{n}|z_t(\theta)|-n\log|\theta|\,1_{\{|\theta|>1\}}. \tag{3.2}$$

We first maximize (3.2) with respect to σ, and obtain its estimator $\widehat{\sigma}=\sum_{t=0}^{n}\frac{|z_t(\theta)|}{n+1}$. Then, by plugging $\widehat{\sigma}$ into (3.2) we get the concentrated Laplacian log-likelihood

$$-(n+1)\log 2-(n+1)\log\sum_{t=0}^{n}|z_t(\theta)|+(n+1)\log(n+1)-(n+1)-n\log|\theta|\,1_{\{|\theta|>1\}}. \tag{3.3}$$

But, maximizing (3.3) is equivalent to minimizing the objective function

$$l_n(\theta,z_{aug})=\begin{cases}\sum_{t=0}^{n}|z_t(\theta)|, & \text{if } |\theta|\le 1,\\ \sum_{t=0}^{n}|\theta z_t(\theta)|, & \text{otherwise.}\end{cases} \tag{3.4}$$

Notice that, in the objective function (3.4), if $|\theta|\le 1$, then we get a standard L_1 loss function; otherwise, if $|\theta|>1$, we need to modify the L_1 loss function by θ. This is a key point; without this modification, the standard L_1 loss function may not produce a consistent estimator.

Now suppose the true model is

$$X_t=Z_t-\theta_0 Z_{t-1}, \tag{3.5}$$

where $\{Z_t\}$ is assumed to be IID with $\mathrm{E}(Z_t)=0$, $\mathrm{E}[\mathrm{sgn}(Z_t)]=0$ (i.e., median of Z_t is zero), $\mathrm{E}(Z_t^4)<\infty$, and probability density function $f_Z(z)=\frac{1}{\sigma}f\left(\frac{z}{\sigma}\right)$, where $\sigma>0$ and $f(\cdot)$ is chosen such that $\int_{-\infty}^{\infty}z^2f(z)dz=1$. Then, it is easy to see that $\mathrm{Var}(Z_t)=\sigma^2$. We exclude the unit root case, which is a separate topic. That is, we assume that $|\theta_0|\neq 1$. Let $\mathbf{X}_n=(X_1,X_2,\ldots,X_n)^T$ be observed data from the true model (3.5). In this chapter we study the asymptotic behavior of the LAD estimator, which is defined as any minimizer of the objective function (3.4).

3.1 Local LAD Estimator

In this section we follow the idea of Davis and Dunsmuir (1997), and build the sample size into the parameterizations (3.6) of θ and z_{aug} (note that z_{aug} is treated as a parameter). We establish the consistency and asymptotic normality of the local LAD estimator, which are based on the following functional limit result.

Theorem 3.1 *Under the parameterizations*

$$\theta = \theta_0 + \frac{\beta}{\sqrt{n}} \qquad \textit{and} \qquad z_{aug} = Z_{aug} + \frac{\alpha\sigma}{\sqrt{n}}, \tag{3.6}$$

we have

$$U_n(\beta, \alpha) \equiv \frac{1}{\sigma}\left[l_n(\theta, z_{aug}) - l_n(\theta_0, Z_{aug})\right] \to U(\beta)$$

in distribution on $C(\mathbb{R}^2)$, where

$$U(\beta) = \beta N + \frac{\beta^2 f(0)}{1-\theta_0^2}, \qquad N \sim N\left(0, \frac{1}{1-\theta_0^2}\right)$$

if the true model is invertible, and

$$U(\beta) = \beta N + \frac{\beta^2 f(0)}{|\theta_0|\left(\theta_0^2 - 1\right)}, \qquad N \sim N\left(0, \frac{1}{\theta_0^2 - 1}\right)$$

if the true model is noninvertible.

Proof. When the true model is invertible, i.e., $|\theta_0| < 1$, under the parameterizations (3.6), we can confine our study to the reduced parameter space $\Theta_I \equiv \{\theta : |\theta| \leq 1 - \delta\}$, where $\delta > 0$ is a small number such that $|\theta_0| \leq 1 - \delta$. Then, minimizing the objective function (3.4) is equivalent to minimizing $\sum_{t=0}^{n} |z_t(\theta)|$, and the result is a special case of Proposition 1 of Davis and Dunsmuir (1997). Note that, the initial innovation z_{aug} in Davis and Dunsmuir (1997) is set to zero. However, as we can see, z_{aug} does not affect the asymptotic results (see Remark 3.1). So the LAD objective function used in Davis and Dunsmuir (1997) is essentially the same as (3.4) for the $|\theta| \leq 1$ case.

When the true model is noninvertible, we can reverse time to get an invertible MA(1) representation of the model and then Proposition 1 of Davis and Dunsmuir (1997) applies. To be specific, we rewrite (3.1) as

$$X_t = \tilde{Z}_t - \frac{1}{\theta}\tilde{Z}_{t+1}, \tag{3.7}$$

where $\tilde{Z}_t = -\theta Z_{t-1}$. Notice that, (3.7) is a reverse-time invertible MA(1) model with parameters $(\tilde{\theta}, \tilde{\sigma})$, where $\tilde{\theta} = 1/\theta$ and $\tilde{\sigma} = \theta\sigma$, and the LAD objective function in Davis and Dunsmuir (1997) becomes $\sum_{t=1}^{n} \left|\tilde{z}_t(\tilde{\theta})\right|$ (that is, $\sum_{t=0}^{n-1} |\theta z_t(\theta)|$) with $\tilde{z}_{n+1}(\tilde{\theta}) = 0$ (that is, $z_n(\theta) = 0$). So, the LAD objective function of Davis and Dunsmuir (1997) is the same as (3.4) for the $|\theta| > 1$ case. ■

REMARK 3.1. The limit process $U(\beta)$ is independent of α, and hence independent of z_{aug}. Actually z_{aug} has no impact on the above functional limit result or the asymptotic results of the LAD estimator, because it only gives rise to an absolutely summable term in $l_n(\theta, z_{aug})$ that can be ignored when $n \to \infty$. So we can set its value depending upon circumstances; for example, we can set z_{aug} to zero to simplify numerical calculations, then the LAD objective function is identical to that in Davis and Dunsmuir (1997) when the true model is invertible.

REMARK 3.2. The limit process $U(\beta)$ is a quadratic function of β and therefore has a unique minimizer $\beta_{\min}$. It is easy to verify that

$$\beta_{\min} = \frac{\sqrt{1-\theta_0^2}}{2f(0)} N_1$$

if the true model is invertible ($|\theta_0| < 1$), and

$$\beta_{\min} = \frac{|\theta_0|\sqrt{\theta_0^2 - 1}}{2f(0)} N_1$$

if the true model is noninvertible ($|\theta_0| > 1$), where $N_1 \sim \mathrm{N}(0,1)$.

The asymptotic results of the local LAD estimator follows immediately from Theorem 3.1 and Remark 1 of Davis, Knight, and Liu (1992).

Theorem 3.2 *There exists a sequence of local LAD estimators* $\widehat{\theta}_{LAD}$ *such that*

$$\sqrt{n}\left(\widehat{\theta}_{LAD}-\theta_0\right)\xrightarrow{d}\beta_{\min}.$$

Suppose we define LAD estimator solely based on the observe data $\mathbf{X}_n$. To be specific, assuming Laplacian underlying noise, we derive the likelihood function $\mathcal{L}(\theta,\sigma|\mathbf{x}_n)$ by integrating out z_{aug} from the joint density $f(\mathbf{x}_n, z_{aug})$ of $\mathbf{X}_n$ and Z_{aug}. And we define LAD estimator to be any maximizer $\widetilde{\theta}$ of

$$\mathcal{L}(\theta,\sigma|\mathbf{x}_n)\equiv\int_{-\infty}^{\infty} f(\mathbf{x}_n, z_{aug})\,dz_{aug},$$

and denote by $\widetilde{\sigma}$ the corresponding estimator of σ. Then, there exists a sequence of local LAD estimators $\widetilde{\theta}$, and corresponding $\widetilde{\sigma}$ such that

$$\begin{aligned}\sqrt{n}\left(\widetilde{\theta}-\theta_0\right) &\rightarrow \beta_{\min},\\ \sqrt{n}\left(\widetilde{\sigma}-\mathrm{E}\left|Z_1\right|\right) &\rightarrow \mathrm{N}\left(0,\sigma^2\right),\end{aligned}$$

in distribution as $n\to\infty$.

We give a heuristic derivation for the case when $|\theta_0|>1$. Under the parameterization $\theta=\theta_0+\beta/\sqrt{n}$, we can confine our study to the reduced parameter space $\Theta_{NI}\equiv\{\theta:|\theta|\geq 1+\delta\}$, where $\delta>0$ is a small number such that $|\theta_0|>1+\delta$. Then the joint density of $\mathbf{X}_n$ and Z_{aug} satisfies

$$\begin{aligned}f(\mathbf{x}_n, z_{aug}) &= |\theta|^{-n}\prod_{t=0}^{n} f_Z(z_t(\theta))\\ &= |\theta|^{-n}\left(\frac{1}{2\sigma}\right)^{n+1}\exp\left(-\frac{\sum_{t=0}^{n}|z_t(\theta)|}{\sigma}\right)\end{aligned}$$

$$
\begin{aligned}
&= |\theta|^{-n}\left(\frac{1}{2\sigma}\right)^{n+1}\exp\left(-\frac{\sum_{t=0}^{n}|\theta_0 Z_t|}{\sigma|\theta|}\right)\exp\left(-\frac{U_n(\beta,\alpha)}{|\theta|}\right)\\
&= |\theta|^{-n}\left(\frac{1}{2\sigma}\right)^{n+1}\exp\left(-\frac{\sum_{t=0}^{n}|\theta_0 Z_t|}{\sigma|\theta|}\right)\\
&\quad\times\exp\left(-\frac{U_n(\beta,0)}{|\theta|}\right)\exp\left(-\frac{U_n(\beta,\alpha)-U_n(\beta,0)}{|\theta|}\right),
\end{aligned}
$$

where $\beta = \sqrt{n}(\theta-\theta_0)$ and $\alpha = \sqrt{n}(z_{aug}-Z_n)/\sigma$. Notice that $dz_{aug} = (\sigma/\sqrt{n})\,d\alpha$. By a change of variables in the integration, we have

$$
\begin{aligned}
f(\mathbf{x}_n) &= \frac{\sigma}{\sqrt{n}}|\theta|^{-n}\left(\frac{1}{2\sigma}\right)^{n+1}\exp\left(-\frac{\sum_{t=0}^{n}|\theta_0 Z_t|}{\sigma|\theta|}\right)\\
&\quad\times\exp\left(-\frac{U_n(\beta,0)}{|\theta|}\right)\int_{-\infty}^{\infty}\exp\left(-\frac{U_n(\beta,\alpha)-U_n(\beta,0)}{|\theta|}\right)d\alpha.
\end{aligned}
$$

Then, the Laplacian log-likelihood of (θ,σ) given $\mathbf{X}_n = \mathbf{x}_n$ satisfies

$$
\begin{aligned}
\mathcal{L}_n^*(\theta,\sigma) &= \log\frac{\sigma}{\sqrt{n}} - n\log|\theta| - (n+1)\log(2\sigma) - \frac{\sum_{t=0}^{n}|\theta_0 Z_t|}{\sigma|\theta|}\\
&\quad -\frac{U_n(\beta,0)}{|\theta|} + \log\int_{-\infty}^{\infty}\exp\left(-\frac{U_n(\beta,\alpha)-U_n(\beta,0)}{|\theta|}\right)d\alpha. \qquad (3.8)
\end{aligned}
$$

Because

$$
U_n(\beta,0)\xrightarrow{d}U(\beta) \qquad \text{and} \qquad U_n(\beta,\alpha)-U_n(\beta,0)\xrightarrow{P}0,
$$

the last two terms on the right-hand side of (3.8) do not depend on σ as $n\to\infty$. By maximizing $\mathcal{L}_n^*(\theta,\sigma)$ with respect to σ, we obtain $\widetilde{\sigma}\approx\sum_{t=0}^{n}|\theta_0 Z_t|/(n|\theta|)$. Then the concentrated Laplacian log-likelihood is given by

$$
\begin{aligned}
\mathcal{L}_n^*(\theta,\widetilde{\sigma}) &\approx -(n+1)\log 2 - n\log\sum_{t=0}^{n}|\theta_0 Z_t| + \left(n-\frac{1}{2}\right)\log n - n\\
&\quad -\frac{U_n(\beta,0)}{|\theta|} + \log\int_{-\infty}^{\infty}\exp\left(-\frac{U_n(\beta,\alpha)-U_n(\beta,0)}{|\theta|}\right)d\alpha.
\end{aligned}
$$

Under the parameterization $\theta = \theta_0 + \beta/\sqrt{n}$, maximizing $\mathcal{L}_n^*(\theta,\widetilde{\sigma})$ with respect to θ to find $\widetilde{\theta}$ is equivalent to maximizing $\mathcal{L}_n^*(\theta_0+\beta/\sqrt{n},\widetilde{\sigma})$ with respect to β to find $\widetilde{\beta}$, the local maximizer of $\mathcal{L}_n^*\left(\theta_0+\beta/\sqrt{n},\widetilde{\sigma}\right)$. But

$$\begin{aligned}&\arg\max_{\beta}\mathcal{L}_n^*\left(\theta_0+\beta/\sqrt{n},\widetilde{\sigma}\right)\\&\qquad\approx\arg\min_{\beta}\left[\frac{U_n(\beta,0)}{|\theta_0+\beta/\sqrt{n}|}-\log\int_{-\infty}^{\infty}\exp\left(-\frac{U_n(\beta,\alpha)-U_n(\beta,0)}{|\theta_0+\beta/\sqrt{n}|}\right)d\alpha\right].\end{aligned}\tag{3.9}$$

Since the second term on the right-hand side of (3.9) does not depend on β as $n\to\infty$, we obtain

$$\widetilde{\beta}\xrightarrow{d}\arg\min_{\beta}\left(U\left(\beta\right)/\left|\theta_0\right|\right)=\beta_{\min}$$

It follows that

$$\sqrt{n}\left(\widetilde{\theta}-\theta_0\right)\xrightarrow{d}\beta_{\min}.$$

Moreover,

$$\widetilde{\sigma}\equiv\widetilde{\sigma}(\widetilde{\theta})=\frac{1}{n}\sum_{t=0}^{n}\left|\theta_0 Z_t\right|/|\widetilde{\theta}|.$$

Since

$$\widetilde{\theta}\xrightarrow{P}\theta_0\qquad\text{and}\qquad\sqrt{n}\left(\frac{1}{n}\sum_{t=0}^{n}|Z_t|-\mathrm{E}\left|Z_1\right|\right)\xrightarrow{d}\mathrm{N}\left(0,\sigma^2\right),$$

we obtain that

$$\sqrt{n}\left(\widetilde{\sigma}-\mathrm{E}\left|Z_1\right|\right)\xrightarrow{d}\mathrm{N}\left(0,\sigma^2\right).$$

3.2 Global LAD Estimator

Having obtained the asymptotic results of the local LAD estimator, a natural question is: what about the global LAD estimator, say $\widehat{\theta}_n$? The problem is that, $\widehat{\theta}_n$ might not even be consistent in the possibly noninvertible MA(1) setup. Actually, by the ergodic theorem,

$$l_n\left(\theta,z_{aug}\right)/n\to\mathfrak{L}(\theta)\equiv\mathrm{E}\left|z_1\left(\theta\right)\right|\left(|\theta|\,1_{\{|\theta|>1\}}+1_{\{|\theta|\le1\}}\right)$$

almost surely. If furthermore the convergence is uniform on compact subsets of $\Theta\equiv\{\theta\in\mathbb{R}:|\theta|\neq1\}$, then $\widehat{\theta}_n$, the global minimizer of $l_n\left(\theta,z_{aug}\right)/n$, is supposed to converge to a global minimizer of $\mathfrak{L}(\theta)$ almost surely. However, the true parameter value θ_0 may not

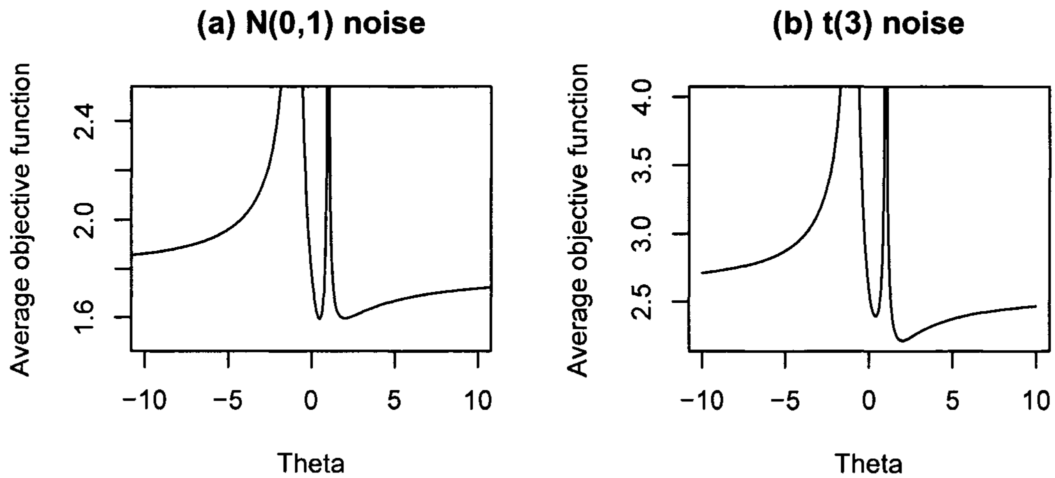

Figure 3.1: $l_n(\theta)/n$ versus θ for MA(1) process $X_t = Z_t + 2Z_{t-1}$: (a) $Z_t \sim$ IID N(0, 1) (b) $Z_t \sim$ IID $t(3)$

be the unique global minimizer of $\mathfrak{L}(\theta)$, and therefore $\widehat{\theta}_n$ may not be consistent. To see this, let us look at an example. Fifty thousand x_t values were generated from the MA(1) model $X_t = Z_t + 2Z_{t-1}, Z_t \sim$ IID N(0, 1). Note that, $l_n(\theta, z_{aug})/n$ approximates $\mathfrak{L}(\theta)$ when n is large enough. So, instead of plotting $\mathfrak{L}(\theta)$ versus θ, we plot $l_n(\theta, z_{aug})/n$ versus θ. Recall that we can set z_{aug} to zero to compute residuals $z_t(\theta)$ and evaluate $l_n(\theta, z_{aug})$. For simplicity we write $l_n(\theta, z_{aug})$ as $l_n(\theta)$, and call $l_n(\theta)/n$ the average objective function. From Figure 3.1 (a) we see that, besides the true model parameter 2, its reciprocal 0.5 is also a global minimizer. In this situation, we have no idea where the global LAD estimator goes; it might go to 2 or 0.5. That is, the global LAD estimator $\widehat{\theta}_n$ may not be consistent.

However, if the underlying noise $\{Z_t\}$ has heavier tails than Gaussian, in the sense that

$$\mathrm{E}\left|\sum_{j=-\infty}^{\infty} c_j Z_{t-j}\right| > \mathrm{E}\left|Z_1\right| \tag{3.10}$$

holds for any sequence $\{c_j\}$ such that $\sum_{j=-\infty}^{\infty}|c_j| < \infty$, $\sum_{j=-\infty}^{\infty} c_j^2 = 1$, and $\{c_j\}$ has at least two non-zero elements. Then the global LAD estimator $\widehat{\theta}_n$ is consistent provided $\mathrm{E}|Z_1| < \infty$ (Huang and Pawitan, 2000). Note that, if the noise is Gaussian, then the two sides in (3.10) are equal. Huang and Pawitan (2000) stated that condition (3.10) is satisfied by many commonly used distributions, such as Student's t, Laplacian, contaminated Gaussian, and other standard heavy-tailed distributions. In these cases, θ_0 is actually the unique global minimizer of $\mathfrak{L}(\theta)$. For illustration, fifty thousand x_t values were generated from MA(1) model $X_t = Z_t + 2Z_{t-1}$, where the underlying noise $\{Z_t\}$ has Student's t distribution with 3 degrees of freedom. We compute $z_t(\theta)$ and then $l_n(\theta)$, and plot $l_n(\theta)/n$ versus θ. From Figure 3.1 (b) we see that, the true parameter 2 is the unique global minimizer; and its reciprocal 0.5 is only a local minimizer. In this case, the global LAD estimator $\widehat{\theta}_n$ is consistent.

In this section, we assume that $\{Z_t\}$ has heavier tails than Gaussian. We also assume that the density function $f_Z(\cdot)$ of $\{Z_t\}$ and the absolute value of its derivative $f_Z'(\cdot)$ are upper-bounded. Then, we can show the strong consistency of $\widehat{\theta}_n$, and its asymptotic normality as well, namely,

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{d} \frac{\sqrt{1-\theta_0^2}}{2f(0)} N_1 \tag{3.11}$$

when $|\theta_0| < 1$, and

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{d} \frac{|\theta_0|\sqrt{\theta_0^2-1}}{2f(0)} N_1, \tag{3.12}$$

when $|\theta_0| > 1$, where $N_1 \sim \mathrm{N}(0,1)$.

3.2.1 Strong Consistency of $\widehat{\theta}_n$

Notice that minimizing $l_n(\theta)/n$ is equivalent to minimizing

$$l_n^{\dagger}(\theta) \equiv \begin{cases} \frac{1}{n}\sum_{t=0}^{n-1} |z_t(\theta)|, & \text{if } |\theta| \le 1, \\ \frac{1}{n}\sum_{t=0}^{n-1} |\theta z_t(\theta)|, & \text{if } |\theta| > 1, \end{cases} \tag{3.13}$$

and $l_n^{\dagger}(\theta) \to \mathfrak{L}(\theta)$ almost surely by the ergodic theorem. Moreover, it can be shown that the convergence is uniform on compact subsets of $\Theta \equiv \{\theta \in \mathbb{R} : |\theta| \ne 1\}$. Without loss of generality, we consider the compact subset $\Theta_1 = \{\theta : 1+\varepsilon \le |\theta| \le M\}$, where $\varepsilon > 0$ is a fixed small number and $M > 0$ is a fixed large number. We first show uniform almost sure equicontinuity of $\left\{l_n^{\dagger}(\theta)\right\}$ on Θ_1. For $\theta_1, \theta_2 \in \Theta_1$, by applying the identity

$$|y| - |z| = (y - z)\,\mathrm{sgn}\,(z) + 2y\left(1_{\{z<0<y\}} - 1_{\{y<0<z\}}\right)$$

for $z \neq 0$, we obtain

$$\begin{aligned}
l_n^{\dagger}(\theta_1) - l_n^{\dagger}(\theta_2) &= \frac{1}{n}\sum_{i=0}^{n-1} |\theta_1 z_i(\theta_1)| - \frac{1}{n}\sum_{i=0}^{n-1} |\theta_2 z_i(\theta_2)| \\
&= \frac{|\theta_1|}{n}\sum_{i=0}^{n-1} \left(|z_i(\theta_1)| - |z_i(\theta_2)|\right) + \frac{1}{n}\sum_{i=0}^{n-1} |z_i(\theta_2)|\left(|\theta_1| - |\theta_2|\right) \\
&= \frac{|\theta_1|}{n}\sum_{i=0}^{n-1} \left(z_i(\theta_1) - z_i(\theta_2)\right)\mathrm{sgn}\left(z_i(\theta_2)\right) \\
&\quad + \frac{2|\theta_1|}{n}\sum_{i=0}^{n-1} z_i(\theta_1)\left(1_{\{z_i(\theta_2)<0<z_i(\theta_1)\}} - 1_{\{z_i(\theta_1)<0<z_i(\theta_2)\}}\right) \\
&\quad + \frac{1}{n}\sum_{i=0}^{n-1} |z_i(\theta_2)|\left(|\theta_1| - |\theta_2|\right) \\
&\equiv A_1 + A_2 + A_3. \qquad (3.14)
\end{aligned}$$

Consider the term A_1 in (3.14) first. By the mean value theorem,

$$|A_1| \leq \frac{|\theta_1|}{n}\sum_{i=0}^{n-1} |z_i(\theta_1) - z_i(\theta_2)| = \frac{|\theta_1 - \theta_2|\,|\theta_1|}{n}\sum_{i=0}^{n-1}\left|\frac{dz_i(\theta^*)}{d\theta}\right|,$$

for some θ^* between θ_1 and θ_2. For any $\theta \in \Theta_1$,

$$z_i(\theta) = -\sum_{t=1}^{\infty} \theta^{-t} X_{i+t} \quad \text{and} \quad \frac{dz_i(\theta)}{d\theta} = \sum_{t=1}^{\infty} t\theta^{-(t+1)} X_{i+t}.$$

Hence,

$$\sup_{\theta\in\Theta_1} |z_i(\theta)| \leq \sum_{t=1}^{\infty} (1+\varepsilon)^{-t} |X_{i+t}|, \qquad (3.15)$$

$$\sup_{\theta\in\Theta_1} \left|\frac{dz_i(\theta)}{d\theta}\right| \leq \sum_{t=1}^{\infty} t(1+\varepsilon)^{-(t+1)} |X_{i+t}|. \qquad (3.16)$$

The coefficient sequences $\{(1+\varepsilon)^{-t}\}$ in (3.15) and $\{t(1+\varepsilon)^{-(t+1)}\}$ in (3.16) decay at a geometric rate, respectively, as $t \to \infty$. So we obtain

$$\begin{aligned} |A_1| &\leq \frac{|\theta_1-\theta_2|\,|\theta_1|}{n}\sum_{i=0}^{n-1}\sum_{t=1}^{\infty} t\,(1+\varepsilon)^{-(t+1)}\,|X_{i+t}| \\ &= |\theta_1-\theta_2|\,O(1) \qquad \text{a.s.} \end{aligned} \tag{3.17}$$

Moving on to the term A_2 in (3.14). Given a fixed $\delta > 0$,

$$\begin{aligned} |A_2| &= \frac{2|\theta_1|}{n}\sum_{i=0}^{n-1} z_i(\theta_1)\left(1_{\{z_i(\theta_2)<0<z_i(\theta_1)\}} - 1_{\{z_i(\theta_1)<0<z_i(\theta_2)\}}\right) \\ &= \frac{2|\theta_1|}{n}\sum_{i=0}^{n-1} z_i(\theta_1)\left(1_{\{z_i(\theta_2)<0<z_i(\theta_1)\leq\delta\}} - 1_{\{-\delta\leq z_i(\theta_1)<0<z_i(\theta_2)\}}\right) \\ &\quad +\frac{2|\theta_1|}{n}\sum_{i=0}^{n-1} z_i(\theta_1)\left(1_{\{z_i(\theta_2)<0,z_i(\theta_1)>\delta\}} - 1_{\{z_i(\theta_1)<-\delta,0<z_i(\theta_2)\}}\right) \\ &\leq \frac{2|\theta_1|}{n}\sum_{i=0}^{n-1} |z_i(\theta_1)|\,1_{\{|z_i(\theta_1)|\leq\delta\}} + \frac{2|\theta_1|}{n}\sum_{i=0}^{n-1} |z_i(\theta_1)|\,1_{\{|z_i(\theta_1)|>\delta\}}1_{\{|z_i(\theta_1)-z_i(\theta_2)|>\delta\}} \\ &\leq 2|\theta_1|\,\delta + \frac{2|\theta_1|}{n}\sum_{i=0}^{n-1}|z_i(\theta_1)|\,|z_i(\theta_1)-z_i(\theta_2)|/\delta \\ &\leq 2|\theta_1|\,\delta + \frac{2|\theta_1-\theta_2|\,|\theta_1|}{\delta n}\sum_{i=0}^{n-1}\left[\sum_{t=1}^{\infty}(1+\varepsilon)^{-t}|X_{i+t}|\sum_{t=1}^{\infty} t\,(1+\varepsilon)^{-(t+1)}\,|X_{i+t}|\right] \\ &= 2|\theta_1|\,\delta + \frac{|\theta_1-\theta_2|}{\delta}O(1) \qquad \text{a.s.} \end{aligned} \tag{3.18}$$

As to the term A_3 in (3.14), we have

$$\begin{aligned} |A_3| &\leq \frac{|\theta_1-\theta_2|}{n}\sum_{i=0}^{n-1}|z_i(\theta_2)| \\ &\leq \frac{|\theta_1-\theta_2|}{n}\sum_{i=0}^{n-1}\sum_{t=1}^{\infty}(1+\varepsilon)^{-t}\,|X_{i+t}| \\ &= |\theta_1-\theta_2|\,O(1) \qquad \text{a.s.} \end{aligned} \tag{3.19}$$

The $O(1)$ terms in (3.17), (3.18), and (3.19) do not depend on θ_1, θ_2, or δ. Therefore, it follows that the sequence $\left\{l_n^{\dagger}(\theta)\right\}$ is uniformly equicontinuous on Θ_1 almost surely.

In addition, it is easy to show that $\left\{l_n^{\dagger}(\theta)\right\}$ is uniformly almost surely bounded on Θ_1. So, by applying the Arzelà-Ascoli theorem, we conclude that the almost sure convergence of $l_n^\dagger(\theta)$ to $\mathfrak{L}(\theta)$ is uniform on Θ_1.

On the other hand, Huang and Pawitan (2000) showed that the true parameter value θ_0 uniquely minimizes $\mathfrak{L}(\theta)$. This, together with the uniform almost sure convergence of $l_n^\dagger(\theta)$ to $\mathfrak{L}(\theta)$ on compact subsets of Θ, establishes the strong consistency of $\widehat{\theta}_n$.

3.2.2 Asymptotic Normality of $\widehat{\theta}_n$

We derive (3.12) for the case when $|\theta_0| > 1$; the derivation for the case when $|\theta_0| < 1$ is similar. Since $\widehat{\theta}_n$ is strongly consistent, when n is large enough, $\widehat{\theta}_n$ globally minimizes $u_n(\theta) \equiv \sum_{i=0}^{n-1} |\theta z_i(\theta)|$ on the reduced parameter space Θ_{NI}. Therefore, we can confine our study to the objective function $u_n(\theta)$ on Θ_{NI}.

To calculate the objective function $u_n(\theta)$, the residuals $z_t(\theta), t = 0, 1, \ldots, n$ are solved backward by $z_{t-1}(\theta) = (z_t(\theta) - X_t)/\theta$ for $t = n, n-1, \ldots, 1$ with $z_n(\theta) = z_{aug}$. By iteration, we have

$$-\theta z_{n-1-i}(\theta) = \sum_{j=0}^{i} \theta^{-(i-j)} X_{n-j} - \theta^{-i} z_{aug}, \qquad i = 0, 1, \ldots, n-1.$$

Since $X_t = Z_t - \theta_0 Z_{t-1}$, we have for $i = 0, 1, \ldots, n-1$,

$$\begin{aligned} -\theta z_{n-1-i}(\theta) &= \sum_{j=0}^{i} \theta^{-(i-j)} (Z_{n-j} - \theta_0 Z_{n-j-1}) - \theta^{-i} z_{aug} \\ &= -\theta_0 Z_{n-i-1} + (1 - \theta_0/\theta) \sum_{j=1}^{i} \theta^{-(i-j)} Z_{n-j} + \theta^{-i}(Z_n - z_{aug}). \end{aligned}$$

Moreover, by setting z_{aug} to be Z_n, we obtain

$$\theta z_i(\theta) = \theta_0 Z_i - (\theta - \theta_0) \sum_{j=1}^{n-1-i} \theta^{-j} Z_{j+i} \equiv r_{ni}(\theta) \tag{3.20}$$

for $i = 0, \ldots, n-1$, and $u_n(\theta) \equiv \sum_{i=0}^{n-1} |r_{ni}(\theta)|$.

The derivation of (3.12) is broken into several steps. First of all, we approximate $|t|$ by a uniformly convergent sequence of twice differentiable functions $h_n(t)$, which is specified by

$$h_n(t) = \begin{cases} \frac{1+(n^p t)^2}{2n^p}, & \text{if } |t| \le n^{-p}, \\ |t|, & \text{if } |t| > n^{-p}, \end{cases} \tag{3.21}$$

for some $p \in (1/2, 1)$. This idea was also used by Bloomfield and Steiger (1983). We define

$$v_n(\theta) \equiv \sum_{i=0}^{n-1} h_n(r_{ni}(\theta)), \tag{3.22}$$

and show that $v_n(\theta) - u_n(\theta)$ converges to zero in probability. Notice that

$$\begin{aligned} v_n(\theta) - u_n(\theta) &= \sum_{i=0}^{n-1} \left[\frac{1 + (n^p r_{ni}(\theta))^2}{2n^p} - |r_{ni}(\theta)| \right] 1_{\{|r_{ni}(\theta)| \le n^{-p}\}} \\ &= \sum_{i=0}^{n-1} \frac{(1 - n^p |r_{ni}(\theta)|)^2}{2n^p} 1_{\{|r_{ni}(\theta)| \le n^{-p}\}} \\ &\ge 0. \end{aligned}$$

For any given θ, the density of $r_{ni}(\theta)$ is upper bounded uniformly in n and i. To be specific, let $y_{ni}(\theta) \equiv \sum_{j=1}^{n-1-i} \theta^{-j} Z_{j+i}$ for $i = 0, \ldots, n-1$, then $r_{ni}(\theta) = \theta_0 Z_i - (\theta - \theta_0) y_{ni}(\theta)$. But $y_{ni}(\theta)$ is independent of Z_i. So, by the convolution formula, the density of $r_{ni}(\theta)$ is given by

$$f_{r_{ni}(\theta)}(y) = \int f_{\theta_0 Z_i}(y - x) \cdot f_{[-(\theta-\theta_0) y_{ni}(\theta)]}(x)\, dx \le \max_z f_{\theta_0 Z_1}(z).$$

Hence,

$$\begin{aligned} \mathrm{E}\,|v_n(\theta) - u_n(\theta)| &= \sum_{i=0}^{n-1} \mathrm{E} \left\{ \frac{(1 - n^p |r_{ni}(\theta)|)^2}{2n^p} 1_{\{|r_{ni}(\theta)| \le n^{-p}\}} \right\} \\ &\le \sum_{i=0}^{n-1} \mathrm{E} \left\{ \frac{1}{2n^p} 1_{\{|r_{ni}(\theta)| \le n^{-p}\}} \right\} \\ &= \frac{1}{2n^p} \sum_{i=0}^{n-1} \left\{ F_{r_{ni}(\theta)}(n^{-p}) - F_{r_{ni}(\theta)}(-n^{-p}) \right\} \\ &= \frac{1}{n^{2p}} \sum_{i=0}^{n-1} \frac{F_{r_{ni}(\theta)}(n^{-p}) - F_{r_{ni}(\theta)}(-n^{-p})}{2n^{-p}} \longrightarrow 0 \end{aligned}$$

for $p > 1/2$, where $F_{r_{ni}(\theta)}(\cdot)$ is the cdf of $r_{ni}(\theta)$. It follows that $v_n(\theta) - u_n(\theta)$ converges to zero in probability.

Next, let $\overline{\theta}_n$ denote the global minimizer of $v_n(\theta)$. We show that $\overline{\theta}_n$ is strongly consistent and hence $\overline{\theta}_n - \widehat{\theta}_n \to 0$ almost surely. Moreover, we show that $\sqrt{n}\left(\overline{\theta}_n - \widehat{\theta}_n\right) \to 0$ in probability.

Notice that $|v_n(\theta) - u_n(\theta)|/n \to 0$ almost surely. Then, the almost sure convergence of $v_n(\theta)/n$ to $\mathfrak{L}(\theta)$ follows from that of $u_n(\theta)/n$ to $\mathfrak{L}(\theta)$. Additionally, the uniform equicontinuity and boundedness of $v_n(\theta)/n$ on compact subsets of Θ_{NI} follow from those of $u_n(\theta)/n$. So, the almost sure convergence of $v_n(\theta)/n$ to $\mathfrak{L}(\theta)$ is actually uniform on compact subsets of Θ_{NI}. Moreover, θ_0 is the unique minimizer of $\mathfrak{L}(\theta)$ on Θ_{NI}. Therefore, the strong consistency of $\overline{\theta}_n$ follows.

As to showing $\sqrt{n}\left(\overline{\theta}_n - \widehat{\theta}_n\right) \to 0$ in probability, we begin with the second derivative of $v_n(\theta)$ at θ_0. The first two derivatives of $h_n(t)$ are

$$\begin{aligned} h_n'(t) &= (n^p t)\, 1_{\{|t|\le n^{-p}\}} + \operatorname{sgn}(t)\, 1_{\{|t|>n^{-p}\}}, \\ h_n''(t) &= n^p 1_{\{|t|\le n^{-p}\}}. \end{aligned}$$

Taking the first derivative of $v_n(\theta)$ with respect to θ yields

$$\frac{dv_n(\theta)}{d\theta} = \sum_{i=0}^{n-1} h_n'(r_{ni}(\theta))\frac{dr_{ni}(\theta)}{d\theta}, \tag{3.23}$$

where

$$h_n'(r_{ni}(\theta)) = n^p r_{ni}(\theta)\, 1_{\{|r_{ni}(\theta)|\le n^{-p}\}} + \operatorname{sgn}(r_{ni}(\theta))\, 1_{\{|r_{ni}(\theta)|>n^{-p}\}}, \tag{3.24}$$

$$\frac{dr_{ni}(\theta)}{d\theta} = -\sum_{j=1}^{n-1-i}\theta^{-j}Z_{j+i} + (\theta-\theta_0)\sum_{j=1}^{n-1-i} j\theta^{-(j+1)}Z_{j+i}. \tag{3.25}$$

Taking the second derivative of $v_n(\theta)$ with respect to θ yields

$$\frac{d^2 v_n(\theta)}{d\theta^2} = \sum_{i=0}^{n-1} \left\{ h_n''(r_{ni}(\theta)) \left[\frac{dr_{ni}(\theta)}{d\theta} \right]^2 + h_n'(r_{ni}(\theta)) \frac{d^2 r_{ni}(\theta)}{d\theta^2} \right\}, \tag{3.26}$$

where

$$\begin{aligned} h_n''(r_{ni}(\theta)) &= n^p 1_{\{|r_{ni}(\theta)| \le n^{-p}\}}, \\ \frac{d^2 r_{ni}(\theta)}{d\theta^2} &= 2 \sum_{j=1}^{n-1-i} j\theta^{-(j+1)} Z_{j+i} - (\theta - \theta_0) \sum_{j=1}^{n-1-i} j(j+1)\theta^{-(j+2)} Z_{j+i}. \end{aligned}$$

Plugging θ_0 into $\frac{d^2 v_n(\theta)}{d\theta^2}$ and multiplying it by $1/n$, we obtain

$$\begin{aligned} n^{-1} \frac{d^2 v_n(\theta_0)}{d\theta^2} &= \frac{1}{n} \sum_{i=0}^{n-1} n^p 1_{\{|\theta_0 Z_i| \le n^{-p}\}} \left(\sum_{j=1}^{n-1-i} \theta_0^{-j} Z_{j+i} \right)^2 \\ &\quad + \frac{2}{n} \sum_{i=0}^{n-1} n^p (\theta_0 Z_i) 1_{\{|\theta_0 Z_i| \le n^{-p}\}} \left(\sum_{j=1}^{n-1-i} j\theta_0^{-(j+1)} Z_{j+i} \right) \\ &\quad + \frac{2}{n} \sum_{i=0}^{n-1} \operatorname{sgn}(\theta_0 Z_i) 1_{\{|\theta_0 Z_i| > n^{-p}\}} \left(\sum_{j=1}^{n-1-i} j\theta_0^{-(j+1)} Z_{j+i} \right) \\ &\equiv T_1 + T_2 + T_3. \end{aligned} \tag{3.27}$$

We show

$$n^{-1} \frac{d^2 v_n(\theta_0)}{d\theta^2} \xrightarrow{P} \frac{2f(0)\sigma}{|\theta_0|(\theta_0^2 - 1)} \tag{3.28}$$

by dealing with the terms T_1, T_2 and T_3 in (3.27), respectively. The term $T_2 \to 0$ in probability since

$$\begin{aligned} \mathrm{E}|T_2| &\le \frac{2}{n} \sum_{i=0}^{n-1} \mathrm{E}\left(n^p |\theta_0 Z_i| 1_{\{|\theta_0 Z_i| \le n^{-p}\}} \right) \cdot \mathrm{E} \left| \sum_{j=1}^{n-1-i} j\theta_0^{-(j+1)} Z_{j+i} \right| \\ &= \left\{ \frac{2}{n} \sum_{i=0}^{n-1} \mathrm{E} \left| \sum_{j=1}^{n-1-i} j\theta_0^{-(j+1)} Z_{j+i} \right| \right\} \mathrm{E}\left(n^p |\theta_0 Z_1| 1_{\{|\theta_0 Z_1| \le n^{-p}\}} \right) \\ &\le \text{Const} \cdot \mathrm{E}\left(1_{\{|\theta_0 Z_1| \le n^{-p}\}} \right) \\ &\to 0. \end{aligned}$$

The term $T_3 \to 0$ in probability too. To see this, we consider

$$T_3' \equiv \frac{2}{n}\sum_{i=0}^{n-1} \operatorname{sgn}(\theta_0 Z_i)\left(\sum\nolimits_{j=1}^{n-1-i} j\theta_0^{-(j+1)} Z_{j+i}\right),$$

and it is easy to show that $T_3' \to 0$ in probability. On the other hand, the result that $T_3 - T_3' \to 0$ in probability is implied by

$$\begin{aligned}
\mathrm{E}\left|T_3 - T_3'\right| &= \frac{2}{n}\mathrm{E}\left|\sum_{i=0}^{n-1} \operatorname{sgn}(\theta_0 Z_i)\, 1_{\{|\theta_0 Z_i| \le n^{-p}\}}\left(\sum\nolimits_{j=1}^{n-1-i} j\theta_0^{-(j+1)} Z_{j+i}\right)\right| \\
&\le \frac{2}{n}\sum_{i=0}^{n-1} \mathrm{E}\left(1_{\{|\theta_0 Z_i| \le n^{-p}\}}\right)\cdot \mathrm{E}\left|\sum\nolimits_{j=1}^{n-1-i} j\theta_0^{-(j+1)} Z_{j+i}\right| \\
&= \left\{\frac{2}{n}\sum_{i=0}^{n-1} \mathrm{E}\left|\sum\nolimits_{j=1}^{n-1-i} j\theta_0^{-(j+1)} Z_{j+i}\right|\right\}\mathrm{E}\left(1_{\{|\theta_0 Z_1| \le n^{-p}\}}\right) \\
&\le \mathrm{Const}\cdot \mathrm{E}\left(1_{\{|\theta_0 Z_1| \le n^{-p}\}}\right) \\
&\to 0.
\end{aligned}$$

To deal with the term T_1, we define

$$\tau_{ni} \equiv n^p 1_{\{|\theta_0 Z_i| \le n^{-p}\}}\left(\sum\nolimits_{j=1}^{n-1-i} \theta_0^{-j} Z_{j+i}\right)^2$$

such that $T_1 = \frac{1}{n}\sum_{i=0}^{n-1} \tau_{ni}$. Then,

$$\begin{aligned}
\mathrm{E}\left(\tau_{ni}|\mathfrak{F}^{i+1}\right) &= \left(\sum\nolimits_{j=1}^{n-1-i} \theta_0^{-j} Z_{j+i}\right)^2 \mathrm{E}\left(n^p 1_{\{|\theta_0 Z_1| \le n^{-p}\}}\right) \\
&= \left(\sum\nolimits_{j=1}^{n-1-i} \theta_0^{-j} Z_{j+i}\right)^2 \frac{F_{\theta_0 Z_1}(n^{-p}) - F_{\theta_0 Z_1}(-n^{-p})}{n^{-p}},
\end{aligned}$$

where $\mathfrak{F}^{i+1} = \sigma(Z_t; t \ge i+1)$, and it can be shown that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=0}^{n-1} \mathrm{E}\left(\tau_{ni}|\mathfrak{F}^{i+1}\right) \\
&= \left\{\frac{1}{n}\sum_{i=0}^{n-1}\left(\sum\nolimits_{j=1}^{n-1-i} \theta_0^{-j} Z_{j+i}\right)^2\right\}\frac{F_{\theta_0 Z_1}(n^{-p}) - F_{\theta_0 Z_1}(-n^{-p})}{n^{-p}} \xrightarrow{P} \frac{2f(0)\sigma}{|\theta_0|\left(\theta_0^2 - 1\right)}.
\end{aligned}$$

On the other hand, it can be shown that

$$T_1 - \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{E}\left(\tau_{ni}|\mathfrak{F}^{i+1}\right) \xrightarrow{P} 0. \tag{3.29}$$

To be specific,

$$\mathrm{E}\left(T_1 - \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{E}\left(\tau_{ni}|\mathfrak{F}^{i+1}\right)\right) = \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{E}\left(\tau_{ni} - \mathrm{E}\left(\tau_{ni}|\mathfrak{F}^{i+1}\right)\right) = 0, \tag{3.30}$$

and

$$\begin{aligned}
\mathrm{Var}\left(T_1 - \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{E}\left(\tau_{ni}|\mathfrak{F}^{i+1}\right)\right) &= \frac{1}{n^2}\sum_{i=0}^{n-1} \mathrm{Var}\left(\tau_{ni} - \mathrm{E}\left(\tau_{ni}|\mathfrak{F}^{i+1}\right)\right) \\
&= \frac{1}{n^2}\sum_{i=0}^{n-1} \mathrm{E}\left\{\mathrm{E}\left[\left(\tau_{ni} - \mathrm{E}\left(\tau_{ni}|\mathfrak{F}^{i+1}\right)\right)^2 |\mathfrak{F}^{i+1}\right]\right\} \\
&= \frac{1}{n^2}\sum_{i=0}^{n-1} \mathrm{E}\left\{\mathrm{E}\left(\tau_{ni}^2|\mathfrak{F}^{i+1}\right) - \left[\mathrm{E}\left(\tau_{ni}|\mathfrak{F}^{i+1}\right)\right]^2\right\} \\
&\leq \frac{1}{n^2}\sum_{i=0}^{n-1} \mathrm{E}\left[\mathrm{E}\left(\tau_{ni}^2|\mathfrak{F}^{i+1}\right)\right].
\end{aligned}$$

But

$$\begin{aligned}
&\sum_{i=0}^{n-1} \mathrm{E}\left[\mathrm{E}\left(\tau_{ni}^2|\mathfrak{F}^{i+1}\right)\right] \\
&\quad = \sum_{i=0}^{n-1} \mathrm{E}\left[n^{2p} 1_{\{|\theta_0 Z_i| \leq n^{-p}\}} \left(\textstyle\sum_{j=1}^{n-1-i} \theta_0^{-j} Z_{j+i}\right)^4\right] \\
&\quad = \left[\sum_{i=0}^{n-1} \mathrm{E}\left(\textstyle\sum_{j=1}^{n-1-i} \theta_0^{-j} Z_{j+i}\right)^4\right] \cdot \mathrm{E}\left[n^{2p} 1_{\{|\theta_0 Z_1| \leq n^{-p}\}}\right] \\
&\quad = 2n^p \left\{\sum_{i=0}^{n-1} \left[\mathrm{E}\left(Z_1^4\right) \textstyle\sum_{j=1}^{n-1-i} \theta_0^{-4j} + 6\sigma^4 \left(\textstyle\sum_{j=1}^{n-1-i} \theta_0^{-2j}\right)^2\right]\right\} \\
&\qquad \times \tfrac{F_{\theta_0 Z_1}(n^{-p}) - F_{\theta_0 Z_1}(-n^{-p})}{2n^{-p}} \\
&\quad \leq \mathrm{Const} \cdot n^{p+1} \cdot \tfrac{F_{\theta_0 Z_1}(n^{-p}) - F_{\theta_0 Z_1}(-n^{-p})}{2n^{-p}}.
\end{aligned}$$

It follows that

$$\mathrm{Var}\left(T_1 - \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{E}\left(\tau_{ni}|\mathfrak{F}^{i+1}\right)\right) \leq \mathrm{Const} \cdot n^{p-1} \cdot \frac{F_{\theta_0 Z_1}(n^{-p}) - F_{\theta_0 Z_1}(-n^{-p})}{2n^{-p}} \to 0 \tag{3.31}$$

for $p < 1$. Therefore, (3.29) follows from (3.30) and (3.31), and hence

$$T_1 \xrightarrow{P} \frac{2f(0)\sigma}{|\theta_0|(\theta_0^2-1)}.$$

The convergence (3.28) follows by combining the results for T_1, T_2, and T_3. Moreover, for any random sequence $\{\theta_n\}$ such that $\theta_n \xrightarrow{\text{a.s.}} \theta_0$,

$$n^{-1}\frac{d^2 v_n(\theta_n)}{d\theta^2} \xrightarrow{P} \frac{2f(0)\sigma}{|\theta_0|(\theta_0^2-1)}, \tag{3.32}$$

and

$$\frac{1}{\sqrt{n}}\left(\frac{dv_n(\theta_n)}{d\theta} - \frac{du_n(\theta_n)}{d\theta}\right) \xrightarrow{P} 0 \tag{3.33}$$

(see Section 3.5 Appendix). It follows from (3.33) that

$$\frac{1}{\sqrt{n}}\frac{dv_n\left(\widehat{\theta}_n\right)}{d\theta} \xrightarrow{P} 0. \tag{3.34}$$

Expanding the first derivative of $v_n(\theta)$ with respect to θ in a first degree Taylor series about $\overline{\theta}_n$, we obtain

$$\begin{aligned}\frac{dv_n\left(\widehat{\theta}_n\right)}{d\theta} &= \frac{dv_n\left(\overline{\theta}_n\right)}{d\theta} + \frac{d^2 v_n\left(\theta_n^{\dagger}\right)}{d\theta^2}\left(\widehat{\theta}_n - \overline{\theta}_n\right) \\ &= \frac{d^2 v_n\left(\theta_n^{\dagger}\right)}{d\theta^2}\left(\widehat{\theta}_n - \overline{\theta}_n\right),\end{aligned}$$

where the random variable $\theta_n^{\dagger}$ lies between $\widehat{\theta}_n$ and $\overline{\theta}_n$. Then

$$\frac{1}{\sqrt{n}}\frac{dv_n\left(\widehat{\theta}_n\right)}{d\theta} = \left[\frac{1}{n}\frac{d^2 v_n\left(\theta_n^{\dagger}\right)}{d\theta^2}\right]\left[\sqrt{n}\left(\widehat{\theta}_n - \overline{\theta}_n\right)\right]. \tag{3.35}$$

Notice that $\theta_n^{\dagger} \xrightarrow{\text{a.s.}} \theta_0$. It follows from (3.35), (3.32), and (3.34) that

$$\sqrt{n}\left(\widehat{\theta}_n - \overline{\theta}_n\right) \xrightarrow{P} 0.$$

Finally, we show

$$\sqrt{n}\left(\overline{\theta}_n-\theta_0\right)\xrightarrow{d}\frac{|\theta_0|\sqrt{\theta_0^2-1}}{2f(0)}N_1 \tag{3.36}$$

and hence (3.12) follows. We expand the first derivative of $v_n(\theta)$ with respect to θ in a first degree Taylor series about $\overline{\theta}_n$:

$$\frac{dv_n(\theta_0)}{d\theta}=\frac{dv_n\left(\overline{\theta}_n\right)}{d\theta}+\frac{d^2v_n\left(\theta_n^{\ddagger}\right)}{d\theta^2}\left(\theta_0-\overline{\theta}_n\right), \tag{3.37}$$

where $\theta_n^{\ddagger}$ lies between θ_0 and $\overline{\theta}_n$. The first term on the right-hand side of (3.37) vanishes because $\overline{\theta}_n$ minimizes $v_n(\theta)$, and thus

$$\sqrt{n}\left(\theta_0-\overline{\theta}_n\right)=\left[\frac{1}{\sqrt{n}}\frac{dv_n(\theta_0)}{d\theta}\right]\left[\frac{1}{n}\frac{d^2v_n\left(\theta_n^{\ddagger}\right)}{d\theta^2}\right]^{-1}. \tag{3.38}$$

However,

$$\frac{1}{n}\frac{d^2v_n\left(\theta_n^{\ddagger}\right)}{d\theta^2}\xrightarrow{P}\frac{2f(0)\sigma}{|\theta_0|\left(\theta_0^2-1\right)} \tag{3.39}$$

since $\theta_n^{\ddagger}\xrightarrow{\text{a.s.}}\theta_0$. So, it suffices to show

$$\frac{1}{\sqrt{n}}\frac{dv_n(\theta_0)}{d\theta}\xrightarrow{d}\mathrm{N}\left(0,\frac{\sigma^2}{\theta_0^2-1}\right). \tag{3.40}$$

From (3.23), (3.24), and (3.25), we have

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{dv_n(\theta_0)}{d\theta} &= -\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}n^p(\theta_0Z_i)\,1_{\{|\theta_0Z_i|\le n^{-p}\}}\left(\sum\nolimits_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i}\right)\\ &\quad -\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\operatorname{sgn}(\theta_0Z_i)\,1_{\{|\theta_0Z_i|>n^{-p}\}}\left(\sum\nolimits_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i}\right)\\ &\equiv S_1+S_2.\end{aligned}$$

Define

$$\eta_{ni}\equiv -n^p(\theta_0Z_i)\,1_{\{|\theta_0Z_i|\le n^{-p}\}}\left(\sum\nolimits_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i}\right)$$

such that $S_1=\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\eta_{ni}$. Then

$$\mathrm{E}\left(\eta_{ni}|\mathcal{F}^{i+1}\right) = -\mathrm{E}\left[n^p\left(\theta_0 Z_1\right) 1_{\{|\theta_0 Z_1|\le n^{-p}\}}\right] \cdot \left(\sum_{j=1}^{n-1-i} \theta_0^{-j} Z_{j+i}\right),$$

and

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\mathrm{E}\left(\eta_{ni}|\mathcal{F}^{i+1}\right) &= -\frac{\mathrm{E}\left[n^p\left(\theta_0 Z_1\right) 1_{\{|\theta_0 Z_1|\le n^{-p}\}}\right]}{\sqrt{n}}\sum_{i=0}^{n-1}\left(\sum_{j=1}^{n-1-i}\theta_0^{-j} Z_{j+i}\right) \\
&= -\frac{\mathrm{E}\left[n^p\left(\theta_0 Z_1\right) 1_{\{|\theta_0 Z_1|\le n^{-p}\}}\right]}{\sqrt{n}}\sum_{j=1}^{n-1}\left(\sum_{i=1}^{j}\theta_0^{-i}\right) Z_j \\
&= -\frac{\mathrm{E}\left[n^p\left(\theta_0 Z_1\right) 1_{\{|\theta_0 Z_1|\le n^{-p}\}}\right]}{\sqrt{n}}\sum_{j=1}^{n-1}\frac{1-\theta_0^{-j}}{\theta_0 - 1} Z_j.
\end{aligned}$$

The summation $\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\mathrm{E}\left(\eta_{ni}|\mathcal{F}^{i+1}\right)$ goes to zero in probability because

$$\begin{aligned}
\mathrm{E}\left[\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\mathrm{E}\left(\eta_{ni}|\mathcal{F}^{i+1}\right)\right] &= 0, \\
\mathrm{Var}\left[\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\mathrm{E}\left(\eta_{ni}|\mathcal{F}^{i+1}\right)\right] &= \frac{\left\{\mathrm{E}\left[n^p\left(\theta_0 Z_1\right) 1_{\{|\theta_0 Z_1|\le n^{-p}\}}\right]\right\}^2}{n\left(\theta_0 - 1\right)^2}\sum_{j=1}^{n-1}\left(1-\theta_0^{-j}\right)\sigma^2 \\
&\le \mathrm{Const}\cdot\left\{\mathrm{E}\left[1_{\{|\theta_0 Z_1|\le n^{-p}\}}\right]\right\}^2 \\
&\to 0.
\end{aligned}$$

On the other hand,

$$\mathrm{E}\left[S_1 - \frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\mathrm{E}\left(\eta_{ni}|\mathcal{F}^{i+1}\right)\right] = \frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\mathrm{E}\left[\eta_{ni} - \mathrm{E}\left(\eta_{ni}|\mathcal{F}^{i+1}\right)\right] = 0,$$

and

$$\begin{aligned}
\mathrm{Var}\left[S_1 - \frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\mathrm{E}\left(\eta_{ni}|\mathcal{F}^{i+1}\right)\right] &= \frac{1}{n}\sum_{i=0}^{n-1}\mathrm{Var}\left[\eta_{ni} - \mathrm{E}\left(\eta_{ni}|\mathcal{F}^{i+1}\right)\right] \\
&= \frac{1}{n}\sum_{i=0}^{n-1}\mathrm{E}\left\{\mathrm{E}\left[\left(\eta_{ni} - \mathrm{E}\left(\eta_{ni}|\mathcal{F}^{i+1}\right)\right)^2|\mathcal{F}^{i+1}\right]\right\} \\
&= \frac{1}{n}\sum_{i=0}^{n-1}\mathrm{E}\left\{\mathrm{E}\left(\eta_{ni}^2|\mathcal{F}^{i+1}\right) - \left[\mathrm{E}\left(\eta_{ni}|\mathcal{F}^{i+1}\right)\right]^2\right\}
\end{aligned}$$

$$\leq \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{E}\left[\mathrm{E}\left(\eta_{ni}^2|\mathfrak{F}^{i+1}\right)\right].$$

But

$$\begin{aligned}
\mathrm{E}\left[\mathrm{E}\left(\eta_{ni}^2|\mathfrak{F}^{i+1}\right)\right] &= \mathrm{E}\left[n^{2p}\left(\theta_0 Z_i\right)^2 1_{\{|\theta_0 Z_i|\leq n^{-p}\}}\left(\sum\nolimits_{j=1}^{n-1-i}\theta_0^{-j} Z_{j+i}\right)^2\right] \\
&= \mathrm{E}\left(\sum_{j=1}^{n-1-i}\theta_0^{-j} Z_{j+i}\right)^2 \cdot \mathrm{E}\left[n^{2p}\left(\theta_0 Z_1\right)^2 1_{\{|\theta_0 Z_1|\leq n^{-p}\}}\right] \\
&\leq \mathrm{Const}\cdot \mathrm{E}\left[1_{\{|\theta_0 Z_1|\leq n^{-p}\}}\right].
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathrm{Var}\left[S_1 - \frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\mathrm{E}\left(\eta_{ni}|\mathfrak{F}^{i+1}\right)\right] &\leq \frac{1}{n}\sum_{i=0}^{n-1}\mathrm{Const}\cdot \mathrm{E}\left[1_{\{|\theta_0 Z_1|\leq n^{-p}\}}\right] \\
&= \mathrm{Const}\cdot \mathrm{E}\left[1_{\{|\theta_0 Z_1|\leq n^{-p}\}}\right] \\
&\to 0.
\end{aligned}$$

It follows that

$$S_1 - \frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\mathrm{E}\left(\eta_{ni}|\mathfrak{F}^{i+1}\right) \xrightarrow{P} 0,$$

and therefore $S_1 \xrightarrow{P} 0$. Moving on to the term S_2. On one hand, it is easy to show that $S_2' \xrightarrow{d} \mathrm{N}\left(0, \frac{\sigma^2}{\theta_0^2-1}\right)$, where

$$S_2' \equiv \frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\mathrm{sgn}\left(\theta_0 Z_i\right)\left(-\sum\nolimits_{j=1}^{n-1-i}\theta_0^{-j} Z_{j+i}\right);$$

on the other hand, following the same lines for the term S_1, it can be shown that

$$S_2 - S_2' = \frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\mathrm{sgn}\left(\theta_0 Z_i\right) 1_{\{|\theta_0 Z_i|\leq n^{-p}\}}\left(-\sum\nolimits_{j=1}^{n-1-i}\theta_0^{-j} Z_{j+i}\right) \xrightarrow{P} 0.$$

Therefore,

$$S_2 \xrightarrow{d} \mathrm{N}\left(0, \frac{\sigma^2}{\theta_0^2 - 1}\right).$$

The convergence (3.40) follows from the results for S_1 and S_2.

3.3 Linearized LAD Estimator

Note that, the objective function (3.4) is not convex in θ, which complicates the search for a minimizer in the sense of numerical computations. To simplify computations, we propose a local linearization procedure to obtain a linearized objective function that is convex in θ. Then, the linearized LAD estimator $\widehat{\theta}_{\text{LLAD}}$, which is defined as the minimizer of the linearized objective function, has the same asymptotic normal law as $\widehat{\theta}_{\text{LAD}}$.

Suppose $\widehat{\theta}_0$ is a set of estimators of θ such that $\widehat{\theta}_0 = \theta_0 + O_p\left(n^{-1/2}\right)$. We linearize the original objective function $l_n\left(\theta, z_{aug}\right)$ in the neighborhood of $\widehat{\theta}_0$ using a Taylor expansion. We discuss cases when $|\theta_0| < 1$ and $|\theta_0| > 1$ respectively. Suppose the true model is invertible, then we confine our study to the reduced parameter space Θ_I. In this case, we actually linearize $z_t\left(\theta\right)$ around $\widehat{\theta}_0$, which yields the objective function

$$\sum_{i=0}^{n-1} \left| z_i\left(\widehat{\theta}_0\right) - D_i\left(\widehat{\theta}_0\right)\left(\theta - \widehat{\theta}_0\right)\right|, \tag{3.41}$$

where $D_i\left(\theta\right) = -d\left(z_i\left(\theta\right)\right)/d\theta$ for $i = 0, \ldots, n-1$. The asymptotic normality of $\widehat{\theta}_{\text{LLAD}}$ is then a special case of Proposition 2 of Davis and Dunsmuir (1997). That is,

$$\sqrt{n}\left(\widehat{\theta}_{\text{LLAD}} - \theta_0\right) \rightarrow \frac{\sqrt{1-\theta_0^2}}{2f\left(0\right)} N_1$$

in distribution, where $\widehat{\theta}_{\text{LLAD}}$ is the minimizer of (3.41).

On the other hand, when the true model is noninvertible, we confine our study to the reduced parameter space Θ_{NI}. Then, we actually linearize $\theta z_t\left(\theta\right)$ around $\widehat{\theta}_0$, which yields the objective function

$$\sum_{i=0}^{n-1} \left| \widehat{\theta}_0 z_i\left(\widehat{\theta}_0\right) - D_i\left(\widehat{\theta}_0\right)\left(\theta - \widehat{\theta}_0\right)\right|, \tag{3.42}$$

where $D_i\left(\theta\right) = -d\left(\theta z_i\left(\theta\right)\right)/d\theta$ for $i = 0, \ldots, n-1$. The asymptotic normality of $\widehat{\theta}_{\text{LLAD}}$ follows from Proposition 2 of Davis and Dunsmuir (1997) by reversing time to obtain an invertible representation. Or one can show the result by dealing with the objective function (3.42) directly. So, when the true model is noninvertible,

$$\sqrt{n}\left(\widehat{\theta}_{\text{LLAD}}-\theta_0\right) \rightarrow \frac{|\theta_0|\sqrt{\theta_0^2-1}}{2f\left(0\right)}N_1$$

in distribution, where $\widehat{\theta}_{\text{LLAD}}$ is the minimizer of (3.42).

3.4 Simulation Study

In this section we undertake a simulation study to evaluate the asymptotic theory. The data $\{x_t\}$ are generated from noninvertible MA(1) model $X_t = Z_t + \theta_0 Z_{t-1}$, with the value of θ_0 taken to be ± 1.1, ± 1.5, and ± 1.9, respectively. We use two different types of the underlying noise $\{Z_t\}$ to simulate data: the Laplacian distribution with the scale parameter $\sigma = 1.0$ and the Student's t distribution with 3 degrees of freedom. We consider two sample sizes of 500 and 5000.

For each case, we simulate 1000 replications, and report the empirical mean, standard deviation, and percent coverage of nominal 95% confidence intervals (CIs) as in the simulation study of Breidt, Davis, and Trindade (2001). To compute the percent coverage of nominal 95% CIs, the empirical CIs are computed from the asymptotic theory for each of 1000 replications. To be specific, once an estimate $\widehat{\theta}_n$ of θ is obtained for each replication, we fit the model and calculate the residuals $\left\{z_t(\widehat{\theta}_n)\right\}$. Then, $\text{Var}(Z_1)$ is estimated by the empirical variance of the residuals, and $f_Z\left(0\right)$ is estimated by the normal kernel density estimation at zero with a default normal scale bandwidth selector. Hence the standard error of $\widehat{\theta}_n$ can be estimated and the 95% confidence interval constructed. We also report 95% confidence intervals for the empirical mean, standard deviation, and percent coverage. The asymptotic mean and standard deviation are based on (3.12).

The method of a combination of golden section search and successive parabolic interpolation is used to search for the minimizer of LAD objective function, which is implemented with the function "optimize" in R package.

The results for the Student's t underlying noise are reported in Tables 3.1 and 3.2, while

Table 3.1: Estimates for MA(1) model $X_t = Z_t + \theta_0 Z_{t-1}, Z_t \sim$ IID $t\,(3)$.

n	Asymptotic		Empirical		
	Mean	Std. dev.	Mean (CI)	Std. dev. (CI)	% Coverage (CI)
500	$\theta_0 = 1.1$	0.0177	1.1067	0.0240	92.7
			(1.1052, 1.1082)	(0.0230, 0.0251)	(91.1, 94.3)
5000	$\theta_0 = 1.1$	0.0056	1.1007	0.0060	94.0
			(1.1003, 1.1011)	(0.0058, 0.0063)	(92.5, 95.5)
500	$\theta_0 = 1.5$	0.0589	1.5074	0.0635	95.6
			(1.5035, 1.5114)	(0.0608, 0.0664)	(94.3, 96.9)
5000	$\theta_0 = 1.5$	0.0186	1.5003	0.0197	95.1
			(1.4989, 1.5013)	(0.0183, 0.0200)	(93.8, 96.4)
500	$\theta_0 = 1.9$	0.1078	1.9128	0.1157	96.4
			(1.9057, 1.9200)	(0.1108, 0.1210)	(95.2, 97.6)
5000	$\theta_0 = 1.9$	0.0341	1.9005	0.0361	94.6
			(1.8982, 1.9027)	(0.0346, 0.0378)	(93.2, 96.0)

Table 3.2: Estimates for MA(1) model $X_t = Z_t + \theta_0 Z_{t-1}, Z_t \sim$ IID $t\,(3)$.

n	Asymptotic		Empirical		
	Mean	Std. dev.	Mean (CI)	Std. dev. (CI)	% Coverage (CI)
500	$\theta_0 =$ -1.1	0.0177	-1.1069	0.0251	92.3
			(-1.1085, -1.1054)	(0.0240, 0.0263)	(90.6, 93.9)
5000	$\theta_0 =$ -1.1	0.0056	-1.1007	0.0059	95.4
			(-1.1011, -1.1004)	(0.0056, 0.0061)	(94.1, 96.7)
500	$\theta_0 =$ -1.5	0.0589	-1.5037	0.0636	94.9
			(-1.5077, -1.4997)	(0.0609, 0.0666)	(93.5, 96.2)
5000	$\theta_0 =$ -1.5	0.0186	-1.5003	0.0193	95.5
			(-1.5015, -1.4991)	(0.0184, 0.0201)	(94.2, 96.8)
500	$\theta_0 =$ -1.9	0.1078	-1.9067	0.1171	95.3
			(-1.9140, -1.8995)	(0.1122, 0.1225)	(94.0, 96.6)
5000	$\theta_0 =$ -1.9	0.0341	-1.9003	0.0357	94.8
			(-1.9025, -1.8981)	(0.0342, 0.0374)	(93.4, 96.2)

Table 3.3: Estimates for MA(1) model $X_t = Z_t + \theta_0 Z_{t-1}, Z_t \sim$ IID Laplacian.

	Asymptotic		Empirical		
n	Mean	Std. dev.	Mean (CI)	Std. dev. (CI)	% Coverage (CI)
500	$\theta_0 = 1.1$	0.0159	1.1045	0.0212	93.3
			(1.1032, 1.1058)	(0.0203, 0.0222)	(91.7, 94.8)
5000	$\theta_0 = 1.1$	0.0050	1.1004	0.0053	96.4
			(1.1000, 1.1007)	(0.0051, 0.0056)	(95.2, 97.6)
500	$\theta_0 = 1.5$	0.0530	1.4999	0.0561	97.0
			(1.4964, 1.5034)	(0.0538, 0.0587)	(95.9, 98.0)
5000	$\theta_0 = 1.5$	0.0168	1.4993	0.0164	97.2
			(1.4983, 1.5003)	(0.0157, 0.0171)	(96.2, 98.2)
500	$\theta_0 = 1.9$	0.0971	1.8989	0.1027	96.4
			(1.8926, 1.9053)	(0.0984, 0.1074)	(95.2, 97.6)
5000	$\theta_0 = 1.9$	0.0307	1.8985	0.0298	97.8
			(1.8966, 1.9003)	(0.0285, 0.0311)	(96.9, 98.7)

Table 3.4: Estimates for MA(1) model $X_t = Z_t + \theta_0 Z_{t-1}, Z_t \sim$ IID Laplacian.

	Asymptotic		Empirical		
n	Mean	Std. dev.	Mean (CI)	Std. dev. (CI)	% Coverage (CI)
500	$\theta_0 = -1.1$	0.0159	-1.1058	0.0197	96.1
			(-1.1070, -1.1046)	(0.0188, 0.0206)	(94.9, 97.3)
5000	$\theta_0 = -1.1$	0.0050	-1.1006	0.0052	97.8
			(-1.1009, -1.1003)	(0.0050, 0.0055)	(96.9, 98.7)
500	$\theta_0 = -1.5$	0.0530	-1.5040	0.0581	96.8
			(-1.5076, -1.5004)	(0.0557, 0.0608)	(95.7, 97.9)
5000	$\theta_0 = -1.5$	0.0168	-1.5008	0.0168	96.8
			(-1.5018, -1.4998)	(0.0161, 0.0176)	(95.7, 97.9)
500	$\theta_0 = -1.9$	0.0971	-1.9076	0.1076	96.6
			(-1.9144, -1.9011)	(0.1031, 0.1126)	(95.5, 97.7)
5000	$\theta_0 = -1.9$	0.0307	-1.9015	0.0304	97.3
			(-1.9034, -1.8996)	(0.0291, 0.0318)	(96.3, 98.3)

those for the Laplacian noise are reported in Tables 3.3 and 3.4. We can see that, the LAD estimates and empirical standard deviations are comparable for the two types of underlying noise. But, while confidence interval coverages for the Student's t noise are close to the nominal 95% level, those for the Laplacian noise tend to be a little higher. In all cases with sample size 5000, the LAD estimates are approximately unbiased and empirical standard deviations are in agreement with asymptotic ones. The boxplots and normal probability plots (Figures 3.3, 3.5, 3.7, and 3.9) look pretty good. For each of the cases with sample size 500, especially for $\theta_0 = \pm 1.1$, there are some bias and more variation as well in the LAD estimates. The boxplots and normal probability plots (Figures 3.2, 3.4, 3.6, and 3.8) indicate that the bias and extra variation are due to a relatively small number of large (in magnitude) outliers. Except these outlying values, most of the LAD estimates follows the asymptotic normal law quite well. The empirical results for $\theta_0 = -1.1$, -1.5, and -1.9, and those for $\theta_0 = 1.1$, 1.5, and 1.9, are roughly symmetric; this is expected because the asymptotic theory is symmetric around zero. The study shows that, when θ_0 gets close to the unit circle, the estimation is more difficult and the convergence is slow.

We also conduct a simulation study regarding the LLAD estimator. The data $\{x_t\}$ are generated from the MA(1) model $X_t = Z_t + \theta_0 Z_{t-1}$, with the value of θ_0 taken to be 1.1, 1.5, and 1.9, respectively. We consider underlying noise following a Student's t distribution with 3 degrees of freedom, and sample size $n = 5000$. A method of moments estimation is used to construct the initial estimate $\widehat{\theta}_0$, and then the linearized objective function is minimized yielding the estimate $\widehat{\theta}_1$. This procedure is iterated 4 times, replacing $\widehat{\theta}_0$ with the newly obtained estimate at each iteration. Results are reported in Table 3.5. The first column gives the asymptotic mean (number at the top) and standard deviation (number in parentheses) based on the theory, and the second column gives the empirical mean and standard deviation of estimates using the method of moments. Other columns state the empirical mean and standard deviation of the LLAD estimates. The results for θ_0 of 1.5 and 1.9 are comparable to corresponding results for the LAD estimates. But for the $\theta_0 = 1.1$ case, the empirical standard deviation of LLAD estimates is larger than the theoretical

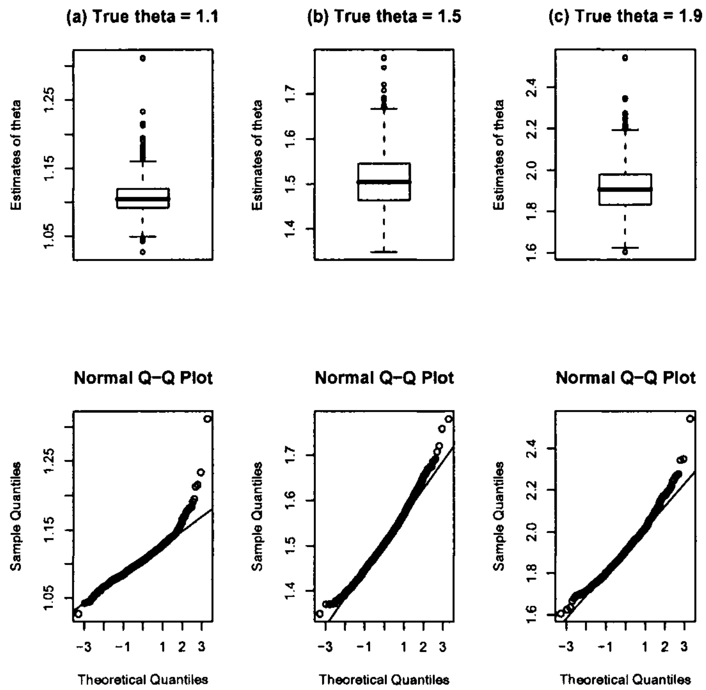

Figure 3.2: Boxplots and Normal probability plots of estimates when $n = 500$ and $\{Z_t\} \sim$ IID $t(3)$. (a) $\theta_0 = 1.1$. (b) $\theta_0 = 1.5$. (c) $\theta_0 = 1.9$.

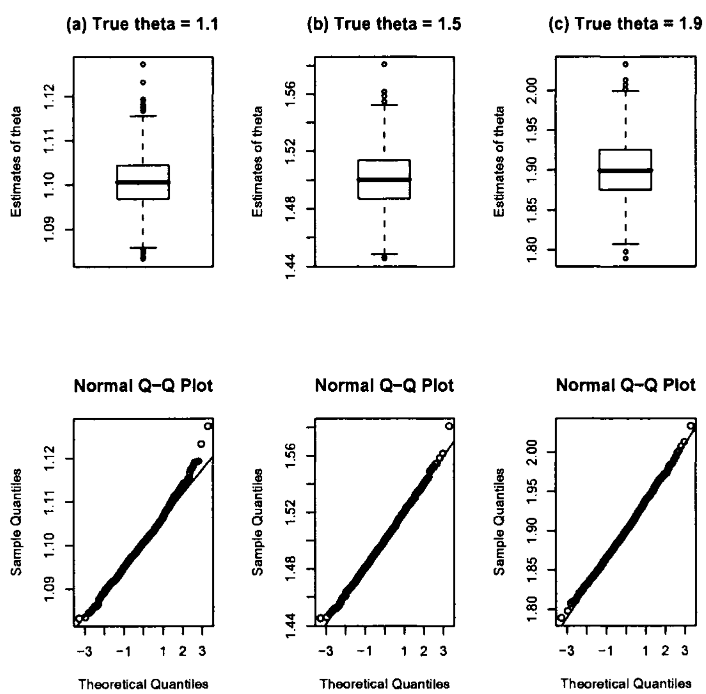

Figure 3.3: Boxplots and Normal probability plots of estimates when $n = 5000$ and $\{Z_t\} \sim$ IID $t(3)$. (a) $\theta_0 = 1.1$. (b) $\theta_0 = 1.5$. (c) $\theta_0 = 1.9$.

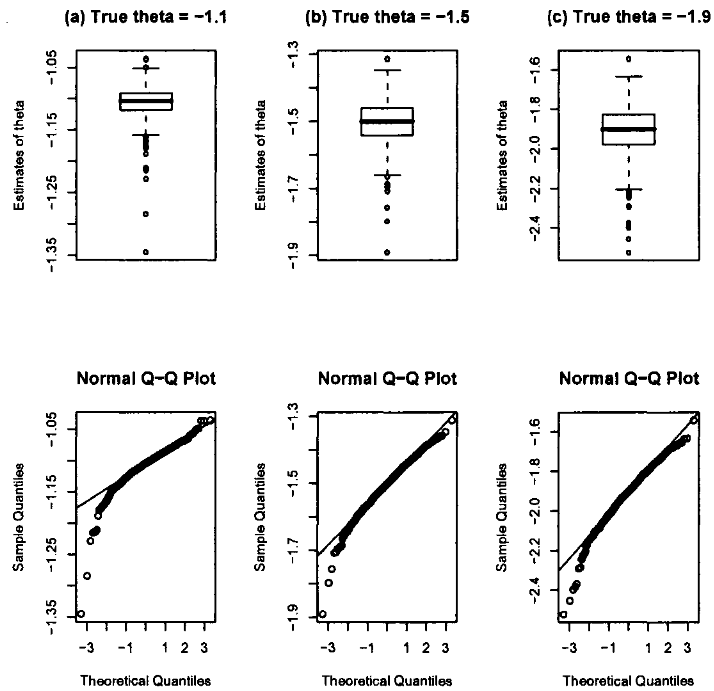

Figure 3.4: Boxplots and Normal probability plots of estimates when n = 500 and $\{Z_t\} \sim$ IID $t(3)$. (a) $\theta_0 = -1.1$. (b) $\theta_0 = -1.5$. (c) $\theta_0 = -1.9$.

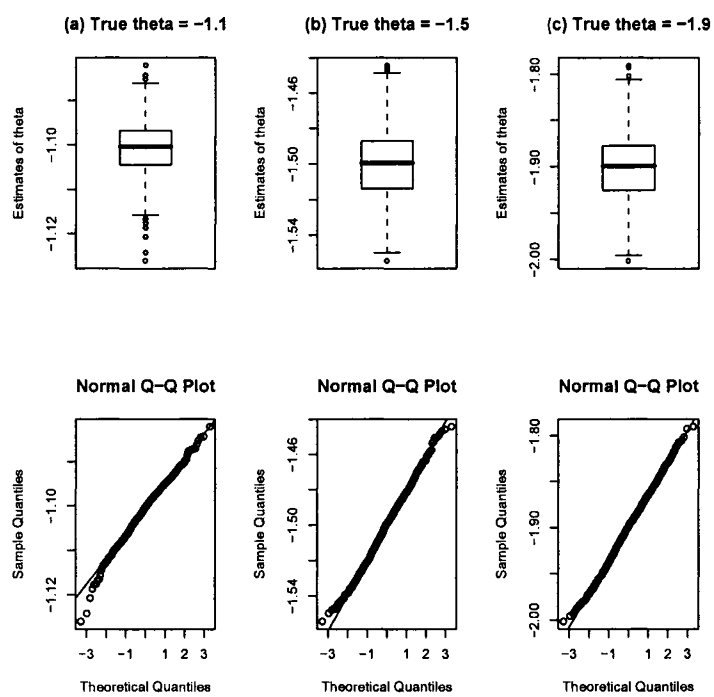

Figure 3.5: Boxplots and Normal probability plots of estimates when n = 5000 and $\{Z_t\} \sim$ IID $t(3)$. (a) $\theta_0 = -1.1$. (b) $\theta_0 = -1.5$. (c) $\theta_0 = -1.9$.

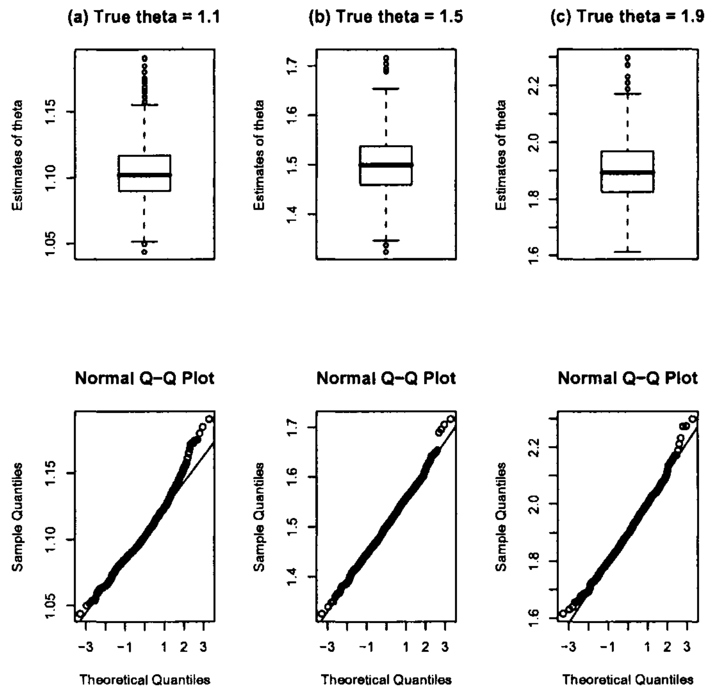

Figure 3.6: Boxplots and Normal probability plots of estimates when $n = 500$ and $\{Z_t\} \sim$ IID Laplacian. (a) $\theta_0 = 1.1$. (b) $\theta_0 = 1.5$. (c) $\theta_0 = 1.9$.

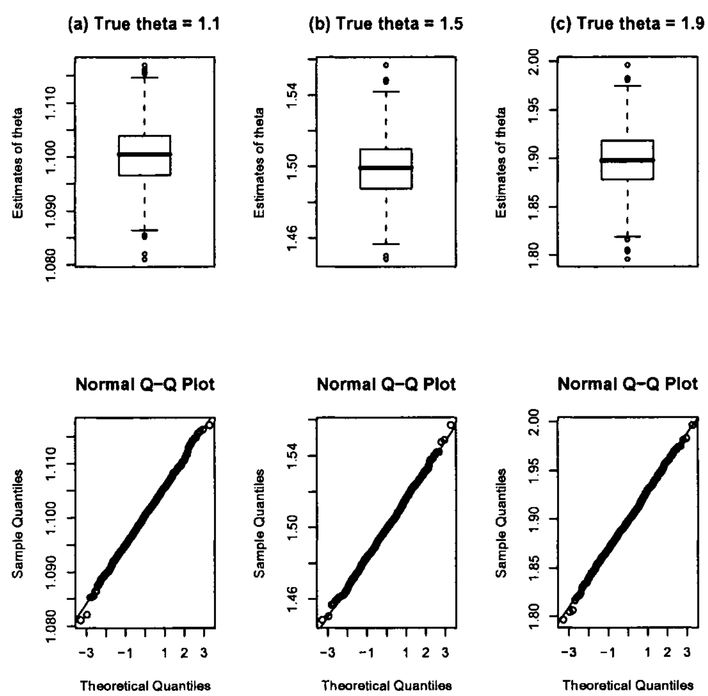

Figure 3.7: Boxplots and Normal probability plots of estimates when $n = 5000$ and $\{Z_t\} \sim$ IID Laplacian. (a) $\theta_0 = 1.1$. (b) $\theta_0 = 1.5$. (c) $\theta_0 = 1.9$.

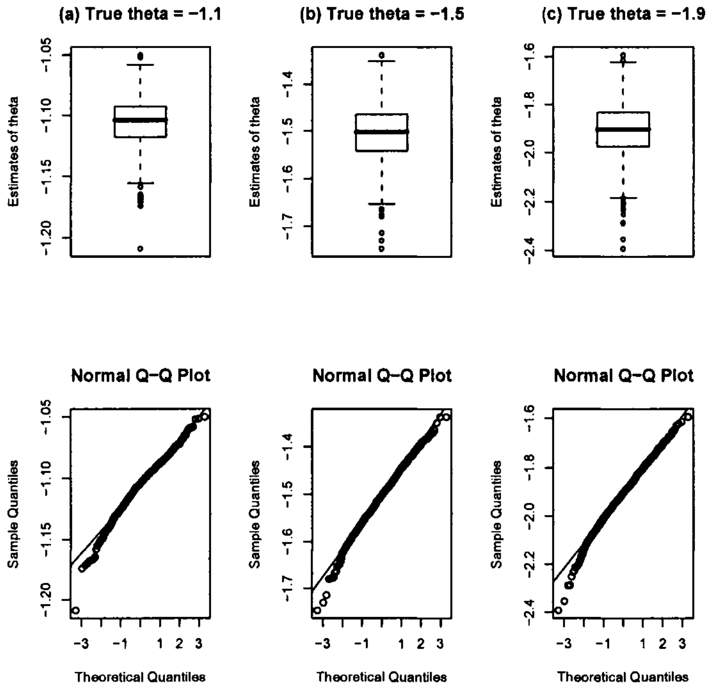

Figure 3.8: Boxplots and Normal probability plots of estimates when $n = 500$ and $\{Z_t\} \sim$ IID Laplacian. (a) $\theta_0 = -1.1$. (b) $\theta_0 = -1.5$. (c) $\theta_0 = -1.9$.

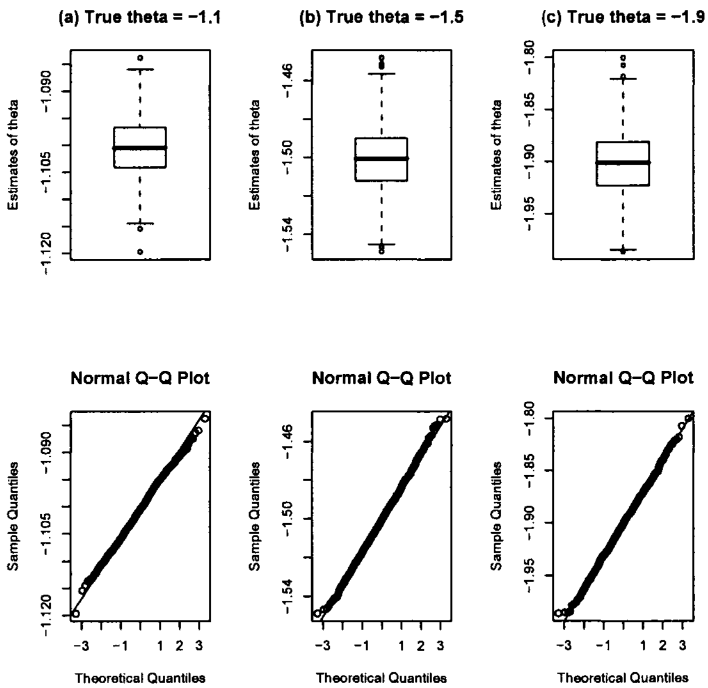

Figure 3.9: Boxplots and Normal probability plots of estimates when $n = 5000$ and $\{Z_t\} \sim$ IID Laplacian. (a) $\theta_0 = -1.1$. (b) $\theta_0 = -1.5$. (c) $\theta_0 = -1.9$.

Table 3.5: Linearized LAD estimates for MA(1) model $X_t = Z_t + \theta_0 Z_{t-1}, Z_t \sim$ IID $t(3)$.

θ_0	$\widehat{\theta_0}$	$\widehat{\theta_1}$	$\widehat{\theta_2}$	$\widehat{\theta_3}$	$\widehat{\theta_4}$
1.1	1.1974	1.0695	1.0857	1.0930	1.0952
(0.0056)	(0.0908)	(0.0289)	(0.0273)	(0.0267)	(0.0282)
1.5	1.5014	1.4912	1.4998	1.5002	1.5002
(0.0186)	(0.0892)	(0.0255)	(0.0195)	(0.0193)	(0.0193)
1.9	1.9039	1.8959	1.9005	1.9005	1.9005
(0.0341)	(0.0898)	(0.0362)	(0.0353)	(0.0353)	(0.0353)

one, showing that there exist more uncertainty and variation when θ_0 gets close to the unit circle. The boxplots and normal probability plots are displayed in Figure 3.10 for $\theta_0 = 1.5$ case.

3.5 Appendix

3.5.1 Derivation of (3.32)

While we are unable to give a complete proof of (3.32) at the present time, we map out one potential argument. We show that the convergence (3.32) holds for the case where $\{\theta_n\}$ is a non-random sequence such that $\theta_n \to \theta_0$, and conjecture that it also holds for a random sequence $\{\theta_n\}$ such that $\theta_n \stackrel{\text{a.s.}}{\to} \theta_0$.

Plugging θ_n into (3.26) yields

$$\begin{aligned}
&\frac{1}{n}\frac{d^2 v_n(\theta_n)}{d\theta^2} \\
&= \frac{1}{n}\sum_{i=0}^{n-1} n^p 1_{\{|r_{ni}(\theta_n)|\le n^{-p}\}}\left(-\sum_{j=1}^{n-1-i}\theta_n^{-j}Z_{j+i} + (\theta_n-\theta_0)\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)}Z_{j+i}\right)^2 \\
&\quad + \frac{1}{n}\sum_{i=0}^{n-1} n^p\left(r_{ni}(\theta_n)\right)1_{\{|r_{ni}(\theta_n)|\le n^{-p}\}} \\
&\qquad\qquad \times\left(2\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)}Z_{j+i} - (\theta_n-\theta_0)\sum_{j=1}^{n-1-i} j(j+1)\theta_n^{-(j+2)}Z_{j+i}\right) \\
&\quad + \frac{1}{n}\sum_{i=0}^{n-1}\operatorname{sgn}\left(r_{ni}(\theta_n)\right)1_{\{|r_{ni}(\theta_n)|> n^{-p}\}} \\
&\qquad\qquad \times\left(2\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)}Z_{j+i} - (\theta_n-\theta_0)\sum_{j=1}^{n-1-i} j(j+1)\theta_n^{-(j+2)}Z_{j+i}\right) \\
&\equiv T_1(\theta_n) + T_2(\theta_n) + T_3(\theta_n).
\end{aligned} \tag{3.43}$$

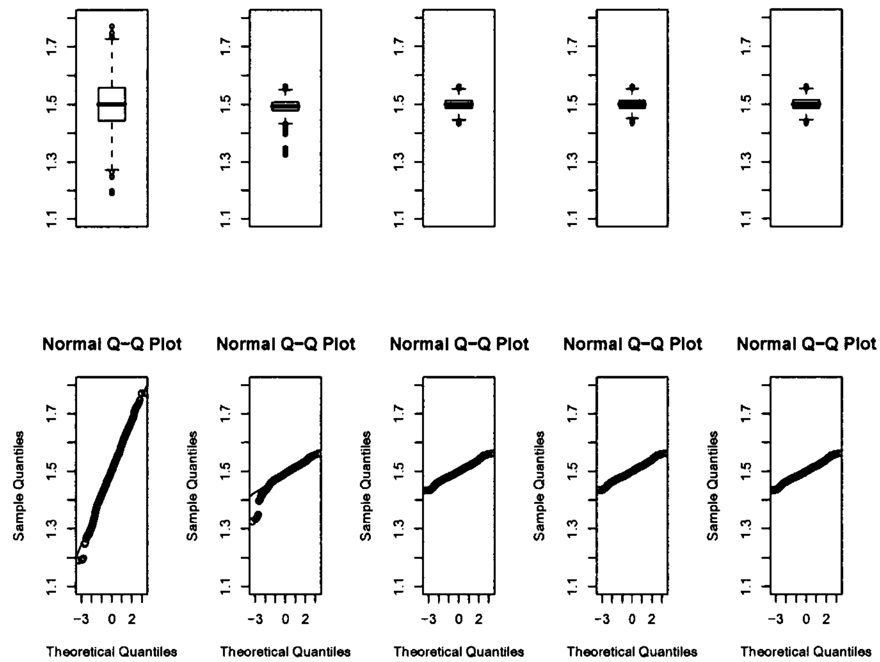

Figure 3.10: Boxplots and Normal probability plots of LLAD estimates when $n = 5000$ and $\{Z_t\} \sim$ IID $t(3)$. (a) $\widehat{\theta}_0$. (b) $\widehat{\theta}_1$. (c) $\widehat{\theta}_2$. (d) $\widehat{\theta}_3$. (e) $\widehat{\theta}_4$.

Let $G(\cdot)$ and $g(\cdot)$ denote the cdf and pdf of $\theta_0 Z_1$, respectively. Firstly we consider the term $T_2(\theta_n)$, which can be written as a sum of following two pieces.

$$\begin{aligned} T_2(\theta_n) &= \frac{2}{n}\sum_{i=0}^{n-1} n^p\left(r_{ni}(\theta_n)\right) 1_{\{|r_{ni}(\theta_n)|\le n^{-p}\}}\left(\sum\nolimits_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i}\right) \\ &\quad -\frac{(\theta_n-\theta_0)}{n}\sum_{i=0}^{n-1} n^p\left(r_{ni}(\theta_n)\right) 1_{\{|r_{ni}(\theta_n)|\le n^{-p}\}}\left(\sum\nolimits_{j=1}^{n-1-i} j(j+1)\theta_n^{-(j+2)} Z_{j+i}\right) \\ &\equiv T_{21}(\theta_n) - T_{22}(\theta_n). \end{aligned}$$

We show

$$T_2(\theta_n) \xrightarrow{P} 0 \qquad (3.44)$$

by showing

$$T_{21}(\theta_n) \xrightarrow{P} 0 \quad \text{and} \quad T_{22}(\theta_n) \xrightarrow{P} 0. \qquad (3.45)$$

For the term $T_{21}(\theta_n)$, we have

$$\begin{aligned}
\mathrm{E}\left|T_{21}\left(\theta_n\right)\right| &\leq \frac{2}{n}\sum_{i=0}^{n-1}\mathrm{E}(n^p\left|r_{ni}\left(\theta_n\right)\right| 1_{\{|r_{ni}(\theta_n)|\leq n^{-p}\}}\left|\sum\nolimits_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i}\right|) \\
&\leq \frac{2}{n}\sum_{i=0}^{n-1}\mathrm{E}\{\mathrm{E}(1_{\{|r_{ni}(\theta_n)|\leq n^{-p}\}}\left|\sum\nolimits_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i}\right||\mathfrak{F}^{i+1})\} \\
&= \frac{2}{n}\sum_{i=0}^{n-1}\mathrm{E}\{\left|\sum\nolimits_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i}\right| \cdot \mathrm{E}(1_{\{|r_{ni}(\theta_n)|\leq n^{-p}\}}|\mathfrak{F}^{i+1})\}.
\end{aligned}$$

Since $g(\cdot)$ is bounded,

$$\begin{aligned}
&\mathrm{E}(1_{\{|r_{ni}(\theta_n)|\leq n^{-p}\}}|\mathfrak{F}^{i+1}) \\
&\quad= \mathrm{E}\left(1_{\{|\theta_0 Z_i-(\theta_n-\theta_0)y_{ni}(\theta_n)|\leq n^{-p}\}}|\mathfrak{F}^{i+1}\right) \\
&\quad= G\left(\left(\theta_n-\theta_0\right) y_{ni}\left(\theta_n\right)+n^{-p}\right)-G\left(\left(\theta_n-\theta_0\right) y_{ni}\left(\theta_n\right)-n^{-p}\right) \qquad (3.46) \\
&\quad= 2n^{-p} g\left(\zeta_{ni}^1\right) \\
&\quad\leq \text{Const}\cdot n^{-p},
\end{aligned}$$

where ζ_{ni}^1 lies between $\left(\theta_n-\theta_0\right) y_{ni}\left(\theta_n\right)-n^{-p}$ and $\left(\theta_n-\theta_0\right) y_{ni}\left(\theta_n\right)+n^{-p}$. Then,

$$\mathrm{E}\left|T_{21}\left(\theta_n\right)\right| \leq \frac{2\text{Const}}{n^{1+p}}\sum_{i=0}^{n-1}\mathrm{E}\left|\sum\nolimits_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i}\right| \leq \frac{\text{Const}}{n^p} \to 0, \qquad (3.47)$$

because $\mathrm{E}\left|\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i}\right| \leq$ Const uniformly in i, which is easy to show. It follows from (3.47) that $T_{21}\left(\theta_n\right) \to 0$ in probability. Likewise, we can show that $T_{22}\left(\theta_n\right) \to 0$ in probability.

Next, we consider the term $T_3\left(\theta_n\right)$.

$$\begin{aligned}
T_3\left(\theta_n\right) = &\frac{2}{n}\sum_{i=0}^{n-1}\operatorname{sgn}\left(r_{ni}\left(\theta_n\right)\right)\left(\sum\nolimits_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i}\right) \\
&-\frac{2}{n}\sum_{i=0}^{n-1}\operatorname{sgn}\left(r_{ni}\left(\theta_n\right)\right) 1_{\{|r_{ni}(\theta_n)|\leq n^{-p}\}}\left(\sum\nolimits_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i}\right) \\
&-\frac{\left(\theta_n-\theta_0\right)}{n}\sum_{i=0}^{n-1}\operatorname{sgn}\left(r_{ni}\left(\theta_n\right)\right)\left(\sum\nolimits_{j=1}^{n-1-i} j\left(j+1\right)\theta_n^{-(j+2)} Z_{j+i}\right) \\
&+\frac{\left(\theta_n-\theta_0\right)}{n}\sum_{i=0}^{n-1}\operatorname{sgn}\left(r_{ni}\left(\theta_n\right)\right) 1_{\{|r_{ni}(\theta_n)|\leq n^{-p}\}}\left(\sum\nolimits_{j=1}^{n-1-i} j\left(j+1\right)\theta_n^{-(j+2)} Z_{j+i}\right)
\end{aligned}$$

$$\equiv T_{31}(\theta_n) - T_{32}(\theta_n) - T_{33}(\theta_n) + T_{34}(\theta_n). \quad (3.48)$$

We show $T_3(\theta_n) \to 0$ in probability by showing $T_{3k}(\theta_n) \to 0$ in probability for $k = 1, 2, 3, 4$. The derivations of results for $T_{32}(\theta_n)$ and $T_{34}(\theta_n)$ pretty much follow the same lines as those for $T_{21}(\theta_n)$ and $T_{22}(\theta_n)$. The result for $T_{33}(\theta_n)$ follows from that

$$\begin{aligned}
\mathrm{E}\left|T_{33}(\theta_n)\right| &\le \frac{|\theta_n - \theta_0|}{n}\sum_{i=0}^{n-1}\mathrm{E}\left|\sum\nolimits_{j=1}^{n-1-i} j(j+1)\,\theta_n^{-(j+2)} Z_{j+i}\right| \\
&\le |\theta_n - \theta_0|\left(\frac{1}{n}\sum_{i=0}^{n-1}\mathrm{Const}\right) \\
&= \mathrm{Const}\cdot|\theta_n - \theta_0| \\
&\to 0.
\end{aligned}$$

Moving on to the term $T_{31}(\theta_n)$. The expectation of $(T_{31}(\theta_n))^2$ can be expressed as

$$\begin{aligned}
&\mathrm{E}\left(T_{31}(\theta_n)\right)^2 \\
&= \tfrac{4}{n^2}\sum_{i=0}^{n-1}\sum_{k=0}^{n-1}\mathrm{E}[\operatorname{sgn}(r_{ni}(\theta_n))\,(\textstyle\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i})\cdot\operatorname{sgn}(r_{nk}(\theta_n))\,(\textstyle\sum_{l=1}^{n-1-k} l\theta_n^{-(l+1)} Z_{l+k})] \\
&= \tfrac{4}{n^2}\sum_{i=0}^{n-1}\mathrm{E}(\textstyle\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i})^2 \\
&\quad + \tfrac{8}{n^2}\sum_{i=1}^{n-1}\sum_{k=0}^{i-1}\mathrm{E}[\operatorname{sgn}(r_{ni}(\theta_n))\,(\textstyle\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i})\cdot\operatorname{sgn}(r_{nk}(\theta_n))\,(\textstyle\sum_{l=1}^{n-1-k} l\theta_n^{-(l+1)} Z_{l+k})] \\
&\equiv T_{31}^1(\theta_n) + T_{31}^2(\theta_n).
\end{aligned}$$

It is easy to show that the term $T_{31}^1(\theta_n)$ goes to zero. The term $T_{31}^2(\theta_n)$ also goes to zero: for $k < i$,

$$\begin{aligned}
&\mathrm{E}[\operatorname{sgn}(r_{ni}(\theta_n))\,(\textstyle\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i})\cdot\operatorname{sgn}(r_{nk}(\theta_n))\,(\textstyle\sum_{l=1}^{n-1-k} l\theta_n^{-(l+1)} Z_{l+k})] \\
&= \mathrm{E}\{\mathrm{E}[\operatorname{sgn}(r_{ni}(\theta_n))\,(\textstyle\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i})\cdot\operatorname{sgn}(r_{nk}(\theta_n))\,(\textstyle\sum_{l=1}^{n-1-k} l\theta_n^{-(l+1)} Z_{l+k})|\mathcal{F}^{k+1}]\} \\
&= \mathrm{E}\{\operatorname{sgn}(r_{ni}(\theta_n))\,(\textstyle\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i})(\textstyle\sum_{l=1}^{n-1-k} l\theta_n^{-(l+1)} Z_{l+k})\cdot\mathrm{E}\left[\operatorname{sgn}(r_{nk}(\theta_n))\,|\mathcal{F}^{k+1}\right]\}.
\end{aligned}$$

By Cauchy-Schwarz inequality we obtain

$$\left|\mathrm{E}[\operatorname{sgn}(r_{ni}(\theta_n))(\textstyle\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i})\cdot \operatorname{sgn}(r_{nk}(\theta_n))(\textstyle\sum_{l=1}^{n-1-k} l\theta_n^{-(l+1)} Z_{l+k})]\right|$$
$$\leq \{\mathrm{E}[(\textstyle\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i})^2(\textstyle\sum_{l=1}^{n-1-k} l\theta_n^{-(l+1)} Z_{l+k})^2]\}^{1/2}\cdot\{\mathrm{E}\{\mathrm{E}\left[\operatorname{sgn}(r_{nk}(\theta_n))\,|\mathcal{F}^{k+1}\right]\}^2\}^{1/2}$$
$$\leq \mathrm{Const}\cdot\{\mathrm{E}\{\mathrm{E}\left[\operatorname{sgn}(r_{nk}(\theta_n))\,|\mathcal{F}^{k+1}\right]\}^2\}^{1/2}.$$

But

$$\begin{aligned}
\mathrm{E}\left[\operatorname{sgn}(r_{nk}(\theta_n))\,|\mathcal{F}^{k+1}\right] &= \mathrm{E}\left[\operatorname{sgn}(\theta_0 Z_k-(\theta_n-\theta_0)\,y_{nk}(\theta_n))\,|\mathcal{F}^{k+1}\right]\\
&= [1-G((\theta_n-\theta_0)\,y_{nk}(\theta_n))]-G((\theta_n-\theta_0)\,y_{nk}(\theta_n))\\
&= 1-2G((\theta_n-\theta_0)\,y_{nk}(\theta_n)),
\end{aligned}$$

and since $(\theta_n-\theta_0)\,y_{nk}(\theta_n)\to 0$ almost surely, by applying a first degree Taylor series to $G((\theta_n-\theta_0)\,y_{nk}(\theta_n))$ about zero we get

$$\begin{aligned}
G((\theta_n-\theta_0)\,y_{nk}(\theta_n)) &= G(0)+g\left(\zeta_{nk}^2\right)(\theta_n-\theta_0)\,y_{nk}(\theta_n)\\
&= \frac{1}{2}+g\left(\zeta_{nk}^2\right)(\theta_n-\theta_0)\,y_{nk}(\theta_n).
\end{aligned}$$

where the random variable ζ_{nk}^2 lies between 0 and $(\theta_n-\theta_0)\,y_{nk}(\theta_n)$. So,

$$\mathrm{E}\left[\operatorname{sgn}(r_{nk}(\theta_n))\,|\mathcal{F}^{k+1}\right]=-2g\left(\zeta_{nk}^2\right)(\theta_n-\theta_0)\,y_{nk}(\theta_n),$$

and then

$$|\mathrm{E}[\operatorname{sgn}(r_{ni}(\theta_n))(\textstyle\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i})\cdot \operatorname{sgn}(r_{nk}(\theta_n))(\textstyle\sum_{l=1}^{n-1-k} l\theta_n^{-(l+1)} Z_{l+k})]|$$
$$\leq \mathrm{Const}\cdot\{\mathrm{E}\{-2g\left(\zeta_{nk}^2\right)(\theta_n-\theta_0)\,y_{nk}(\theta_n)\}^2\}^{1/2}$$
$$\leq \mathrm{Const}\cdot|\theta_n-\theta_0|\cdot[\mathrm{E}(y_{nk}(\theta_n))^2]^{1/2}$$
$$= \mathrm{Const}\cdot|\theta_n-\theta_0|.$$

It follows that

$$\left|T_{31}^2(\theta_n)\right|\leq\frac{8}{n^2}\sum_{i=1}^{n-1}\sum_{k=0}^{i-1}\mathrm{Const}\cdot|\theta_n-\theta_0|\leq\mathrm{Const}\cdot|\theta_n-\theta_0|\to 0.$$

The convergences of $T_{31}^{1}(\theta_n)$ and $T_{31}^{2}(\theta_n)$ to zero yield that $\mathrm{E}(T_{31}(\theta_n))^2 \to 0$, which in turn implies that $T_{31}(\theta_n) \to 0$ in probability.

Hence, in order to show (3.32), it suffices to show

$$T_1(\theta_n) \xrightarrow{P} \frac{2f(0)\sigma}{|\theta_0|\left(\theta_0^2-1\right)}. \tag{3.49}$$

Rewrite $T_1(\theta_n)$ as

$$T_1(\theta_n) = \left(T_1(\theta_n) - T_1^{\dagger}(\theta_n)\right) + T_1^{\dagger}(\theta_n), \tag{3.50}$$

where $T_1^{\dagger}(\theta_n) = \frac{1}{n}\sum_{i=0}^{n-1} n^p 1_{\{|r_{ni}(\theta_n)|\le n^{-p}\}}(\sum_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i})^2$. We derive (3.49) by showing that

$$T_1(\theta_n) - T_1^{\dagger}(\theta_n) \xrightarrow{P} 0, \tag{3.51}$$

and

$$T_1^{\dagger}(\theta_n) \xrightarrow{P} \frac{2f(0)\sigma}{|\theta_0|\left(\theta_0^2-1\right)}. \tag{3.52}$$

To show (3.51), we further split the term $T_1(\theta_n) - T_1^{\dagger}(\theta_n)$ into three pieces:

$$\begin{aligned}
&T_1(\theta_n) - T_1^{\dagger}(\theta_n)\\
&= \tfrac{1}{n}\sum_{i=0}^{n-1} n^p 1_{\{|r_{ni}(\theta_n)|\le n^{-p}\}}[-\textstyle\sum_{j=1}^{n-1-i}\theta_n^{-j}Z_{j+i} + (\theta_n-\theta_0)\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)}Z_{j+i}]^2\\
&\quad - \tfrac{1}{n}\sum_{i=0}^{n-1} n^p 1_{\{|r_{ni}(\theta_n)|\le n^{-p}\}}(\textstyle\sum_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i})^2\\
&= \tfrac{1}{n}\sum_{i=0}^{n-1} n^p 1_{\{|r_{ni}(\theta_n)|\le n^{-p}\}}[(\textstyle\sum_{j=1}^{n-1-i}\theta_n^{-j}Z_{j+i})^2 - (\sum_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i})^2]\\
&\quad - \tfrac{2(\theta_n-\theta_0)}{n}\sum_{i=0}^{n-1} n^p 1_{\{|r_{ni}(\theta_n)|\le n^{-p}\}}(\textstyle\sum_{j=1}^{n-1-i}\theta_n^{-j}Z_{j+i})(\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)}Z_{j+i})\\
&\quad + \tfrac{(\theta_n-\theta_0)^2}{n}\sum_{i=0}^{n-1} n^p 1_{\{|r_{ni}(\theta_n)|\le n^{-p}\}}(\textstyle\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)}Z_{j+i})^2\\
&\equiv \Delta_1 - \Delta_2 + \Delta_3.
\end{aligned}$$

We show that $\Delta_j, j = 1,2,3$, goes to zero in probability, respectively, and then the convergence (3.51) follows. To be specific, firstly the expectation of $|\Delta_1|$ is upper-bounded as follows.

$$
\begin{aligned}
\mathrm{E}\left|\Delta_1\right| &\leq \tfrac{1}{n}\sum_{i=0}^{n-1} n^p \mathrm{E}\{|(\textstyle\sum_{j=1}^{n-1-i}\theta_n^{-j}Z_{j+i})^2-(\sum_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i})^2|\cdot\mathrm{E}\left[1_{\{|r_{ni}(\theta_n)|\leq n^{-p}\}}|\mathcal{F}^{i+1}\right]\} \\
&\leq \tfrac{1}{n}\sum_{i=0}^{n-1} n^p \mathrm{E}\{|(\textstyle\sum_{j=1}^{n-1-i}\theta_n^{-j}Z_{j+i})^2-(\sum_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i})^2|\cdot(\mathrm{Const}\cdot n^{-p})\} \\
&= \mathrm{Const}\cdot\tfrac{1}{n}\sum_{i=0}^{n-1}\mathrm{E}|(\textstyle\sum_{j=1}^{n-1-i}\theta_n^{-j}Z_{j+i})^2-(\sum_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i})^2|.
\end{aligned}
$$

But

$$
\begin{aligned}
&\mathrm{E}\left|(\textstyle\sum_{j=1}^{n-1-i}\theta_n^{-j}Z_{j+i})^2-(\sum_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i})^2\right| \\
&\quad= \mathrm{E}\left|[\textstyle\sum_{j=1}^{n-1-i}(\theta_n^{-j}-\theta_0^{-j})Z_{j+i}][\sum_{j=1}^{n-1-i}(\theta_n^{-j}+\theta_0^{-j})Z_{j+i}]\right| \\
&\quad\leq \{\mathrm{E}[\textstyle\sum_{j=1}^{n-1-i}(\theta_n^{-j}-\theta_0^{-j})Z_{j+i}]^2\}^{1/2}\cdot\{\mathrm{E}[\sum_{j=1}^{n-1-i}(\theta_n^{-j}+\theta_0^{-j})Z_{j+i}]^2\}^{1/2} \\
&\quad= [\textstyle\sum_{j=1}^{n-1-i}(\theta_n^{-j}-\theta_0^{-j})^2\sigma^2]^{1/2}\cdot[\sum_{j=1}^{n-1-i}(\theta_n^{-j}+\theta_0^{-j})^2\sigma^2]^{1/2} \\
&\quad\leq \mathrm{Const}\cdot[\textstyle\sum_{j=1}^{n-1-i}(\theta_n^{-j}-\theta_0^{-j})^2]^{1/2}
\end{aligned}
$$

by applying Cauchy-Schwarz inequality. Recall that $\theta_n \to \theta_0$ and $|\theta_0| > 1$, so there exists a small $\delta > 0$ such that $1 < |\theta_0| - \delta \leq |\theta_n|$ when n is large enough. Applying the inequality $|a^l - b^l| \leq |l|\,(a \wedge b)^{l-1}\,|a-b|$ for a negative integer number l, we have

$$
\left|\theta_n^{-j}-\theta_0^{-j}\right| \leq j\,|\theta_n-\theta_0|\,(\theta_n\wedge\theta_0)^{-(j+1)} = j\,|\theta_n-\theta_0|\,(|\theta_0|-\delta)^{-(j+1)}.
$$

Then,

$$
\begin{aligned}
&\textstyle\sum_{j=1}^{n-1-i}(\theta_n^{-j}-\theta_0^{-j})^2 \\
&\quad\leq (\theta_n-\theta_0)^2\textstyle\sum_{j=1}^{n-1-i}j^2\,(|\theta_0|-\delta)^{-2(j+1)} \\
&\quad\leq (\theta_n-\theta_0)^2\cdot d[\tfrac{2d^2(1-d^{(n-1-i)})}{(1-d)^3}+\tfrac{d-(n-i)d^{(n-i)}-(n-1-i)d^{(n+1-i)}}{(1-d)^2}-\tfrac{(n-1-i)(n-i)d^{(n-i)}}{1-d}] \\
&\quad\leq (\theta_n-\theta_0)^2\,[\tfrac{2d^3}{(1-d)^3}+\tfrac{d^2}{(1-d)^2}] \\
&\quad= \mathrm{Const}\cdot(\theta_n-\theta_0)^2,
\end{aligned}
$$

where $d = (|\theta_0|-\delta)^{-2} \in (0,1)$. Therefore,

$$
\mathrm{E}\left|\Delta_1\right| \leq \mathrm{Const}\cdot\frac{1}{n}\sum_{i=0}^{n-1}[\sum_{j=1}^{n-1-i}(\theta_n^{-j}-\theta_0^{-j})^2]^{1/2}
$$

$$\le \text{Const} \cdot \frac{1}{n} \sum_{i=0}^{n-1} [\text{Const} \cdot (\theta_n - \theta_0)^2]^{1/2}$$

$$= \text{Const} \cdot |\theta_n - \theta_0|$$

$$\to 0,$$

which implies that $\Delta_1 \to 0$ in probability. Likewise, we can show $\Delta_2 \to 0$ in probability by showing that the expectation of $|\Delta_2|$ goes to zero:

$$\begin{aligned}
&\mathrm{E}\,|\Delta_2| \\
&\le \tfrac{2|\theta_n - \theta_0|}{n} \sum_{i=0}^{n-1} n^p \cdot \mathrm{E}\{|(\textstyle\sum_{j=1}^{n-1-i} \theta_n^{-j} Z_{j+i})(\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i})|\mathrm{E}\left[1_{\{|r_{ni}(\theta_n)| \le n^{-p}\}}|\mathcal{F}^{i+1}\right]\} \\
&\le \text{Const} \cdot |\theta_n - \theta_0| \cdot \tfrac{1}{n} \sum_{i=0}^{n-1} \mathrm{E}|(\textstyle\sum_{j=1}^{n-1-i} \theta_n^{-j} Z_{j+i})(\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i})| \\
&\le \text{Const} \cdot |\theta_n - \theta_0| \\
&\to 0
\end{aligned}$$

because $\mathrm{E}|(\sum_{j=1}^{n-1-i} \theta_n^{-j} Z_{j+i})(\sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i})|$ is upper-bounded by a constant uniformly in i. As to Δ_3, following the same lines for Δ_2 we obtain

$$\mathrm{E}\,|\Delta_3| \le \text{Const} \cdot (\theta_n - \theta_0)^2 \to 0,$$

from which the convergence of Δ_3 to zero in probability follows.

In order to show (3.52), we define

$$\varpi_{ni} \equiv n^p 1_{\{|r_{ni}(\theta_n)| \le n^{-p}\}} \left(\sum_{j=1}^{n-1-i} \theta_0^{-j} Z_{j+i}\right)^2$$

such that $T_1^\dagger(\theta_n) = \frac{1}{n} \sum_{i=0}^{n-1} \varpi_{ni}$. Then the conditional expectation

$$\mathrm{E}\left(\varpi_{ni}|\mathcal{F}^{i+1}\right) = n^p \left(\sum_{j=1}^{n-1-i} \theta_0^{-j} Z_{j+i}\right)^2 \mathrm{E}\left(1_{\{|r_{ni}(\theta_n)| \le n^{-p}\}}|\mathcal{F}^{i+1}\right).$$

But

$$
\begin{aligned}
&\mathrm{E}\left(1_{\{|r_{ni}(\theta_n)|\le n^{-p}\}}|\mathcal{F}^{i+1}\right)\\
&= \mathrm{E}\left(1_{\{|\theta_0 Z_i-(\theta_n-\theta_0)y_{ni}(\theta_n)|\le n^{-p}\}}|\mathcal{F}^{i+1}\right)\\
&= G\left((\theta_n-\theta_0)\,y_{ni}(\theta_n)+n^{-p}\right)-G\left((\theta_n-\theta_0)\,y_{ni}(\theta_n)-n^{-p}\right)\\
&= G\left((\theta_n-\theta_0)\,y_{ni}(\theta_n)\right)+g\left((\theta_n-\theta_0)\,y_{ni}(\theta_n)\right)\cdot n^{-p}+\tfrac{1}{2}g'\left(\zeta_{ni}^3\right)\cdot n^{-2p}\\
&\quad -\left[G\left((\theta_n-\theta_0)\,y_{ni}(\theta_n)\right)-g\left((\theta_n-\theta_0)\,y_{ni}(\theta_n)\right)\cdot n^{-p}+\tfrac{1}{2}g'\left(\zeta_{ni}^4\right)\cdot n^{-2p}\right]\\
&= 2n^{-p}g\left((\theta_n-\theta_0)\,y_{ni}(\theta_n)\right)+\tfrac{n^{-2p}}{2}\left[g'\left(\zeta_{ni}^3\right)-g'\left(\zeta_{ni}^4\right)\right]
\end{aligned}
$$

where the third equality follows from second degree Taylor expansions about $(\theta_n-\theta_0)\,y_{ni}(\theta_n)$, and random variables ζ_{ni}^3 and ζ_{ni}^4 are between $(\theta_n-\theta_0)\,y_i^0(\theta_n)+n^{-p}$ and $(\theta_n-\theta_0)\,y_i^0(\theta_n)$, and $(\theta_n-\theta_0)\,y_i^0(\theta_n)-n^{-p}$ and $(\theta_n-\theta_0)\,y_i^0(\theta_n)$, respectively. Then,

$$
\begin{aligned}
\tfrac{1}{n}\sum_{i=0}^{n-1}\mathrm{E}\left(\varpi_{ni}|\mathcal{F}^{i+1}\right) &= \tfrac{2}{n}\sum_{i=0}^{n-1}\left(\textstyle\sum_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i}\right)^2\cdot g\left((\theta_n-\theta_0)\,y_{ni}(\theta_n)\right)\\
&\quad+\tfrac{1}{2n^{1+p}}\sum_{i=0}^{n-1}\left(\textstyle\sum_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i}\right)^2\cdot\left[g'(\zeta_{ni}^3)-g'\left(\zeta_{ni}^4\right)\right]\\
&= \tfrac{2g(0)}{n}\sum_{i=0}^{n-1}\left(\textstyle\sum_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i}\right)^2\\
&\quad+\tfrac{2}{n}\sum_{i=0}^{n-1}\left(\textstyle\sum_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i}\right)^2\cdot\left[g\left((\theta_n-\theta_0)\,y_{ni}(\theta_n)\right)-g(0)\right]\\
&\quad+\tfrac{1}{2n^{1+p}}\sum_{i=0}^{n-1}\left(\textstyle\sum_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i}\right)^2\cdot\left[g'\left(\zeta_{ni}^3\right)-g'\left(\zeta_{ni}^4\right)\right]\\
&\equiv 2g(0)\,A_1+2A_2+A_3.
\end{aligned}
\tag{3.53}
$$

It is easy to show

$$
A_1\equiv\frac{1}{n}\sum_{i=0}^{n-1}\left(\sum\nolimits_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i}\right)^2\xrightarrow{P}\frac{\sigma^2}{\theta_0^2-1}. \tag{3.54}
$$

With a first degree Taylor expansion of $g(\cdot)$ about zero, we have

$$
g\left((\theta_n-\theta_0)\,y_{ni}(\theta_n)\right)-g(0)=g'\left(\zeta_{ni}^5\right)(\theta_n-\theta_0)\,y_{ni}(\theta_n)
$$

where ζ_{ni}^5 is a random variable lying between 0 and $(\theta_n-\theta_0)\,y_{ni}(\theta_n)$. It follows that

$$
\mathrm{E}\,|A_2|\le\frac{1}{n}\sum_{i=0}^{n-1}\mathrm{E}[(\sum\nolimits_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i})^2\cdot|g'\left(\zeta_{ni}^5\right)(\theta_n-\theta_0)\,y_{ni}(\theta_n)\,|]
$$

$$
\begin{aligned}
&\leq \text{Const} \cdot |\theta_n - \theta_0| \cdot \frac{1}{n} \sum_{i=0}^{n-1} \mathrm{E}[(\sum\nolimits_{j=1}^{n-1-i} \theta_0^{-j} Z_{j+i})^2 \, |y_{ni}(\theta_n)|] \\
&\leq \text{Const} \cdot |\theta_n - \theta_0| \cdot \frac{1}{n} \sum_{i=0}^{n-1} [\mathrm{E}(\sum\nolimits_{j=1}^{n-1-i} \theta_0^{-j} Z_{j+i})^4]^{1/2} \cdot [\mathrm{E}(y_{ni}(\theta_n))^2]^{1/2} \\
&\leq \text{Const} \cdot |\theta_n - \theta_0| \cdot \frac{1}{n} \sum_{i=0}^{n-1} \text{Const} \\
&= \text{Const} \cdot |\theta_n - \theta_0| \\
&\to 0,
\end{aligned}
$$

which implies

$$A_2 \xrightarrow{P} 0. \tag{3.55}$$

Moreover,

$$\mathrm{E}\,|A_3| \leq \frac{\text{Const}}{n^{1+p}} \sum_{i=0}^{n-1} \mathrm{E}(\sum\nolimits_{j=1}^{n-1-i} \theta_0^{-j} Z_{j+i})^2 \leq \frac{\text{Const}}{n^p} \to 0. \tag{3.56}$$

Notice that $g(0) = f(0)/(\sigma\,|\theta_0|)$. By putting (3.53)-(3.56) together, we establish

$$\frac{1}{n} \sum_{i=0}^{n-1} \mathrm{E}\left(\varpi_{ni} | \mathfrak{F}^{i+1}\right) \xrightarrow{P} \frac{2f(0)\,\sigma}{|\theta_0|\left(\theta_0^2 - 1\right)}. \tag{3.57}$$

Now consider the difference between $T_1^{\dagger}(\theta_n)$ and $\frac{1}{n}\sum_{i=0}^{n-1}\mathrm{E}(\varpi_{ni}|\mathfrak{F}^{i+1})$. We have

$$\mathrm{E}[T_1^{\dagger}(\theta_n) - \frac{1}{n} \sum_{i=0}^{n-1} \mathrm{E}\left(\varpi_{ni}|\mathfrak{F}^{i+1}\right)] = \frac{1}{n} \sum_{i=0}^{n-1} \mathrm{E}\left[\varpi_{ni} - \mathrm{E}\left(\varpi_{ni}|\mathfrak{F}^{i+1}\right)\right] = 0$$

and

$$
\begin{aligned}
\mathrm{Var}[T_1^{\dagger}(\theta_n) - \frac{1}{n} \sum_{i=0}^{n-1} \mathrm{E}\left(\varpi_{ni}|\mathfrak{F}^{i+1}\right)] &= \frac{1}{n^2} \sum_{i=0}^{n-1} \mathrm{E}\left[\varpi_{ni} - \mathrm{E}\left(\varpi_{ni}|\mathfrak{F}^{i+1}\right)\right]^2 \\
&= \frac{1}{n^2} \sum_{i=0}^{n-1} \mathrm{E}\{\mathrm{E}\{\left[\varpi_{ni} - \mathrm{E}\left(\varpi_{ni}|\mathfrak{F}^{i+1}\right)\right]^2 |\mathfrak{F}^{i+1}\}\} \\
&= \frac{1}{n^2} \sum_{i=0}^{n-1} \mathrm{E}\{\mathrm{E}\left(\varpi_{ni}^2|\mathfrak{F}^{i+1}\right) - \left[\mathrm{E}\left(\varpi_{ni}|\mathfrak{F}^{i+1}\right)\right]^2\}
\end{aligned}
$$

$$\leq \frac{1}{n^2}\sum_{i=0}^{n-1} \mathrm{E}\left[\mathrm{E}\left(\varpi_{ni}^2|\mathfrak{F}^{i+1}\right)\right].$$

But

$$\begin{aligned}
\mathrm{E}\left(\varpi_{ni}^2|\mathfrak{F}^{i+1}\right) &= n^{2p}(\sum_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i})^4\mathrm{E}\left[1_{\{|r_{ni}(\theta_n)|\leq n^{-p}\}}|\mathfrak{F}^{i+1}\right] \\
&\leq \text{Const}\cdot n^p\cdot(\sum_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i})^4.
\end{aligned}$$

Hence

$$\begin{aligned}
\mathrm{Var}[T_1^{\dagger}(\theta_n) - \frac{1}{n}\sum_{i=0}^{n-1}\mathrm{E}\left(\varpi_{ni}|\mathfrak{F}^{i+1}\right)] &\leq \text{Const}\cdot\frac{1}{n^{2-p}}\sum_{i=0}^{n-1}\mathrm{E}(\sum_{j=1}^{n-1-i}\theta_0^{-j}Z_{j+i})^4 \\
&\leq \text{Const}\cdot\frac{1}{n^{1-p}} \\
&\to 0
\end{aligned}$$

for $p \in (1/2, 1)$. It follows that

$$T_1^{\dagger}(\theta_n) - \frac{1}{n}\sum_{i=0}^{n-1}E\left(\varpi_{ni}|\mathfrak{F}^{i+1}\right) \xrightarrow{P} 0. \tag{3.58}$$

The convergence (3.52) follows from (3.58) and (3.57).

This completes the derivation of (3.32). ■

3.5.2 Derivation of (3.33)

The first derivative of $u_n(\theta)$ with respect to θ is

$$\frac{du_n(\theta)}{d\theta} = \sum_{i=0}^{n-1}\mathrm{sgn}(r_{ni}(\theta))\frac{dr_{ni}(\theta)}{d\theta}. \tag{3.59}$$

Plugging θ_n into (3.23) and (3.59) yields

$$\frac{dv_n(\theta_n)}{d\theta} = \sum_{i=0}^{n-1}n^p(r_{ni}(\theta_n))1_{\{|r_{ni}(\theta_n)|\leq n^{-p}\}}(-\sum_{j=1}^{n-1-i}\theta_n^{-j}Z_{j+i} + (\theta_n-\theta_0)\sum_{j=1}^{n-1-i}j\theta_n^{-(j+1)}Z_{j+i})$$

$$+\sum_{i=0}^{n-1} \operatorname{sgn}\left(r_{ni}\left(\theta_n\right)\right) 1_{\{|r_{ni}(\theta_n)|>n^{-p}\}}(-\sum_{j=1}^{n-1-i} \theta_n^{-j} Z_{j+i} + \left(\theta_n - \theta_0\right) \sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i}),$$

$$\frac{du_n\left(\theta_n\right)}{d\theta} = \sum_{i=0}^{n-1} \operatorname{sgn}\left(r_{ni}\left(\theta_n\right)\right)(-\sum_{j=1}^{n-1-i} \theta_n^{-j} Z_{j+i} + \left(\theta_n - \theta_0\right) \sum_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i}).$$

Hence

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\left(\frac{dv_n\left(\theta_n\right)}{d\theta} - \frac{du_n\left(\theta_n\right)}{d\theta}\right) \\
&= -\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} n^p \left(r_{ni}\left(\theta_n\right)\right) 1_{\{|r_{ni}(\theta_n)|\le n^{-p}\}} \left(\sum\nolimits_{j=1}^{n-1-i} \theta_n^{-j} Z_{j+i}\right) \\
&\quad + \frac{\left(\theta_n - \theta_0\right)}{\sqrt{n}} \sum_{i=0}^{n-1} n^p \left(r_{ni}\left(\theta_n\right)\right) 1_{\{|r_{ni}(\theta_n)|\le n^{-p}\}} \left(\sum\nolimits_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i}\right) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \operatorname{sgn}\left(r_{ni}\left(\theta_n\right)\right) 1_{\{|r_{ni}(\theta_n)|\le n^{-p}\}} \left(\sum\nolimits_{j=1}^{n-1-i} \theta_n^{-j} Z_{j+i}\right) \\
&\quad - \frac{\left(\theta_n - \theta_0\right)}{\sqrt{n}} \sum_{i=0}^{n-1} \operatorname{sgn}\left(r_{ni}\left(\theta_n\right)\right) 1_{\{|r_{ni}(\theta_n)|\le n^{-p}\}} \left(\sum\nolimits_{j=1}^{n-1-i} j\theta_n^{-(j+1)} Z_{j+i}\right) \\
&\equiv S_1\left(\theta_n\right) + S_2\left(\theta_n\right) + S_3\left(\theta_n\right) + S_4\left(\theta_n\right).
\end{aligned}$$

We show (3.33) by showing $S_k\left(\theta_n\right) \to 0$ in probability for $k = 1, 2, 3, 4$. Actually, it suffices to show that the term $S_1\left(\theta_n\right)$ goes to zero in probability, and results for other terms can be shown by following the same lines. Similar to establishing (3.47) for the term $T_{21}\left(\theta_n\right)$, we have

$$\mathrm{E}\left|S_1\left(\theta_n\right)\right| \le \frac{\text{Const}}{n^{p+1/2}} \sum_{i=0}^{n-1} \mathrm{E}\left|\sum\nolimits_{j=1}^{n-1-i} \theta_n^{-(j+1)} Z_{j+i}\right| \le \frac{\text{Const}}{n^{p-1/2}} \to 0 \tag{3.60}$$

because $\mathrm{E}\left|\sum_{j=1}^{n-1-i} \theta_n^{-(j+1)} Z_{j+i}\right| \le$ Const uniformly in i, and $p > 1/2$. Therefore, $S_1\left(\theta_n\right)$ goes to zero in probability. ■

4 LAD ESTIMATION FOR NONCAUSAL AND/OR NONINVERTIBLE ARMA(p, q) MODELS

In this chapter we study LAD estimation for noncausal and/or noninvertible ARMA(p, q) models. Let $\{X_t\}$ be an ARMA(p, q) process satisfying

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \tag{4.1}$$

where the underlying noise $\{Z_t\}$ is IID, $\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j$ and $\theta(z) = 1 + \sum_{j=1}^{q} \theta_j z^j$ have no roots on the unit circle, and $\phi(z)$ and $\theta(z)$ have no common roots. Let $\boldsymbol{\lambda} = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)^T \in \mathbb{R}^{p+q}$ be the parameter vector of the model (4.1) with $\boldsymbol{\lambda}_0 = (\phi_{01}, \ldots, \phi_{0p}, \theta_{01}, \ldots, \theta_{0q})^T$ denoting the true parameter vector. With the AR polynomial $\phi(\cdot)$ and the MA polynomial $\theta(\cdot)$, the model (4.1) can be written as

$$\phi(B) X_t = \theta(B) Z_t, \tag{4.2}$$

where B is the backshift operator. Under the assumption that the AR polynomial $\phi(z)$ has no roots on the unit circle, there exists a unique stationary solution of the ARMA(p, q) model (4.2).

We begin with deconstruction of the model (4.2). We factor the AR polynomial $\phi(\cdot)$ into its causal component $\phi^+(\cdot)$ and purely noncausal component $\phi^*(\cdot)$. To be specific, suppose $\phi(z) = \phi^+(z)\phi^*(z)$, where

$$\phi^+(z) = 1 - \phi_1^+ z - \cdots - \phi_{r'}^+ z^{r'} \neq 0 \qquad \text{for } |z| \leq 1,$$
$$\phi^*(z) = 1 - \phi_1^* z - \cdots - \phi_{s'}^* z^{s'} \neq 0 \qquad \text{for } |z| \geq 1,$$

with $r', s' \geq 0$ and $r' + s' = p$. Let $\zeta_1, \ldots, \zeta_{r'}$ be the r' roots of $\phi^+(z)$ (the roots of $\phi(z)$ outside the unit circle), and $\zeta_1^*, \ldots, \zeta_{s'}^*$ be the s' roots of $\phi^*(z)$ (the roots of $\phi(z)$ inside the unit circle). Then

$$\phi^+(z) = \prod_{j=1}^{r'} (1 - z/\zeta_j) \qquad \text{and} \qquad \phi^*(z) = \prod_{j=1}^{s'} (1 - z/\zeta_j^*),$$

where $|\zeta_j| > 1$, $j = 1, \ldots, r'$, and $|\zeta_j^*| < 1$, $j = 1, \ldots, s'$. Note that $-\phi_{r'}^+ = \prod_{j=1}^{r'} (-\zeta_j)^{-1}$ and $-\phi_{s'}^* = \prod_{j=1}^{s'} (-\zeta_j^*)^{-1}$. With the factorization of $\phi(\cdot)$, the model (4.2) is equivalent to the model

$$\phi^+(B)\, \phi^*(B)\, X_t = \theta(B)\, Z_t. \tag{4.3}$$

Moreover, by defining

$$U_t^+ \equiv \phi^+(B)\, X_t \qquad \text{and} \qquad U_t^* \equiv \phi^*(B)\, X_t,$$

we obtain

$$\phi^*(B)\, U_t^+ = \theta(B)\, Z_t \qquad \text{and} \qquad \phi^+(B)\, U_t^* = \theta(B)\, Z_t. \tag{4.4}$$

Likewise, we factor the MA polynomial $\theta(\cdot)$ into its invertible component $\theta^+(\cdot)$ and purely noninvertible component $\theta^*(\cdot)$. That is, suppose $\theta(z) = \theta^+(z)\, \theta^*(z)$, where

$$\begin{aligned} \theta^+(z) &= 1 + \theta_1^+ z + \cdots + \theta_r^+ z^r \neq 0 \qquad \text{for } |z| \leq 1, \\ \theta^*(z) &= 1 + \theta_1^* z + \cdots + \theta_s^* z^s \neq 0 \qquad \text{for } |z| \geq 1, \end{aligned}$$

with $r, s \geq 0$ and $r + s = q$. Let $\xi_1, \ldots, \xi_r$ be the r roots of $\theta^+(z)$, and $\xi_1^*, \ldots, \xi_s^*$ be the s roots of $\theta^*(z)$. Then

$$\theta^+(z) = \prod_{j=1}^{r} (1 - z/\xi_j) \qquad \text{and} \qquad \theta^*(z) = \prod_{j=1}^{s} (1 - z/\xi_j^*),$$

where $|\xi_j| > 1$, $j = 1, \ldots, r$, and $|\xi_j^*| < 1$, $j = 1, \ldots, s$. Note that $\theta_r^+ = \prod_{j=1}^{r} (-\xi_j)^{-1}$ and $\theta_s^* = \prod_{j=1}^{s} (-\xi_j^*)^{-1}$. Now the model (4.3) is further equivalent to the model

$$\phi^{+}(B)\,\phi^{*}(B)\,X_t = \theta^{+}(B)\,\theta^{*}(B)\,Z_t. \tag{4.5}$$

By defining

$$V_t^{+} \equiv \theta^{+}(B)\,Z_t \qquad \text{and} \qquad V_t^{*} \equiv \theta^{*}(B)\,Z_t,$$

we obtain

$$\begin{aligned} \phi^{+}(B)\,U_t^{*} &= \theta^{+}(B)\,V_t^{*}, & \phi^{*}(B)\,U_t^{+} &= \theta^{+}(B)\,V_t^{*}, \\ \phi^{+}(B)\,U_t^{*} &= \theta^{*}(B)\,V_t^{+}, & \phi^{*}(B)\,U_t^{+} &= \theta^{*}(B)\,V_t^{+}. \end{aligned} \tag{4.6}$$

In the causal-invertible case, $s' = s = 0$, $r' = p$, and $r = q$. On the other hand, $s' > 0$ in the noncausal case, and $s > 0$ in the noninvertible case. If the model (4.2) is noncausal, then X_t can be expressed in terms of $\{Z_t\}$, namely $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$, where ψ_j is the coefficient of z^j in the Laurent series expansion of $\theta(z)/\phi(z)$, which exists in some annulus $d < |z| < d^{-1}$, $d < 1$. Similarly, if the model is noninvertible, then we can write the noise Z_t in terms of $\{X_t\}$ by $Z_t = \sum_{j=-\infty}^{\infty} \pi_j X_{t-j}$, where π_j is the coefficient of z^j in the Laurent series expansion of $\phi(z)/\theta(z)$.

For causal-invertible ARMA(p,q) models, Davis and Dunsmuir (1997) proved the asymptotic normality of the LAD estimator of parameter vector. Following the strategy of Davis and Dunsmuir (1997), Pan, Wang, and Yao (2007) proposed a weighted LAD estimation for the case where the underlying noise has infinite variance, and showed that the proposed weighted LAD estimator is asymptotically normal with the standard root-n convergence rate. The idea of weighted LAD estimation is to give less weigh to outlying observations, which has also been used for IVAR models by Ling (2005), and for ARCH models with heavy-tailed innovations by Horvath and Liese (2004). Breidt, Davis, and Trindade (2001) established the asymptotic normality of LAD estimator for all-pass models, which are a special class of noncausal and/or noninvertible ARMA models. In this chapter, we establish the asymptotic behavior of LAD estimator for general ARMA(p,q) models, which can be noncausal and/or noninvertible.

We will focus on the deconstructed model (4.5) and study the LAD estimator $\widehat{\boldsymbol{\kappa}}$ of the model parameter vector $\boldsymbol{\kappa}$ defined as

$$\boldsymbol{\kappa} = \left(\phi_1^+, \ldots, \phi_{r'}^+, \phi_1^*, \ldots, \phi_{s'}^*, \theta_1^+, \ldots, \theta_r^+, \theta_1^*, \ldots, \theta_s^*\right)^T.$$

Let $\boldsymbol{\kappa}_0 = \left(\phi_{01}^+, \ldots, \phi_{0r'}^+, \phi_{01}^*, \ldots, \phi_{0s'}^*, \theta_{01}^+, \ldots, \theta_{0r}^+, \theta_{01}^*, \ldots, \theta_{0s}^*\right)^T$ denote the true parameter vector. We assume that s' and s are fixed; the asymptotic results are also valid when the LAD objective function depends on s' and s (see Remark 4.2).

4.1 LAD Criterion

The LAD criterion is derived by a likelihood approximation assuming that the underlying noise $\{Z_t\}$ is Laplacian, like in the MA(1) setup of Chapter 3. The derivation is based on Lii and Rosenblatt (1996) and is similar to that in Breidt, Davis, and Trindade (2001).

Let $X_1, \ldots, X_n$ be observed data from the model (4.5). We consider the augmented data vector

$$\mathbf{x} \equiv \left(V_{1-r}^*, \ldots, V_0^*, X_{1-p}, \ldots, X_{n+p}, V_{n+p+1-s}^+, \ldots, V_{n+p}^+\right)^T$$

and the augmented noise vector

$$\mathbf{z} \equiv \left(U_{1-r'}^*, \ldots, U_0^*, Z_{1-q}, \ldots, Z_{n+p}, U_{n+p+1-s'}^+, \ldots, U_{n+p}^+\right)^T.$$

Note that, $\left(U_{1-r'}^*, \ldots, U_0^*\right)^T$ is independent of $\{Z_1, Z_2, \ldots\}$. To see this, recall from (4.4) that $\phi^+(B)\, U_t^* = \theta(B)\, Z_t$. Since $\phi^+(z)$ is the causal component of $\phi(z)$, we can express its reciprocal as $\frac{1}{\phi^+(z)} = \sum_{i=0}^{\infty} \eta_i^+ z^i$, where $\sum_{i=0}^{\infty} \left|\eta_i^+\right| < \infty$. Then,

$$U_t^* = \left(\sum_{i=0}^{\infty} \eta_i^+ B^i\right)\left(1 + \sum_{j=1}^{q} \theta_j B^j\right) Z_t.$$

Hence, $U_t^* \in \sigma(Z_t, Z_{t-1}, \ldots)$, from which it follows that $\left(U_{1-r'}^*, \ldots, U_0^*\right)^T \in \sigma(Z_0, Z_{-1}, \ldots)$. Likewise, $(U_{n+p+1-s'}^+, \ldots, U_{n+p}^+)^T$ is independent of $\{Z_{n+p-q}, Z_{n+p-q-1}, \ldots\}$. To be specific, since $\phi^*(z)$ is the purely noncausal component of $\phi(z)$, we can express its reciprocal as $\frac{1}{\phi^*(z)} = \sum_{j=s'}^{\infty} \eta_j^* z^{-j}$, where $\sum_{j=s'}^{\infty} |\eta_j^*| < \infty$. Then, $\phi^*(B)\, U_t^+ = \theta(B)\, Z_t$ implies

$$U_t^+ = (\sum_{j=s'}^{\infty} \eta_j^* B^{-j})(1 + \sum_{j=1}^{q} \theta_j B^j) Z_t.$$

So, $U_t^+ \in \sigma(Z_{t+s'-q}, Z_{t+s'-q+1}, \ldots)$, from which it follows that $(U_{n+p+1-s'}^+, \ldots, U_{n+p}^+)^T \in \sigma(Z_{n+p-q+1}, Z_{n+p-q+2}, \ldots)$.

The joint probability density of $\mathbf{z}$ can be written as

$$\begin{aligned} h(\mathbf{z}) &= h_1\left(u_{1-r'}^*, \ldots, u_0^*, z_{1-q}, \ldots, z_0\right) \prod_{t=1}^{n+p-q} f_Z(z_t) \\ &\quad \times h_2\left(z_{n+p-q+1}, \ldots, z_{n+p}, u_{n+p+1-s'}^+, \ldots, u_{n+p}^+\right), \end{aligned}$$

where $f_Z(\cdot)$ is the density of the underlying noise $\{Z_t\}$. Then, the joint probability density of $\mathbf{x}$ is given by

$$h(\mathbf{x}) = h_1 \prod_{t=1}^{n+p-q} f_Z(z_t) \left|\frac{\phi_{s'}^*}{\theta_s^*}\right|^{n+p-q} c(\boldsymbol{\kappa})\, h_2,$$

where $c(\boldsymbol{\kappa})$ is a function of $\boldsymbol{\kappa}$ (see Lii and Rosenblatt, 1996). Because h_1, h_2, and $c(\boldsymbol{\kappa})$ do not depend on n, we can ignore these terms when n is large enough, and approximate the log-likelihood function of $\boldsymbol{\kappa}$ by

$$\mathcal{L}(\boldsymbol{\kappa}) = \sum_{t=1}^{n+p-q} \log f_Z(z_t) + (n+p-q)\log\left|\phi_{s'}^*/\theta_s^*\right|,$$

where the residual sequence $\{z_t\}$ is a function of $\boldsymbol{\kappa}$ and initial values of a set of random variables. Given these initial values, $\{z_t\}$ can be calculated from a sequence of transformations in Lii and Rosenblatt (1996). In practice, we can simply set these involved initial values to zero for computing $\{z_t\}$. To be specific, let $\{X_t : t = 1, 2, \ldots, n\}$ be observed data from the true model with the parameter vector $\boldsymbol{\kappa}_0$. We write $\{z_t\}$ as $\{z_t(\boldsymbol{\kappa})\}$ to indicate that z_t is a function of $\boldsymbol{\kappa}$. By defining $U_t(\boldsymbol{\kappa}) \equiv \phi(B) X_t$, we obtain

$$U_t(\boldsymbol{\kappa}) = \theta^*(B)\, V_t^+(\boldsymbol{\kappa}) \qquad \text{and} \qquad V_t^+(\boldsymbol{\kappa}) = \theta^+(B)\, z_t(\boldsymbol{\kappa}). \tag{4.7}$$

Setting

$$X_{1-p} = \cdots = X_0 = 0 \qquad \text{and} \qquad X_{n+1} = \cdots = X_{n+p} = 0,$$

we can calculate $U_t(\boldsymbol{\kappa})$ for $t = 1, \ldots, n+p$. Then, we compute $z_t(\boldsymbol{\kappa})$ for $t = 1, \ldots, n+p-q$ via the following two steps. Firstly, it follows from the first equation in (4.7) that

$$U_t(\boldsymbol{\kappa}) = V_t^+(\boldsymbol{\kappa}) + \theta_1^* V_{t-1}^+(\boldsymbol{\kappa}) + \cdots + \theta_{s-1}^* V_{t-s+1}^+(\boldsymbol{\kappa}) + \theta_s^* V_{t-s}^+(\boldsymbol{\kappa}),$$

or

$$V_t^+(\boldsymbol{\kappa}) = \frac{1}{\theta_s^*}\left[U_{t+s}(\boldsymbol{\kappa}) - V_{t+s}^+(\boldsymbol{\kappa}) - \theta_1^* V_{t+s-1}^+(\boldsymbol{\kappa}) - \cdots - \theta_{s-1}^* V_{t+1}^+(\boldsymbol{\kappa})\right]. \tag{4.8}$$

So, we compute $V_t^+(\boldsymbol{\kappa})$ for $t = -s+1, \ldots, n+p-s$ backwards by setting $V_t^+(\boldsymbol{\kappa}) = 0$ for $t \geq n+p-s+1$ and using the formula (4.8) for $t = n+p-s, \ldots, -s+1$. Secondly, it follows from the second equation in (4.7) that

$$z_t(\boldsymbol{\kappa}) = V_t^+(\boldsymbol{\kappa}) - \theta_1^+ z_{t-1}(\boldsymbol{\kappa}) - \cdots - \theta_r^+ z_{t-r}(\boldsymbol{\kappa}). \tag{4.9}$$

So, we compute $z_t(\boldsymbol{\kappa})$ for $t = 1, \ldots, n+p-q$ forwards by setting $z_t(\boldsymbol{\kappa}) = 0$ for $t \leq 0$ and using the formula (4.9).

Now, if we assume that the distribution of $\{Z_t\}$ is Laplacian, i.e., $f_Z(z) = \frac{1}{2\sigma}\exp\left(-\frac{|z|}{\sigma}\right)$, then the Laplacian log-likelihood function is approximated by

$$-(n+p-q)\log\sigma - \frac{1}{\sigma}\sum_{t=1}^{n+p-q}|z_t(\boldsymbol{\kappa})| + (n+p-q)\log|\phi_{s'}^*/\theta_s^*|. \tag{4.10}$$

By maximizing (4.10) with respect to σ, we obtain $\widehat{\sigma} = \frac{1}{n+p-q}\sum_{t=1}^{n+p-q}|z_t(\boldsymbol{\kappa})|$. Then, substituting $\widehat{\sigma}$ for σ in (4.10) we obtain the concentrated Laplacian log-likelihood function

$$\text{Constant} - (n+p-q)\log\left(\sum_{t=1}^{n+p-q}|z_t(\boldsymbol{\kappa})|\right) + (n+p-q)\log|\phi_{s'}^*/\theta_s^*|. \tag{4.11}$$

Note that, maximizing (4.11) is equivalent to minimizing the objective function

$$l_n(\boldsymbol{\kappa}) \equiv \sum_{t=1}^{n+p-q} \left| \frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa}) \right|, \tag{4.12}$$

and the LAD estimator $\widehat{\boldsymbol{\kappa}}_{\mathrm{LAD}}$ of $\boldsymbol{\kappa}$ is defined as any minimizer of (4.12).

4.2 Local Approximation

The objective function (4.12) is not convex in $\boldsymbol{\kappa}$, which complicates the study of the asymptotic behavior of $\widehat{\boldsymbol{\kappa}}_{\mathrm{LAD}}$. This is also the case in Davis and Dunsmuir (1997) for causal-invertible ARMA processes; a local linearization technique was used there to circumvent the difficulty. For noncausal and/or noninvertible ARMA processes, however, in order to establish the asymptotic results of $\widehat{\boldsymbol{\kappa}}_{\mathrm{LAD}}$, an extra quadratic term is needed for local approximation using the Taylor expansion. To be specific, for $t = 1, \ldots, n+p-q$, we approximate $\frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa})$ by

$$\frac{\theta_{0s}^*}{\phi_{0s'}^*} z_t(\boldsymbol{\kappa}_0) - \mathbf{D}_t^T(\boldsymbol{\kappa}_0)(\boldsymbol{\kappa} - \boldsymbol{\kappa}_0) - \frac{1}{2}(\boldsymbol{\kappa} - \boldsymbol{\kappa}_0)^T \mathbf{H}_t(\boldsymbol{\kappa}_0)(\boldsymbol{\kappa} - \boldsymbol{\kappa}_0),$$

where

$$\mathbf{D}_t(\boldsymbol{\kappa}) = -\frac{\partial\left(\frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa})\right)}{\partial \boldsymbol{\kappa}} = \left(D_{t,1}(\boldsymbol{\kappa}), \ldots, D_{t,p+q}(\boldsymbol{\kappa})\right)^T,$$
$$\mathbf{H}_t(\boldsymbol{\kappa}) = -\frac{\partial^2\left(\frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa})\right)}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\kappa}^T} = \left[H_{t,lm}(\boldsymbol{\kappa})\right]_{l,m=1}^{p+q}.$$

We evaluate $D_{t,l}(\boldsymbol{\kappa}) = -\frac{\partial\left(\frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l}$ for each $l = 1, \ldots, p+q$ as follows. Firstly, for $l = 1, \ldots, r'$, we consider the first equation in (4.6), from which we have

$$\left(1 - \phi_1^+ B - \cdots - \phi_{r'}^+ B^{r'}\right) U_t^*(\boldsymbol{\phi}^*) = \left(1 + \theta_1^+ B + \cdots + \theta_r^+ B^r\right) V_t^*(\boldsymbol{\kappa}),$$

and

$$V_t^*(\boldsymbol{\kappa}) = U_t^*(\boldsymbol{\phi}^*) - \phi_1^+ U_{t-1}^*(\boldsymbol{\phi}^*) - \cdots - \phi_{r'}^+ U_{t-r'}^*(\boldsymbol{\phi}^*) - \theta_1^+ V_{t-1}^*(\boldsymbol{\kappa}) - \cdots - \theta_r^+ V_{t-r}^*(\boldsymbol{\kappa}),$$

where $\boldsymbol{\phi}^* = \left(\phi_1^*, \ldots, \phi_{s'}^*\right)^T$. Then

$$\frac{\partial V_t^*(\boldsymbol{\kappa})}{\partial \kappa_l} = -U_{t-l}^*(\boldsymbol{\phi}^*) - \theta_1^+ \frac{\partial V_{t-1}^*(\boldsymbol{\kappa})}{\partial \kappa_l} - \cdots - \theta_r^+ \frac{\partial V_{t-r}^*(\boldsymbol{\kappa})}{\partial \kappa_l}.$$

That is, for $l = 1, \ldots, r'$,

$$\theta^+(B) \frac{\partial V_t^*(\boldsymbol{\kappa})}{\partial \kappa_l} = -U_{t-l}^*(\boldsymbol{\phi}^*). \tag{4.13}$$

On the other hand, it follows from $V_t^*(\boldsymbol{\kappa}) = \theta^*(B) z_t(\boldsymbol{\kappa})$ that

$$\frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa}) = \frac{1}{\phi_{s'}^*} V_{t+s}^*(\boldsymbol{\kappa}) - \frac{1}{\theta_s^*}\left[\left(\frac{\theta_s^*}{\phi_{s'}^*} z_{t+s}(\boldsymbol{\kappa})\right) + \theta_1^*\left(\frac{\theta_s^*}{\phi_{s'}^*} z_{t+s-1}(\boldsymbol{\kappa})\right) + \cdots + \theta_{s-1}^*\left(\frac{\theta_s^*}{\phi_{s'}^*} z_{t+1}(\boldsymbol{\kappa})\right)\right].$$

So,

$$\begin{aligned}\frac{\partial\left(\frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l} &= \frac{1}{\phi_{s'}^*}\frac{\partial V_{t+s}^*(\boldsymbol{\kappa})}{\partial \kappa_l} \\ &\quad -\frac{1}{\theta_s^*}\left[\frac{\partial\left(\frac{\theta_s^*}{\phi_{s'}^*} z_{t+s}(\boldsymbol{\kappa})\right)}{\partial \kappa_l} + \theta_1^*\frac{\partial\left(\frac{\theta_s^*}{\phi_{s'}^*} z_{t+s-1}(\boldsymbol{\kappa})\right)}{\partial \kappa_l} + \cdots + \theta_{s-1}^*\frac{\partial\left(\frac{\theta_s^*}{\phi_{s'}^*} z_{t+1}(\boldsymbol{\kappa})\right)}{\partial \kappa_l}\right].\end{aligned}$$

That is,

$$\theta^*(B)\frac{\partial\left(\frac{\theta_s^*}{\phi_{s'}^*} z_{t+s}(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = \frac{\theta_s^*}{\phi_{s'}^*}\frac{\partial V_{t+s}^*(\boldsymbol{\kappa})}{\partial \kappa_l},$$

or equivalently

$$\theta^*(B)\frac{\partial\left(\frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = \frac{\theta_s^*}{\phi_{s'}^*}\frac{\partial V_t^*(\boldsymbol{\kappa})}{\partial \kappa_l}. \tag{4.14}$$

This together with (4.13) yields

$$\theta^+(B)\,\theta^*(B)\,D_{t,l}(\boldsymbol{\kappa}) = \frac{\theta_s^*}{\phi_{s'}^*} U_{t-l}^*(\boldsymbol{\phi}^*). \tag{4.15}$$

Secondly, for $l = r'+1, \ldots, p$, we consider the equation

$$\phi^*(B)\,U_t^+(\boldsymbol{\phi}^+) = \theta^+(B)\,V_t^*(\boldsymbol{\kappa})$$

in (4.6), where $\boldsymbol{\phi}^+ = \left(\phi_1^+, \ldots, \phi_{s'}^+\right)^T$. By rearranging the terms, we obtain

$$\frac{1}{\phi^*_{s'}} V^*_t(\boldsymbol{\kappa}) = \frac{1}{\phi^*_{s'}} \left[U^+_t(\boldsymbol{\phi}^+) - \phi^*_1 U^+_{t-1}(\boldsymbol{\phi}^+) - \cdots - \phi^*_{s'-1} U^+_{t-s'+1}(\boldsymbol{\phi}^+)\right] - U^+_{t-s'}(\boldsymbol{\phi}^+)$$
$$-\theta^+_1\left(\frac{1}{\phi^*_{s'}} V^*_{t-1}(\boldsymbol{\kappa})\right) - \cdots - \theta^+_r\left(\frac{1}{\phi^*_{s'}} V^*_{t-r}(\boldsymbol{\kappa})\right).$$

Then, for $l = r'+1, \ldots, p-1$, we have

$$\frac{\partial\left(\frac{1}{\phi^*_{s'}} V^*_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = -\frac{1}{\phi^*_{s'}} U^+_{t+r'-l}(\boldsymbol{\phi}^+) - \theta^+_1 \frac{\partial\left(\frac{1}{\phi^*_{s'}} V^*_{t-1}(\boldsymbol{\kappa})\right)}{\partial \kappa_l} - \cdots - \theta^+_r \frac{\partial\left(\frac{1}{\phi^*_{s'}} V^*_{t-r}(\boldsymbol{\kappa})\right)}{\partial \kappa_l},$$

i.e.,

$$\theta^+(B) \frac{\partial\left(\frac{1}{\phi^*_{s'}} V^*_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = -\frac{1}{\phi^*_{s'}} U^+_{t+r'-l}(\boldsymbol{\phi}^+). \tag{4.16}$$

It follows from (4.14) and (4.16) that

$$\theta^+(B)\theta^*(B) \frac{\partial\left(\frac{\theta^*_s}{\phi^*_{s'}} z_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = \theta^*_s \theta^+(B) \frac{\partial\left(\frac{1}{\phi^*_{s'}} V^*_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = -\frac{\theta^*_s}{\phi^*_{s'}} U^+_{t+r'-l}(\boldsymbol{\phi}^+);$$

that is,

$$\theta^+(B)\theta^*(B) D_{t,l}(\boldsymbol{\kappa}) = \frac{\theta^*_s}{\phi^*_{s'}} U^+_{t+r'-l}(\boldsymbol{\phi}^+). \tag{4.17}$$

For $l = p$, we have

$$\frac{\partial\left(\frac{1}{\phi^*_{s'}} V^*_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = -\frac{1}{(\phi^*_{s'})^2}\left[U^+_t(\boldsymbol{\phi}^+) - \phi^*_1 U^+_{t-1}(\boldsymbol{\phi}^+) - \cdots - \phi^*_{s'-1} U^+_{t-s'+1}(\boldsymbol{\phi}^+)\right]$$
$$-\theta^+_1 \frac{\partial\left(\frac{1}{\phi^*_{s'}} V^*_{t-1}(\boldsymbol{\kappa})\right)}{\partial \kappa_l} - \cdots - \theta^+_r \frac{\partial\left(\frac{1}{\phi^*_{s'}} V^*_{t-r}(\boldsymbol{\kappa})\right)}{\partial \kappa_l},$$

or

$$\theta^+(B) \frac{\partial\left(\frac{1}{\phi^*_{s'}} V^*_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = -\frac{1}{(\phi^*_{s'})^2} \phi^*(B) U^+_t(\boldsymbol{\phi}^+) - \frac{1}{\phi^*_{s'}} U^+_{t-s'}(\boldsymbol{\phi}^+). \tag{4.18}$$

It follows from (4.14) and (4.18) that

$$\theta^+(B)\theta^*(B) \frac{\partial\left(\frac{\theta^*_s}{\phi^*_{s'}} z_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = \theta^*_s \theta^+(B) \frac{\partial\left(\frac{1}{\phi^*_{s'}} V^*_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l}$$

$$= -\frac{\theta_s^*}{\left(\phi_{s'}^*\right)^2}\phi^*(B)\,U_t^+\left(\boldsymbol{\phi}^+\right) - \frac{\theta_s^*}{\phi_{s'}^*}U_{t-s'}^+\left(\boldsymbol{\phi}^+\right);$$

that is,

$$\theta^+(B)\,\theta^*(B)\,D_{t,l}(\boldsymbol{\kappa}) = \frac{\theta_s^*}{\left(\phi_{s'}^*\right)^2}\phi^*(B)\,U_t^+\left(\boldsymbol{\phi}^+\right) + \frac{\theta_s^*}{\phi_{s'}^*}U_{t-s'}^+\left(\boldsymbol{\phi}^+\right). \tag{4.19}$$

Next, for $l = p+1, \ldots, p+r$, by considering the equation $\phi(B)\,X_t = \theta^+(B)\,V_t^*(\boldsymbol{\kappa})$ in (4.6) we have

$$\frac{\partial V_t^*(\boldsymbol{\kappa})}{\partial \kappa_l} = -V_{t+p-l}^*(\boldsymbol{\kappa}) - \theta_1^+\frac{\partial V_{t-1}^*(\boldsymbol{\kappa})}{\partial \kappa_l} - \cdots - \theta_r^+\frac{\partial V_{t-r}^*(\boldsymbol{\kappa})}{\partial \kappa_l},$$

i.e.,

$$\theta^+(B)\,\frac{\partial V_t^*(\boldsymbol{\kappa})}{\partial \kappa_l} = -V_{t+p-l}^*(\boldsymbol{\kappa}). \tag{4.20}$$

It follows from (4.14) and (4.20) that

$$\theta^+(B)\,\theta^*(B)\,\frac{\partial\left(\frac{\theta_s^*}{\phi_{s'}^*}z_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = \frac{\theta_s^*}{\phi_{s'}^*}\theta^+(B)\,\frac{\partial\left(V_t^*(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = -\frac{\theta_s^*}{\phi_{s'}^*}V_{t+p-l}^*(\boldsymbol{\kappa}).$$

That is,

$$\theta^+(B)\,\theta^*(B)\,D_{t,l}(\boldsymbol{\kappa}) = \frac{\theta_s^*}{\phi_{s'}^*}V_{t+p-l}^*(\boldsymbol{\kappa}). \tag{4.21}$$

Finally, for $l = p+r+1, \ldots, p+q$, it follows the equation $\phi(B)\,X_t = \theta^*(B)\,V_t^+(\boldsymbol{\kappa})$ that

$$\begin{aligned}\theta_s^* V_t^+(\boldsymbol{\kappa}) &= \phi(B)\,X_{t+s} - V_{t+s}^+(\boldsymbol{\kappa}) - \theta_1^* V_{t+s-1}^+(\boldsymbol{\kappa}) - \cdots - \theta_{s-1}^* V_{t+1}^+(\boldsymbol{\kappa}) \\ &= \phi(B)\,X_{t+s} - \frac{1}{\theta_s^*}\left[\left(\theta_s^* V_{t+s}^+(\boldsymbol{\kappa})\right) + \theta_1^*\left(\theta_s^* V_{t+s-1}^+(\boldsymbol{\kappa})\right) + \cdots + \theta_{s-1}^*\left(\theta_s^* V_{t+1}^+(\boldsymbol{\kappa})\right)\right].\end{aligned}$$

For $l = p+r+1, \ldots, p+q-1$, we have

$$\begin{aligned}\frac{\partial\left(\theta_s^* V_t^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l} &= -V_{t+p+q-l}^+(\boldsymbol{\kappa}) \\ &\quad - \frac{1}{\theta_s^*}\left[\frac{\partial\left(\theta_s^* V_{t+s}^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l} + \theta_1^*\frac{\partial\left(\theta_s^* V_{t+s-1}^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l} + \cdots + \theta_{s-1}^*\frac{\partial\left(\theta_s^* V_{t+1}^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l}\right],\end{aligned}$$

and

$$\theta^*(B)\frac{\partial\left(\theta_s^* V_{t+s}^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = -\theta_s^* V_{t+p+q-l}^+(\boldsymbol{\kappa}).$$

So,

$$\theta^*(B)\frac{\partial\left(\theta_s^* V_t^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = -\theta_s^* V_{t+p+r-l}^+(\boldsymbol{\kappa}). \tag{4.22}$$

On the other hand, it follows from $V_t^+(\boldsymbol{\kappa}) = \theta^+(B)\, z_t(\boldsymbol{\kappa})$ that

$$\frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa}) = \frac{\theta_s^*}{\phi_{s'}^*} V_t^+(\boldsymbol{\kappa}) - \left[\theta_1^+\left(\frac{\theta_s^*}{\phi_{s'}^*} z_{t-1}(\boldsymbol{\kappa})\right) + \cdots + \theta_r^+\left(\frac{\theta_s^*}{\phi_{s'}^*} z_{t-r}(\boldsymbol{\kappa})\right)\right],$$

and hence

$$\frac{\partial\left(\frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = \frac{1}{\phi_{s'}^*}\frac{\partial\left(\theta_s^* V_t^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l} - \left[\theta_1^+\frac{\partial\left(\frac{\theta_s^*}{\phi_{s'}^*} z_{t-1}(\boldsymbol{\kappa})\right)}{\partial \kappa_l} + \cdots + \theta_r^+\frac{\partial\left(\frac{\theta_s^*}{\phi_{s'}^*} z_{t-r}(\boldsymbol{\kappa})\right)}{\partial \kappa_l}\right].$$

That is,

$$\theta^+(B)\frac{\partial\left(\frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = \frac{1}{\phi_{s'}^*}\frac{\partial\left(\theta_s^* V_t^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l}. \tag{4.23}$$

It follows from (4.22) and (4.23) that

$$\theta^+(B)\,\theta^*(B)\frac{\partial\left(\frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = \frac{1}{\phi_{s'}^*}\theta^*(B)\frac{\partial\left(\theta_s^* V_t^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = -\frac{\theta_s^*}{\phi_{s'}^*} V_{t+p+r-l}^+(\boldsymbol{\kappa}),$$

or

$$\theta^+(B)\,\theta^*(B)\, D_{t,l}(\boldsymbol{\kappa}) = \frac{\theta_s^*}{\phi_{s'}^*} V_{t+p+r-l}^+(\boldsymbol{\kappa}). \tag{4.24}$$

For $l = p+q$, we have

$$\begin{aligned}\frac{\partial\left(\theta_s^* V_t^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l} &= \frac{1}{\theta_s^*}\left[V_{t+s}^+(\boldsymbol{\kappa}) + \theta_1^* V_{t+s-1}^+(\boldsymbol{\kappa}) + \cdots + \theta_{s-1}^* V_{t+1}^+(\boldsymbol{\kappa})\right] \\ &\quad -\frac{1}{\theta_s^*}\left[\frac{\partial\left(\theta_s^* V_{t+s}^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l} + \theta_1^*\frac{\partial\left(\theta_s^* V_{t+s-1}^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l} + \cdots + \theta_{s-1}^*\frac{\partial\left(\theta_s^* V_{t+1}^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l}\right],\end{aligned}$$

i.e.,

$$\theta^*(B)\frac{\partial\left(\theta_s^* V_{t+s}^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = \theta^*(B)\, V_{t+s}^+(\boldsymbol{\kappa}) - \theta_s^* V_t^+(\boldsymbol{\kappa}).$$

So,

$$\theta^*(B)\frac{\partial\left(\theta_s^* V_t^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = \theta^*(B) V_t^+(\boldsymbol{\kappa}) - \theta_s^* V_{t-s}^+(\boldsymbol{\kappa}). \tag{4.25}$$

It follows from (4.25) and (4.23) that

$$\theta^+(B)\theta^*(B)\frac{\partial\left(\frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = \frac{1}{\phi_{s'}^*}\theta^*(B)\frac{\partial\left(\theta_s^* V_t^+(\boldsymbol{\kappa})\right)}{\partial \kappa_l} = \frac{1}{\phi_{s'}^*}\theta^*(B) V_t^+(\boldsymbol{\kappa}) - \frac{\theta_s^*}{\phi_{s'}^*} V_{t-s}^+(\boldsymbol{\kappa}),$$

or

$$\theta^+(B)\theta^*(B) D_{t,l}(\boldsymbol{\kappa}) = -\frac{1}{\phi_{s'}^*}\theta^*(B) V_t^+(\boldsymbol{\kappa}) + \frac{\theta_s^*}{\phi_{s'}^*} V_{t-s}^+(\boldsymbol{\kappa}). \tag{4.26}$$

In summary, for $t = 1, \ldots, n+p-q$, and $l = 1, \ldots, p+q$, $D_{t,l}(\boldsymbol{\kappa})$ satisfies the following difference equations

$$\begin{aligned}
&\theta^+(B)\theta^*(B) D_{t,l}(\boldsymbol{\kappa}) = \tfrac{\theta_s^*}{\phi_{s'}^*} U_{t-l}^*(\boldsymbol{\phi}^*), && \text{for } l = 1, \ldots, r', \\
&\theta^+(B)\theta^*(B) D_{t,l}(\boldsymbol{\kappa}) = \tfrac{\theta_s^*}{\phi_{s'}^*} U_{t+r'-l}^+(\boldsymbol{\phi}^+), && \text{for } l = r'+1, \ldots, p-1, \\
&\theta^+(B)\theta^*(B) D_{t,l}(\boldsymbol{\kappa}) = \tfrac{\theta_s^*}{(\phi_{s'}^*)^2}\phi^*(B) U_t^+(\boldsymbol{\phi}^+) + \tfrac{\theta_s^*}{\phi_{s'}^*} U_{t-s'}^+(\boldsymbol{\phi}^+), && \text{for } l = p, \\
&\theta^+(B)\theta^*(B) D_{t,l}(\boldsymbol{\kappa}) = \tfrac{\theta_s^*}{\phi_{s'}^*} V_{t+p-l}^*(\boldsymbol{\kappa}), && \text{for } l = p+1, \ldots, p+r, \\
&\theta^+(B)\theta^*(B) D_{t,l}(\boldsymbol{\kappa}) = \tfrac{\theta_s^*}{\phi_{s'}^*} V_{t+p+r-l}^+(\boldsymbol{\kappa}), && \text{for } l = p+r+1, \ldots, p+q-1, \\
&\theta^+(B)\theta^*(B) D_{t,l}(\boldsymbol{\kappa}) = -\tfrac{1}{\phi_{s'}^*}\theta^*(B) V_t^+(\boldsymbol{\kappa}) + \tfrac{\theta_s^*}{\phi_{s'}^*} V_{t-s}^+(\boldsymbol{\kappa}), && \text{for } l = p+q.
\end{aligned}$$

In particular, $D_{t,l}(\boldsymbol{\kappa}_0)$ is well approximated by

$$\begin{cases}
\frac{\theta_{0s}^*}{\phi_{0s'}^*} W_{1,t-l}, & \text{for } l = 1, \ldots, r', \\
\frac{\theta_{0s}^*}{\phi_{0s'}^*} W_{2,t-l}, & \text{for } l = r'+1, \ldots, p-1, \\
\frac{\theta_{0s}^*}{\phi_{0s'}^*}\left(W_{2,t-l} + \frac{1}{\phi_{0s'}^*} Z_t\right), & \text{for } l = p, \\
\frac{\theta_{0s}^*}{\phi_{0s'}^*} W_{3,t-l}, & \text{for } l = p+1, \ldots, p+r, \\
\frac{\theta_{0s}^*}{\phi_{0s'}^*} W_{4,t-l}, & \text{for } l = p+r+1, \ldots, p+q-1, \\
\frac{\theta_{0s}^*}{\phi_{0s'}^*}\left(W_{4,t-l} - \frac{1}{\theta_{0s}^*} Z_t\right), & \text{for } l = p+q,
\end{cases}$$

where $W_{j,t}$, $j = 1, 2, 3, 4$, are AR processes defined as follows.

$$
\begin{aligned}
\phi_0^+(B)\,W_{1,t} &= Z_t,\\
\phi_0^*(B)\,W_{2,t} &= Z_{t+r'},\\
\theta_0^+(B)\,W_{3,t} &= Z_{t+p},\\
\theta_0^*(B)\,W_{4,t} &= Z_{t+p+r},
\end{aligned}
\tag{4.27}
$$

where $\phi_0^+(z)$ and $\phi_0^*(z)$ are the causal and purely noncausal components of $\phi_0(z)$, and $\theta_0^+(z)$ and $\theta_0^*(z)$ are the invertible and purely noninvertible components of $\theta_0(z)$. We can express the reciprocals of these polynomials by

$$
\begin{aligned}
\tfrac{1}{\phi_0^+(z)} &= \textstyle\sum_{i=0}^{\infty}\beta_i^+ z^i, & \tfrac{1}{\phi_0^*(z)} &= \textstyle\sum_{j=s'}^{\infty}\beta_j^* z^{-j},\\
\tfrac{1}{\theta_0^+(z)} &= \textstyle\sum_{i=0}^{\infty}\alpha_i^+ z^i, & \tfrac{1}{\theta_0^*(z)} &= \textstyle\sum_{j=s}^{\infty}\alpha_j^* z^{-j},
\end{aligned}
\tag{4.28}
$$

where $\beta_0^+ = \alpha_0^+ = 1$, $\beta_{s'}^* = -1/\phi_{0s'}^*$, and $\alpha_s^* = 1/\theta_{0s}^*$, which are easy to check; for example, $\beta_{s'}^* = -1/\phi_{0s'}^*$ is obtained from

$$
1 = \phi_0^*(z)\cdot\frac{1}{\phi_0^*(z)} = \left(1-\phi_{01}^* z-\cdots-\phi_{0s'}^* z^{s'}\right)\sum_{j=s'}^{\infty}\beta_j^* z^{-j}.
$$

Now we define $\widetilde{\mathbf{Q}}_t \equiv \left(\mathbf{W}_1^T,\mathbf{W}_2^T,\mathbf{W}_3^T,\mathbf{W}_4^T\right)^T$, where

$$
\begin{aligned}
\mathbf{W}_1 &= \left(W_{1,t-1},\ldots,W_{1,t-r'}\right)^T,\\
\mathbf{W}_2 &= \left(W_{2,t-r'-1},\ldots,W_{2,t-p+1},\left(W_{2,t-p}+Z_t/\phi_{0s'}^*\right)\right)^T,\\
\mathbf{W}_3 &= \left(W_{3,t-p-1},\ldots,W_{3,t-p-r}\right)^T,\\
\mathbf{W}_4 &= \left(W_{4,t-p-r-1},\ldots,W_{4,t-p-q+1},\left(W_{4,t-p-q}-Z_t/\theta_{0s}^*\right)\right)^T.
\end{aligned}
$$

Further, let $\mathbf{Q}_t = \frac{\theta_{0s}^*}{\phi_{0s'}^*}\widetilde{\mathbf{Q}}_t$. Then $\mathbf{D}_t(\boldsymbol{\kappa}_0)$ is well approximated by $\mathbf{Q}_t$. It is easy to see from (4.27) and (4.28) that $\mathbf{W}_1,\mathbf{W}_3 \in \sigma(Z_{t-1},Z_{t-2},\ldots)$ and $\mathbf{W}_2,\mathbf{W}_4 \in \sigma(Z_{t+1},Z_{t+2},\ldots)$. It follows that $\mathbf{Q}_t$ (or $\widetilde{\mathbf{Q}}_t$) is independent of Z_t, and moreover $\mathbf{W}_1$ and $\mathbf{W}_3$ are independent of $\mathbf{W}_2$ and $\mathbf{W}_4$.

Let $\mathbf{\Gamma}_{\widetilde{\mathbf{Q}}} = \mathrm{Cov}\left(\widetilde{\mathbf{Q}}_t\right) = [\widetilde{\sigma}_{l,m}]_{l,m=1}^{p+q}\mathrm{Var}(Z_1)$, then $\widetilde{\sigma}_{l,m}$ is given by

$$
\widetilde{\sigma}_{l,m} = \begin{cases}
\sum_{j=0}^{\infty} \beta_j^+ \beta_{j+|l-m|}^+, & \text{if } l,m = 1,\ldots,r', \\
\sum_{j=0}^{\infty} \beta_j^+ \alpha_{j+p-|l-m|}^+, & \text{if } l = 1,\ldots,r'; m = p+1,\ldots,p+r, \\
& \quad \text{or } l = p+1,\ldots,p+r; m = 1,\ldots,r', \\
\sum_{j=s'}^{\infty} \beta_j^* \beta_{j+|l-m|}^*, & \text{if } l,m = r'+1,\ldots,p, \text{ but } (l,m) \neq (p,p), \\
\sum_{j=s'+1}^{\infty} \left(\beta_j^*\right)^2, & \text{if } (l,m) = (p,p), \\
\sum_{j=s'}^{\infty} \beta_j^* \alpha_{j-s'-r+|l-m|}^*, & \text{if } l = r'+1,\ldots,p; m = p+r+1,\ldots,p+q, \\
& \quad \text{but } (l,m) \neq (p,p+q), \\
& \text{or } l = p+r+1,\ldots,p+q; m = r'+1,\ldots,p, \\
& \quad \text{but } (l,m) \neq (p+q,p), \\
\sum_{j=s'+1}^{\infty} \beta_j^* \alpha_{j+s-s'}^*, & \text{if } (l,m) = (p,p+q) \text{ or } (p+q,p), \\
\sum_{j=0}^{\infty} \alpha_j^+ \alpha_{j+|l-m|}^+, & \text{if } l,m = p+1,\ldots,p+r, \\
\sum_{j=s}^{\infty} \alpha_j^* \alpha_{j+|l-m|}^*, & \text{if } l,m = p+r+1,\ldots,p+q, \\
& \quad \text{but } (l,m) \neq (p+q,p+q), \\
\sum_{j=s+1}^{\infty} \left(\alpha_j^*\right)^2, & \text{if } (l,m) = (p+q,p+q), \\
0, & \text{otherwise,}
\end{cases}
$$

where $\alpha_j^+ = \beta_j^+ = 0$ for $j < 0$, $\alpha_j^* = 0$ for $j < s$, and $\beta_j^* = 0$ for $j < s'$.

In addition, further derivations show that, $\mathbf{H}_t(\boldsymbol{\kappa}_0)$ is well approximated by

$$
\mathbf{R}_t \equiv \left[\sum_{j=-\infty}^{\infty} h_{j,lm} Z_{t-j} \right]_{l,m=1}^{p+q},
$$

where $\{h_{j,lm}\}$ is a sequence of constants such that $\sum_{j=-\infty}^{\infty} |h_{j,lm}| < \infty$ for each (j,m) pair. Notice that, each element of matrix $\mathbf{R}_t$ is a linear combination of all Z_t's. But, as we will see in the next section, among these constants $h_{j,lm}$ we only need to deal with the ones associated with Z_t. Matrices $\mathbf{\Psi} \equiv [h_{0,lm}]_{l,m=1}^{p+q}$ and $\widetilde{\mathbf{\Psi}} \equiv \left[\widetilde{\psi}_{l,m}\right]_{l,m=1}^{p+q} = -\frac{\phi_{0s'}^*}{\theta_{0s}^*}\mathbf{\Psi}$ are defined for this purpose. That is, instead of evaluating $\mathbf{R}_t$ explicitly, we only need $\mathbf{\Psi}$ (or $\widetilde{\mathbf{\Psi}}$) to establish asymptotic results of the LAD estimator. The element $\widetilde{\psi}_{l,m}$ is given by

$$\widetilde{\psi}_{l,m} = \begin{cases} \sum_{i=0}^{\infty} \beta_i^+ \beta_{l+m-r'+i}^*, & \text{if } l = 1, \ldots, r'; m = r'+1, \ldots, p, \\ & \quad \text{or } l = r'+1, \ldots, p; m = 1, \ldots, r', \\ \sum_{i=0}^{\infty} \beta_i^+ \alpha_{l+m-p-r+i}^*, & \text{if } l = 1, \ldots, r'; m = p+r+1, \ldots, p+q, \\ & \quad \text{or } l = p+r+1, \ldots, p+q; m = 1, \ldots, r', \\ \sum_{i=0}^{\infty} \alpha_i^+ \beta_{l+m-p-r'+i}^*, & \text{if } l = r'+1, \ldots, p; m = p+1, \ldots, p+r, \\ & \quad \text{or } l = p+1, \ldots, p+r; m = r'+1, \ldots, p, \\ \sum_{i=0}^{\infty} \alpha_i^+ \alpha_{l+m-2p-r+i}^*, & \text{if } l = p+1, \ldots, p+r; m = p+r+1, \ldots, p+q, \\ & \quad \text{or } l = p+r+1, \ldots, p+q; m = p+1, \ldots, p+r, \\ 0, & \text{otherwise.} \end{cases}$$

4.3 Asymptotic Properties of LAD Estimator

Now we assume that the underlying noise $\{Z_t\}$ is from a scale family with probability density function $f_Z(z) = \frac{1}{\sigma} f\left(\frac{z}{\sigma}\right)$, where $\sigma > 0$ and $f(\cdot)$ is chosen such that $\int_{-\infty}^{\infty} z^2 f(z) dz = 1$ and hence $\mathrm{Var}(Z_t) = \sigma^2$. We assume that $\mathrm{E}(Z_t) = 0$, $\mathrm{E}[\mathrm{sgn}(Z_t)] = 0$, and $\mathrm{E}(Z_t^2) < \infty$.

In order to establish asymptotic results of the LAD estimator $\widehat{\boldsymbol{\kappa}}_{\mathrm{LAD}}$, we build sample size into a new parameterization, namely $\boldsymbol{\kappa} = \boldsymbol{\kappa}_0 + \boldsymbol{\nu}/\sqrt{n}$. Then, under this parameterization, minimizing (4.12) with respect to $\boldsymbol{\kappa}$ is equivalent to minimizing

$$S_n(\boldsymbol{\nu}) = \sum_{t=1}^{n+p-q} \left(\left| \frac{\theta_s^*}{\phi_{s'}^*} z_t \left(\boldsymbol{\kappa}_0 + \boldsymbol{\nu}/\sqrt{n}\right) \right| - \left| \frac{\theta_{0s}^*}{\phi_{0s'}^*} z_t (\boldsymbol{\kappa}_0) \right| \right)$$

with respect to $\boldsymbol{\nu}$. We first prove a functional limit result for $S_n(\boldsymbol{\nu})$, which states the convergence of $S_n(\boldsymbol{\nu})$ in distribution on $C(\mathbb{R}^{p+q})$ to a limit process $S(\boldsymbol{\nu})$ that is a quadratic function in $\boldsymbol{\nu}$. Let $\widehat{\boldsymbol{\nu}}_{\mathrm{LAD}}$ denote any minimizer of $S_n(\boldsymbol{\nu})$. Then, it follows from Remark 1 of Davis, Knight, and Liu (1992) that there exists a sequence of minimizers $\widehat{\boldsymbol{\nu}}_{\mathrm{LAD}}$ that converges in distribution to the unique minimizer of $S(\boldsymbol{\nu})$. But, the weak convergence of $\widehat{\boldsymbol{\nu}}_{\mathrm{LAD}}$ is equivalent to that of $\sqrt{n}(\widehat{\boldsymbol{\kappa}}_{\mathrm{LAD}} - \boldsymbol{\kappa}_0)$. So, asymptotic results of $\widehat{\boldsymbol{\kappa}}_{\mathrm{LAD}}$ are readily obtained.

We first state a result to be used in the derivation, which is a modified Lemma 1 of Breidt, Davis, and Trindade (2001).

Proposition 4.1 *(Andrews, 2003) Suppose $\{Y_t\}$ and $\{V_t\}$ are linear processes given by*

$$Y_t = \sum_{j=-\infty}^{\infty} c_j Z_{t-j} \qquad \text{and} \qquad V_t = \sum_{j=-\infty}^{\infty} d_j Z_{t-j},$$

where $c_0 = 0$, $\sum_{j=-\infty}^{\infty} |c_j| < \infty$, $\sum_{j=-\infty}^{\infty} |d_j| < \infty$, and $\{Z_t\} \sim IID(0, \tau^2)$ with median zero and common distribution function F_Z that is continuously differentiable in a neighborhood of zero. Then

$$S_n \equiv \sum_{t=1}^{n} \left(\left| Z_t - Y_t/\sqrt{n} - V_t/n \right| - |Z_t| \right) \xrightarrow{d} Var(Y_t)\, f_Z(0) - d_0 E|Z_t| + N,$$

where $Var(Y_t) = \sum_{j=-\infty}^{\infty} c_j^2 \tau^2$, f_Z is the density function corresponding to F_Z, and N is a random variable such that

$$\begin{aligned} N &\sim N\left(0, \gamma^*(0) + 2\sum_{h=1}^{\infty} \gamma^*(h)\right), \\ \gamma^*(h) &= E\left[Y_t sgn(Z_t)\, Y_{t+h} sgn(Z_{t+h})\right]. \end{aligned}$$

The functional limit result for $S_n(\boldsymbol{\nu})$ is proved via several steps.

Lemma 4.2 *For $\boldsymbol{\nu} \in \mathbb{R}^{p+q}$, define*

$$S_n^{\ddagger}(\boldsymbol{\nu}) = \sum_{t=1}^{n+p-q} \left(\left| \frac{\theta_{0s}^*}{\phi_{0s'}^*} Z_t - \frac{\mathbf{Q}_t^T \boldsymbol{\nu}}{\sqrt{n}} - \frac{\boldsymbol{\nu}^T \mathbf{R}_t \boldsymbol{\nu}}{2n} \right| - \left| \frac{\theta_{0s}^*}{\phi_{0s'}^*} Z_t \right| \right).$$

Then $S_n^{\ddagger}(\boldsymbol{\nu}) \xrightarrow{d} S(\boldsymbol{\nu})$ on $C(\mathbb{R}^{p+q})$ where

$$S(\boldsymbol{\nu}) = \left| \frac{\theta_{0s}^*}{\phi_{0s'}^*} \right| \left[\boldsymbol{\nu}^T \left(\boldsymbol{\Gamma}_{\tilde{\mathbf{Q}}} f_Z(0) + \frac{1}{2} \tilde{\boldsymbol{\Psi}} E|Z_1| \right) \boldsymbol{\nu} + \boldsymbol{\nu}^T \mathbf{N} \right],$$

with $\mathbf{N} \sim N\left(\mathbf{0}, \mathbf{\Gamma}_{\widetilde{\mathbf{Q}}} + 2\sum_{h=1}^{\infty} \mathbf{\Sigma}(h)\right)$ *and* $\mathbf{\Sigma}(h) \equiv E\left[\widetilde{\mathbf{Q}}_t \widetilde{\mathbf{Q}}_{t+h}^T sgn(Z_t)\, sgn(Z_{t+h})\right]$.

Proof. Note that

$$S_n^{\ddagger}(\boldsymbol{\nu}) = \left|\frac{\theta_{0s}^*}{\phi_{0s'}^*}\right| \sum_{t=1}^{n+p-q} \left(\left| Z_t - \frac{\widetilde{\mathbf{Q}}_t^T \boldsymbol{\nu}}{\sqrt{n}} - \frac{1}{2n} \boldsymbol{\nu}^T \left(\frac{\phi_{0s'}^*}{\theta_{0s}^*} \mathbf{R}_t \right) \boldsymbol{\nu} \right| - |Z_t| \right).$$

With $Y_t = \widetilde{\mathbf{Q}}_t^T \boldsymbol{\nu}$ and $V_t = \frac{1}{2}\boldsymbol{\nu}^T \left(\frac{\phi_{0s'}^*}{\theta_{0s}^*}\mathbf{R}_t\right)\boldsymbol{\nu}$, we have

$$\begin{aligned}
\operatorname{Var}(Y_t) &= \operatorname{Var}\left(\widetilde{\mathbf{Q}}_t^T \boldsymbol{\nu}\right) = \boldsymbol{\nu}^T \operatorname{Cov}\left(\widetilde{\mathbf{Q}}_t\right)\boldsymbol{\nu} = \boldsymbol{\nu}^T \mathbf{\Gamma}_{\widetilde{\mathbf{Q}}} \boldsymbol{\nu}, \\
d_0 &= \frac{1}{2}\boldsymbol{\nu}^T \left(\frac{\phi_{0s'}^*}{\theta_{0s}^*}\mathbf{\Psi}\right)\boldsymbol{\nu} = -\frac{1}{2}\boldsymbol{\nu}^T \widetilde{\mathbf{\Psi}} \boldsymbol{\nu},
\end{aligned}$$

and

$$\gamma^*(h) = \mathrm{E}\left[Y_t \operatorname{sgn}(Z_t)\, Y_{t+h} \operatorname{sgn}(Z_{t+h})\right] = \boldsymbol{\nu}^T \mathrm{E}\left[\widetilde{\mathbf{Q}}_t \widetilde{\mathbf{Q}}_{t+h}^T \operatorname{sgn}(Z_t) \operatorname{sgn}(Z_{t+h})\right]\boldsymbol{\nu} = \boldsymbol{\nu}^T \mathbf{\Sigma}(h)\boldsymbol{\nu},$$

for $h = 0, \pm 1, \ldots$. Now applying Proposition 4.1 to

$$\sum_{t=1}^{n+p-q} \left(\left| Z_t - \frac{\widetilde{\mathbf{Q}}_t^T \boldsymbol{\nu}}{\sqrt{n}} - \frac{1}{2n} \boldsymbol{\nu}^T \left(\frac{\phi_{0s'}^*}{\theta_{0s}^*} \mathbf{R}_t \right) \boldsymbol{\nu} \right| - |Z_t| \right),$$

we obtain

$$\begin{aligned}
&\sum_{t=1}^{n+p-q} \left(\left| Z_t - \tfrac{\widetilde{\mathbf{Q}}_t^T \boldsymbol{\nu}}{\sqrt{n}} - \tfrac{1}{2n} \boldsymbol{\nu}^T \left(\tfrac{\phi_{0s'}^*}{\theta_{0s}^*} \mathbf{R}_t \right) \boldsymbol{\nu} \right| - |Z_t| \right) \\
&\quad \xrightarrow{d} \boldsymbol{\nu}^T \left(\mathbf{\Gamma}_{\widetilde{\mathbf{Q}}} f_Z(0) + \tfrac{1}{2}\widetilde{\mathbf{\Psi}} \mathrm{E}|Z_1| \right)\boldsymbol{\nu} + \boldsymbol{\nu}^T \mathrm{N}\left(\mathbf{0}, \mathbf{\Sigma}(0) + 2\sum_{h=1}^{\infty}\mathbf{\Sigma}(h)\right).
\end{aligned} \tag{4.29}$$

Note that $\mathbf{\Sigma}(0) = \mathbf{\Gamma}_{\widetilde{\mathbf{Q}}}$. Then it follows from (4.29) that, for any given $\boldsymbol{\nu}$,

$$S_n^{\ddagger}(\boldsymbol{\nu}) \xrightarrow{d} S(\boldsymbol{\nu}).$$

Since $S_n^{\ddagger}$ has convex sample paths, the weak convergence is in fact on $C(\mathbb{R}^{p+q})$ (Davis, Knight, and Liu, 1992). ■

REMARK 4.1. Let $\mathbf{\Sigma}(h) = [\sigma_{l,m}(h)]_{l,m=1}^{p+q} \mathrm{E}^2 |Z_1|$, then, for $h > 0$,

$$\sigma_{l,m}(h) = \begin{cases} \beta^*_{h-r'+l}\beta^+_{h-m}, & \text{if } l = r'+1,\dots,p; m = 1,\dots,r', \\ \beta^*_{h-r'+l}\alpha^+_{h+p-m}, & \text{if } l = r'+1,\dots,p; m = p+1,\dots,p+r, \\ \alpha^*_{h-p-r+l}\beta^+_{h-m}, & \text{if } l = p+r+1,\dots,p+q; m = 1,\dots,r', \\ \alpha^*_{h-p-r+l}\alpha^+_{h+p-m}, & \text{if } l = p+r+1,\dots,p+q; m = p+1,\dots,p+r, \\ 0, & \text{otherwise.} \end{cases}$$

Lemma 4.3 *For $\boldsymbol{\nu} \in \mathbb{R}^{p+q}$, define*

$$S_n^{\dagger}(\boldsymbol{\nu}) = \sum_{t=1}^{n+p-q} \left(\left| \frac{\theta^*_{0s}}{\phi^*_{0s'}} z_t(\boldsymbol{\kappa}_0) - \frac{\mathbf{D}_t^T(\boldsymbol{\kappa}_0)\boldsymbol{\nu}}{\sqrt{n}} - \frac{\boldsymbol{\nu}^T \mathbf{H}_t(\boldsymbol{\kappa}_0)\boldsymbol{\nu}}{2n} \right| - \left| \frac{\theta^*_{0s}}{\phi^*_{0s'}} z_t(\boldsymbol{\kappa}_0) \right| \right).$$

Then,

1. *$S_n^{\dagger}(\boldsymbol{\nu}) - S_n^{\ddagger}(\boldsymbol{\nu}) \xrightarrow{P} 0$ on $C(\mathbb{R}^{p+q})$, and*

2. *$S_n^{\dagger}(\boldsymbol{\nu}) \xrightarrow{d} S(\boldsymbol{\nu})$ on $C(\mathbb{R}^{p+q})$.*

Proof. 1. Note that, when n is large enough,

$$Z_t = \sum_{i=0}^{\infty} \alpha_i^+ V_{t-i}^+ \qquad \text{and} \qquad z_t(\boldsymbol{\kappa}_0) = \sum_{i=0}^{t-1} \alpha_i^+ V_{t-i}^+(\boldsymbol{\kappa}_0), \tag{4.30}$$

for $t = 1,\dots,n+p-q$, and

$$V_t^+ = \sum_{j=0}^{\infty} \alpha^*_{s+j} U_{t+s+j} \qquad \text{and} \qquad V_t^+(\boldsymbol{\kappa}_0) = \sum_{j=0}^{n+p-s-t} \alpha^*_{s+j} U_{t+s+j}, \tag{4.31}$$

for $t = n+p-q,\dots,\max(p-s,0)+1$, where $U_t = \phi_0(B) X_t$. Therefore, for $t = \max(p-s,0)+1,\dots,n+p-q$,

$$|Z_t - z_t(\boldsymbol{\kappa}_0)| = \left| \sum_{i=0}^{t-1} \alpha_i^+ \left(V_{t-i}^+ - V_{t-i}^+(\boldsymbol{\kappa}_0)\right) + \sum_{i=t}^{\infty} \alpha_i^+ V_{t-i}^+ \right|$$

$$
\begin{aligned}
&\le \left|\sum_{i=0}^{t-1} \alpha_i^+ \left(V_{t-i}^+ - V_{t-i}^+(\boldsymbol{\kappa}_0)\right)\right| + \left|\sum_{i=t}^{\infty} \alpha_i^+ V_{t-i}^+\right| \\
&= \left|\sum_{i=0}^{t-1} \alpha_i^+ \sum_{j=n+p-s-(t-i)+1}^{\infty} \alpha_{s+j}^* U_{t-i+s+j}\right| + \left|\sum_{i=t}^{\infty} \alpha_i^+ V_{t-i}^+\right|,
\end{aligned}
$$

and hence

$$
\limsup_{n\to\infty} \mathrm{E} \sum_{t=M}^{n+p-q} |Z_t - z_t(\boldsymbol{\kappa}_0)| \le \text{Const} \sum_{t=M}^{\infty}\sum_{i=t}^{\infty} \left|\alpha_i^+\right| \to 0, \qquad \text{as } M \to \infty.
$$

One can see that similar results hold for $\mathrm{E}\Big(\limsup_{n\to\infty} \sum_{t=M}^{n+p-q} |Q_{t,l} - D_{t,l}(\boldsymbol{\kappa}_0)|\Big)$ and $\mathrm{E}\Big(\limsup_{n\to\infty} \sum_{t=M}^{n+p-q} |R_{t,lm} - H_{t,lm}(\boldsymbol{\kappa}_0)|\Big)$ for $l, m = 1, \ldots, p+q$, where $Q_{t,l}$ is the l^{th} component of $\mathbf{Q}_t$, and $R_{t,lm}$ is the (l,m) element of $\mathbf{R}_t$. It follows from the triangle inequality as well as these results that $S_n^{\dagger}(\boldsymbol{\nu}) - S_n^{\ddagger}(\boldsymbol{\nu}) \to 0$ in probability on $C(\mathbb{R}^{p+q})$.

2. The result follows from Part 1 together with Lemma 4.2. ■

Theorem 4.4 *$S_n(\boldsymbol{\nu}) \xrightarrow{d} S(\boldsymbol{\nu})$ on $C(\mathbb{R}^{p+q})$.*

Proof. We have

$$
\begin{aligned}
S_n(\boldsymbol{\nu}) &= \sum_{t=1}^{n+p-q} \left(\left|\frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa}_0 + \boldsymbol{\nu}/\sqrt{n})\right| - \left|\frac{\theta_{0s}^*}{\phi_{0s'}^*} z_t(\boldsymbol{\kappa}_0)\right| \right) \\
&= \sum_{t=1}^{n+p-q} \left(\left|\frac{\theta_s^*}{\phi_{s'}^*} z_t(\boldsymbol{\kappa}_0 + \boldsymbol{\nu}/\sqrt{n})\right| - \left|\frac{\theta_{0s}^*}{\phi_{0s'}^*} z_t(\boldsymbol{\kappa}_0) - \frac{\mathbf{D}_t^T(\boldsymbol{\kappa}_0)\boldsymbol{\nu}}{\sqrt{n}} - \frac{\boldsymbol{\nu}^T \mathbf{H}_t(\boldsymbol{\kappa}_0)\boldsymbol{\nu}}{2n}\right| \right) + S_n^{\dagger}(\boldsymbol{\nu}).
\end{aligned}
$$

It follows that

$$
\left|S_n(\boldsymbol{\nu}) - S_n^{\dagger}(\boldsymbol{\nu})\right| \le \frac{1}{2n} \sum_{t=1}^{n+p-q} \left|\boldsymbol{\nu}^T \left[\mathbf{H}_t(\boldsymbol{\kappa}_0) - \mathbf{H}_t\left(\boldsymbol{\kappa}_t^{\dagger}\right)\right] \boldsymbol{\nu}\right|,
$$

where $\boldsymbol{\kappa}_t^{\dagger}$ is between $\boldsymbol{\kappa}_0$ and $\boldsymbol{\kappa}_0 + \boldsymbol{\nu}/\sqrt{n}$. Since $\boldsymbol{\kappa}_t^{\dagger} \to \boldsymbol{\kappa}_0$, the term on the right-hand side converges to zero in probability on $C(\mathbb{R}^{p+q})$. Therefore, $S_n(\boldsymbol{\nu})$ have the same limit as $S_n^{\dagger}(\boldsymbol{\nu})$. That is, $S_n(\boldsymbol{\nu}) \xrightarrow{d} S(\boldsymbol{\nu})$ on $C(\mathbb{R}^{p+q})$. ■

Note that, the limit process $S(\boldsymbol{\nu})$ is a quadratic function in $\boldsymbol{\nu}$ and hence has a unique minimizer $-\left(2\boldsymbol{\Gamma}_{\tilde{\mathbf{Q}}} f_Z(0) + \widetilde{\boldsymbol{\Psi}} \mathrm{E}|Z_1|\right)^{-1} \mathbf{N}$, which is a normal random vector. Then, asymptotic results of $\widehat{\boldsymbol{\kappa}}_{\text{LAD}}$ immediately follow from Remark 1 of Davis, Knight, and Liu (1992).

Theorem 4.5 *There exists a sequence of local LAD estimators* $\widehat{\boldsymbol{\kappa}}_{LAD}$ *such that*

$$\sqrt{n}\left(\widehat{\boldsymbol{\kappa}}_{LAD}-\boldsymbol{\kappa}_0\right) \xrightarrow{d} -\left(2\boldsymbol{\Gamma}_{\tilde{\mathbf{Q}}} f_Z(0)+\widetilde{\boldsymbol{\Psi}} E|Z_1|\right)^{-1} \mathbf{N}, \tag{4.32}$$

where $\mathbf{N} \sim N\left(\mathbf{0}, \boldsymbol{\Gamma}_{\tilde{\mathbf{Q}}}+2 \sum_{h=1}^{\infty} \boldsymbol{\Sigma}(h)\right)$.

Remark 4.2. The asymptotic results are also valid when the LAD objective function depends on s' and s. This is because that, under the parameterization $\boldsymbol{\kappa}=\boldsymbol{\kappa}_0+\boldsymbol{\nu}/\sqrt{n}$, $\boldsymbol{\kappa}$ converges to $\boldsymbol{\kappa}_0$ and hence s' and s converge to their corresponding true value.

The asymptotic normality of the LAD estimator $\widehat{\boldsymbol{\lambda}}_{\text{LAD}}$ of the original parameter vector $\boldsymbol{\lambda}=(\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q)^T$ can be obtained from that of $\widehat{\boldsymbol{\kappa}}_{\text{LAD}}$ by standard techniques.

There exists a sequence of local LAD estimators $\widehat{\boldsymbol{\lambda}}_{\text{LAD}}$ such that

$$\sqrt{n}\left(\widehat{\boldsymbol{\lambda}}_{\text{LAD}}-\boldsymbol{\lambda}_0\right) \xrightarrow{d} -\mathbf{R}\left(2\boldsymbol{\Gamma}_{\tilde{\mathbf{Q}}} f_Z(0)+\widetilde{\boldsymbol{\Psi}} \mathrm{E}|Z_1|\right)^{-1} \mathbf{N},$$

where $\mathbf{N} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Gamma}_{\tilde{\mathbf{Q}}}+2 \sum_{h=1}^{\infty} \boldsymbol{\Sigma}(h)\right)$ and $\mathbf{R} \equiv \left(\frac{\partial \lambda_j}{\partial \kappa_i}\right)_{j,i=1}^{p+q} |_{\boldsymbol{\kappa}=\boldsymbol{\kappa}_0}$ is computed from the transformation between $\boldsymbol{\kappa}$ and $\boldsymbol{\lambda}$. To be specific, since $\phi(z)=\phi^{+}(z)\phi^{*}(z)$, or

$$1-\phi_1 z-\cdots-\phi_p z^p=\left(1-\phi_1^{+} z-\cdots-\phi_{r'}^{+} z^{r'}\right)\left(1-\phi_1^{*} z-\cdots-\phi_{s'}^{*} z^{s'}\right),$$

we obtain

$$\phi_j=\begin{cases}\phi_j^{+}+\phi_j^{*}-\sum_{k=1}^{j-1} \phi_{j-k}^{+} \phi_k^{*}, & j=1,\ldots,r', \\ \phi_j^{*}-\sum_{k=j-r'}^{j-1} \phi_{j-k}^{+} \phi_k^{*}, & j=r'+1,\ldots,p,\end{cases} \tag{4.33}$$

where $\phi_0^{+}=\phi_0^{*}=1$, $\phi_j^{+}=0$ for $j>r'$, and $\phi_j^{*}=0$ for $j>s'$. Likewise, since $\theta(z)=\theta^{+}(z)\theta^{*}(z)$, or

$$1+\theta_1 z+\cdots+\theta_q z^q=\left(1+\theta_1^{+} z+\cdots+\theta_r^{+} z^r\right)\left(1+\theta_1^{*} z+\cdots+\theta_s^{*} z^s\right),$$

we obtain

$$\theta_j = \begin{cases} \theta_j^+ + \sum_{k=1}^{j} \theta_{j-k}^+ \theta_k^*, & j = 1, \ldots, r, \\ \sum_{k=j-r}^{j} \theta_{j-k}^+ \theta_k^*, & j = r+1, \ldots, q, \end{cases} \tag{4.34}$$

where $\theta_0^+ = \theta_0^* = 1$, $\theta_j^+ = 0$ for $j > r$, and $\theta_j^* = 0$ for $j > s$. It follows from (4.33) and (4.34) that the transformation between $\boldsymbol{\kappa}$ and $\boldsymbol{\lambda}$ is given by

$$\lambda_j = \begin{cases} \kappa_j + \kappa_{r'+j} 1_{\{j \le s'\}} - \sum_{k=1}^{j-1} \kappa_{j-k} \kappa_{r'+k} 1_{\{k \le s'\}}, & j = 1, \ldots, r', \\ \kappa_{r'+j} 1_{\{j \le s'\}} - \sum_{k=j-r'}^{j-1} \kappa_{j-k} \kappa_{r'+k} 1_{\{k \le s'\}}, & j = r'+1, \ldots, p. \\ \kappa_j + \kappa_{r+j} 1_{\{j \le s\}} + \sum_{k=1}^{j-p-1} \kappa_{j-k} \kappa_{p+r+k} 1_{\{k \le s\}}, & j = p+1, \ldots, p+r, \\ \kappa_{r+j} 1_{\{j \le s\}} + \sum_{k=j-p-r}^{j-p-1} \kappa_{j-k} \kappa_{p+r+k} 1_{\{k \le s\}}, & j = p+r+1, \ldots, p+q. \end{cases} \tag{4.35}$$

REMARK 4.3. We conjecture that, if the global LAD estimator $\widehat{\boldsymbol{\kappa}}_n$ (or $\widehat{\boldsymbol{\lambda}}_n$) is consistent, then it has the same asymptotic normal law as $\widehat{\boldsymbol{\kappa}}_{\text{LAD}}$ (or $\widehat{\boldsymbol{\lambda}}_{\text{LAD}}$). The following simulation study supports our conjecture.

4.4 Numerical Study

4.4.1 Simulation Study

In this subsection we conduct a simulation study to evaluate the asymptotic theory. The data $\{x_t\}$ are generated from the ARMA(1,1) model $X_t - \phi_0 X_{t-1} = Z_t + \theta_0 Z_{t-1}$. The value of true parameter vector (ϕ_0, θ_0) is taken to be $(0.9, 1.1)$, $(0.5, 1.5)$, and $(0.1, 1.9)$, and we use two different types of the underlying noise $\{Z_t\}$ to simulate data: the Laplacian distribution with the scale parameter $\sigma = 1.0$ and the Student's t distribution with 3 degrees of freedom. We consider two sample sizes of 500 and 5000.

For each case, we simulate 1000 replications, and report the empirical mean, standard deviation, and percent coverage of nominal 95% confidence intervals (CIs) as in the simulation study of Breidt, Davis, and Trindade (2001). To compute the percent coverage of nominal 95% CIs, the empirical CIs are computed from the asymptotic theory for each of 1000 replications. To be specific, once the estimates $\widehat{\phi}_n$ of ϕ and $\widehat{\theta}_n$ of θ are obtained for each replication, we fit the model and calculate the residuals $\left\{z_t(\widehat{\phi}_n, \widehat{\theta}_n)\right\}$. Then, $\text{Var}(Z_1)$ is estimated by the empirical variance of the residuals, $\text{E}|Z_1|$ is estimated by the empirical mean of the absolute values of residuals, and $f_Z(0)$ is estimated by the normal kernel density estimation at zero with a default normal scale bandwidth selector. Hence the standard errors of $\widehat{\phi}_n$ and $\widehat{\theta}_n$ can be estimated and the 95% confidence intervals constructed. We also report 95% confidence intervals for the empirical mean, standard deviation, and percent coverage. The asymptotic mean and standard deviation are based on (4.32).

The method of Nelder and Mead (1965) is used to search for the minimizer of the LAD objective function, which uses only function values and is robust but relatively slow. It works reasonably well for non-differentiable functions. In order to reduce the chance of being trapped in a local minimum, we use 10 starting values for each replication and find the optimized value with each of them. Then, among the 10 values, we choose the one that gives the smallest evaluation of the LAD objective function as the estimate.

The results when the underlying noise follows a Student's t distribution are reported in Table 4.1, while those when the noise follows a Laplacian distribution are reported in Table 4.2. We can see that, the two sets of results are comparable regarding the LAD estimates and empirical standard deviations. The confidence interval coverages with the Laplacian noise tend to be a little higher than those with the Student's t noise. But, except for the LAD estimates of θ when $(\phi_0, \theta_0) = (0.9, 1.1)$, confidence interval coverages are close to the nominal 95% level.

Empirical standard deviations are in agreement with asymptotic ones when sample size is 5000, but a bit larger than asymptotic ones when sample size is 500. The LAD estimates of ϕ are approximately unbiased except a few cases: $(\phi_0, \theta_0) = (0.5, 1.5)$ with the Laplacian noise and $(\phi_0, \theta_0) = (0.9, 1.1)$ with both types of noise, when sample size is 500. In these cases, there is some negative bias in the LAD estimates. The boxplots and normal probability plots indicate that the bias is due to a relatively small number of small outliers; except the outlying values, most of the LAD estimates follows the asymptotic normal law quite well.

Table 4.1: Estimates for ARMA(1,1) model $X_t - \phi_0 X_{t-1} = Z_t + \theta_0 Z_{t-1}, Z_t \sim$ IID $t\,(3)$.

	Asymptotic		Empirical		
n	Mean	Std. dev.	Mean (CI)	Std. dev. (CI)	% Coverage (CI)
500	$\phi_0 = 0.9$	0.0153	0.8970 (0.8960, 0.8980)	0.0166 (0.0159, 0.0174)	95.8 (94.6, 97.0)
	$\theta_0 = 1.1$	0.0177	1.1644 (1.1604, 1.1685)	0.0655 (0.0627, 0.0685)	47.9 (44.8, 51.0)
5000	$\phi_0 = 0.9$	0.0048	0.8995 (0.8992, 0.8998)	0.0048 (0.0046, 0.0050)	95.9 (94.7, 97.1)
	$\theta_0 = 1.1$	0.0056	1.1070 (1.1064, 1.1076)	0.0093 (0.0089, 0.0098)	74.6 (71.9, 77.3)
500	$\phi_0 = 0.5$	0.0317	0.4981 (0.4960, 0.5001)	0.0333 (0.0319, 0.0349)	96.0 (94.8, 97.2)
	$\theta_0 = 1.5$	0.0614	1.5113 (1.5070, 1.5156)	0.0689 (0.0660, 0.0720)	95.2 (93.9, 96.5)
5000	$\phi_0 = 0.5$	0.0100	0.4992 (0.4986, 0.4998)	0.0099 (0.0095, 0.0104)	95.4 (94.1, 96.7)
	$\theta_0 = 1.5$	0.0194	1.5005 (1.4992, 1.5018)	0.0206 (0.0197, 0.0216)	94.9 (93.5, 96.3)
500	$\phi_0 = 0.1$	0.0395	0.0999 (0.0973, 0.1026)	0.0425 (0.0408, 0.0445)	96.0 (94.8, 97.2)
	$\theta_0 = 1.9$	0.1219	1.9185 (1.9097, 1.9274)	0.1434 (0.1374, 0.1500)	96.1 (94.9, 97.5)
5000	$\phi_0 = 0.1$	0.0125	0.0993 (0.0985, 0.1001)	0.0124 (0.0119, 0.0130)	95.4 (94.1, 96.7)
	$\theta_0 = 1.9$	0.0386	1.8997 (1.8972, 1.9023)	0.0410 (0.0393, 0.0429)	95.4 (94.1, 96.7)

Table 4.2: Estimates for ARMA(1,1) model $X_t - \phi_0 X_{t-1} = Z_t + \theta_0 Z_{t-1}$, $Z_t \sim$ IID Laplacian.

	Asymptotic		Empirical		
n	Mean	Std. dev.	Mean (CI)	Std. dev. (CI)	% Coverage (CI)
500	$\phi_0 = 0.9$	0.0138	0.8972 (0.8962, 0.8981)	0.0154 (0.0148, 0.0161)	96.8 (95.7, 97.9)
	$\theta_0 = 1.1$	0.0160	1.1610 (1.1574, 1.1647)	0.0589 (0.0564, 0.0616)	49.6 (46.5, 52.7)
5000	$\phi_0 = 0.9$	0.0044	0.8997 (0.8994, 0.9000)	0.0045 (0.0043, 0.0047)	97.6 (96.7, 98.5)
	$\theta_0 = 1.1$	0.0050	1.1065 (1.1060, 1.1070)	0.0087 (0.0083, 0.0091)	75.8 (73.1, 78.5)
500	$\phi_0 = 0.5$	0.0282	0.4965 (0.4946, 0.4985)	0.0314 (0.0301, 0.0328)	97.1 (96.1, 98.1)
	$\theta_0 = 1.5$	0.0547	1.5024 (1.4986, 1.5062)	0.0607 (0.0581, 0.0635)	97.0 (95.9, 98.1)
5000	$\phi_0 = 0.5$	0.0089	0.5000 (0.4994, 0.5005)	0.0091 (0.0087, 0.0095)	97.1 (96.1, 98.1)
	$\theta_0 = 1.5$	0.0173	1.4997 (1.4986, 1.5007)	0.0171 (0.0164, 0.0179)	98.0 (97.1, 98.9)
500	$\phi_0 = 0.1$	0.0344	0.0976 (0.0951, 0.1000)	0.0396 (0.0379, 0.0414)	98.3 (97.5, 99.1)
	$\theta_0 = 1.9$	0.1060	1.8985 (1.8909, 1.9060)	0.1219 (0.1168, 0.1275)	97.0 (95.9, 98.1)
5000	$\phi_0 = 0.1$	0.0109	0.1000 (0.0993, 0.1007)	0.0112 (0.0107, 0.0117)	97.8 (96.9, 98.7)
	$\theta_0 = 1.9$	0.0335	1.8987 (1.8966, 1.9007)	0.0329 (0.0315, 0.0344)	98.2 (97.4, 99.0)

Table 4.3: Estimates for ARMA(1,1) model $X_t - \phi_0 X_{t-1} = Z_t + \theta_0 Z_{t-1}, Z_t \sim$ IID $t(3)$.

n	Asymptotic Mean	Asymptotic Std. dev.	Empirical Mean (CI)	Empirical Std. dev. (CI)	Empirical % Coverage (CI)
500	$\phi_0 = 0.9$	0.0153	0.8970	0.0166	95.8
			(0.8960, 0.8980)	(0.0159, 0.0174)	(94.6, 97.0)
	$\theta_0 = 1.1$	0.0177	1.1644	0.0655	47.9
			(1.1604, 1.1685)	(0.0627, 0.0685)	(44.8, 51.0)
5000	$\phi_0 = 0.9$	0.0048	0.8995	0.0048	95.9
			(0.8992, 0.8998)	(0.0046, 0.0050)	(94.7, 97.1)
	$\theta_0 = 1.1$	0.0056	1.1070	0.0093	74.6
			(1.1064, 1.1076)	(0.0089, 0.0098)	(71.9, 77.3)
500	$\phi_0 =$ -0.9	0.0153	-0.8963	0.0173	94.9
			(-0.8973, -0.8952)	(0.0166, 0.0181)	(93.5, 96.3)
	$\theta_0 =$ -1.1	0.0177	-1.1611	0.0691	52.3
			(-1.1654, -1.1568)	(0.0662, 0.0723)	(49.2, 55.4)
5000	$\phi_0 =$ -0.9	0.0048	-0.8996	0.0048	96.0
			(-0.8999, -0.8993)	(0.0046, 0.0051)	(94.8, 97.2)
	$\theta_0 =$ -1.1	0.0056	-1.1072	0.0097	75.9
			(-1.1078, -1.1066)	(0.0093, 0.0101)	(73.2, 78.6)
500	$\phi_0 =$ -0.9	0.0188	-0.8935	0.0164	97.2
			(-0.8953, -0.8917)	(0.0153, 0.0178)	(95.3, 99.0)
	$\theta_0 = 1.1$	0.0217	1.1058	0.0208	96.5
			(1.1036, 1.1081)	(0.0193, 0.0226)	(94.5, 98.5)
5000	$\phi_0 =$ -0.9	0.0059	-0.8996	0.0059	97.0
			(-0.8999, -0.8992)	(0.0057, 0.0062)	(95.9, 98.1)
	$\theta_0 = 1.1$	0.0069	1.1001	0.0074	95.1
			(1.0997, 1.1006)	(0.0070, 0.0077)	(93.8, 96.4)
500	$\phi_0 = 0.9$	0.0188	0.8946	0.0176	96.3
			(0.8926, 0.8966)	(0.0163, 0.0192)	(94.1, 98.4)
	$\theta_0 =$ -1.1	0.0217	-1.1051	0.0212	97.0
			(-1.1075, -1.1027)	(0.0196, 0.0231)	(95.0, 98.9)
5000	$\phi_0 = 0.9$	0.0059	0.8994	0.0063	95.2
			(0.8990, 0.8998)	(0.0060, 0.0065)	(93.9, 96.5)
	$\theta_0 =$ -1.1	0.0069	-1.1003	0.0075	95.8
			(-1.1007, -1.0998)	(0.0072, 0.0078)	(94.6, 97.0)

Overall, the LAD estimates of θ exhibit a similar pattern to those of ϕ. However, the estimation of θ seems a little harder than that of ϕ, especially when θ_0 is close to the unit circle. In the cases where bias exists in the estimates of θ, it is a positive rather than negative bias, which is due to a relatively small number of large outliers. Other combinations of (ϕ_0, θ_0) value are also considered, showing that the estimation becomes more difficult and the convergence becomes slower when θ_0 or ϕ_0 gets close to the unit circle. Moreover, given a θ_0 value, a change of ϕ_0 value does not affect simulation results much; in contrast, for a given ϕ_0 value, a change of θ_0 value affects simulation results to some extent. Further investigations are needed to explore more on this.

We also investigate the symmetry of the LAD estimates for the case when $(\phi_0, \theta_0) = (\pm 0.9, \pm 1.1)$ and the underlying noise follows a Student's t distribution. The results are reported in Table 4.3. It shows that, the empirical results are symmetric about the origin, which is in agreement with asymptotic theory. Note that, with true value of $(-0.9, 1.1)$ or $(0.9, -1.1)$, the performance of θ estimates improves a lot compared with that with true value of $(0.9, 1.1)$ or $(-0.9, -1.1)$. Actually in former cases, the ARMA(1, 1) model is approximately an all-pass model with $\phi_0 = -0.9$ or 0.9, and the simulation results are comparable to those in Breidt, Davis, and Trindade (2001).

Figure 4.13 is the contour plot of $l_n(\phi, \theta)/n$ when $(\phi_0, \theta_0) = (0.5, 1.5)$ and the underlying noise follows a Student's t distribution. From the figure we can see that the true value of $(0.5, 1.5)$ is the unique global minimizer of $l_n(\phi, \theta)/n$.

4.4.2 Applications

1. Microsoft stock trading volumes

Figure 4.14 shows the natural logarithms of the volumes of Microsoft stock traded over 755 transaction days from 06/03/96 to 05/27/99. The sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of the log-volume series are displayed in Figure 4.15, indicating that an AR model of order 3 or 1 might be appropriate. The data were also studied by Breidt, Davis, and Trindade (2001); they fitted an AR(1) model

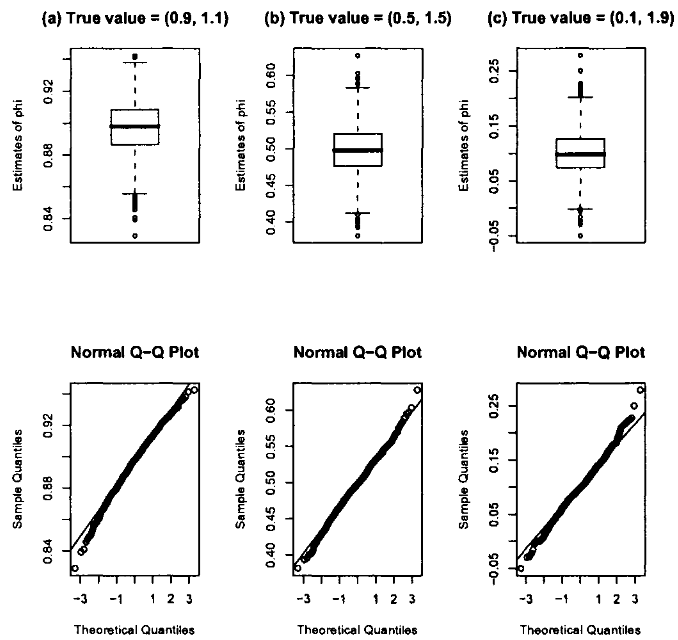

Figure 4.1: Boxplots and Normal probability plots of ϕ estimates when $n = 500$ and $\{Z_t\} \sim$ IID $t(3)$. (a) $(\phi_0, \theta_0) = (0.9, 1.1)$. (b) $(\phi_0, \theta_0) = (0.5, 1.5)$. (c) $(\phi_0, \theta_0) = (0.1, 1.9)$.

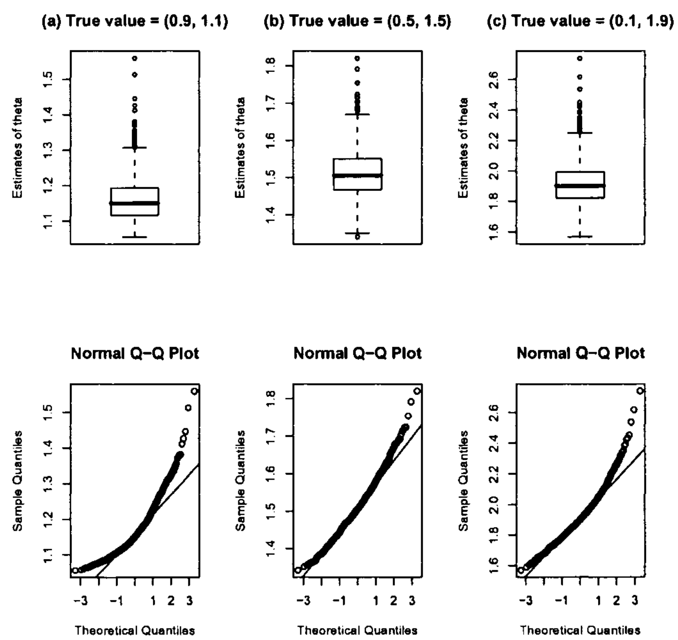

Figure 4.2: Boxplots and Normal probability plots of θ estimates when $n = 500$ and $\{Z_t\} \sim$ IID $t(3)$. (a) $(\phi_0, \theta_0) = (0.9, 1.1)$. (b) $(\phi_0, \theta_0) = (0.5, 1.5)$. (c) $(\phi_0, \theta_0) = (0.1, 1.9)$.

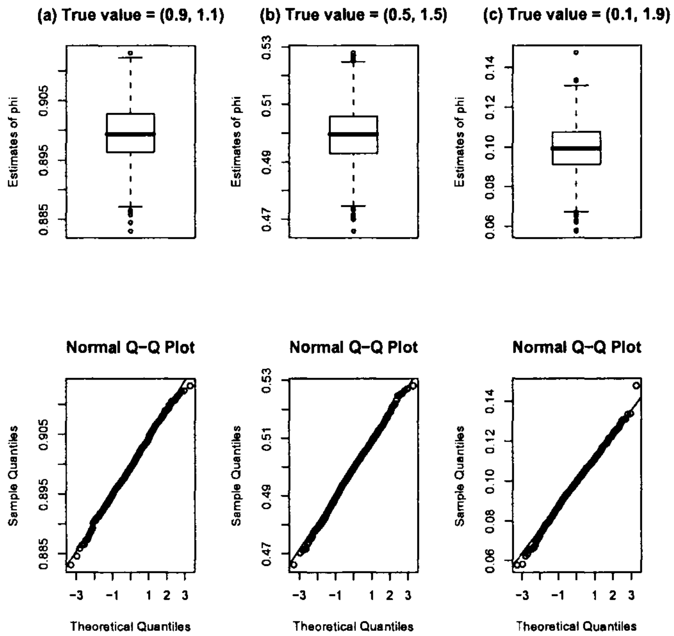

Figure 4.3: Boxplots and Normal probability plots of ϕ estimates when $n = 5000$ and $\{Z_t\} \sim$ IID $t(3)$. (a) $(\phi_0, \theta_0) = (0.9, 1.1)$. (b) $(\phi_0, \theta_0) = (0.5, 1.5)$. (c) $(\phi_0, \theta_0) = (0.1, 1.9)$.

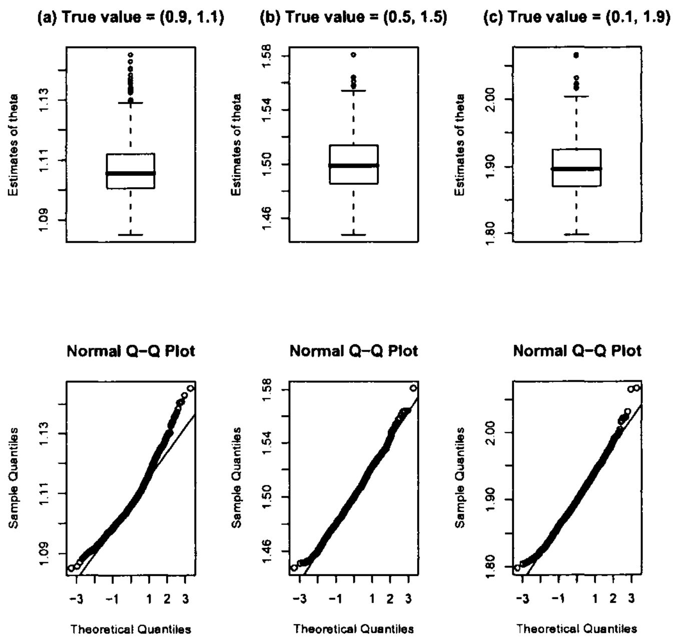

Figure 4.4: Boxplots and Normal probability plots of θ estimates when $n = 5000$ and $\{Z_t\} \sim$ IID $t(3)$. (a) $(\phi_0, \theta_0) = (0.9, 1.1)$. (b) $(\phi_0, \theta_0) = (0.5, 1.5)$. (c) $(\phi_0, \theta_0) = (0.1, 1.9)$.

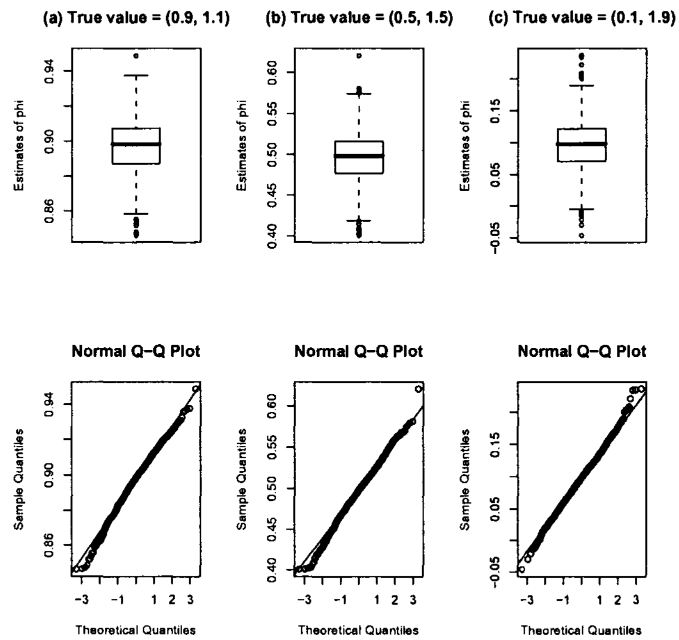

Figure 4.5: Boxplots and Normal probability plots of ϕ estimates when $n = 500$ and $\{Z_t\} \sim$ IID Laplacian. (a) $(\phi_0, \theta_0) = (0.9, 1.1)$. (b) $(\phi_0, \theta_0) = (0.5, 1.5)$. (c) $(\phi_0, \theta_0) = (0.1, 1.9)$.

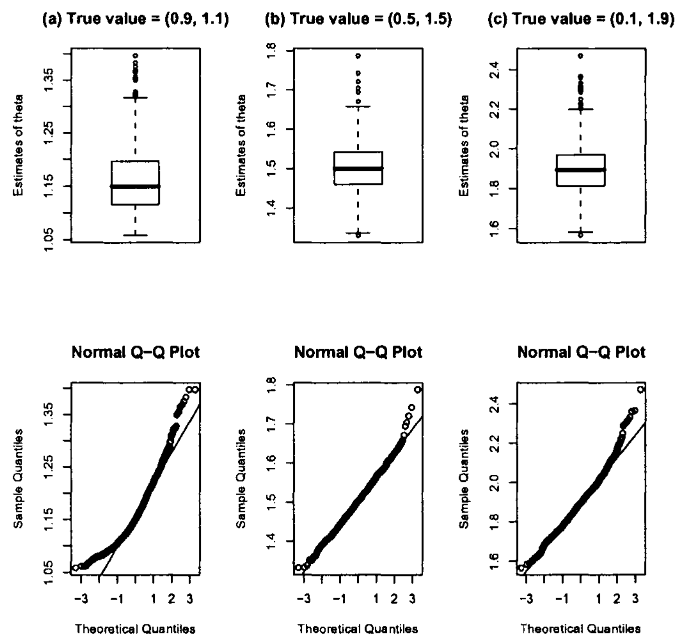

Figure 4.6: Boxplots and Normal probability plots of θ estimates when $n = 500$ and $\{Z_t\} \sim$ IID Laplacian. (a) $(\phi_0, \theta_0) = (0.9, 1.1)$. (b) $(\phi_0, \theta_0) = (0.5, 1.5)$. (c) $(\phi_0, \theta_0) = (0.1, 1.9)$.

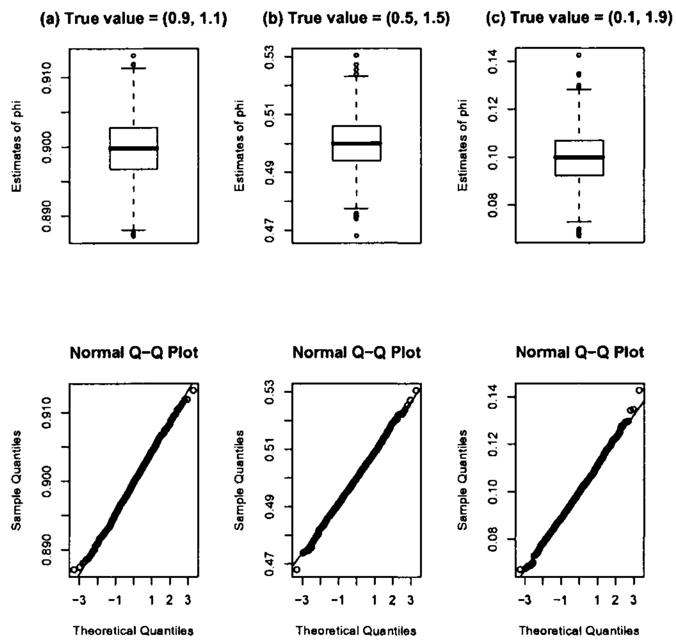

Figure 4.7: Boxplots and Normal probability plots of ϕ estimates when $n = 5000$ and $\{Z_t\} \sim$ IID Laplacian. (a) $(\phi_0, \theta_0) = (0.9, 1.1)$. (b) $(\phi_0, \theta_0) = (0.5, 1.5)$. (c) $(\phi_0, \theta_0) = (0.1, 1.9)$.

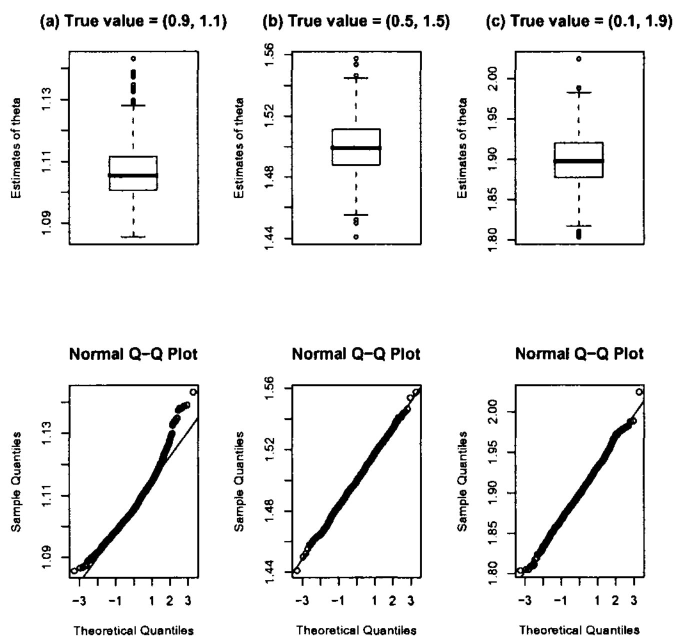

Figure 4.8: Boxplots and Normal probability plots of θ estimates when $n = 5000$ and $\{Z_t\} \sim$ IID Laplacian. (a) $(\phi_0, \theta_0) = (0.9, 1.1)$. (b) $(\phi_0, \theta_0) = (0.5, 1.5)$. (c) $(\phi_0, \theta_0) = (0.1, 1.9)$.

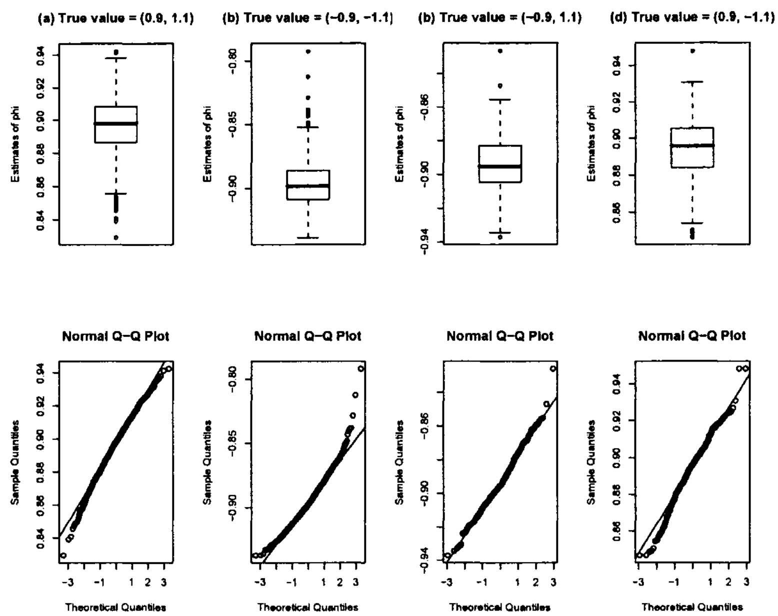

Figure 4.9: Boxplots and Normal probability plots of ϕ estimates when $n = 500$ and $\{Z_t\} \sim$ IID $t(3)$. (a) $(\phi_0, \theta_0) = (0.9, 1.1)$. (b) $(\phi_0, \theta_0) = (-0.9, -1.1)$. (c) $(\phi_0, \theta_0) = (-0.9, 1.1)$. (d) $(\phi_0, \theta_0) = (0.9, -1.1)$.

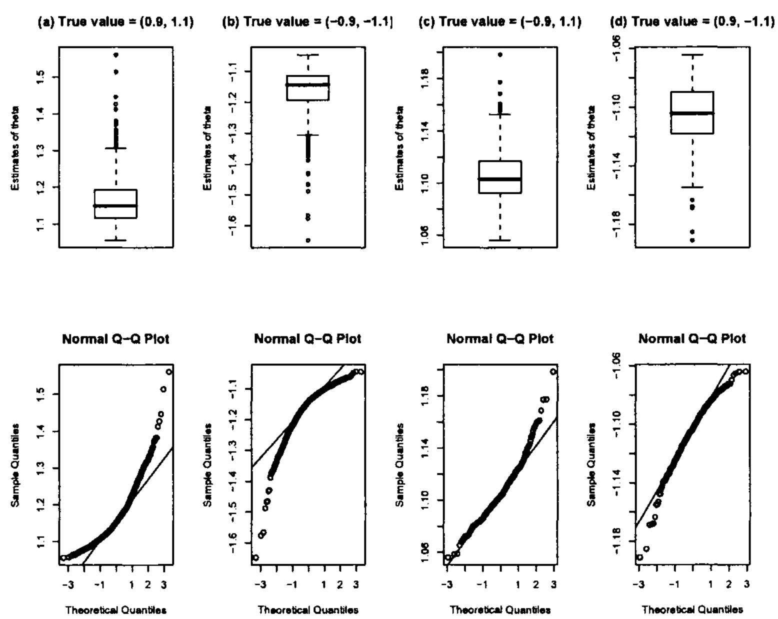

Figure 4.10: Boxplots and Normal probability plots of θ estimates when $n = 500$ and $\{Z_t\} \sim$ IID $t(3)$. (a) $(\phi_0, \theta_0) = (0.9, 1.1)$. (b) $(\phi_0, \theta_0) = (-0.9, -1.1)$. (c) $(\phi_0, \theta_0) = (-0.9, 1.1)$. (d) $(\phi_0, \theta_0) = (0.9, -1.1)$.

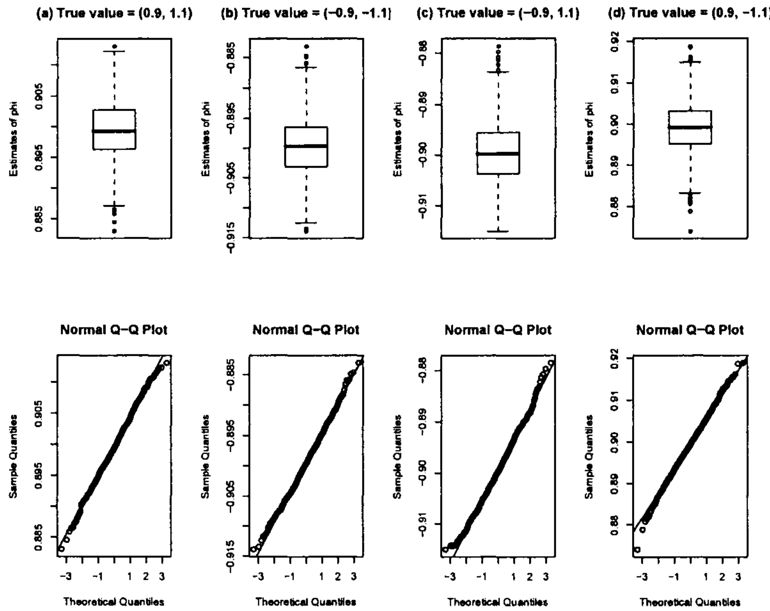

Figure 4.11: Boxplots and Normal probability plots of ϕ estimates when $n = 5000$ and $\{Z_t\} \sim$ IID $t(3)$. (a) $(\phi_0, \theta_0) = (0.9, 1.1)$. (b) $(\phi_0, \theta_0) = (-0.9, -1.1)$. (c) $(\phi_0, \theta_0) = (-0.9, 1.1)$. (d) $(\phi_0, \theta_0) = (0.9, -1.1)$.

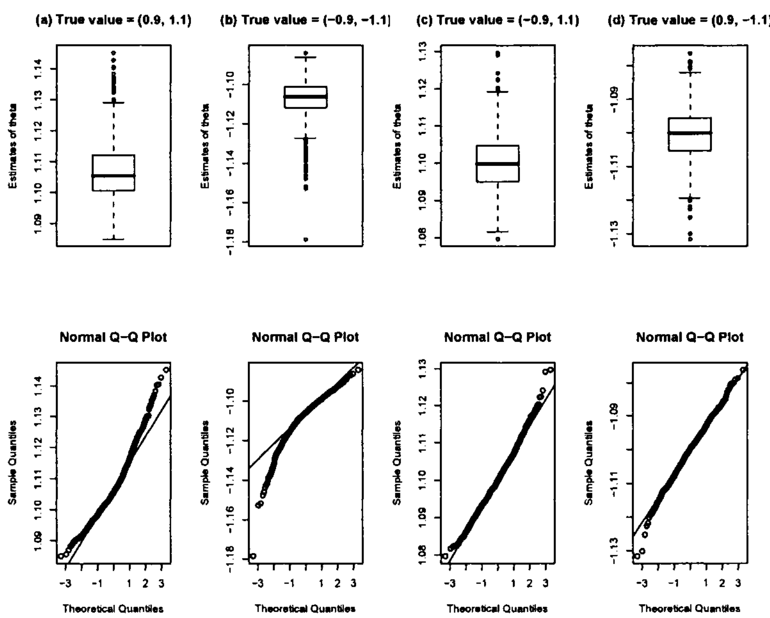

Figure 4.12: Boxplots and Normal probability plots of θ estimates when $n = 5000$ and $\{Z_t\} \sim$ IID $t(3)$. (a) $(\phi_0, \theta_0) = (0.9, 1.1)$. (b) $(\phi_0, \theta_0) = (-0.9, -1.1)$. (c) $(\phi_0, \theta_0) = (-0.9, 1.1)$. (d) $(\phi_0, \theta_0) = (0.9, -1.1)$.

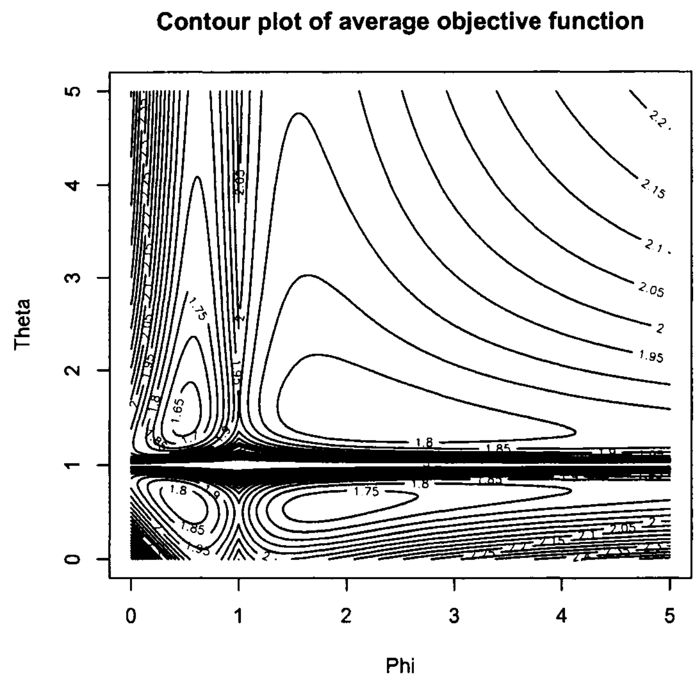

Figure 4.13: Contour Plot of $l_n(\phi, \theta)/n$ when $(\phi_0, \theta_0) = (0.5, 1.5)$, and noise $\{Z_t\} \sim$ IID $t(3)$.

to the log-volume series for the illustration of noncausal AR model fitting using all-pass models.

We firstly use the function "ar" in R package to automatically select order and fit an AR model to the log-volume series, which results in the following causal AR(3) model:

$$X_t = 0.5139X_{t-1} + 0.0237X_{t-2} + 0.1378X_{t-3} + W_t. \tag{4.36}$$

The ACF of resulting residuals $\{\hat{W}_t\}$ (Figure 4.16 (a)) indicates that $\{\hat{W}_t\}$ are white noise. However, the ACFs of $\{\hat{W}_t^2\}$ and $\{|\hat{W}_t|\}$ (Figure 4.16 (b) and (c)) suggest that $\{\hat{W}_t\}$ are not likely to be independent, as both $\{\hat{W}_t^2\}$ and $\{|\hat{W}_t|\}$ have significant lag 1 autocorrelation.

Now, a direct AR(3) fit using LAD estimation yields a purely noncausal AR(3) model

$$X_t = -0.0644X_{t-1} - 3.3959X_{t-2} + 6.4165X_{t-3} + Z_t, \tag{4.37}$$

with all roots 0.7910 and $-0.1309 \pm 0.4241i$ inside the unit circle. The residuals $\{\hat{Z}_t\}$ appear

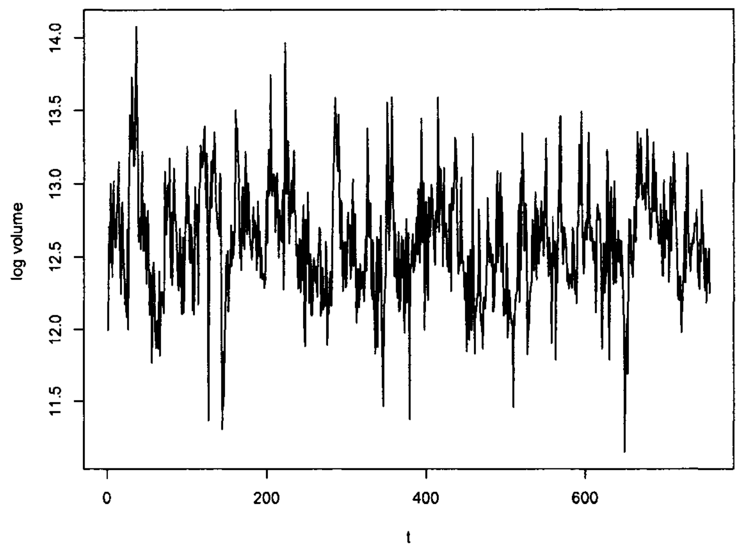

Figure 4.14: Log-volumes of Microsoft stock.

independent based on the ACF plots in Figure 4.17. Therefore, the model (4.37) seems more appropriate for this log-volume series.

One also could check all possible AR(3) model fits by flipping the roots of AR polynomial of the model (4.36). The purely noncausal representation of the process $\{\hat{X}_t\}$ in (4.36) is

$$X_t = -0.172X_{t-1} - 3.7292X_{t-2} + 7.2564X_{t-3} + U_t.$$

The ACF plots of the residuals $\{\hat{U}_t\}$ in Figure 4.18 are almost identical to their counterparts in Figure 4.17.

2. Wal-Mart stock trading volumes

In this example we study the volumes of Wal-Mart stock traded daily on the New York Stock Exchange from 1/1/03 to 12/31/04, which are shown in Figure 4.19. The sample ACF and PACF plots of the volume series are given in Figure 4.20. The data appear heavy-tailed instead of Gaussian. Moreover, the first half $\{X_t\}_{t=1}^{218}$ and the second half $\{X_t\}_{t=219}^{504}$ of the series display different mean structures. One could model the two halves separately. Nevertheless, we consider the whole series here, and use ITSM package (Brockwell and

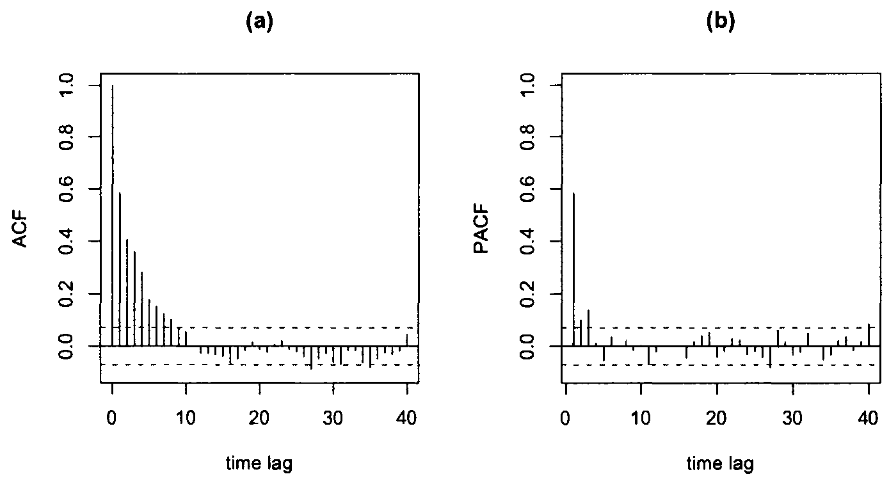

Figure 4.15: (a) ACF of log-volumes. (b) PACF of log-volumes.

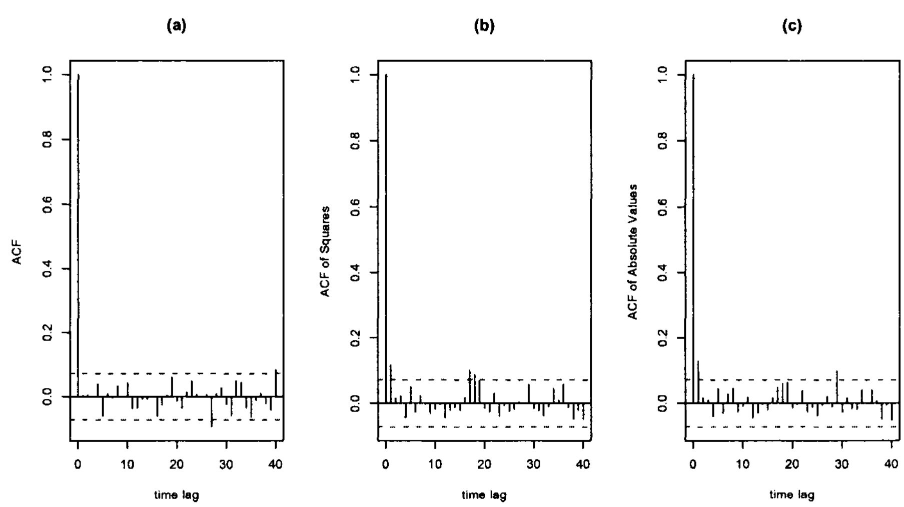

Figure 4.16: (a) ACF of $\{\hat{W}_t\}$. (b) ACF of $\{\hat{W}_t^2\}$. (c) ACF of $\{|\hat{W}_t|\}$.

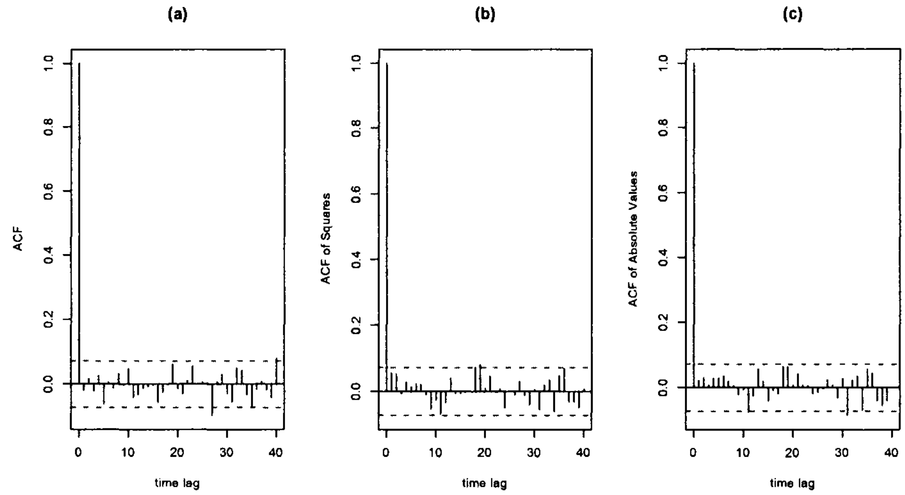

Figure 4.17: (a) ACF of $\{\hat{Z}_t\}$. (b) ACF of $\{\hat{Z}_t^2\}$. (c) ACF of $\{|\hat{Z}_t|\}$.

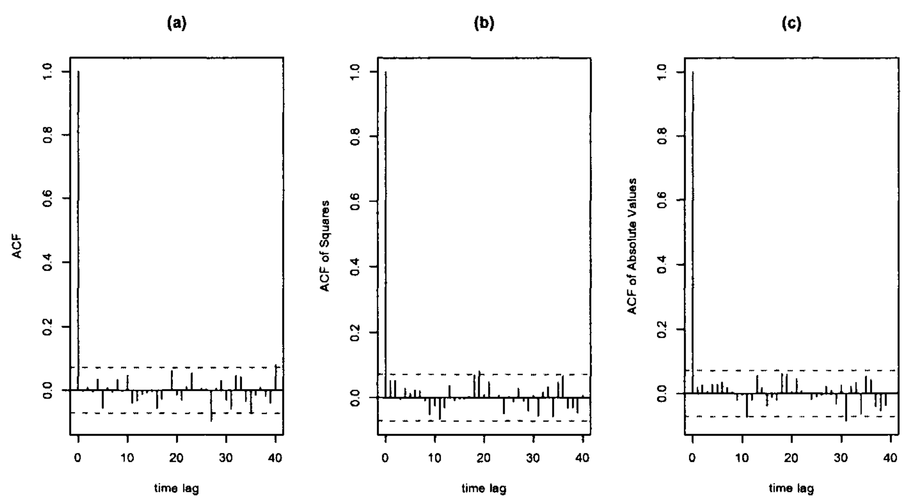

Figure 4.18: (a) ACF of $\{\hat{U}_t\}$. (b) ACF of $\{\hat{U}_t^2\}$. (c) ACF of $\{|\hat{U}_t|\}$.

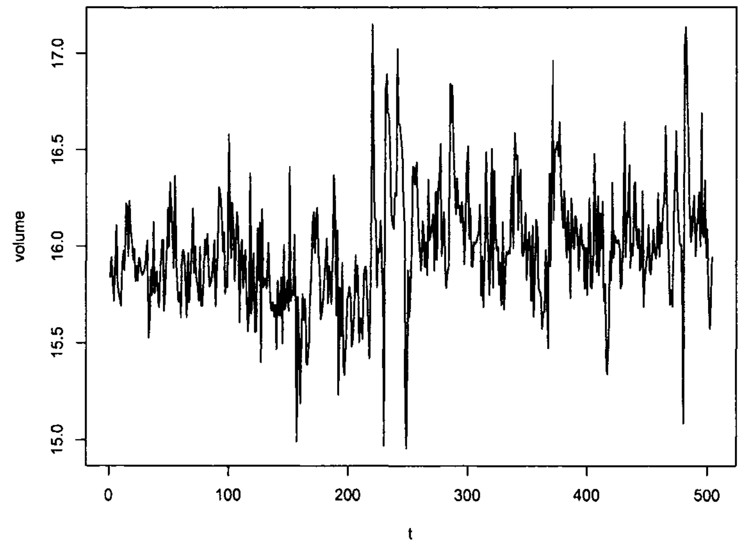

Figure 4.19: The volumes of Wal-Mart stock.

Davis, 1996) to automatically select orders and fit an ARMA model to the data, which yields the following causal-invertible ARMA(3,1) model:

$$X_t = 1.357X_{t-1} - 0.2842X_{t-2} - 0.08316X_{t-3} + W_t - 0.9512W_{t-1}. \tag{4.38}$$

The ACF of resulting residuals $\{\hat{W}_t\}$ (Figure 4.21 (a)) indicates that $\{\hat{W}_t\}$ are white noise. However, they are not likely to be independent because both $\{\hat{W}_t^2\}$ and $\{|\hat{W}_t|\}$ have significant lag 1 autocorrelation (Figure 4.21 (b) and (c)).

Now we fit an ARMA(3,1) model using LAD estimation, yielding

$$X_t = 0.8496X_{t-1} - 0.0491X_{t-2} - 0.0114X_{t-3} + Z_t - 2.2800Z_{t-1}. \tag{4.39}$$

This is a causal-noninvertible ARMA(3,1) model with roots $-11.4453, 5.8495, 1.3054$ of the AR polynomial $\phi(z)$ outside the unit circle, and root 0.4386 of the MA polynomial $\theta(z)$ inside the unit circle. The residuals $\{\hat{Z}_t\}$ appear independent based on the ACF plots in Figure 4.22. Therefore, the model (4.37) is more appropriate for this series of the Wal-Mart stock trading volumes.

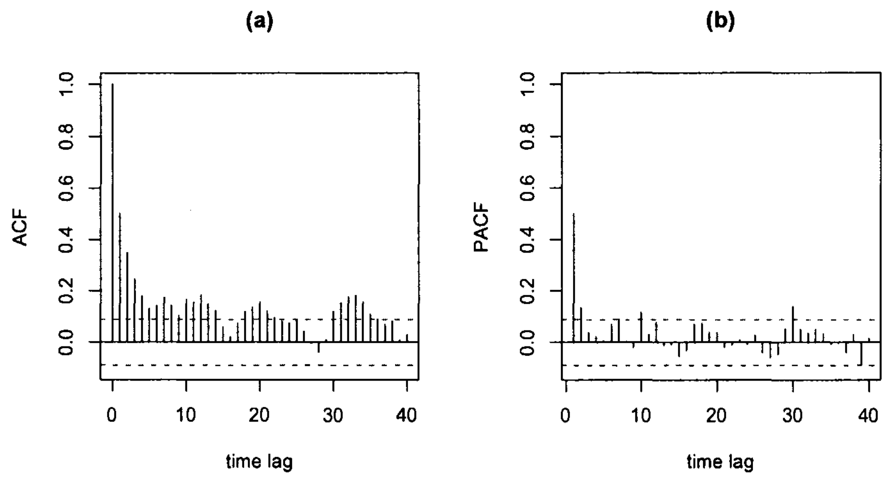

Figure 4.20: (a) ACF of Wal-Mart stock volumes. (b) PACF of Wal-Mart stock volumes.

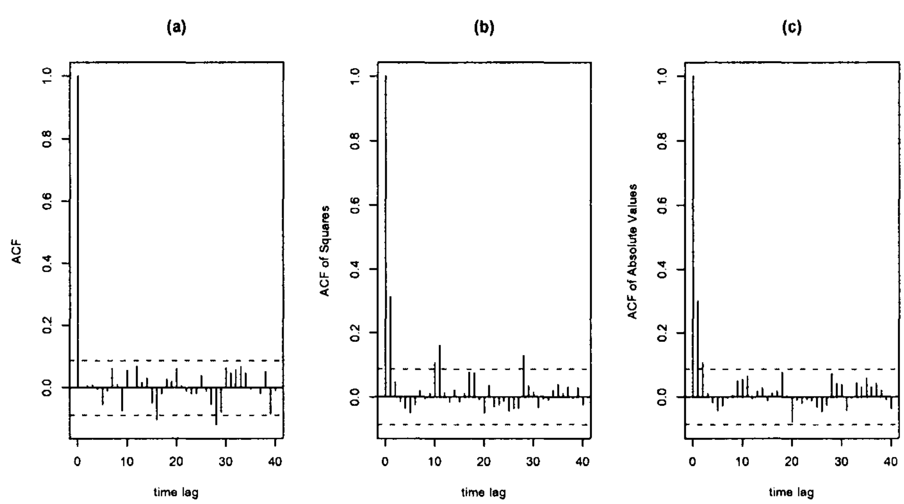

Figure 4.21: (a) ACF of $\{\hat{W}_t\}$. (b) ACF of $\{\hat{W}_t^2\}$. (c) ACF of $\{|\hat{W}_t|\}$.

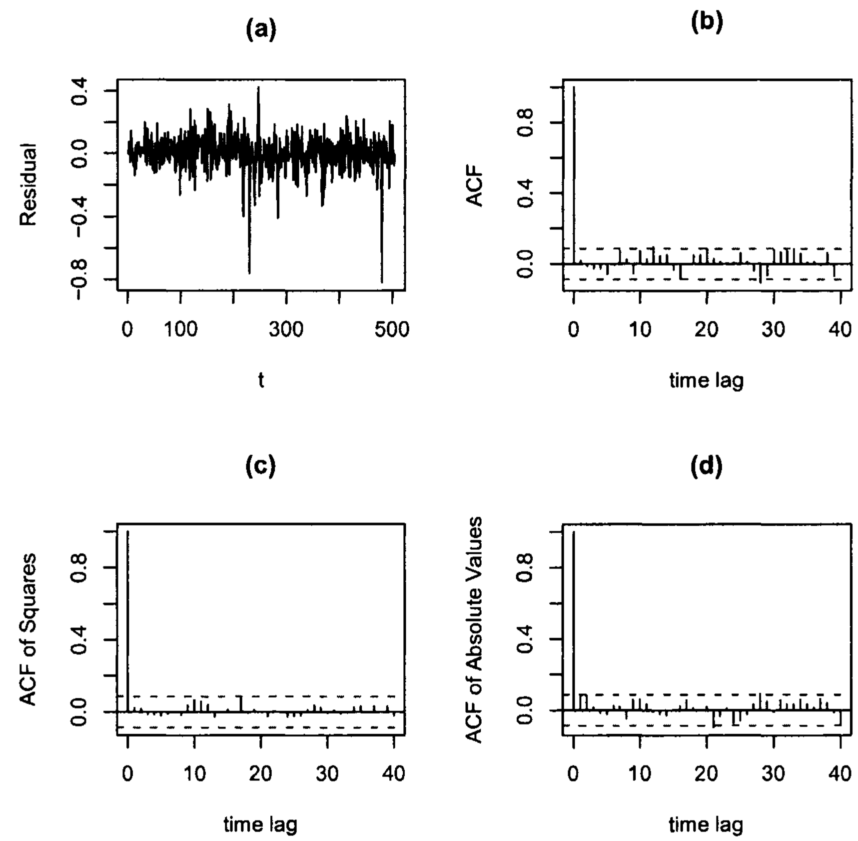

Figure 4.22: (a) Residuals. (b) ACF of $\{\hat{Z}_t\}$. (c) ACF of $\{\hat{Z}_t^2\}$. (d) ACF of $\{|\hat{Z}_t|\}$.

5 SUMMARY AND FUTURE WORK

In this dissertation, we studied parameter estimation for two different classes of models: parameter-driven GLMs for time series, and noncausal and/or noninvertible ARMA models. We studied GLM estimation for the former class of models, and LAD estimation for the latter. Chapter 1 gave the motivation of our research as well as an introduction to the two classes of models.

In Chapter 2, we first considered regression analysis of time series of count data. The time series of counts $\{Y_t : t = 1, \ldots, n\}$, conditional on regressors $\mathbf{x}_t$ and a latent process $\{\alpha_t\}$, was assumed to be independent, and the conditional distribution was specified by a negative binomial distribution. On the other hand, the logit function $\log \frac{p_t}{1-p_t} = \mathbf{x}_t^T \boldsymbol{\beta} + \alpha_t$ incorporated the autocorrelated latent process $\{\alpha_t\}$ into the model, introducing serial dependence among the observed data $\{Y_t\}$. In this dissertation, the latent process $\{\alpha_t\}$ was assumed to evolve independently of the observed data, whence the model was a type of parameter-driven models. The GLM estimator $\widehat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$ was obtained by ignoring the latent process and maximizing the resulting pseudo-likelihood function. We showed the consistency and asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$ under two cases: where the latent process was assumed to be a stationary Gaussian process, and where it was assumed to be a stationary strongly mixing process.

We also studied parameter-driven GLMs for general time series, where the random component of a GLM was specified by a distribution from the one-parameter exponential family, and link function was chosen such that the log-likelihood function was concave and $\mathrm{E}\left[\mathcal{H}\left(\mathbf{x}_t^T \boldsymbol{\beta} + \alpha_t\right)\right] = \mathcal{H}\left(\mathbf{x}_t^T \boldsymbol{\beta}\right)$, where $\mathcal{H}(\cdot)$ was the inverse of link function. We generalized the asymptotic results about the GLM estimator under suitable conditions, and thus unified in a common framework the results for Poisson log-linear regression models (Davis, Dunsmuir, and Wang, 2000) and negative binomial logit regression models. In addition, asymptotic results for a Gaussian setup were given as an example.

The simulation study for evaluating the asymptotic theory included a model with Bernoulli setting as well as two models with negative binomial setting. In all cases, empirical GLM estimates were in agreement with corresponding asymptotic theory. We also applied the theory and methods to the monthly numbers of poliomyelitis cases in the U.S.A. from the year 1970 to 1983 as reported by the Centers for Disease Control.

As to future work, we are interested in statistical inference of parameter-driven GLMs for time series based on the exact likelihood. Although the likelihood cannot be calculated explicitly, asymptotic properties of the MLE can be investigated via the framework of generalized state-space models (see Brockwell and Davis (1996) for background on generalized state-space models).

Inference for generalized state-space models traces back to Baum and Petrie (1966), who studied inference for hidden Markov models (HMMs), which are a special class of generalized state-space models. In HMMs, the observed variables $\{Y_k\}$ are conditionally independent given a sequence of unobservable state variables $\{S_k\}$, and the distribution of Y_t at time t depends on $\{S_k\}$ only through the state S_t. On the other hand, the evolution of $\{S_k\}$ is specified by a discrete-time finite-state homogeneous Markov chain. That is, state variables $\{S_k\}$ take values in a finite set. In generalized state-space models, however, the state process $\{S_k\}$ is not necessarily to be Markovian and may follow both continuous and discrete distributions.

Baum and Petrie (1966) considered the case where both $\{Y_k\}$ and $\{S_k\}$ take values in a finite set, and established consistency and asymptotic normality of the MLE under certain conditions. Leroux (1992) proved the strong consistency of the MLE for a general stationary ergodic HMM, where $\{Y_k\}$ may take infinitely-many values. Consistency of the MLE was also proved for several extensions of standard HMMs using a corresponding ergodic theorem. The asymptotic normality for general HMMs was proved by Bickel, Ritov and Rydén (1998) after being an open problem for more than thirty years (Ephraim and Merhav, 2002).

The proof relied on a central limit theorem for the score function $D_\theta \mathfrak{L}_n(\theta)$, where $\mathfrak{L}_n(\theta)$ denoted the log-likelihood function based on $Y_1, \ldots, Y_n$, and a law of large numbers for the observed information $-D_\theta^2 \mathfrak{L}_n(\theta)$. The idea was applied by Jensen and Petersen (1999) to state-space models with a separable compact state space that is not necessarily finite. Le Gland and Mevel (2000) developed a different approach for proving the consistency and asymptotic normality of the MLE for HMMs with finite state space. The idea is to express log-likelihood function as an additive function of an extended Markov chain and use the geometric ergodicity of the extended chain. Douc and Matias (2001), following the approach of Le Gland and Mevel (2000), proved the consistency and asymptotic normality of the MLE for general HMMs, where the latent Markov process took values in a topological space. Douc, Moulines and Rydén (2004) proved the asymptotic normality of conditional MLE for an AR process with Markov regime having a separable compact state space that is not necessarily finite.

Two key assumptions in Jensen and Petersen (1999) and Douc and Matias (2001) do not hold in our setup, that state space is separable compact and transition density of states is uniformly bounded below by a small positive number. However, the aforementioned papers pave a way for studying asymptotic properties of the MLE for parameter-driven GLMs for time series.

In Chapters 3 and 4, we first investigated LAD estimation for the MA(1) model $X_t = Z_t - \theta Z_{t-1}$, where $|\theta| \neq 1$. The derivation of LAD criterion was based on Breidt, Davis, Hsu, and Rosenblatt (2006). By building sample size into new parameterizations, we showed the existence of a sequence of local LAD estimators $\widehat{\theta}_{\text{LAD}}$ that is consistent and asymptotic normal, which followed from a functional limit theorem for random processes. Moreover, strong consistency and asymptotic normality were established for the global LAD estimator $\widehat{\theta}_n$, with the additional assumptions that the underlying noise $\{Z_t\}$ has upper-bounded density and absolute value of the first derivative of density, and that it has heavier tails than Gaussian, in the sense that $\mathrm{E}\left|\sum_{j=-\infty}^{\infty} c_j Z_{t-j}\right| > \mathrm{E}\left|Z_1\right|$ holds for any sequence $\{c_j\}$ such that $\sum_{j=-\infty}^{\infty} |c_j| < \infty$, $\sum_{j=-\infty}^{\infty} c_j^2 = 1$, and $\{c_j\}$ has at least two non-zero elements.

To simplify numerical computations in searching for a minimizer of the objective function that is not convex in θ, we proposed a local linearization procedure, approximating the objective function in a neighborhood of $\widehat{\theta}_0$ using a Taylor expansion, where $\widehat{\theta}_0$ is a set of initial estimators of θ such that $\widehat{\theta}_0 = \theta_0 + O_p\left(n^{-1/2}\right)$. Then, the linearized LAD estimator $\widehat{\theta}_{\text{LLAD}}$, which was defined as minimizer of the linearized objective function, followed the same asymptotic normal law as $\widehat{\theta}_{\text{LAD}}$.

Then, LAD estimation for noncausal and/or noninvertible ARMA(p, q) models was investigated. To facilitate the analysis, we deconstructed an ARMA(p, q) model by factoring its AR part into causal and purely noncausal components, and its MA part into invertible and purely noninvertible components. The study was focused on the deconstructed model. The LAD criterion was derived by a likelihood approximation assuming a Laplacian underlying noise. To get around the difficulty caused by the non-convexity of objective function in parameter vector $\boldsymbol{\kappa}$, we applied a local approximation technique. We built sample size into a new parameterization, and established a functional limit theorem for a random process, from which the consistency and asymptotic normality of the LAD estimator were identified.

Simulation study was conducted for MA(1) models to evaluate the asymptotic theory about the LAD and linearized LAD estimators, and for ARMA(1, 1) models as well. We also applied the theory and methods to two real-life data sets: the volumes of Microsoft stock traded over 755 transaction days from 06/03/96 to 05/27/99, and the volumes of Wal-Mart stock traded daily on the New York Stock Exchange from 1/1/03 to 12/31/04.

As to future work, we will complete the proof of (3.32) for a random sequence $\{\theta_n\}$ such that $\theta_n \xrightarrow{\text{a.s.}} \theta_0$; one possible approach is to partition the interval specified by $|\cdot| \leq n^{-p}$ into subintervals, and bound relevant random terms in each subinterval such that the methods used for non-random sequence $\{\theta_n\}$ may apply. On the other hand, we are interested in studying other commonly used estimation methods, such as M-estimation or R-estimation, for noncausal and/or noninvertible ARMA models.

REFERENCES

1. Andrews, B., Davis, R.A., and Breidt, F.J. (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis* 97, 1638-1659.

2. Andrews, M.E. (2003). Parameter Estimation for All-Pass Time Series Models. Ph.D. Dissertation, Department of Statistics, Colorado State University, Fort Collins, CO.

3. Baum, L. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics* 37, 1554-1563.

4. Benjamin, M.A., Rigby, R.A., Stasinopoulos, D.M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association* 98, 214-223.

5. Bickel, P.J., Ritov, Y., and Rydén, T. (1998). Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *Annals of Statistics* 26, 1614-1635.

6. Billingsley, P. (1995). *Probability and Measure.* 3rd edition. John Wiley and Sons, New York.

7. Blais, M., MacGibbon, B., and Roy, R. (2000). Limit theorem for regression models of time series of counts. *Statistics and Probability Letters* 46, 161-168.

8. Bloomfield, P. and Steiger, W. (1983). *Least Absolute Deviations: Theory, Applications and Algorithms.* Birkhauser, Boston, MA.

9. Booth, J.G., Casella, G., Friedl, H., and Hobert, J.P. (2003). Negative binomial loglinear mixed models. *Statistical Modeling* 3 (3), 179-191.

10. Brännäs, K. and Johansson, P. (1994). Time series count data regression. *Communications in Statistics: Theory and Methods* 23, 2907-2925.

11. Breidt, F.J., Davis, R.A., Hsu, N.-J., and Rosenblatt, M. (2006). Pile-up probabilities for the Laplace likelihood estimator of a non-invertible first order moving average. In *IMS Lecture Notes-Monograph Series: Time Series and Related Topics.* Vol. 52, 1-19.

12. Breidt, F.J., Davis, R.A., and Trindade, A.A. (2001). Least absolute deviation estimation for all-pass time series models, *Annals of Statistics* 29, 919-946.

13. Breidt, F.J., Davis, R.A., Lii, K.S., and Rosenblatt, M. (1991). Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis* 36, 175-198.

14. Breidt, F.J. and Hsu, N.-J. (2005). Best mean square prediction for moving averages. *Statistica Sinica* 15, 427-446.

15. Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods.* 2nd edition. Springer-Verlag, New York.

16. Brockwell, P.J. and Davis, R.A. (1996). *Introduction to Time Series and Forecasting.* Springer-Verlag, New York.

17. Byers, A.L., Heather, A., Gill, T.M., and Peduzzi, P.N. (2003). Application of negative binomial modeling for discrete outcomes: A case study in aging research. *Journal of Clinical Epidemiology* 56, 559-564.

18. Cameron, A.C. and Trivedi, P.K. (1996). Count data models for financial data. In *Handbook of Statistics, Volume 14, Statistical Methods in Finance* (Maddala, G.S. and C.R. Rao, ed.), pp 363-392. Elsevier Science: Amsterdam, North-Holland.

19. Cameron, A.C. and Trivedi, P.K. (1998). *Regression Analysis of Count Data.* Cambridge University Press.

20. Campbell, M.J. (1994). Time series regression for counts: an investigation into the relationship between sudden infant death syndrome and environmental temperature. *Journal of the Royal Statistical Society,* Series A 157, 191-208.

21. Chan, K.S. and Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association* 90, 242-252.

22. Chien, H.-M., Yang, H.-L., and Chi, C.-Y. (1997). Parametric cumulant based phase estimation of 1-d and 2-d nonminimum phase systems by allpass filtering. *IEEE Trans. Signal Process* 45, 1742-1762.

23. Cox, D.R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics* 8, 93-115.

24. Davidson, J. (1992). A central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes. *Econometric Theory* 8, 313-329.

25. Davis, R.A. and Dunsmuir, W.T.M. (1996). Maximum likelihood estimation for MA(1) processes with a root on or near the unit circle. *Econometric Theory* 12, 1-29.

26. Davis, R.A. and Dunsmuir, W.T.M. (1997). Least Absolute Deviation Estimation for Regression with ARMA Errors. *Journal of Theoretical Probability* 10, 481-497.

27. Davis, R.A., Dunsmuir, W.T.M., and Wang, Y. (1999). Modelling time series of count data. In *Asymptotics, Nonparametrics and Time Series* (Ghosh, S. ed.), pp 63-114. Marcel Dekker, New York.

28. Davis, R.A., Dunsmuir, W.T.M., and Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika* 87, 491-505.

29. Davis, R.A., Knight, K., and Liu, J. (1992). M-estimation for autoregressions with infinite variance. *Stochastic Processes and their Applications* 40, 145-180.

30. Davis, R.A. and Resnick, S.I. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *Annals of Probability* 13, 179-195.

31. Douc, R. and Matias, C. (2001). Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli* 7, 381-420.

32. Douc, R., Moulines, É., and Rydén, T. (2004). Asymptotic properties of the maximum likelihood estimation in autoregressive models with Markov regime. *Annals of Statistics* 32, 2254-2304.

33. Durbin, J. and Koopman, S.J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika* 84, 669-684.

34. Durbin, J. and Koopman, S.J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society*, Series B 62, 3-29.

35. Ephraim, Y. and Merhav, N. (2002). Hidden Markov processes. *IEEE Transactions on Information Theory* 48 (6), 1518-1569.

36. Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics* 13, 342-368.

37. Geyer, C.J. (1994). On the convergence of Monte Carlo maximum likelihood calculations. *Journal of the Royal Statistical Society*, Series B 56, 261-274.

38. Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: applications to Poisson models. *Econometrica* 52, 701-720.

39. Haberman, S.J. (1977). Maximum likelihood estimates in exponential response models. *Annals of Statistics* 5, 815-841.

40. Hannan, E.J. (1970). *Multiple Time Series.* John Wiley and Sons, New York.

41. Harvey, A.C. and Fernandes, C. (1989). Time series models for count or qualitative observations. *Journal of Business and Economic Statistics* 7 (4), 407-417.

42. Horvath, L. and Liese, F. (2004). L_p-estimators in ARCH models. *Journal of Statistical Planning and Inference* 119, 227-309.

43. Huang, J. and Pawitan, Y. (2000). Quasi-likelihood estimation of non-invertible moving average processes. *Scandinavian Journal of Statistics* 27, 689-702.

44. Jensen, J.L. and Petersen, N.V. (1999). Asymptotic normality of the maximum likelihood estimator in state space models. *Annals of Statistics* 27, 514-535.

45. Johansson, P. (1996). Speed limitation and motorway casualties: a time series count data regression approach. *Accident Analysis and Prevention* 28, 73-87.

46. Johnson, N.L., Kemp, A.W., and Kotz, S. (2005). *Univariate Discrete Distributions.* 3rd edition. John Wiley and Sons, Hoboken, NJ.

47. Jørgensen, B., Lundbye-Christensen, S., Song, X.-K., and Sun, L. (1996). A longitudinal study of emergency room visits and air pollution for Prince George, British Columbia. *Statistics in Medicine* 15, 823-836.

48. Jørgensen, B., Lundbye-Christensen, S., Song, P.X.-K., and Sun, L. (1999). A state space model for multivariate longitudinal count data. *Biometrika* 86, 169-181.

49. Kang, K.M. (1975). A comparison of estimators for moving average processes. Unpublished manuscript, Australian Bureau of Statistics.

50. Kelsall, J.E., Zeger, S.L., and Samet, J.M. (1999). Frequency domain log-linear models; air pollution and mortality. *Journal of the Royal Statistical Society*, Series C 48, 331-344.

51. Kuk, A.Y.C. (1999). The use of approximating models in Monte Carlo maximum likelihood estimation. *Statistics and Probability Letters* 45, 325-333.

52. Kuk, A.Y.C. and Cheng, Y.W. (1997). The Monte Carlo Newton-Raphson algorithm. *Journal of Statistical Computation and Simulation* 59, 233-250.

53. Lawless, J.F. (1987). Regression methods for Poisson process data. *Journal of the American Statistical Association* 82, 808-815.

54. Le Gland, F. and Mevel, L. (2000). Exponential forgetting and geometric ergodicity in hidden Markov models. *Mathematics of control, Signals, and Systems* 13, 63-93.

55. Leroux, B.G. (1992). Maximum likelihood estimation for hidden Markov models. *Stochastic Processes and their Applications* 40, 127-143.

56. Lii, K.S. and Rosenblatt, M. (1992). An approximate maximum likelihood estimation for non-Gaussian non-minimum phase moving average Processes.*Journal of Multivariate Analysis* 43, 272-299.

57. Lii, K.S. and Rosenblatt, M. (1996). Maximum likelihood estimation for nonGaussian nonminimum phase ARMA sequences. *Statistica Sinica* 6, 1-22.

58. Ling, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society*, Series B 67, 381-393.

59. MacDonald, I.L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series.* Chapman and Hall, London.

60. McCullagh, P. (1983). Quasi-likelihood functions. *Annals of Statistics* 11, 59-67.

61. McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models.* 2nd edition. Chapman and Hall, London.

62. Nelder, J.A. and Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal* 7, 308-313.

63. Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, Series A 135, 370-384.

64. Pan, J., Wang, H., and Yao, Q. (2007). Weighted least absolute deviations estimation for ARMA models with infinite variance. To be appeared in *Econometric Theory.*

65. Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, 186-199.

66. Rockafellar, R.T. (1970). *Convex Analysis.* Princeton University Press, Princeton, NJ.

67. Safi, S. and Zeroual, A. (2002). MA system identification using higher order cumulants application to modelling solar radiation. *Journal of Statistical Computation and Simulation* 72, 533-548.

68. Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika* 78, 719-727.

69. Wang, Y. (2002). Modeling Time Series of Count Data. Ph.D. Dissertation, Department of Statistics, Colorado State University, Fort Collins, CO.

70. Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* 61, 439-447.

71. Wiggins, R.A. (1978). Minimum entropy deconvolution. *Geoexploration* 16, 21-35.

72. Xue, D. and Deddens, J.A. (1992). Overdispersion negative binomial regression models. *Communications in Statistics: Theory and Methods* 21, 2215-2226.

73. Zeger, S.L. (1988). A regression model for time series of counts. *Biometrika* 75, 621-629.