2006

ASYMPTOTIC PROPERTIES OF ARMA-(I)GARCH MODELS

Pingguo Lu
Western University

ASYMPTOTIC PROPERTIES OF ARMA-(I)GARCH MODELS

(Spine title: ARMA-(I)GARCH Models)

(Thesis format: Monograph)

by
Pingguo Lu

Graduate Program
in
Statistics

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
FACULTY OF GRADUATE STUDIES

CERTIFICATE OF EXAMINATION

Supervisor	Examiners
Hao Yu	Peter Song
	Ian McLeod
	Reg Kulperger
	Mark Ressor

The thesis by

Pingguo Lu

entitled:

Asymptotic Properties of ARMA-(I)GARCH Models

is accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Date ______________________ ______________________
Chair of the Thesis Examination Board

Abstract

This thesis is motivated to investigate distribution theory of a quasi maximum likelihood estimator (QMLE) and test of goodness-of-fitting of an ARMA-(I)GARCH model.

We obtain asymptotic consistency and normality of the QMLEs based on an arbitrary likelihood kernel. It shows that the moment conditions of errors in the ARMA part and innovations in the GARCH part depend on the choice of likelihood kernel. For example, the asymptotic normality of QMLEs based on student t likelihood kernel holds with arbitrary small positive moment on error term and $2 - \iota$ moment on innovation term, where $0 \leq \iota < 1$. It also shows that the asymptotic efficiency of QMLEs depends on the choice of likelihood kernel and the distribution of innovation. For the pure GARCH model with nonzero constant mean, we show that the common practice of using the sample mean to center financial data is workable if the error term has finite variance. Consequently, we study some processes based on residuals of an ARMA-(I)GARCH model. We show that the k-th power partial sum process converges to a Brownian process plus two correction terms, where the correction terms always depend on ARMA-GARCH parameters. We also show that the

CUSUM and the self-normalized processes (standardized by the residual sample mean and variance) behave as if the residuals were asymptotically IID. Finally, applications of these results are exhibited with numerical examples.

Chapter 1 gives a brief introduction of financial return, econometric models such as ARMA, GARCH and their extensions, as well as model estimation and diagnosis.

Chapter 2 focuses on the distribution theory of one step QMLEs and two step QMLEs of an ARMA-(I)GARCH model. Special cases like pure ARMA and pure GARCH are considered too. Three specific examples with varied kernels are presented.

Chapter 3 deals with the high moment partial sum processes, the CUSUM and the self-normalized processes based on residuals of an ARMA-(I)GARCH model, originally proposed by Kulperger and Yu (2005) for a pure GARCH model.

In Chapter 4, we present some numerical examples of the applications of Chapter 2 & 3, for instance, efficiency of QMLEs based on different kernels, CUSUM statistic for testing ARMA-GARCH model structural changes, Jarque-Bera omnibus statistic for testing normality of the unobservable innovation of an ARMA-GARCH model. Finally some conclusions and discussions are put forward.

Keywords: ARMA-GARCH, ARMA-IGARCH, quasi-maximum likelihood estimation, two-step estimation, asymptotic consistency, asymptotic normality, asymptotic efficiency, residuals, high moment partial sum process, weak convergence, CUSUM, omnibus, skewness, kurtosis, $\sqrt{n}$ consistency.

To my family

(my mother Jiannan Li, my father Yafu (Lu) He,

my wife Luqiong Jiao and my son David Lu)

Acknowledgements

I would like to express my gratitude to all those who gave me support and encouragement to complete this thesis.

I am deeply indebted to my supervisor Dr. Hao Yu, for his advice and supervision during both my master and Ph.D study. His stimulating suggestions and encouragement helped me in all the time of research for and writing of this thesis. Without his guidance, support and good nature, I would never have been able to complete my thesis.

I greatly appreciate Dr. Ian McLeod and Dr. Bruce Jones for their offering me the chance to work at Statlab, where I gained lots of precious practical experience. I also would like to express my sincere thanks to the former Stalab manager, Dr. Ying Zhang for her help and encouragement.

I also want to thank Dr. David Bellhouse, Dr. Reg Kulperger, Dr. John Braun, Dr. Duncan Murdoch, Dr. Serge Provost, Dr. Ricardas Zitikis for their instructions. I learned a lot from their teachings.

I would like to thank Jennifer, Lisa and Suzanne, who have been kind enough to help me in their respective roles.

I greatly appreciate the examiners Dr. Peter Song, Dr. Mark Reesor, Dr. Ian McLeod and Dr. Reg Kulperger for their precious comments on this thesis.

I acknowledge the Department of Statistical & Actuarial Sciences and Faculty of Graduate Study at the University of Western Ontario for providing me scholarships to pursue the degree.

I am grateful to all my friends: Xiaoqiang, Jingming, Caixia, Andrei, Alex, Yulin, Douglas and so many others for having lots of discussions and fun with them while I stay in London, Ontario.

Especially, I would like to dedicate this thesis to my family, my mother Jiannan Li, my father Yafu Lu, my wife Luqiong Jiao and my son David, for their love, patience and understanding - they allowed me to spend most of the time on this thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter is devoted to some brief introduction of financial return, econometric models such as ARMA, GARCH and their extensions, as well as the model estimation and diagnosis.

1.1 Financial time series

1.1.1 Financial returns

To meet and satisfy the commercial and productive needs of various of investors and markets, many financial tools and derivatives such as stocks, options, forwards and bonds have been produced. We call these tools as financial assets.

The financial world is full of uncertainty and events take place every minute. Nonetheless, there are regularities and patterns to be identified. The fast expansion of financial markets and increasing variety and complexity of financial products give impetus to the development of econometrics. The aim is to make use of data, statistical inference methods and structural or descriptive modelling to deal with uncertainty and guide decisions in economics.

To investigate the regularities and patterns, we turn to the return, instead of the asset price itself. In econometric analysis, the return is conventionally defined as the logarithmic price changes:

Definition 1.1.1. Denote a financial asset with price p_t at time t (t is an integer) and price p_{t-1} at time $t-1$, the return is defined as:

$$R(t-1,t) = \log \frac{p_t}{p_{t-1}}.$$

We suppose the asset price includes the dividends if it has a dividend payment during the period.

1.1.2 Time series

Definition 1.1.2. Time series is a discrete stochastic process where the time index takes on a finite or countably infinite set of values, e.g. $\{X_t, -\infty < t < \infty\}$.

With respect to financial data, the price or return process of any asset naturally gives rise to a time series.

In the rest of this thesis, all quoted sequences like $\{Y_t, -\infty < t < \infty\}$, $\{\varepsilon_t, -\infty < t < \infty\}$, $\{\eta_t, -\infty < t < \infty\}$ are time series.

Generally an observed time series can be decomposed into three components: the trend (long term direction), the seasonal (periodic related movements) and the irregular or residuals (unsystematic, short term fluctuations). The trend and seasonal effects are deterministic and can be removed by regression, smoothing, difference or other methods. In the thesis, we focus on the nondeterministic part. From now on, without specification, by saying time series we mean the purely nondeterministic series with the deterministic components being removed from original series.

Weak stationary time series have time independent first and second moments. Define $\gamma_x(k) = \mathbf{E}[(X_t - \mathbf{E}(X_t))(X_{t-k} - \mathbf{E}(X_t))]$ as lag k autocovariance of X_t. The lag k autocorrelation function (ACF) of X_t is defined by $\rho_x(k) = \gamma_x(k)/\gamma_x(0)$. Intuitively, a stationary time series is characterized by its mean, variance and ACF. The lag k sample autocovariance and lag k sample autocorrelation function (SACF) are given by: $\hat{\gamma}_x(k) = n^{-1}\sum_{t=i+1}^{n}(X_t - \bar{X})(X_{t-i} - \bar{X})$, $\hat{\rho}_x(k) = \hat{\gamma}_x(k)/\hat{\gamma}_x(0)$, where $\bar{X} = n^{-1}\sum_{t=1}^{n} X_t$ is the sample mean.

Time series analysis accounts for the fact that data points taken over time may have a serial (such as autocorrelation) that should be accounted for. It plays an important role in evaluating any investment strategy, risk modelling and arbitrage. Analysis of a given asset's price or return time series could forecast its future price movements. A wide variety of mathematical and statistical tools have been developed for dealing with time series data.

A fundamental theorem in time series analysis is Wold's decomposition (c.f. Fuller (1996) pg. 96), which states that every weakly stationary and purely nondeterministic time series can be written as a linear combination of a sequence of uncorrelated random variables. The general Wold form of a stationary and ergodic time series is handy for theoretical analysis but is not practically useful for estimation purposes.

1.2 Econometric modelling of financial data

Portfolio mean-variance optimizing investors are assumed to evaluate the performance of their investment in terms of two summary statistics that represent the expected gain of a portfolio and its expected risk determined from asset volatility. These statistics correspond to the first two conditional moments of asset price or return. In statistical terms, volatility or risk is usually measured by variance, or standard deviation. Risk from an individual company is diversifiable, while a market component cannot be diversified.

In the following we will introduce several econometric models that are broadly utilized in exploring financial return time series.

1.2.1 ARMA model

Definition 1.2.1. $\{Y_t, -\infty < t < \infty\}$ is an $ARMA(P,Q)$ process if Y_t is stationary and, for every t,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_P Y_{t-P} + \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + ... + \varphi_Q \varepsilon_{t-Q} ,$$

where ε_t is white noise$(0, \sigma^2)$ and the polynomials $\mathcal{A}_\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - ... - \phi_P z^P$ and $\mathcal{B}_\varphi(z) = 1 + \varphi_1 z + \varphi_2 z^2 + ... + \varphi_Q z^Q$ have no common factors.

Define L as the back-shift operator such that $LY_t = Y_{t-1}$, $L^k Y_t = Y_{t-k}$. Then ARMA process can be written as $\mathcal{A}_\phi(L)Y_t = \mathcal{B}_\varphi(L)\varepsilon_t$. The process is a Moving Average (MA) process if $\mathcal{A}_\phi(z) \equiv 1$, or an Autoregressive (AR) process if $\mathcal{B}_\varphi(z) \equiv 1$.

Definition 1.2.2. An $ARMA(P,Q)$ process $\{Y_t, -\infty < t < \infty\}$ is causal if for all t, Y_t can be written as $Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$ with $\sum_{i=0}^{\infty} |\psi_i| < \infty$.

Definition 1.2.3. An $ARMA(P,Q)$ process $\{Y_t, -\infty < t < \infty\}$ is invertible if for all t, ε_t can be written as $\varepsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i}$ with $\sum_{i=0}^{\infty} |\pi_i| < \infty$.

Proposition 1.2.1. *An $ARMA(P,Q)$ process defined by $\mathcal{A}_\phi(L)Y_t = \mathcal{B}_\varphi(L)\varepsilon_t$ is causal if $\mathcal{A}_\phi(z) = 0$ has no roots inside or on the unit circle.*

Proposition 1.2.2. *An ARMA(P, Q) process defined by* $\mathcal{A}_\phi(L)Y_t = \mathcal{B}_\varphi(L)\varepsilon_t$ *is invertible if* $\mathcal{B}_\varphi(z) = 0$ *has no roots inside or on the unit circle.*

1.2.2 ARCH model and its applications

Empirically financial asset returns tend to be leptokurtotic and show volatility clustering: large changes tend to be followed by large changes, and small changes tend to be followed by small changes (Mandelbrot 1963). Volatility clustering and heavy tailed returns are closely related. Usually financial asset returns also show strong autocorrelation among squared returns. If such patterns are present in a time series, We say the data has ARCH effect.

The ARMA models successfully captures the movements of conditional mean. But it assumes that the conditional variance is time-invariant and contains no past information. The measure of the unconditional variance does not recognize that there may be predictable patterns in stock market volatility.

Predictable volatility implies investors can predict the risk and uncertainty based on current and past information. An important role of this prediction is that for periods where an investor has forecasted prices to be very volatile, he/she should either exit the market or require a large premium as a compensation for bearing an unusual high risk. To assess the variation of risk, an approach of involving conditional heteroscedasticity is required. Engle (1982) proposed Autoregressive Conditional Heteroscedasticity (ARCH) model, which plays a revolutionary role in modelling of time series variances. Because of his contribution, Engle won the 2003 Nobel Prize in Economics.

Definition 1.2.4. An ARCH(p) process $\{Y_t, -\infty < t < \infty\}$ with constant mean term c takes the form

$$Y_t - c = \eta_t \sigma_t, \tag{1.2.1}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i (Y_{t-i} - c)^2, \tag{1.2.2}$$

where $\alpha_0 > 0, \alpha_i \geq 0, 1 \leq i \leq p, c \in \mathbb{R}$, are constants. We also assume that

$$\{\eta_t, -\infty < t < \infty\} \tag{1.2.3}$$

is a sequence of random variables identically and independently distributed with $\mathbf{E}(\eta_0) = 0$ and $\mathbf{E}(\eta_0^2) = 1$ (IID(0,1)).

ARCH specifies the conditional variance as a linear function of past squared returns. It explains the volatility clustering and heavy-tailed non-Gaussian distribution of the returns.

Volatility has become a very important concept in different areas in financial theory and practice. It has been used in risk management, portfolio selection, derivative pricing, etc. As pointed out by Gouriéroux (1997), there are two main categories of potential applications of ARCH. The first category involves examining several economic or financial theories concerning the stock or other financial assets. The second one is basically operational and related to the intervention of banks on the market, such as risk management, choice of optimal portfolios, hedging portfolios, value at risk, sizes and times of block trading. The second category is often subject to some confidentiality restrictions, contrary to the first one, which is of a more global use.

1.2.3 GARCH model

If ARCH effect is present, we fit time series with an ARCH model. In practice it is often found that a large number p of lags is needed, and thus a large number of parameters is required to obtain a good model fit. Inspired by the idea of the ARCH model and the ARMA model, ARCH was generalized (GARCH) by Bollerslev (1986) by adding the past conditional variance to the conditional variance term.

Definition 1.2.5. A GARCH(p, q) process $\{Y_t, -\infty < t < \infty\}$ with constant mean term c is of the form:

$$Y_t - c = \varepsilon_t, \tag{1.2.4}$$

$$\varepsilon_t = \eta_t \sigma_t \ \text{ and } \ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{1.2.5}$$

where $\alpha_0 > 0, \alpha_i \geq 0, 1 \leq i \leq p, \beta_j \geq 0, 1 \leq j \leq q, c \in \mathbb{R}$, are constants. We also assume that

$$\{\eta_t, -\infty < t < \infty\} \tag{1.2.6}$$

is a sequence of IID(0,1) random variables.

The process reduces to Engle's ARCH(p) process if $q = 0$.

When $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1$, model (1.2.4)-(1.2.5) is called integrated GARCH (IGARCH), due to the fact: $\mathbf{E}\sigma_t^2 = \infty$.

Define

$$A_t = \left(\begin{array}{ccc|ccc} \alpha_1 \eta_t^2 & \cdots & \alpha_p \eta_t^2 & \beta_1 \eta_t^2 & \cdots & \beta_q \eta_t^2 \\ & I_{p-1} & O_{(p-1)\times 1} & & O_{(p-1)\times q} & \\ \hline \alpha_1 & \cdots & \alpha_p & \beta_1 & \cdots & \beta_q \\ & O_{(q-1)\times p} & & & I_{q-1} & O_{(q-1)\times 1} \end{array} \right),$$

where I_k is $k \times k$ identity matrix.

Nelson (1990) shows that the model (1.2.4)-(1.2.5) with $p = q = 1$ has a unique stationary solution of ε_t if and only if $\mathbf{E}\log(\beta_1 + \alpha_1\eta_0) < 0$. The general case was investigated by Bougerol and Picard (1992a, b). They showed that a unique strictly stationary ε_t sequence exists if and only if

$$\gamma(A_t) = \inf_{1\leq t<\infty} \frac{1}{t}\mathbf{E}\left(\log|A_tA_{t-1}\cdots A_1|\right) < 0 \quad a.s, \tag{1.2.7}$$

where we use $|\cdot|$ to denote the absolute value of a scalar, or maximum norm of vectors or matrices. The definition of $\gamma(A_t)$ does not depend on the choice of a norm on the space of the $(p+q) \times (p+q)$ matrices.

Ling (2005) shows that for $0 < \iota \leq 1$, if there exists an integer i_0 such that

$$\mathbf{E}\left|\prod_{k=0}^{i_0-1} A_k\right|^{\iota} < 1, \tag{1.2.8}$$

then $\{\varepsilon_t\}$ is strictly stationary and ergodic with $\mathbf{E}|\varepsilon_t|^{2\iota} < \infty$.

Define $v_t = \varepsilon_t^2 - \sigma_t^2$. By rearranging (1.2.5), we have

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{m}(\alpha_i + \beta_i)\varepsilon_{t-i}^2 + v_t - \sum_{j=1}^{q}\beta_j v_{t-j}, \tag{1.2.9}$$

where $m = \max(p,q)$, $\alpha_i = 0$ for $i > p$, and $\beta_i = 0$ for $i > q$. Thus a GARCH model can be expressed as an ARMA model with ε_t^2. Given the ARMA representation of the GARCH model, many properties of the GARCH model follow easily from those of the corresponding ARMA model. With this representation, the GARCH model is capable of explaining many stylized facts like: volatility clustering, fat tails, and volatility mean reversion.

1.2.4 Extensions of GARCH

In many cases, the basic GARCH model provides a reasonably good model for analyzing financial time series and estimating conditional volatility. However, there are some aspects of the model which can be improved to better capture the characteristics and dynamics of a particular time series.

This section introduces several extensions of the basic GARCH model.

GARCH (p, q) models successfully capture heavy tailed returns and volatility clustering. But positive and negative shocks have the same effect on volatility since the model depends only on the squared previous shocks. It fails to capture the "leverage effect", which means volatility responds more rapidly to falls (bad news) in financial market than to corresponding rises (good news). Extended models like Exponential GARCH or EGARCH (Nelson, 1991), and Threshold ARCH or TARCH (Zakoian (1990), Glosten, Jaganathan, and Runkle (1993)) capture this asymmetric responding mechanism.

In the EGARCH model, conditional variance in (1.2.5) is substituted with

$$\log \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i (\varpi_i \eta_{t-i} + |\eta_{t-i}|) + \sum_{j=1}^{q} \beta_j \log \sigma_{t-j}^2. \tag{1.2.10}$$

The asymmetric news impact (leverage effect) is potential since it allows volatility to respond more rapidly to bad news. Note that when η_{t-i} is positive or there is good news, the total effect of η_{t-i} is $\alpha_i(1 + \varpi_i)$. In contrast, when η_{t-i} is negative or there is bad news, the total effect of η_{t-i} is $\alpha_i(\varpi_i - 1)$. Bad news can have a larger impact on volatility, and the value of ϖ_i would be expected to be negative. EGARCH also releases the nonnegativity constraints in the GARCH model parameters, which are too restrictive.

In the TARCH model, conditional variance in (1.2.5) is substituted with

$$\sigma_t^s = \alpha_0 + \sum_{i=1}^{p}(\alpha_i^+ I\{\eta_{t-i} > 0\}|\varepsilon_t|^s + \alpha_i^- I\{\eta_{t-i} \leq 0\}|\varepsilon_t|^s) + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^s,$$

where I$\{\cdot\}$ denotes the indicator function and $s > 0$. This model allows response of volatility to news with different coefficients for good and bad news, but maintains the assertion that the minimum volatility will result when there is no news. That is, depending on whether η_{t-i} is above or below the threshold value of zero, η_{t-i} has different effects on the conditional variance. So one would expect that the α_i^- is bigger than α_i^+.

Ding, Granger and Engle (1993) proposed the a Power GARCH (PGARCH) model. In PGARCH, the conditional variance in (1.2.5) is substituted by

$$\sigma_t^d = \alpha_0 + \sum_{i=1}^{p}\alpha_i\big(|\varepsilon_t| + \varpi_i\varepsilon_t\big)^d + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^d,$$

where d is a positive exponent, and ϖ_i denotes the coefficient of leverage effects. Note that when d = 2, PGARCH reduces to the basic GARCH model with leverage effects.

In response to the finding that squares of return series tend to have very slowly decaying autocorrelations, Baillie, Bollerslev and Mikkelson (1996) proposed Fractionally Integrated GARCH (FIGARCH). The main characterization of a FIGARCH model is that conditional variances exhibit not only short-run dynamics of the ARMA type, as in the standard GARCH model, but also the long-run persistence that decays slowly at hyperbolic rates (instead of the usual exponential rates as of the GARCH model). A Fractionally Integrated GARCH (p,d,q) process ε_t is defined as

$$\left(1-\sum_{i=1}^{p}\alpha_i L^i\right)(1-L)^d\varepsilon_t^2=\alpha_0+\left(1-\sum_{j=1}^{q}\beta_j L^j\right)(\varepsilon_t^2-\sigma_t^2), \tag{1.2.11}$$

where $0 \leq d \leq 1$. The corresponding conditional variance σ_t^2 can be expressed more explicitly as:

$$\left(1-\sum_{j=1}^{q}\beta_j L^j\right)\sigma_t^2=\alpha_0+\left(1-\sum_{j=1}^{q}\beta_j L^j\right)\varepsilon_t^2-\left(1-\sum_{i=1}^{p}\alpha_i L^i\right)(1-L)^d\varepsilon_t^2.$$

The fractional differencing operator $(1-L)^d$ that allows the process ε_t^2 to have a long memory. Baillie (1996) argues that the presence of FIGARCH may explain the common findings of IGARCH in modelling high-frequency financial data.

In financial investment, high risk is often associated with a expected high return. Engle, Lilien and Robins (1987) proposed to extend the basic GARCH model so that the conditional volatility can generate a risk premium which is part of the expected returns. This extended GARCH model is often referred to as GARCH-in-mean (GARCH-M) model.

The GARCH-M model extends the conditional mean equation (1.2.4) as follows:

$$Y_t = c + mf(\sigma_t) + \varepsilon_t,$$

where m is a constant and f can be any arbitrary function of volatility σ_t, i.e. $f(\sigma_t) = \sigma_t$, $f(\sigma_t) = \sigma_t^2$, or $f(\sigma_t) = \ln \sigma_t$.

Other extensions of GARCH are skipped here.

1.2.5 ARMA(P,Q)-GARCH(p, q)

The GARCH model successfully captures the movements of conditional volatility. Empirically the conditional mean is dynamic rather than zero or constant. As pointed out by Francq and Zakoïan (2004), in economic applications, it is a common practice to fit financial return series by an autoregressive moving average (ARMA) model with GARCH errors. The ARMA-GARCH model combines an ARMA model for modelling the dynamic conditional mean and a GARCH model for modelling the dynamic conditional volatility. An ARMA(P,Q)-GARCH(p, q) sequence $\{Y_t, -\infty < t < \infty\}$ is of the form:

$$Y_t - c = \sum_{i=1}^{P} \phi_i (Y_{t-i} - c) + \varepsilon_t + \sum_{j=1}^{Q} \varphi_j \varepsilon_{t-j}, \tag{1.2.12}$$

$$\varepsilon_t = \sigma_t \eta_t \text{ and } \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{1.2.13}$$

where the innovations $\{\eta_t, -\infty < t < \infty\}$ is a sequence of non-degenerate IID(0,1) random variables, $c \in \mathbb{R}, \phi_l \in \mathbb{R}, 1 \le l \le P, \varphi_k \in \mathbb{R}, 1 \le k \le Q, \alpha_0 > 0, \alpha_i \ge 0, 1 \le i \le p, \beta_j \ge 0, 1 \le j \le q$ are constants. In this model, we refer to $\{\varepsilon_t, -\infty < t < \infty\}$ as the sequence of GARCH errors for the ARMA model and $\{\eta_t, -\infty < t < \infty\}$ as the sequence of GARCH innovations.

The ARMA-GARCH model can be extended by adding exogenous explanatory variables in the conditional mean and (or) conditional variance equations. For instance, the Capital Asset Pricing Model (CAPM) implies that stock returns should be related to the returns of a market index. And it is widely believed that trading volume affects the volatility. Then both market index and trading volume could be potential explanatory variables candidates.

1.3 Testing the ARCH effect

To check ARCH effect, we plot the ACF of the series itself versus the ACF of squared series. It is usually the case that there is little serial correlation in the time series itself, while the squared series exhibits strong autocorrelation. Since the squared series measures the second order moment of the original time series, it indicates that the variance of the series based on its past history may change over time.

After we fit a financial time series, it is usually a good practice to test for the presence of ARCH effect in the residuals. Suppose we fit the data with an ARMA model by assuming the error term is white noise. If the white noise assumption of error term does not hold, an ARCH effect is present in the residuals. It leads to serious model mis-specification and results in inappropriate standard error of parameter estimator. On the other hand, If there is no ARCH effect in the residuals, then the ARCH model is unnecessary.

We assume that linear serial dependence inside the original series is removed and any remaining serial dependence must be due to conditional heteroscedasticity, which is not captured by the model. Detection of ARCH effect in a series is actually a joint test for heteroscedasticity of the residuals e_t.

1.3.1 Engle's Lagrange multiplier test

Since the ARCH model has the form of an autoregressive model, Engle (1982) proposed the Lagrange Multiplier (LM) test for ARCH effect. LM test is defined as a test of H_0 : ε_t has a constant variance versus H_a : The conditional variance is an ARCH(p) process, that is to test whether the ARCH parameters are all zeros.

Let residuals be $e_t = \hat{\varepsilon}_t/\hat{\sigma}_t$, where $\hat{\varepsilon}_t$ and $\hat{\sigma}_t$ are estimators of ε_t and σ_t with finite sample respectively. The test is based on the regression of e_t^2 on $e_{t-1}^2, e_{t-2}^2, ...e_{t-p}^2$. The test statistic is nR^2, where n is the sample size, and R^2 is the sample multiple correlation coefficient computed from the regression. Under the null hypothesis that there is no ARCH effect, the test statistic is asymptotically distributed as χ_p^2. From this test, it can be seen if the data is homoscedastic, then the variance cannot be predicted and variation in e_t is purely random. If the ARCH effect is present, then the variation can be predicted by lagged values of squared residuals. It should be mentioned that the test rejects if the residuals themselves contain some remaining autocorrelations or other form of non-linearity. So we can not simply assume the ARCH effect is necessarily present when the test rejects.

1.3.2 McLeod-Li test

Since ε_t^2 in (1.2.5) can be written as:

$$\varepsilon_t^2 = \alpha_0 + \left(\sum_{i=1}^{p} \alpha_i(L^i) + \sum_{j=1}^{q} \beta_j(L^j)\right)\varepsilon_t^2 + \left(1 - \sum_{j=1}^{q} \beta_j(L^j)\right) v_t,$$

where $v_t = \varepsilon_t^2 - \sigma_t^2$ is not autocorrelated, so $\{\varepsilon_t^2, -\infty < t < \infty\}$ follows a ARMA (max(p,q), q) model.

McLeod and Li (1983) proposed a test for diagnostic checking of possible departures from the linear ARMA model assumption. They used the autocorrelation of the squares of the residuals rather than the residuals themselves as in the Ljung-Box test. They showed the sample autocorrelation of e_t^2 have asymptotic variance $1/n$. McLeod and Li statistic tests whether the first k autocorrelations for the squared residuals are collectively small in magnitude. The statistic is defined as:

$$Q_{ML} = n(n+2)\sum_{i=1}^{k} \frac{\hat{\rho}_{e^2}^2(i)}{(n-i)} ,$$

where n is the sample size, $\hat{\rho}_{e^2}(i) = \sum_{t=i+1}^{n} e_t^2 e_{t-i}^2 / \sum_{t=1}^{n} e_t^4$ is sample autocorrelation of the squared residual series at lag i, and k is the number of lags being tested. Under the null hypothesis of no ARCH effect in the data, McLeod-Li statistic is asymptotically χ_k^2 distributed.

Luukkonen, Saikkonen and Terasvirta (1988) pointed out that McLeod-Li test is asymptotically equivalent to Engle's Lagrange multiplier test.

For more tests, see Li (2004, p100-p112).

1.4 Model estimation

1.4.1 Quasi maximum likelihood estimation

Usually model parameters are estimated by maximum likelihood method. But maximum likelihood estimation cannot be applied to the model if we do not know the exact distribution of the random variables. Sometimes we can still estimate the model parameters by presuming the variables come from a particular distribution.

Definition 1.4.1. A maximum likelihood estimator based on a likelihood function with misspecified density is called a QMLE.

1.4.2 Large sample estimation properties

For data with a large sample size, we have some established asymptotic consistency and normality results for the estimators. Let $\hat{Q}_n(\theta)$ be an objective function, such that

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} \hat{Q}_n(\theta), \tag{1.4.1}$$

where Θ is the parameter space and is assumed to be compact. Usually $\hat{Q}_n(\theta)$ has the form:

$$\hat{Q}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} f(y_i, \theta),$$

where $y_1, \cdots, y_n$ are a realizations of an IID sequence.

With different functions f, the resulting estimators can be (quasi) maximum likelihood estimator, least-squares estimator, or generalized moment estimator.

Theorem 1.4.1. *(Consistency Theorem, Newey and Mcfadden 1994)*

If there is a function $Q_0(\theta) = \mathbf{E}f(y_t, \theta)$ such that (i) $Q_0(\theta)$ is uniquely maximized at the θ_0; (ii) Θ is compact; (iii) $Q_0(\theta)$ is continuous; (iv) $\hat{Q}_n(\theta)$ converge uniformly in Θ in probability to $Q_0(\theta)$, then

$$\hat{\theta}_n \xrightarrow{p} \theta_0.$$

Theorem 1.4.2. *(Normality Theorem, Newey and Mcfadden 1994)*

Suppose that $\hat{\theta}_n$ satisfies (1.4.1), $\hat{\theta}_n \xrightarrow{p} \theta_0$, and (i) θ_0 is in the interior of Θ; (ii) $\hat{Q}_n(\theta)$ is twice continuously differentiable in an open neighborhood Θ_0 of θ_0; (iii) $\sqrt{n}\partial\hat{Q}_n(\theta)/\partial\theta|_{\theta=\theta_0} \xrightarrow{d} N(0,\Sigma)$; (iv) there is $H(\theta)$ that is continuous at θ_0 and

$\sup_{\theta\in\Theta_0} |\partial^2 \hat{Q}_n(\theta)\partial\theta^2 - H(\theta)| \xrightarrow{p} 0$; *(v)* $H = H(\theta_0)$ *is nonsingular. Then*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, H^{-1}\Sigma H^{-1}).$$

1.5 Model diagnostic checks

Mis-specification may result in inconsistency and loss of efficiency in estimating parameters. Consequently it will lead to poor forecasts. It should be noted that in many financial econometric models the conditional variance equations play a major role. Reliable estimation and inference of the conditional variance depend on well-specified conditional variance models. Therefore testing goodness-of-fit after fitting the model becomes an important issue. The following tests give some routines of testing various features of ARMA-GARCH models.

(i) ARCH effect or randomness of residuals

Diagnostic test for conditional heteroscedasticity models applied in the literature can be divided into three categories: Portmanteau test of the Box-Pierce-Ljung type, Lagrange multiplier (LM) test and other residual-based diagnostics. The Box-Pierce-Ljung portmanteau statistic is perhaps the most widely used diagnostic test. It is readily computable from the standardized residuals and has been applied in many empirical works for model diagnostic checks (see, for example, the papers by Bollerslev (1990), Baillie and Myers (1991) and Karolyi (1995)). Ling and Li (1997) further developed this work and derived the asymptotic distribution of the portmanteau statistic in the multivariate case. The Ling-Li statistic is based on the serial correlation coefficients of the transformed vector of residuals.

If the model is successful at modelling the serial correlation structure in the conditional mean and conditional variance, then there should be no autocorrelation left in the standardized residuals and squared standardized residuals. This can be checked by using the Ljung-Box test with standardized residuals and McLeod-Li test with squared standardized residuals. In both cases, we will reject the null hypothesis (that there is no autocorrelation left) if the statistic is large.

(ii) Distribution of innovation

The normal distribution for the innovations is usually assumed. If the model is correctly specified then the estimated standardized residuals should behave like standard normal random variables. To evaluate the normality assumption, a QQ-plot of the standardized residuals or Jarque-Bera normality test can be performed.

(iii) Change point problem

Another key assumption is that sequence is stationary or the model parameters stay constant through time. Parameter instability is evidence of model misspecification and standard econometric theory no-longer applies. Robust estimation requires at a minimum that the conditional mean and variance be correctly specified. However, GARCH models are rarely tested for structural breaks.

Though model diagnostic checks based on standardized residuals can be used to compare the effectiveness of different econometric models. Selecting the best model for a particular data set still can be a daunting task. Since GARCH models can be treated as ARMA models of squared residuals, traditional model selection criteria such as Akaike information criterion (AIC) and Bayesian information criterion (BIC) can also be used for selecting models.

1.6 Objectives

Later in this thesis, I will focus on an ARMA(P, Q)-GARCH(p, q) model. The general conditions of distribution theory of QMLE for a general model have been set up as in Section 1.4.2. As for an ARMA(P,Q)-GARCH(p,q) model, it is an open problem on the distribution theory of QMLE when GARCH innovation η_t has no $4th$ moment. I will solve this problem by applying an arbitrary likelihood kernel to build the likelihood function. The details will be given in Chapter 2.

Standard goodness-of-fit tests (such as Kolmogorov-Smirnov test) and other model diagnostic tests based on the empirical process of an ARMA-GARCH residuals have been found to be invalid. Kulperger and Yu (2005) studied the high moment partial sum processes, the CUSUM and the self-normalized processes based on residuals of an (I)GARCH model. The results are applied to the goodness-of-fit tests and model diagnostic test. Can we extend their results to an ARMA(P,Q)-GARCH(p,q) model? This question will be answered in Chapter 3.

More tests (such as scaling issue in S-plus Finmetrics module, efficiency of QMLE, structural change problems, distribution of innovations) and numerical examples are presented in Chapter 4.

Chapter 2

QMLE of ARMA-GARCH

In this Chapter, we obtain asymptotic consistency and normality of a class of global QMLEs based on arbitrary likelihood kernels and weak moment conditions on both ε_t and η_t. Two step estimation is also studied.

This chapter is organized as follows: Section 1 exhibits some existing distribution theorems of QMLE. Section 2 presents the assumptions and results. Some examples are given in Section 3. Section 4 is devoted to the proofs. Some lengthy expansion of $\varepsilon_t(\gamma)$, $\sigma_t^2(\lambda)$ and etc., as well as proof of Proposition 2.1 are given in the end of this Chapter as an appendix.

2.1 Existing Distribution Theories of QMLE

Without particular specification, the QMLEs mentioned in this section are based on normal density. The asymptotic properties of QMLE for ARMA-ARCH were first presented by Weiss (1986) under assumption of finite fourth moment on ε_t. The problem of finding weaker conditions for asymptotic properties of QMLE has attracted much attention in the literature. Lee and Hansen (1994) and Lumsdaine

(1996) obtained the asymptotic consistency and normality of QMLE for GARCH(1, 1) and IGARCH(1, 1) with nonzero constant mean. They require a strict condition on the distribution of η_t and the values of parameters. The former requires $\mathbf{E}\eta_0^4 < \infty$, and the later requires $\mathbf{E}\eta_0^{32} < \infty$. Linton (1997) studied an asymptotic expansion of QMLE for GARCH(1,1) and IGARCH(1,1) with nonzero constant mean c_0. He showed if η_0 is symmetric about 0 and has more than 6th finite moment, then QMLE of $(\alpha_0, \alpha_1, \beta_1)$ are asymptotically independent of any $\sqrt{n}$-consistent estimator of c_0.

Berkes, Horváth, and Kokoszka (2003) extended the above results to a general GARCH(p, q) model with mean term zero and relaxed the conditions in Lee and Hansen (1994) and Lumsdaine (1996). They showed that the asymptotic normality of QMLE for a GARCH(p, q) holds with $\sum_{j=1}^{q} \beta_j < 1$ and $\mathbf{E}|\eta_0|^{4+\zeta} < \infty$ for some $\zeta > 0$. Hall and Yao (2003) obtained the asymptotic normality of QMLE for a GARCH(p, q) with mean term zero under $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$ and $\mathbf{E}\eta_0^4 < \infty$. Berkes and Horváth (2003) showed that QMLE of GARCH(p, q) parameters based on a normal kernel cannot be $\sqrt{n}$-consistent if $\mathbf{E}|\eta_0|^4 = \infty$. Berkes and Horváth (2004) proposed a class of estimators bases on an arbitrary likelihood kernel for a GARCH(p, q) model and showed that the QMLE based on double exponential density are better than those based on the standard normal density if the tail of the distribution of η_0 is polynomial. They showed the moment requirement of η_0 depends on the choice of likelihood kernel. Asymptotic normality of QMLE based on double exponential likelihood kernel holds if $\mathbf{E}|\eta_0|^2 < \infty$.

Francq and Zakoïan (2004) obtained the asymptotic consistency and normality results of QMLE for (I)GARCH(p, q). They removed the condition of $\lim_{x \to 0} x^{-\zeta} P\{\epsilon_0^2 \leq$

$x\} = 0$, $\zeta > 0$ in Berkes, Horváth, and Kokoszka (2003). They also relaxed the parameter restrictions of requiring all parameters to be in the interior of the parameter space for asymptotic consistency result. This is essential to handle situations of over-identification. They showed asymptotic normality holds with $\mathbf{E}\eta_0^4 < \infty$.

The above results have been extended to ARMA-GARCH. Ling and Li (1997) obtained the consistency and normality of local QMLE for ARMA-GARCH under $\mathbf{E}\varepsilon_0^4 < \infty$. Francq and Zakoïan (2004) obtained the global QMLE of an ARMA-GARCH model with a weak condition. Asymptotic normality result holds with both $\mathbf{E}\varepsilon_0^4 < \infty$ and $\mathbf{E}\eta_0^4 < \infty$. Ling (2005) proposed a self-weighted QMLE (SWQMLE), which is asymptotically normally distributed under only a fractional moment of ε_0. By using the SWQMLE as an initial value, he obtained the local QMLE for ARMA-(I)GARCH. In both global and local cases, asymptotic normality requires $\mathbf{E}\eta_0^4 < \infty$. In general, it is hard to compare the efficiency of SWQMLE and QMLE. However, Ling showed that SWQMLE is less efficient than QMLE when $\eta_t \sim N(0,1)$. For additional related works, see Li and Ling (1997, 1998, 2003), Li, Ling and McAleer (2002) and Ling and McAleer (2003).

The parameters in the conditional volatility are restricted by the moment conditions on ε_0. To make this clear, we present two examples. The example of ARCH(1) was given by Ling (2005). In ARCH(1) model with $\eta_0 \sim N(0,1)$, the parameter space of α_1 is $(0,1)$ if $\mathbf{E}\varepsilon_0^2 < \infty$, or $(0,\ 1/\sqrt{3})$ if $\mathbf{E}\varepsilon_0^4 < \infty$. For GARCH(1,1) with $\eta_0 \sim N(0,1)$, if $\mathbf{E}\varepsilon_0^2 < \infty$, then $\alpha_1 + \beta_1 < 1$. If $\mathbf{E}\varepsilon_0^4 < \infty$, then $\alpha_1 \in (0,\ 1/\sqrt{3})$ and $\beta_1 \in (0, \sqrt{1-2\alpha_1^2} - \alpha_1)$, where $\sqrt{1-2\alpha_1^2} - \alpha_1$ goes to 0 quickly as α_1 is close to $1/\sqrt{3}$. The space is more restrictive for higher order of GARCH. It can be seen that

$\mathbf{E}\varepsilon_0^4 < \infty$ is a strong condition.

In addition, the parameter space depends on the distribution of η_0. For example, in ARCH(1) model with $\sqrt{2}\eta_0$ being double exponential distribution, the parameter space of α_1 is $(0, 1)$ if $\mathbf{E}\varepsilon_0^2 < \infty$, or $(0,\ 0.408\cdots)$ if $\mathbf{E}\varepsilon_0^4 < \infty$. In addition, there may be some connections between the moments of ε_t and η_t. Ling (2005) showed if $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$, $\mathbf{E}|\eta_0|^{2+\zeta} < \infty$ implies $\mathbf{E}|\varepsilon_0|^{2+\zeta^*} < \infty$, where $0 < \zeta < 1$ and $0 < \zeta^* < \zeta$.

2.2 Assumptions and Theorems of QMLE

Denote $\gamma = (c, \phi_1, ..., \phi_P, \varphi_1, ..., \varphi_Q)^T, \delta = (\alpha_0, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q)^T, \lambda = (\gamma^T, \delta^T)^T$. Denote parameter space $\Theta = (\Theta_\gamma, \Theta_\delta) = ((\Theta_\phi, \Theta_\varphi), (\Theta_\alpha, \Theta_\beta)) \subset \mathbb{R}^{P+Q+1} \times \mathbb{R}^+ \times \mathbb{R}_0^{p+q}$, where $\mathbb{R}^+ = (0, +\infty)$, $\mathbb{R}_0 = [0, +\infty)$.

We write the model (1.2.12) and (1.2.13) in parametric from as:

$$\varepsilon_t(\gamma) = (Y_t - c) - \sum_{i=1}^{P} \phi_i (Y_{t-i} - c) - \sum_{j=1}^{Q} \varphi_j \varepsilon_{t-j}(\gamma), \tag{2.2.1}$$

$$\eta_t(\lambda) = \frac{\varepsilon_t(\gamma)}{\sigma_t(\lambda)} \quad \text{and} \quad \sigma_t^2(\lambda) = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2(\gamma) + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2(\lambda)\ . \tag{2.2.2}$$

The true parameter values are unknown and denoted as $\lambda_0 = (\gamma_0^T, \delta_0^T)^T$. Throughout the rest of this Chapter, we assume $\lambda_0 \in \Theta$. Clearly $\varepsilon_t = \varepsilon_t(\gamma_0)$, $\sigma_t^2 = \sigma_t^2(\lambda_0)$, and $\eta_t = \varepsilon_t(\gamma_0)/\sigma_t(\lambda_0) = \varepsilon_t/\sigma_t$.

To make the model be identified and stationary, we introduce following assumptions:

Assumption 1. Θ is compact.

Assumption 2. For each $\gamma \in \Theta_\gamma$, $\mathcal{A}_\gamma(z)$ and $\mathcal{B}_\gamma(z)$ have no common roots, the roots of $\mathcal{A}_\gamma(z)\mathcal{B}_\gamma(z) = 0$ are outside of unit circle. $\alpha_P \neq 0$ or $\beta_Q \neq 0$.

Assumption 3. For each $\delta \in \Theta_\delta$, $\sum_{j=1}^{q} \beta_j < 1$, $\mathcal{A}_\delta(z)$ and $\mathcal{B}_\delta(z)$ have no common roots. $\mathcal{A}_\delta(1) \neq 0$ and $\alpha_p + \beta_q \neq 0$.

Remark 2.2.1. *Assumption 2 implies the stationarity, invertibility and identifiability of model (2.2.1). Assumption 3 implies that model (2.2.2) is minimal in the sense that there is no pair (p^*, q^*) such that $p^* < p$, $q^* < q$ and (2.2.2) holds. In particular, if $P = 0$, then $\mathcal{A}_\gamma(z) = 1$; if $Q = 0$, then $\mathcal{B}_\gamma(z) = 1$; if $q = 0$, then $\mathcal{B}_\beta(z) = 1$.*

In Chapter 1, we have defined $\mathcal{A}_\phi(z) = 1 - \sum_{i=1}^{P} \phi_i z^i$, $\mathcal{B}_\varphi(z) = 1 + \sum_{j=1}^{Q} \varphi_j z^j$. Similarly we define $\mathcal{A}_\alpha(z) = \sum_{i=1}^{p} \alpha_i z^i$, $\mathcal{B}_\beta(z) = 1 - \sum_{j=1}^{q} \beta_j z^j$, and

$$\mathcal{C}_\gamma(z) = \mathcal{B}_\varphi{}^{-1}(z)\mathcal{A}_\phi(z) = \sum_{i=0}^{\infty} a_\gamma(i) z^i, \quad \mathcal{C}_\delta(z) = \mathcal{B}_\beta{}^{-1}(z)\mathcal{A}_\alpha(z) = \sum_{i=0}^{\infty} a_\delta(i) z^i,$$

$$\mathcal{A}_\phi^{-1}(z) = \sum_{i=0}^{\infty} a_\phi(i) z^i, \ \mathcal{B}_\varphi^{-1}(z) = \sum_{i=0}^{\infty} a_\varphi(i) z^i, \ \text{ and } \ \mathcal{B}_\beta^{-1}(z) = \sum_{i=0}^{\infty} a_\beta(i) z^i,$$

where the expressions of $a_\gamma(i)$ and $a_\delta(i)$ are given in Appendix A.2. Lemma 2.4.1 shows that the absolute summation of $a_\gamma(i)$, $a_\delta(i)$, $a_\phi(i)$, $a_\varphi(i)$ and $a_\beta(i)$ are finite respectively. Thus (2.2.1) and (2.2.2) can be rewritten as:

$$\varepsilon_t(\gamma) = \mathcal{B}_\varphi{}^{-1}(L)\mathcal{A}_\phi(L)(Y_t - c) = \sum_{i=0}^{\infty} a_\gamma(i)(Y_{t-i} - c) \ , \tag{2.2.3}$$

$$\begin{aligned} \sigma_t^2(\lambda) &= \mathcal{B}_\beta{}^{-1}(1)\alpha_0 + \mathcal{B}_\beta{}^{-1}(L)\mathcal{A}_\alpha(L)\varepsilon_t^2(\gamma) \\ &= \mathcal{B}_\beta{}^{-1}(1)\alpha_0 + \sum_{i=0}^{\infty} a_\delta(i)\varepsilon_{t-i}^2(\gamma) \ . \end{aligned} \tag{2.2.4}$$

In an application, it is impossible to have an infinite number of observations of Y_t. Hence the initial values are replaced with some fixed constants, which are neither random nor functions of the parameters. However this does not affect the asymptotic results (Ling and McAleer, 2003).

Given initial value $Y_0, ..., Y_{1-P}, \tilde{\varepsilon}_0, ..., \tilde{\varepsilon}_{1-\max(p,Q)}, \tilde{\sigma}_0, ..., \tilde{\sigma}_{1-q}$, then $\tilde{\varepsilon}_t(\gamma)$, $\tilde{\sigma}_t^2(\lambda)$ and $\tilde{\eta}_t(\lambda)$ for $t = 1, ..., n$ can be computed from following equations:

$$\tilde{\varepsilon}_t(\gamma) = (Y_t - c) - \sum_{i=1}^{P} \phi_i (Y_{t-i} - c) - \sum_{j=1}^{Q} \varphi_j \tilde{\varepsilon}_{t-j}(\gamma) , \tag{2.2.5}$$

$$\tilde{\eta}_t(\lambda) = \frac{\tilde{\varepsilon}_t(\gamma)}{\tilde{\sigma}_t(\lambda)} \text{ and } \tilde{\sigma}_t^2(\lambda) = \alpha_0 + \sum_{i=1}^{p} \alpha_j \tilde{\varepsilon}_{t-i}^2(\gamma) + \sum_{j=1}^{q} \beta_j \tilde{\sigma}_{t-j}^2(\lambda) . \tag{2.2.6}$$

As shown in (A.2.2) and (A.2.3) of Appendix A.2, (2.2.5) and (2.2.6) can be rewritten as:

$$\tilde{\varepsilon}_t(\gamma) = \sum_{i=0}^{t-1} a_\gamma(i)(Y_{t-i} - c) + O(\rho^t) \;\; a.s. , \tag{2.2.7}$$

$$\tilde{\sigma}_t^2(\lambda) = \mathcal{B}_{\beta,t}^{-1} \alpha_0 + \sum_{i=0}^{t-1} a_\delta(i) \tilde{\varepsilon}_{t-i}^2(\gamma) + O(\rho^t) \;\; a.s. , \tag{2.2.8}$$

where $0 < \rho < 1$ and $\mathcal{B}_{\beta,t}^{-1}(z) = \sum_{i=0}^{t-1} a_\beta(i) z^i$.

The conditional likelihood function based on a finite sample is defined as:

$$\tilde{L}_n(\lambda) = \frac{1}{n} \sum_{t=1}^{n} \tilde{l}_t(\lambda), \text{ and } \tilde{l}_t(\lambda) = \log \frac{h(\tilde{\eta}_t(\lambda))}{\tilde{\sigma}_t(\lambda)}, \tag{2.2.9}$$

where h is a continuous positive function satisfying properties of a probability density function with third-order continuous derivative.

Corresponding to $\tilde{L}_n(\lambda)$, define

$$L_n(\lambda) = \frac{1}{n} \sum_{t=1}^{n} l_t(\lambda), \text{ and } l_t(\lambda) = \log \frac{h(\eta_t(\lambda))}{\sigma_t(\lambda)}. \tag{2.2.10}$$

Based on h, define a function g and its derivatives as:

$$g(x) = \frac{\partial \log h(x)}{\partial x} = \frac{\partial h(x)/\partial x}{h(x)}, \;\; g'(x) = \frac{\partial g(x)}{\partial x}, \;\; g''(x) = \frac{\partial^2 g(x)}{\partial x^2}.$$

Some other conditions on h are assumed.

Assumption 4. For some $\iota_1 \geq 0$ and some constant $C > 0$,

(i) $|\log h(x)|$ is bounded by $C|x|^{2\iota_1}$,

(ii) $|g(x)|$ is bounded by $C\big(\max\{|x|,1\}\big)^{2\iota_1-1}$.

(iii) $|g'(x)|$ is bounded by $C\big(\max\{|x|,1\}\big)^{2(\iota_1-1)}$.

(iv) $|g''(x)|$ is bounded by $C\big(\max\{|x|,1\}\big)^{2\iota_1-3}$.

Remark 2.2.2. *The value of ι_1 is completely determined by h. Assumption 4 implies $\log h(x)$ has order 3 continuous derivatives. It also implies $\mathbf{E}|\log h(\eta_t)| < C\mathbf{E}|\eta_t|^{2\iota_1}$ and expressions like $\mathbf{E}|g(\eta_t)|$, $\mathbf{E}|g'(\eta_t)\eta_t|$ and $\mathbf{E}|g''(\eta_t)\eta_t^2|$ are bounded by $C\mathbf{E}|\eta_t|^{2\iota_1-1}$. These expressions may be used in Section 2.3 and 2.4 for the proofs.*

Assumption 5. For any $w > 0$ and $v \in \mathbb{R}$, functions

$$\begin{aligned} \mathbf{E}g(w\eta_t + v) &= 0, \\ \frac{1}{w} + \mathbf{E}[g(w\eta_t + v)\eta_t] &= 0, \end{aligned}$$

have a unique solution at $w = 1$ and $v = 0$.

Proposition 2.2.1. *Suppose that η_t is symmetrically distributed about zero and* $\mathbf{E}[g(\eta_t)\eta_t] = -1$.

(Case i) *If g is an odd function with $g(0) = 0$, $|g(x)x| \leq Cx^2$, $g(x) \leq 0$ but not always 0 for $x > 0$, $g'(x) \leq 0$ for $x > 0$, then $\mathbf{E}\log[wh(w\eta_t + v)] < \mathbf{E}\log h(\eta_t)$ for any $w \neq 1$ and $v \neq 0$.*

(Case ii) *If η_t is not uniformly distributed and its density is decreasing on right side, $g(x)$ is an odd function and $g(x)x$ is a strictly monotone function for $x > 0$, then $\mathbf{E}\log[wh(w\eta_t + v)] < \mathbf{E}\log h(\eta_t)$ for any $w \neq 1$ and $v \neq 0$.*

(Case iii) *If η_t is not uniformly distributed and its density is decreasing on right side, $g(x)$ is an odd function, then* $\mathbf{E}\log[h(\eta_t + v)] < \mathbf{E}\log h(\eta_t)$ *for any* $v \neq 0$.

(Case iv) *If $g(x)x$ is a strictly monotone function for $x > 0$, then* $\mathbf{E}\log[wh(w\eta_t)] < \mathbf{E}\log h(\eta_t)$ *for any* $w \neq 1$.

The proof of Proposition 2.2.1 will be given in Appendix A.1.

Remark 2.2.3. *Assumption 5 implies* $\mathbf{E}g(\eta_t) = 0$ *and* $\mathbf{E}[g(\eta_t)\eta_t] = -1$. *It also implies* $\mathbf{E}\log[wh(w\eta_t + v)] \leq \mathbf{E}\log h(\eta_t)$ *for any* $w > 0$ *and* $v \in \mathbb{R}$. *The equality holds if and only if* $w = 1$ *and* $v = 0$. *By Proposition 2.2.1, if h is nicely defined and η_t is symmetrically distributed about 0, then* $\mathbf{E}[g(\eta_t)\eta_t] = -1$ *implies* $\mathbf{E}\log[wh(w\eta_t + v)] < \mathbf{E}\log h(\eta_t)$ *for any* $w \neq 1$ *and* $v \neq 0$.

Remark 2.2.4. *Assumption 5 guarantees* $\mathbf{E}l_0(\lambda)$ *is maximized at true value* λ_0 *and connects the distribution of η_0 with h. By Remark 2.2.3, when h is normal kernel,* $-\mathbf{E}[g(\eta_t)\eta_t] = \mathbf{E}\eta_t^2 = 1$ *guarantees* $\mathbf{E}l_0(\lambda)$ *is maximized at* λ_0. *In addition,* $\mathbf{E}\eta_0^2 = 1$ *is usually assumed to identify model (1.2.13). When fitting data by a likelihood kernel other than the standard normal density, we may have to scale η_t such that $\eta_t^* = a\eta_t$ for some constant $a > 0$ to satisfy* $\mathbf{E}[g(\eta_t^*)\eta_t^*] = -1$ *(see Remark 2.2.3). This makes only the parameters of $\alpha_0, \alpha_1, \cdots, \alpha_p$ scaled, while ϕ_i, φ_i and β_i stay unchanged. After fitting the scaled model, we have to scale the QMLE of ARCH part of the scaled model back to get the QMLE of the original model. The algorithm of fitting ARMA-GARCH model in Splus module S+FinMetrics does not scale back the estimators. A numeric example is given in Chapter 4.*

QMLE of λ_0 is defined as:

$$\tilde{\lambda}_n = \arg\max_{\lambda \in \Theta} \tilde{L}_n(\lambda) \ . \tag{2.2.11}$$

Theorem 2.2.1. *Let $\tilde{\lambda}_n$ be defined in (2.2.11). Under Assumptions 1-5, if* $\mathbf{E}|\eta_t|^{2\iota_1} < \infty$ *(or* $\mathbf{E}|\eta_t|^s < \infty$ *for some* $s > 0$*, if* $\iota_1 = 0$ *), then*

$$\tilde{\lambda}_n \longrightarrow \lambda_0, \quad a.s. \quad as\ n \to \infty.$$

Remark 2.2.5. *Consistency result imposes no moment requirement on* ε_0*. The moment requirement on* η_0 *depends on the choice of* h*. For example, when* h *is the student's t density function, asymptotic consistency requires only* $\mathbf{E}|\eta_0|^s < \infty$ *for some* $s > 0$*. When* h *is the smoothed double exponential (to be introduced in Section 2.3.2) density function, consistency requires* $\mathbf{E}|\eta_0| < \infty$*. However,* $\mathbf{E}|\eta_0|^2 < \infty$ *required by GARCH model will surpass the assumption. When* h *is the standard normal density function, consistency requires* $\mathbf{E}|\eta_0|^2 < \infty$*.*

For the normality result, we need two additional assumptions:

Assumption 6. λ_0 is in the interior of Θ.

Assumption 7. $\mathbf{E}^2\left(g^2(\eta_t)\eta_t\right) < \mathbf{E}g^2(\eta_t)\mathbf{E}(g(\eta_t)\eta_t + 1)^2$.

Remark 2.2.6. *With* $\mathbf{E}g(\eta_t) = 0$*, by Cauchy inequality, it is readily to show that Assumption 7 holds if and only if* $P[g(\eta_t)\eta_t + 1 = kg(\eta_t)] < 1$ *for any constant* k*. As mentioned in Ling (2005), a simple condition for this is that* η_t *has a positive density on some interval provided that* $g(x)x + 1 = kg(x)$ *has finite roots.*

By (2.2.10), the first derivative of $l_t(\lambda)$ is:

$$\frac{\partial l_t(\lambda)}{\partial \lambda} = -\frac{1}{2}\Big\{1 + g\big(\eta_t(\lambda)\big)\eta_t(\lambda)\Big\}\frac{\partial \sigma_t^2(\lambda)/\partial\lambda}{\sigma_t^2(\lambda)} + g\big(\eta_t(\lambda)\big)\frac{\partial \varepsilon_t(\gamma)/\partial\lambda}{\sigma_t(\lambda)} \,. \qquad (2.2.12)$$

Based on (2.2.12), we have the second derivative of $l_t(\lambda)$:

$$\begin{aligned}\frac{\partial^2 l_t(\lambda)}{\partial\lambda\partial\lambda^T} &= -\frac{1}{2}\Big(1+g(\eta_t(\lambda))\eta_t(\lambda)\Big)\frac{\partial^2\sigma_t^2(\lambda)/\partial\lambda\partial\lambda^T}{\sigma_t^2(\lambda)} \qquad (2.2.13)\\ &+\frac{1}{4}\Big(2+3g(\eta_t(\lambda))\eta_t(\lambda)+g'(\eta_t(\lambda))\eta_t^2(\lambda)\Big)\frac{\partial\sigma_t^2(\lambda)/\partial\lambda}{\sigma_t^2(\lambda)}\frac{\partial\sigma_t^2(\lambda)/\partial\lambda^T}{\sigma_t^2(\lambda)}\\ &-\frac{1}{2}\Big(g'(\eta_t(\lambda))\eta_t(\lambda)+g(\eta_t(\lambda))\Big)\frac{\partial\sigma_t^2(\lambda)/\partial\lambda}{\sigma_t^2(\lambda)}\frac{\partial\varepsilon_t(\gamma)/\partial\lambda^T}{\sigma_t(\lambda)}\\ &-\frac{1}{2}\left(g'(\eta_t(\lambda))\eta_t(\lambda)+g(\eta_t(\lambda))\right)\frac{\partial\varepsilon_t(\gamma)/\partial\lambda}{\sigma_t(\lambda)}\frac{\partial\sigma_t^2(\lambda)/\partial\lambda^T}{\sigma_t^2(\lambda)}\\ &+g'(\eta_t(\lambda))\frac{\partial\varepsilon_t(\gamma)/\partial\lambda}{\sigma_t(\lambda)}\frac{\partial\varepsilon_t(\lambda)/\partial\lambda^T}{\sigma_t(\lambda)}+g(\eta_t(\lambda))\frac{\partial^2\varepsilon_t(\gamma)/\partial\lambda\partial\lambda^T}{\sigma_t(\lambda)}\ .\end{aligned}$$

The first and second derivatives of $\varepsilon_t(\gamma)$ and $\sigma_t(\lambda)$ are presented in Appendix A.3.

Throughout the following, for simplicity, we denote

$$\left.\frac{\partial z(\lambda)}{\partial\lambda}\right|_{\lambda=\lambda_0}=\frac{\partial z(\lambda_0)}{\partial\lambda}\quad and\quad \left.\frac{\partial^2 z(\lambda)}{\partial\lambda\partial\lambda^T}\right|_{\lambda=\lambda_0}=\frac{\partial^2 z(\lambda_0)}{\partial\lambda\partial\lambda^T},$$

for any function z.

Define matrix

$$\mathcal{I}=\mathbf{E}\left(\frac{\partial l_0(\lambda_0)}{\partial\lambda}\frac{\partial l_0(\lambda_0)}{\partial\lambda^T}\right),\quad \mathcal{J}=-\mathbf{E}\left(\frac{\partial^2 l_0(\lambda_0)}{\partial\lambda\partial\lambda^T}\right).$$

As to be shown in Lemma 2.4.9, under conditions $\mathbf{E}g^2(\eta_0)\eta_0=0$ and $\mathbf{E}g'(\eta_0)\eta_0=0$, $\mathcal{I}$ and $\mathcal{J}$ can be written as block-diagonal respectively (due to the partitioning of parameter space) with following forms:

$$\mathcal{I}=\begin{pmatrix}\mathcal{I}_1 & 0\\ 0 & \mathcal{I}_2\end{pmatrix},\ \mathcal{J}=\begin{pmatrix}\mathcal{J}_1 & 0\\ 0 & \mathcal{J}_2\end{pmatrix},$$

where

$$\begin{aligned}\mathcal{I}_1 &= \frac{1}{4}\Big(\mathbf{E}(g(\eta_t)\eta_t)^2-1\Big)\mathbf{E}\left(\frac{\partial\sigma_t^2(\lambda_0)/\partial\gamma}{\sigma_t^2}\frac{\partial\sigma_t^2(\lambda_0)/\partial\gamma^T}{\sigma_t^2}\right) \qquad (2.2.14)\\ &+\mathbf{E}g^2(\eta_t)\mathbf{E}\left(\frac{\partial\varepsilon_t(\gamma_0)/\partial\gamma}{\sigma_t}\frac{\partial\varepsilon_t(\gamma_0)/\partial\gamma^T}{\sigma_t}\right),\end{aligned}$$

$$\mathcal{I}_2 = \frac{1}{4}\Big(\mathbf{E}\big(g(\eta_t)\eta_t\big)^2 - 1\Big)\mathbf{E}\left(\frac{\partial\sigma_t^2(\lambda_0)/\partial\delta}{\sigma_t^2}\frac{\partial\sigma_t^2(\lambda_0)/\partial\delta^T}{\sigma_t^2}\right), \quad (2.2.15)$$

$$\mathcal{J}_1 = -\frac{1}{4}\mathbf{E}\Big(g'(\eta_t)\eta_t^2 - 1\Big)\mathbf{E}\left(\frac{\partial\sigma_t^2(\lambda_0)/\partial\gamma}{\sigma_t^2}\frac{\partial\sigma_t^2(\lambda_0)/\partial\gamma^T}{\sigma_t^2}\right) \quad (2.2.16)$$
$$-\mathbf{E}g'(\eta_t)\mathbf{E}\left(\frac{\partial\varepsilon_t(\gamma_0)/\partial\gamma}{\sigma_t}\frac{\partial\varepsilon_t(\gamma_0)/\partial\gamma^T}{\sigma_t}\right),$$

$$\mathcal{J}_2 = -\frac{1}{4}\mathbf{E}\Big(g'(\eta_t)\eta_t^2 - 1\Big)\mathbf{E}\left(\frac{\partial\sigma_t^2(\lambda_0)/\partial\delta}{\sigma_t^2}\frac{\partial\sigma_t^2(\lambda_0)/\partial\delta^T}{\sigma_t^2}\right). \quad (2.2.17)$$

Theorem 2.2.2. *Let $\tilde{\lambda}_n$ be defined in (2.2.11). Under Assumptions 1-7, if $\mathcal{J}$ is non-singular, there exist $0 \le \iota_2 < 1$ such that $\mathbf{E}|\varepsilon_t|^{2(\iota_1+1)(1-\iota_2)} < \infty$ and $\mathbf{E}|\eta_0|^{4\iota_1} < \infty$ (or $\mathbf{E}|\eta_t|^s < \infty$ for some $s > 0$, if $\iota_1 = 0$), then $\sqrt{n}(\tilde{\lambda}_n - \lambda_0)$ is asymptotically distributed as $N(0, \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1})$.*

Remark 2.2.7. *By Lemma 2.4.9, $\mathcal{J}$ is nonsingular under some regularity conditions together with $\mathbf{E}g'(\eta_0) \le 0$, $\mathbf{E}g'(\eta_0)\eta_0^2 < 1$ and $\mathbf{E}g'(\eta_0)\eta_0 = 0$. In particular, if η_0 is symmetric about 0 and h is the normal, or the smoothed generalized error distribution ((SGED), to be introduced in Section 2.3.2), or the student's t density function, then $\mathcal{J}$ is non-singular.*

Remark 2.2.8. *The value of ι_1 is completely determined by h. When h is the standard normal density function, asymptotic normality result holds if $\mathbf{E}|\varepsilon_0|^{4(1-\iota_2)} < \infty$ and $\mathbf{E}|\eta_0|^4 < \infty$. When h is the smoothed double exponential distribution density function, asymptotic normality holds if $\mathbf{E}|\varepsilon_0|^{3(1-\iota_2)} < \infty$ and $\mathbf{E}|\eta_0|^2 < \infty$. Again when h is the student's t density function, asymptotic normality holds if $\mathbf{E}|\varepsilon_0|^{2(1-\iota_2)} < \infty$ and $\mathbf{E}|\eta_0|^s < \infty$ for some $s > 0$. Thus by choosing an appropriate h, we can relax the moment requirements on ε_0 and η_0 in Francq and Zakoïan (2004) who need $\mathbf{E}|\varepsilon_0|^4 < \infty$ and $\mathbf{E}|\eta_0|^4 < \infty$, and the moment requirement on η_0 in Ling (2005) who needs $\mathbf{E}|\eta_0|^4 < \infty$.*

Remark 2.2.9. *Which value of ι_2 to choose depends on the parameter space Θ_δ, in particular, β_{01} (true value of β_1). In general, we choose $\iota_2 = 0$. If we use the condition $\beta_{01} > 0$ specifically, we can find a positive ι_2 so the moment of ε_0 will be reduced slightly. If h is the student's t density, it enables us to find a global QMLE for an ARMA-IGARCH model but not an ARMA-IARCH model. With $\iota_2 = 0$, Theorem 2.2.2 holds for ARMA-ARCH, and ARCH models only if we remove the redundant parameters and the corresponding components in the covariance matrix.*

Remark 2.2.10. *In general, $\tilde{\gamma}_n$ and $\tilde{\delta}_n$ may be asymptotically correlated. If η_0 is symmetric about 0, $\mathbf{E}(g^2(\eta_0)\eta_0) = 0$ and $\mathbf{E}(g'(\eta_0)\eta_0) = 0$, then $\tilde{\gamma}_n$ and $\tilde{\delta}_n$ are asymptotically independent with*

$$\begin{aligned}\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1} &= \begin{pmatrix} \mathcal{J}_1^{-1} & 0 \\ 0 & \mathcal{J}_2^{-1} \end{pmatrix} \begin{pmatrix} \mathcal{I}_1 & 0 \\ 0 & \mathcal{I}_2 \end{pmatrix} \begin{pmatrix} \mathcal{J}_1^{-1} & 0 \\ 0 & \mathcal{J}_2^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{J}_1^{-1}\mathcal{I}_1\mathcal{J}_1^{-1} & 0 \\ 0 & \mathcal{J}_2^{-1}\mathcal{I}_2\mathcal{J}_2^{-1} \end{pmatrix}.\end{aligned}$$

Remark 2.2.11. *After adopting similar notations used by Ling (2005), we can write*

$$\mathcal{I} = \mathbf{E}[U_t(\lambda_0) I U_t^T(\lambda_0)], \ \mathcal{J} = \mathbf{E}[U_t(\lambda_0) J U_t^T(\lambda_0)],$$

where

$$U_t(\lambda) = \left(\frac{\partial\sigma_t^2(\lambda)/\partial\lambda}{2\sigma_t^2(\lambda)}, \frac{\partial\varepsilon_t(\lambda)/\partial\lambda}{\sigma_t(\lambda)} \right),$$

$$I = \begin{bmatrix} \mathbf{E}\big(g(\eta_t)\eta_t + 1\big)^2 & -\mathbf{E}\big(g^2(\eta_t)\eta_t\big) \\ -\mathbf{E}\big(g^2(\eta_t)\eta_t\big) & \mathbf{E}g^2(\eta_t) \end{bmatrix}, J = \begin{bmatrix} -\mathbf{E}\big(g'(\eta_t)\eta_t^2 - 1\big) & \mathbf{E}\big(g'(\eta_t)\eta_t\big) \\ \mathbf{E}\big(g'(\eta_t)\eta_t\big) & -\mathbf{E}g'(\eta_t) \end{bmatrix}.$$

When $P = Q = 0$, $c_0 \neq 0$, model (1.2.12)-(1.2.13) reduces to pure GARCH with nonzero constant mean term. The redundant parameters ϕ_i and φ_j are removed and the QMLE is reduced to $(\tilde{c}_n, \tilde{\delta}_n)$.

Theorem 2.2.3. *When $P = Q = 0$ and $c_0 \neq 0$, under Assumption 1, and Assumptions 3- 4, if $\mathbf{E}|\eta_t|^{2\iota_1} < \infty$ (or $\mathbf{E}|\eta_t|^s < \infty$ for some $s > 0$, if $\iota_1 = 0$), then $(\tilde{c}_n, \tilde{\delta}_n) \longrightarrow (c_0, \delta_0)$ almost surely. Further with Assumption 6, if $\mathbf{E}|\varepsilon_t|^{2\iota_1(1-\iota_2)} < \infty$ (or $\mathbf{E}|\varepsilon_t|^s < \infty$ for some $s > 0$, if $\iota_1 = 0$) and $\mathbf{E}|\eta_t|^{\max(1,4\iota_1)}$, then $\sqrt{n}(\tilde{c}_n - c_0, \tilde{\delta}_n - \delta_0)$ is asymptotically distributed as $N(0, \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1})$.*

Remark 2.2.12. *Results given in Theorem 2.2.3 are not covered by Berkes, Horváth and Kokoszka (2003) and Berkes and Horváth (2003), and are not discussed in detail by Francq and Zakoïan (2004) and Ling (2005). Theorem 2.2.3 implies that for any h satisfying the assumptions, consistency of $(\tilde{c}_n, \tilde{\delta}_n)$ holds without any moment requirement on ε_t, and normality of $(\tilde{c}_n, \tilde{\delta}_n)$ holds with only $\mathbf{E}|\varepsilon_t|^{2\iota_1(1-\iota_2)} < \infty$. If $\beta_{01} > 0$, then $\iota_2 > 0$, and hence IGARCH with nonzero constant mean can be dealt with even when h is the normal density.*

Corollary 2.2.1. *Let $\sigma_c^2 = \mathcal{J}_1^{-1}\mathcal{I}_1\mathcal{J}_1^{-1}$. For pure GARCH with nonzero constant mean term, under the same conditions as in Theorem 2.2.3, if $\varepsilon_t \sim N(0,1)$ and h takes standard normal density, then*

$$\frac{Var(\bar{\varepsilon}_n)}{\sigma_c^2} > 1$$

as $n \to \infty$, where $\bar{\varepsilon}_n = \sum_i^n \varepsilon_i / n$ is the sample mean estimator of c_0. The ratio does not depend on α_{00} (true value of α_0).

A simulation result of this Corollary is given in Chapter 4.

Corollary 2.2.2. *When $P = Q = 0$ and $c_0 = 0$, under Assumptions 1, 3 - 4 and Assumption 5 with $v = 0$, if $\mathbf{E}|\eta_t|^{2\iota_1} < \infty$ (or $\mathbf{E}|\eta_t|^s < \infty$ for some $s > 0$, if $\iota_1 = 0$), then $\tilde{\delta}_n \longrightarrow \delta_0$ almost surely. Further with Assumption 6, if $\mathbf{E}|\varepsilon_t|^{2\iota_1(1-\iota_2)} < \infty$ and $\mathbf{E}|\eta_t|^{4\iota_1} < \infty$ (or $\mathbf{E}|\varepsilon_t|^s < \infty$ and $\mathbf{E}|\eta_t|^s < \infty$ for some $s > 0$ if $\iota_1 = 0$), then $\sqrt{n}(\tilde{\delta}_n - \delta_0)$ is asymptotically distributed as $N(0, 4\tau^2\mathcal{D}^{-1})$, where*

$$\tau^2 = \frac{\mathbf{E}\big(g(\eta_t)\eta_t + 1\big)^2}{\Big(\mathbf{E}\big(g'(\eta_t)\eta_t^2\big) - 1\Big)^2}, \quad \mathcal{D} = \mathbf{E}\left(\frac{\partial\sigma_t^2(\delta_0)/\partial\delta}{\sigma_t^2}\frac{\partial\sigma_t^2(\delta_0)/\partial\delta^T}{\sigma_t^2}\right).$$

Remark 2.2.13. *Corollary 2.2.2 reduces to Theorems 1.1 and 1.2 of Horváth and Kokoszka (2003) with weaker conditions, or Theorems 2.1 and 2.2 of Francq and Zakoïan (2004) when h is the standard normal density. For pure GARCH with mean zero, h can be double exponential density. Again if h is the student t density function, both consistency and normality results require only* $\mathbf{E}|\eta_0|^s < \infty$ *for some* $s > 0$*. As Berkes and Horváth (2004) had shown: for a given series of* η_t*, efficiency of QMLE for pure GARCH depends only on* τ^2*, which is determined by the distribution of* η_0 *and choice of h.*

Theorem 2.2.4. *When* $p = q = 0$*, under Assumptions 1 - 2, Assumption 4 with* $\iota_1 = 1$ *and Assumption 5 with* $w = 1$*, if* $\mathbf{E}|\varepsilon_t|^2 < \infty$*, then* $\tilde{\gamma}_n \to \gamma_0$ *almost surely. Further with* γ_0 *being in interior of* Θ_γ*, then* $\sqrt{n}(\tilde{\gamma}_n - \gamma_0)$ *is asymptotically distributed as* $N(0, \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1})$*, where* $\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1} = \tau^2_{arma}\mathcal{D}_\varepsilon^{-1}$*,*

$$\tau^2_{arma} = \frac{\mathbf{E}g^2(\varepsilon_t/\sqrt{\alpha_{00}})}{(\mathbf{E}g'(\varepsilon_t/\sqrt{\alpha_{00}}))^2}, \quad \text{and } \mathcal{D}_\varepsilon = -\mathbf{E}\frac{\partial\varepsilon_t(\gamma_0)/\partial\gamma}{\sqrt{\alpha_{00}}}\frac{\partial\varepsilon_t(\gamma_0)/\partial\gamma^T}{\sqrt{\alpha_{00}}}.$$

Remark 2.2.14. *Here for pure ARMA,* α_{00} *(the variance of* ε_t*) is taken as nuisance parameter. Since the sequence* $\{\varepsilon_t, -\infty < t < \infty\}$ *is IID, the likelihood kernel does not affect the moment requirement on* ε_t*, however it does affect the efficiency of QMLE.*

Let $\hat{\gamma}_n$ be any $\sqrt{n}$ consistent estimator (i.e, (Q)MLE, LSE, weighted or self-weighted LSE) of γ_0 and $\tilde{\varepsilon}_t = \tilde{\varepsilon}_t(\hat{\gamma}_n)$ be the corresponding residuals calculated by replacing γ with $\hat{\gamma}_n$ in (2.2.5) from the ARMA part. Then we use $\tilde{\varepsilon}_t$ as artificial observations of ε_t and fit them with a GARCH(p,q) model to obtain the estimator of δ.

Define

$$\tilde{\sigma}_t^2(\hat{\gamma}_n, \delta) = \alpha_0 + \sum_{i=1}^{p} \alpha_j \tilde{\varepsilon}_{t-i}^2 + \sum_{j=1}^{q} \beta_j \tilde{\sigma}_{t-j}^2(\hat{\gamma}_n, \delta), \qquad (2.2.18)$$

$$\tilde{\eta}_t(\hat{\gamma}_n, \delta) = \frac{\tilde{\varepsilon}_t}{\tilde{\sigma}_t(\hat{\gamma}_n, \delta)}. \tag{2.2.19}$$

And define the QMLE of δ as

$$\hat{\delta}_n = \arg\max_{\delta \in \Theta_\delta} \tilde{L}_n(\hat{\gamma}_n, \delta),$$

where

$$\tilde{L}_n(\hat{\gamma}_n, \delta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\hat{\gamma}_n, \delta) = \frac{1}{n}\sum_{t=1}^{n} \log \frac{h(\tilde{\eta}_t(\hat{\gamma}_n, \delta))}{\tilde{\sigma}_t(\hat{\gamma}_n, \delta)}. \tag{2.2.20}$$

Theorem 2.2.5. *With Assumptions 1-5, if* $\mathbf{E}|\eta_t|^{\max(1, 2\iota_1)} < \infty$*, then* $\hat{\delta}_n \longrightarrow \delta_0$ *in probability. Further with Assumptions 6 and 7,* $\mathbf{E}|\varepsilon_t|^{2\iota_1(1-\iota_2)} < \infty$ *(or* $\mathbf{E}|\varepsilon_t|^s < \infty$ *for some* $s > 0$*, if* $\iota_1 = 0$*) and* $\mathbf{E}|\eta_t|^{\max(1, 4\iota_1)} < \infty$*, if* η_t *is symmetric about zero and* $\mathbf{E}g'(\eta_t)\eta_t = 0$*, then*

$$n^{1/2}(\hat{\delta}_n - \delta_0) \to N(0, 4\tau^2 \mathcal{D}^{-1}).$$

Remark 2.2.15. $g'(x)$ *is an even function when* h *is the normal or student's t, or smoothed generalized double exponential density. Thus* $\mathbf{E}g'(\eta_t)\eta_t = 0$ *when* η_0 *is symmetric about 0.*

Remark 2.2.16. *When* $P = Q = 0$*,* $c_0 \neq 0$*, if* $\mathbf{E}\varepsilon_t^2 < \infty$*, by CLT,* $\bar{\varepsilon}_n$ *is a* $\sqrt{n}$ *consistent estimator of* c*. Theorem 2.2.5 implies the common practice of using the sample mean to center financial data is workable provided that* η_0 *is symmetric about 0.*

Remark 2.2.17. *Under conditions of* η_t *being symmetric about zero and* $\mathbf{E}g'(\eta_t)\eta_t = 0$*, the variance of* $\hat{\gamma}_n$ *has no effect on the asymptotic variance of* $\hat{\delta}_n$*.*

2.3 Examples

2.3.1 QMLE based on the student's t density

When h is the student's t probability density function, we have:

$$h(x) = \left(1 + \frac{x^2}{d}\right)^{-\frac{d+1}{2}},$$

where $d > 1$ is the degrees of freedom and we ignore the constant term. Then

$$h'(x) = \frac{\partial h(x)}{\partial x} = -\frac{(d+1)x}{d}\left(1 + \frac{x^2}{d}\right)^{-\frac{d+1}{2}-1},$$

$$g(x) = -(d+1)\frac{x}{d+x^2},$$

$$g'(x) = -(d+1)\left\{\frac{1}{d+x^2} - \frac{2x^2}{(d+x^2)^2}\right\},$$

$$g''(x) = 2(d+1)\left\{\frac{3x}{(d+x^2)^2} - \frac{4x^3}{(d+x^2)^3}\right\}.$$

Obviously $g(x)$ is odd and $g(x)x$ is strictly monotone decreasing for $x > 0$. Then by (Case ii) in Proposition 2.2.1, Assumption 5 is satisfied when η_t is symmetrically distributed (except uniform distribution) about zero with density function decreasing on right side and $\mathbf{E}(1/(d + \eta_0^2)) = 1/(d+1)$. Clearly Assumption 4 is satisfied with $\iota_1 = 0$.

With the fact that $\mathbf{E}X^2 \geq (\mathbf{E}X)^2$, we have $\mathbf{E}(1/(d+\eta_0^2)^2) \geq 1/(d+1)^2$. Then for $d > 1$, we have

$$\begin{aligned}
\mathbf{E}g'(\eta_0) &= -(d+1)\mathbf{E}\left\{\frac{1}{d+\eta_0^2} - \frac{2\eta_0^2}{(d+\eta_0^2)^2}\right\} \\
&= -(d+1)\left\{-\mathbf{E}\frac{1}{d+\eta_0^2} + \mathbf{E}\frac{2d}{(d+\eta_0^2)^2}\right\} \\
&\leq -(d+1)\left\{-\frac{1}{d+1} + \frac{2d}{(d+1)^2}\right\} \\
&= -\frac{d-1}{d+1} < 0\,,
\end{aligned}$$

and

$$
\begin{aligned}
\mathbf{E}g'(\eta_0)\eta_0^2 &= -(d+1)\mathbf{E}\left\{\frac{\eta_0^2}{d+\eta_0^2} - \frac{2\eta_0^4}{(d+\eta_0^2)^2}\right\} \\
&= -(d+1)\left\{1 - \mathbf{E}\frac{d}{d+\eta_0^2} - 2 + \mathbf{E}\frac{4d\eta_0^2}{(d+\eta_0^2)^2} + \mathbf{E}\frac{2d^2}{(d+\eta_0^2)^2}\right\} \\
&= -(d+1)\left\{1 - \mathbf{E}\frac{d}{d+\eta_0^2} - 2 + \mathbf{E}\frac{4d}{d+\eta_0^2} - \mathbf{E}\frac{4d^2}{(d+\eta_0^2)^2} + \mathbf{E}\frac{2d^2}{(d+\eta_0^2)^2}\right\} \\
&= -(d+1)\left\{\frac{2d-1}{d+1} - \mathbf{E}\frac{2d^2}{(d+\eta_0^2)^2}\right\} \\
&< (d+1)\mathbf{E}\frac{2d}{d+\eta_0^2} - 2d + 1 = 1\ .
\end{aligned}
$$

If η_0 is symmetric about 0, then it is obvious that $\mathbf{E}g'(\eta_0)\eta_0 = 0$ (since $\mathbf{E}|g'(\eta_0)\eta_0| < C\max(\mathbf{E}|\eta_0|^{2\iota_1-1}, 1) < \infty$). Thus by Lemmas 2.4.9 and 2.4.10, $\mathcal{J}$ is nonsingular and block-diagonal.

Assumption 7 is satisfied, since

$$
(\mathbf{E}g^2(\eta_0)\eta_0)^2 = (d+1)^2\left(\mathbf{E}\frac{\eta_0^3}{(d+\eta_0^2)^2}\right)^2 = 0 < \mathbf{E}g^2(\eta_0)\mathbf{E}(g(\eta_0)\eta_0+1)^2.
$$

Then by Lemmas 2.4.9 and 2.4.10, $\mathcal{I}$ is nonsingular and block-diagonal.

Thus when the student's t density function is used for the quasi-likelihood function, the QMLE is asymptotically consistent for any small positive moment on η_t. The asymptotic normality result holds if $\mathbf{E}|\varepsilon_t|^{2(1-\iota_2)} < \infty$ and $\mathbf{E}|\eta_t|^s < \infty$ for some $s > 0$.

2.3.2 QMLE based on smoothed generalized error density

The density of the generalized error distribution (GED) (Nelson 1991) has the form of $f(x) \sim e^{-|x|^d}$, where d is a certain positive constant. When d is an odd integer, the density is not smooth at $x=0$ and thus has no derivatives at $x=0$. Adopting the idea of Hitomi (1997) for double exponential distribution, we smooth the GED density as

$$h(x) = e^{-(x^2+b^2)^{d/2}},$$

where $b \neq 0$ is the smooth parameter and $d > 0$ (b could be 0 when d is an even integer), and we ignore the constant term. We call such a distribution as smoothed generalized error distribution (SGED(d)). When $d=1$, it is smoothed double exponential distribution (SDE). Then

$$h'(x) = -de^{-(x^2+b^2)^{d/2}}(x^2+b^2)^{d/2-1}x, \quad g(x) = -d(x^2+b^2)^{d/2-1}x,$$

$$g'(x) = -d(d-2)(x^2+b^2)^{d/2-2}x^2 - d(x^2+b^2)^{d/2-1},$$

$$g''(x) = -d(d-2)(d-4)(x^2+b^2)^{d/2-3}x^3 - 3d(d-2)(x^2+b^2)^{d/2-2}x,$$

It is obvious that $g(x)$ is odd with $g(0)=0$, $|g(x)x| \leq Kx^2$ when $d \leq 2$, $g(x) < 0$ for $x > 0$. For $d \geq 1$, we have

$$\begin{aligned} g'(x) &= -d(x^2+b^2)^{d/2-1}\left(1-(2-d)\frac{x^2}{x^2+b^2}\right) \\ &= -d(x^2+b^2)^{d/2-1}\frac{(d-1)x^2+b^2}{x^2+b^2} \\ &\leq 0 \end{aligned}$$

Thus by (Case i) in Proposition 2.2.1, for $1 \leq d \leq 2$, Assumption 5 is satisfied if η_t is symmetric about zero with $\mathbf{E}\big(d(\eta_0^2+b^2)^{d/2-1}\eta_0^2\big) = 1$.

By the expressions of $g(x)$, $g'(x)$ and $g''(x)$, it is obvious that Assumption 4 is satisfied with $\iota_1 = d/2$.

For $1 \leq d \leq 2$, we have $g'(x) \leq 0$, and $g'(x)x^2 \leq 0$. If η_0 is symmetric about 0, it is clearly $\mathbf{E}g'(\eta_0)\eta_0 = 0$. Thus by Lemmas 2.4.9 and 2.4.10, $\mathcal{J}$ is nonsingular and block-diagonal.

Also when η_0 is symmetric about 0, Assumption 7 is satisfied, since

$$(\mathbf{E}g^2(\eta_0)\eta_0)^2 = d^4(\mathbf{E}(\eta_0^2 + b^2)^{d-2}\eta_0^3)^2 = 0 < \mathbf{E}g^2(\eta_0)\mathbf{E}(g(\eta_0)\eta_0 + 1)^2.$$

Therefore by Lemmas 2.4.9 and 2.4.10, $\mathcal{I}$ is nonsingular and block-diagonal.

Thus when the SGED (with $1 \leq d \leq 2$) density function is used for the Quasi-likelihood function, the QMLE is asymptotically consistent when $\mathbf{E}|\eta_t|^d < \infty$. The asymptotic normality result holds if $\mathbf{E}|\varepsilon_t|^{(d+2)(1-\iota_2)} < \infty$ and $\mathbf{E}|\eta_t|^{2d} < \infty$.

2.3.3 QMLE based on the normal density

When h is the standard normal probability density function, such that:

$$h(x) = e^{-\frac{x^2}{2}},$$

where we ignore the constant term.

Then we have

$$h'(x) = \partial h(x)/\partial x = -xe^{-\frac{x^2}{2}},$$

$$g(x) = -x, \quad g'(x) = -1, \quad g''(x) = 0,$$

If $\mathbf{E}\eta_0 = 0$, and $\mathbf{E}\eta_0^2 = 1$, then

$$\log w - \mathbf{E}\frac{(w\eta_0 + v)^2}{2} - \frac{1}{2}\log 2\pi = \log w - \frac{w^2 + v^2}{2} - \frac{1}{2}\log 2\pi.$$

By setting the partial differential equations with respect to w and v of above equation as 0 and solving them, we have $w = 1,\ v = 0$. Thus $l_t(\lambda)$ is uniquely maximized at $\lambda = \lambda_0$. Assumption 4 is satisfied with $\iota_1 = 1$. Thus normality holds with $\mathbf{E}|\varepsilon_0|^{4(1-\iota_2)} < \infty$ and $\mathbf{E}|\eta_0|^4 < \infty$. It is obvious that $\mathbf{E}g'(\eta_0)\eta_0^2 = -1 < 1$, $\mathbf{E}g'(\eta_0) = -1 < 0$, $\mathbf{E}g'(\eta_0)\eta_0 = 0$, so that $\mathcal{J}$ is nonsingular. Assumption 7 is satisfied, since $g(x) = -x$ and $\mathbf{E}\eta_0 = 0$, by Cauchy-Schwarz inequality, we have

$$(\mathbf{E}g^2(\eta_0)\eta_0)^2 = [\mathbf{E}g(\eta_0)(g(\eta_0)\eta_0 + 1)]^2 < \mathbf{E}g^2(\eta_0)\mathbf{E}(g(\eta_0)\eta_0 + 1)^2.$$

Thus by Lemmas 2.4.9 and 2.4.10, $\mathcal{I}$ is nonsingular.

So when the standard normal probability density is used for the quasi-likelihood function and $\beta_{01} > 0$, the asymptotic normality result holds even if $\mathbf{E}|\varepsilon_t|^4 = \infty$.

Remark 2.3.1. *It can been seen that if h is a student's t density and $\beta_{01} > 0$, QMLE is asymptotically normally distributed even if $\mathbf{E}|\varepsilon_t|^2 = \infty$. From the above examples of the normal kernel and the student t kernel, we see, when the degrees of freedom go to infinity, $g(x)$, $g'(x)$ and $g''(x)$ computed from the student t kernel converges to those computed from the normal kernel. Thus $\mathcal{I}$ and $\mathcal{J}$ computed from the student t kernel converges to those computed from normal kernel. Thus the variance of the QMLE based on the student t kernel converges to that based on the normal kernel. This implies we can obtain QMLE by the student t kernel with a large degree of freedom so as to reach almost the same efficiency of those obtained by the standard normal kernel. This is very useful when the variance of GARCH error ε_t is infinite, a case where the validity of the asymptotic normality for QMLE based on the normal kernel is unclear.*

2.3.4 Efficiency of QMLE

Assumption 5 connects the distribution of η_0 with h. Usually $\mathbf{E}\eta_0^2 = 1$ is assumed for model identifiability (1.2.12)-(1.2.13). If h is the normal kernel, Assumption 5 implies

$\mathbf{E}\eta_0 = 0$ and $\mathbf{E}\eta_0^2 = 1$. When we fit data by likelihood kernel other than the standard normal density, we may have to scale η_t to meet Assumption 5. Let $\eta_t^{**} = a\eta_t$ and $\sigma_t^{**} = \sigma_t/a$ for some positive constant a such that Assumption 5 is satisfied. Then model (2.2.1) -(2.2.2) are modified as:

$$\varepsilon_t(\gamma) = (Y_t - c) - \sum_{i=1}^{P} \phi_i(Y_{t-i} - c) - \sum_{j=1}^{Q} \varphi_j \varepsilon_{t-j}(\gamma), \tag{2.3.1}$$

$$\eta_t(\lambda^{**}) = \frac{\varepsilon_t(\gamma)}{\sigma_t(\lambda^{**})} \text{ and } \sigma_t{}^2(\lambda^{**}) = \alpha_0^{**} + \sum_{i=1}^{p} \alpha_i^{**}\varepsilon_{t-i}^2(\gamma) + \sum_{j=1}^{q} \beta_j^{**}\sigma_{t-j}^2(\lambda^{**}) \,. \tag{2.3.2}$$

The parameters of the scaled model are denoted as $\lambda^{**} = (\gamma, \delta^{**})$. Comparing to model (2.2.1) -(2.2.2), we have $(\alpha_0^{**}, \cdots, \alpha_p^{**})^T = (\alpha_0/a^2, \cdots, \alpha_p/a^2)^T$, $(\beta_1^{**}, \cdots, \beta_q^{**})^T = (\beta_1, \cdots, \beta_q)^T$.

Let $M_\lambda = diag(1, \cdots, 1, 1/a^2, \cdots, 1/a^2, 1, \cdots, 1)$ with $(P+Q+2)$-th to $(P+Q+p+2)$-th elements $1/a^2$ and all other elements 1. Thus $\lambda^{**} = M_\lambda \lambda$. From the derivatives of $\sigma_t^2(\lambda)$ and $\varepsilon_t(\gamma)$ as in (A.3.1)-(A.3.8) of Appendix 3, we have that

$$\frac{\partial \sigma_t{}^2(\lambda^{**})}{\partial \lambda^{**}} = \frac{1}{a^2} M_\lambda^{-1} \frac{\partial \sigma_t^2(\lambda)}{\partial \lambda}, \quad \frac{\partial \sigma_t{}^2(\lambda^{**})/\partial \lambda^{**}}{\sigma_t{}^2(\lambda^{**})} = M_\lambda^{-1} \frac{\partial \sigma_t^2(\lambda)/\partial \lambda}{\sigma_t^2(\lambda)},$$

$$\frac{\partial \varepsilon_t(\gamma)}{\partial \lambda^{**}} = \frac{\partial \varepsilon_t(\lambda)}{\partial \lambda}, \quad \frac{\partial \varepsilon_t(\gamma)/\partial \lambda^{**}}{\sigma_t(\lambda^{**})} = a \frac{\partial \varepsilon_t(\lambda)/\partial \lambda}{\sigma_t(\lambda)}.$$

Denote the QMLE of the modified model (2.3.1)-(2.3.2) with $\tilde{\lambda}_n^{**}$. By Theorem 2.2.2, we have:

$$\sqrt{n}(\tilde{\lambda}_n^{**} - \lambda_0^{**}) \longrightarrow N(0, \mathcal{J}^{**-1}\mathcal{I}^{**}\mathcal{J}^{**-1}) \,,$$

where

$$\mathcal{I}^{**} = \mathbf{E}[U_t(\lambda_0^{**}) I^{**} U_t{}^T(\lambda_0^{**})], \ \mathcal{J}^{**} = \mathbf{E}[U_t(\lambda_0^{**}) J^{**} U_t{}^T(\lambda_0^{**})],$$

$$U_t(\lambda^{**}) = \left(\frac{\partial \sigma_t{}^2(\lambda^{**})/\partial \lambda^{**}}{2\sigma_t{}^2(\lambda^{**})}, \ \frac{\partial \varepsilon_t(\gamma)/\partial \lambda^{**}}{\sigma_t(\lambda^{**})} \right),$$

I^{**}, J^{**} are 2×2 matrix obtained from I, J by replacing η_t with η_t^{**}.

Let $M_1 = \begin{pmatrix} M_\lambda^{-1} & 0 \\ 0 & a \end{pmatrix}$, then $U_t(\lambda^{**}) = U_t(\lambda)M_1$. Thus

$$\mathcal{I}^{**} = \mathbf{E}[U_t(\lambda_0)M_1 I^{**} M_1^T U_t^T(\lambda_0)], \ \mathcal{J}^{**} = \mathbf{E}[U_t(\lambda_0)M_1 J^{**} M_1^T U_t^T(\lambda_0)] .$$

By Re-scaling $\tilde{\lambda}_n^{**}$, we obtain the QMLE of original model. Thus

$$\sqrt{n}(\tilde{\lambda}_n - \lambda_0) = \sqrt{n}(M_\lambda^{-1}\tilde{\lambda}_n^{**} - M_\lambda^{-1}\lambda_0^{**}) \longrightarrow N(0, M_\lambda^{-1}\mathcal{J}^{**-1}\mathcal{I}^{**}\mathcal{J}^{**-1}M_\lambda^{-1}) .$$

It can be seen the asymptotic variance covariance matrix of the QMLE of original model depend on I^{**}, J^{**}, and M_λ, which depend on choice of h and the distribution of η_0. Thus given the distribution of η_0, the variance of QMLE is decided by the choice of h.

By Remark 2.2.10, if η_0 is symmetric about 0, $\mathbf{E}(g^2(\eta_0^{**})\eta_0^{**}) = 0$ and $\mathbf{E}(g'(\eta_0^{**})\eta_0^{**}) = 0$, then $\tilde{\delta}_n^{**}$ is asymptotically independent with $\tilde{\gamma}_n^{**}$. In this case, we can find a nice variance form of $\tilde{\delta}_n$.

Define $M_\delta = \{M(i,j), 1 \le i,j \le 1+p+q\}$, where $M_\delta(i,j) = 0$, if $i \neq j$, $M_\delta(i,i) = 1/a^2$, if $i \le 1+p$, and $M_\delta(i,i) = 1$, if $i > 1+p$. Then $\delta^{**} = M_\delta \delta$.

by (A.3.5)-(A.3.8) in Appendix 3, we have

$$\frac{\partial {\sigma_t}^2(\lambda^{**})}{\partial \delta^{**}} = \frac{1}{a^2} M_\delta^{-1} \frac{\partial \sigma_t^2(\lambda)}{\partial \delta}, \quad \frac{\partial {\sigma_t}^2(\lambda^{**})/\partial \delta^{**}}{{\sigma_t}^2(\lambda^{**})} = M_\delta^{-1} \frac{\partial \sigma_t^2(\lambda)/\partial \delta}{\sigma_t^2(\lambda)}.$$

By Theorem 2.2.2, we have:

$$\sqrt{n}(\delta_n^{**} - \delta^{**}) \longrightarrow N(0, {\mathcal{J}^{**}}_2^{-1}{\mathcal{I}^{**}}_2{\mathcal{J}^{**}}_2^{-1}),$$

where

$$\mathcal{J}^{**}{}_2^{-1}\mathcal{I}^{**}{}_2\mathcal{J}^{**}{}_2^{-1}$$
$$= \frac{4\Big(\mathbf{E}\big(g(\eta_t^{**})\eta_t^{**}\big)^2-1\Big)}{\Big\{\mathbf{E}g(\eta_t^{**})\eta_t^{**}+g'(\eta_t^{**})\eta_t^{**2}\Big)\Big\}^2}\left\{\mathbf{E}\left(\frac{\partial\sigma_t{}^2(\lambda_0^{**})/\partial\delta^{**}}{\sigma_t{}^2(\lambda_0^{**})}\frac{\partial\sigma_t{}^2(\lambda^{**})/\partial\delta^{**T}}{\sigma_t{}^2(\lambda_0^{**})}\right)\right\}^{-1}$$
$$= 4\tau^{**2}\mathcal{D}^{**-1}\,.$$

Hence

$$\sqrt{n}(\tilde{\delta}_n-\delta_0)=\sqrt{n}(M_\delta^{-1}\tilde{\delta}_n^{**}-M_\delta^{-1}\delta_0^{**})\longrightarrow N(0,4\tau^{**2}M_\delta^{-1}\mathcal{D}^{**-1}M_\delta^{-1}),$$

and

$$\begin{aligned} M_\delta^{-1}\mathcal{D}^{**-1}M_\delta^{-1} &= \left\{\mathbf{E}\left(M_\delta\frac{\partial\sigma_t{}^2(\lambda_0^{**})/\partial\delta^{**}}{\sigma_t{}^2(\lambda_0^{**})}\frac{\partial\sigma_t{}^2(\lambda_0^{**})/\partial^T\delta^{**}}{\sigma_t{}^2(\lambda_0^{**})}M_\delta\right)\right\}^{-1} \\ &= \left\{\mathbf{E}\left(\frac{\partial\sigma_t^2(\lambda_0)/\partial\delta}{\sigma_t^2(\lambda_0)}\frac{\partial\sigma_t^2(\lambda_0)/\partial^T\delta}{\sigma_t^2(\lambda_0)}\right)\right\}^{-1} \\ &= \mathcal{D}^{-1}\,. \end{aligned}$$

It can been seen that when distribution of η_t is fixed, the covariance matrix of QMLE of δ is decided only by τ^{**2}, which depends on choice of h. When $P = Q = 0$, it reduces to pure GARCH, which coincides with the result of Berkes and Horváth (2004). Some numeric computation of effect are left in Section 4.1.2 in Chapter 4.

In general, it is hard to compare the variance of QMLE of γ for ARMA-GARCH.

2.4 Proofs

2.4.1 Proof of Theorem 2.2.1

Proof of Theorem 2.2.1: Theorem 2.2.1 can be proved by standard compactness argument with following results:

(i) $\lim_{n\to\infty} \sup_{\lambda\in\Theta} |L_n(\lambda) - \tilde{L}_n(\lambda)| = 0$. a.s;

(ii) $\mathbf{E}|l_t(\lambda_0)| < \infty$, and if $\lambda \neq \lambda_0$, then $\mathbf{E}l_t(\lambda_0) > \mathbf{E}l_t(\lambda)$,

where (i) is established in Lemma 2.4.3 in Section 2.4.6 and (ii) is established in Lemma 2.4.5.

Since for any $\lambda \in \Theta$, $l_t(\lambda)$ is a stationary sequence with finite mean, which is also ergodic by theorem 3.5.8 of Stout(1974). This implies that

$$\sup_{\lambda\in\Theta} |\mathbf{E}l_t(\lambda) - L_n(\lambda)| \to 0 \quad a.s.$$

Together with (i), (ii) and $\tilde{L}_n(\lambda)$ being maximized at $\tilde{\lambda}_n$, we have:

$$\begin{aligned} 0 &\leq \mathbf{E}l_t(\lambda_0) - \mathbf{E}l_t(\tilde{\lambda}_n) \\ &= \big(\mathbf{E}l_t(\lambda_0) - L_n(\lambda_0)\big) + \big(L_n(\lambda_0) - \tilde{L}_n(\lambda_0)\big) + \big(\tilde{L}_n(\lambda_0) - \tilde{L}_n(\tilde{\lambda}_n)\big) \\ &\quad + \big(\tilde{L}_n(\tilde{\lambda}_n) - L_n(\tilde{\lambda}_n)\big) + \big(L_n(\tilde{\lambda}_n) - \mathbf{E}l_t(\tilde{\lambda}_n)\big) \\ &\leq 2\sup_{\lambda\in\Theta} \big|\mathbf{E}l_t(\lambda) - L_n(\lambda)\big| + 2\sup_{\lambda\in\Theta} \big|L_n(\lambda) - \tilde{L}_n(\lambda)\big| \\ &\to 0\ ,\ a.s. \end{aligned}$$

Thus,

$$\Big|\mathbf{E}l_t(\lambda_0) - \mathbf{E}l_t(\tilde{\lambda}_n)\Big| \to 0\ .$$

Since $\mathbf{E}l_t(\lambda)$ is continuous and has a unique maximum at λ_0 by (ii), we have that

$$\tilde{\lambda}_n \to \lambda_0\ ,\ a.s.$$

This completes the proof of Theorem 2.2.1. END

2.4.2 Proof of Theorem 2.2.2

Proof of Theorem 2.2.2: We adopt the same approach as that in the proof of Theorem 3.2 by Francq and Zakoïan (2004). Theorem 2.2.2 can be proved with following results:

(i) $\mathbf{E}\left|(\partial l_t(\lambda_0)/\partial\lambda)(\partial l_t(\lambda_0)/\partial\lambda^T)\right| < \infty, \ \ \mathbf{E}\left|\partial^2 l_t(\lambda_0)/\partial\lambda\partial\lambda^T\right| < \infty$;

(ii) $\left|n^{-1/2}\sum_{i=1}^n\left(\partial l_t(\lambda_0)/\partial\lambda - \partial\tilde{l}_t(\lambda_0)/\partial\lambda\right)\right| \to 0$
and $\sup_{\lambda\in\Theta}\left|n^{-1/2}\sum_{i=1}^n\left(\partial^2 l_t(\lambda)/\partial\lambda\partial\lambda^T - \partial^2\tilde{l}_t(\lambda)/\partial\lambda\partial\lambda^T\right)\right| \to 0$, in probability as $n\to\infty$;

(iii) $n^{-1}\sum_{t=1}^n(\partial^2 l_t(\lambda^*)/\partial\lambda_i\partial\lambda_j) \to \mathcal{J}(i,j)$, a.s., where λ^* between $\tilde{\lambda}_n$ and λ_0;

(iv) $\mathcal{I}$ is nonsingular and $n^{-1/2}\sum_{i=1}^n(\partial l_t(\lambda_0)/\partial\lambda) \Rightarrow N(0,\mathcal{I})$ ($\Rightarrow$ means converge in distribution);

where (i) is established in Lemma 2.4.8; (ii) is established in Lemma 2.4.11; (iii) is established in Lemma 2.4.12 and (iv) is established in Lemma 2.4.13.

Since by definition of $\tilde{\lambda}_n$, $\sum_{t=1}^n(\partial\tilde{l}_t(\tilde{\lambda}_n)/\partial\lambda)/\sqrt{n} = 0$, then by the mean value theorem, for some λ^* such that $|\lambda^*-\lambda_0| \le |\tilde{\lambda}_n-\lambda_0|$, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n\frac{\partial\tilde{l}_t(\lambda_0)}{\partial\lambda} = \frac{1}{\sqrt{n}}\sum_{t=1}^n\frac{\partial\tilde{l}_t(\lambda_0)}{\partial\lambda} - \frac{1}{\sqrt{n}}\sum_{t=1}^n\frac{\partial\tilde{l}_t(\tilde{\lambda}_n)}{\partial\lambda} = -\sqrt{n}(\tilde{\lambda}_n-\lambda_0)\left(\frac{1}{n}\sum_{t=1}^n\frac{\partial^2\tilde{l}_t(\lambda^*)}{\partial\lambda\partial\lambda^T}\right).$$

Together with (ii), we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n\frac{\partial l_t(\lambda_0)}{\partial\lambda} = -\sqrt{n}(\tilde{\lambda}_n-\lambda_0)\left(\frac{1}{n}\sum_{t=1}^n\frac{\partial^2 l_t(\lambda^*)}{\partial\lambda\partial\lambda^T} + o_p(n^{-1/2})\right) + o_p(1)\ .$$

Thus with (i), (iii) and (iv), by ergodic theorem and Slutsky lemma, we can prove Theorem 2.2.2 if $\mathcal{J}$ is nonsingular. This completes the proof. END

2.4.3 Proofs of Theorem 2.2.3, Corollaries 2.2.1 and 2.2.2

Proof of Theorem 2.2.3: For pure GARCH with nonzero c, after dropping the redundant parameters, the log-likelihood function and its first and second derivatives are simplified. Thus Assumptions 2 and 7 are removed.

In particular, as showed in Appendix A.4,

$$\sup_{\lambda\in\Theta}\left|\frac{\partial\varepsilon_t(c)/\partial\lambda}{\sigma_t(\lambda)}\right|, \sup_{\lambda\in\Theta}\left|\frac{\partial^2\varepsilon_t(c)/\partial\lambda\partial^T\lambda}{\sigma_t(\lambda)}\right|, \sup_{\lambda\in\Theta}\left|\frac{\partial\sigma_t^2(\lambda)/\partial\lambda}{\sigma_t^2(\lambda)}\right|, \sup_{\lambda\in\Theta}\left|\frac{\partial^2\sigma_t^2(\lambda)/\partial\lambda\partial^T\lambda}{\sigma_t^2(\lambda)}\right|$$

have any moments. Thus as a result, the moment condition on ε_t in Lemma 2.4.12 is reduced to $2\iota_1(1-\iota_2)$. Moment condition of η_t in Lemmas 2.4.8-2.4.12 is reduced to $\mathbf{E}|\eta_t|^{\max(1,4\iota_1)}$.

Then by Lemmas 2.4.3-2.4.5 with the reduced moment conditions, consistency in Theorem 2.2.3 can be established by the same approach as that in the proof of Theorem 2.2.1.

By Lemmas 2.4.8-2.4.12 with the reduced moment conditions, normality in Theorem 2.2.3 can be established by the same approach as that in the proof of Theorem 2.2.2. END

Proof of Corollary 2.2.1: Under the given condition, by Lemma 2.4.10, both $\mathcal{I}$ and $\mathcal{J}$ are block-diagonal. By Theorem 2.2.3 and Remark 2.2.10, we have the variance form of $\tilde{c}_n$:

$$\begin{aligned}
\sigma_c^2 &= \mathcal{J}_1^{-1}\mathcal{I}_1\mathcal{J}_1^{-1} = \frac{1}{n}\left\{\frac{1}{2}\mathbf{E}\Big(\frac{\partial\sigma_0^2(\lambda_0)/\partial c}{\sigma_0^2}\Big)^2 + E\frac{1}{\sigma_0^2}\right\}^{-1} \\
&< \frac{1}{n}\left\{\mathbf{E}\frac{1}{\sigma_0^2}\right\}^{-1} \le \frac{1}{n}\mathbf{E}\sigma_0^2 = Var(\bar{\varepsilon}_n) \\
&= \frac{1}{n}\frac{\alpha_{00}}{1-(\alpha_{01}+\ldots+\alpha_{0p})-(\beta_{01}+\ldots+\beta_{0q})}.
\end{aligned}$$

Refer to proof of Theorem 2.1 in Ling (2005), $\sigma_t^2 = \alpha_{00}\left(1 + \sum_{j=1}^{\infty} \underline{1}^T \prod_{i=0}^{j-1} A_{t-i}\varsigma_{t-j}\right)$ a.s, where $\varsigma_t = (\eta_t^2, 0, \cdots, 0, 1, 0, \cdots, 0)^T_{(p+q)\times 1}$ with first component η_t^2 and $(p+1)th$ component 1, and $\underline{1} = (0, \cdots, 0, 1, 0, \cdots, 0)^T_{(p+q)\times 1}$ with $(p+1)th$ component 1. Thus $\sigma_t^2 = \alpha_{00} f_1(\cdot)$ for some function f_1. And

$$\partial\sigma_t^2(\lambda_0)/\partial c = -2\mathcal{B}_\delta^{-1}(L)\mathcal{A}_\delta(L)\mathcal{A}_\gamma(1)\mathcal{B}_\gamma^{-1}(1)\eta_t\sigma_t = \sqrt{\alpha_{00}} f_2(\cdot)$$

for some function f_2. So there is some function f_3, such that $\sigma_c^2 = \alpha_{00} f_3(\cdot)$. All f_1, f_2 and f_3 are not functions of α_{00}. Thus $Var(\bar{\varepsilon}_n)/\sigma_c^2$ is independent of α_{00}. This completes the proof. END

Proof of Corollary 2.2.2:

Since $c_0 = 0$, the parameters are reduced to δ. Assumptions 2 and 7 are removed as in Theorem 2.2.3. Assumption 5 is modified by letting $v = 0$, which implies $\mathbf{E}g(\eta_t)\eta_t + 1 = 0$.

As a result, Lemma 2.4.4 is adjusted as: Under Assumption 3, if there exists some t such that $\sigma_t^2(\delta) = \sigma_t^2(\delta_0)$ almost surely, then $\delta = \delta_0$.

For this special case, the first and second derivatives of $l_t(\delta)$ are further reduced to:

$$\begin{aligned}
\frac{\partial l_t(\delta)}{\partial \delta} &= -\frac{1}{2}\left\{1 + g\big(\eta_t(\delta)\big)\eta_t(\delta)\right\}\frac{\partial\sigma_t^2(\delta)/\partial\delta}{\sigma_t^2(\delta)}, \\
\frac{\partial^2 l_t(\delta)}{\partial\delta\partial\delta^T} &= -\frac{1}{2}\left(1 + g\big(\eta_t(\delta)\big)\eta_t(\delta)\right)\frac{\partial^2\sigma_t^2(\delta)/\partial\delta\partial\delta^T}{\sigma_t^2(\delta)} \\
&\quad + \frac{1}{4}\left(2 + 3g\big(\eta_t(\delta)\big)\eta_t(\delta) + g'\big(\eta_t(\delta)\big)\eta_t^2(\delta)\right)\frac{\partial\sigma_t^2(\delta)/\partial\delta}{\sigma_t^2(\delta)}\frac{\partial\sigma_t^2(\delta)/\partial\delta^T}{\sigma_t^2(\delta)}.
\end{aligned}$$

Based on above forms of the first and second derivatives of $l_t(\delta)$, Lemma 2.4.9 is modified as: Under Assumptions 1, 3 to 6, if $\mathbf{E}|\eta_t|^{4\iota_1} < \infty$, then $\mathcal{I}$ is nonsingular.

Further with $\mathbf{E}g'(\eta_0)\eta_0^2 \leq 1$, then $\mathcal{J}$ is nonsingular.

Lemma 2.4.10 is even not necessary in this case.

Since $\mathbf{E}g(\eta_t)\eta_t = -1$, we have $\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1} = 4\tau^2\mathcal{D}^{-1}$ with forms of τ^2 and $\mathcal{D}$ as given in Corollary 2.2.2. For this special case, the asymptotic covariance matrix has a nice form, which is helpful in comparing efficiencies of different QMLEs based on different likelihood kernels.

END

2.4.4 Proof of Theorem 2.2.4

Proof of Theorem 2.2.4: For pure ARMA, some conditions and assumptions required for ARMA-GARCH are removed. Redundant Assumptions 3 and 7 are dropped off. Assumption 5 is modified by letting $w = 1$. Assumption 4 is modified by letting $\iota_1 = 1$.

Due to simplification of the log-likelihood function and its first and second derivatives, Lemmas 2.4.3-2.4.12 are modified as that shown in Appendix A.5. In particular, Lemma 2.4.12 holds with $\mathbf{E}|\varepsilon_t|^2 < \infty$.

The consistency in Theorem 2.2.4 can be established by the same approach as that in the proof of Theorem 2.2.1 by Lemma 2.4.3-2.4.5 with modified conditions.

The normality in Theorem 2.2.4 can be established by the same approach as that in the proof of Theorem 2.2.2 by Lemma 2.4.8-2.4.12 with modified conditions. END

2.4.5 Proof of Theorem 2.2.5

Proof of Theorem 2.2.5: First, by definition, $\mathbf{E}l_t(\gamma_0, \delta)$ is continuous and defined on a compact space. Second, by Lemma 2.4.5, $\mathbf{E}l_t(\gamma_0, \delta)$ exists and is uniquely maximized at δ_0. $L_n(\gamma_0, \delta)$ is stationary and ergodic, by Lemma 2.4.15 and ergodic theorem, for any $\delta \in \Theta_\delta$, $L_n(\gamma_0, \delta) \to \mathbf{E}l_t(\gamma_0, \delta)$ a.s.. Third, $\tilde{L}_n(\hat{\gamma}_n, \delta)$ is continuous on $\delta \in \Theta_\delta$ and is a measurable function of $\{Y_t, Y_{t-1}, ...\}$ for all $\delta \in \Theta_\delta$. By Lemma 2.4.14, $\tilde{L}_n(\hat{\gamma}_n, \delta)$ converge to $L_n(\gamma_0, \delta)$ in probability uniformly for all $\delta \in \Theta_\delta$. Hence $\tilde{L}_n(\gamma_n, \delta) \to \mathbf{E}l_t(\gamma_0, \delta)$ uniformly for all $\delta \in \Theta_\delta$. These meet the conditions in Theorem 1.4.1 (Newey and Mcfadden, 1994). Thus $\hat{\delta}_n \to \delta_0$ in probability.

δ_0 is in the interior of Θ_δ. $\mathbf{E}\partial^2 l_t(\gamma_0, \delta)/\partial\delta\partial\delta^T$ exists and is continuous in Θ_δ. Since $\partial^2 L_n(\gamma_0, \delta)/\partial\delta\partial\delta^T$ is stationary and ergodic, by ergodic theorem, for each $\delta \in \Theta_\delta$, $\partial^2 L_n(\gamma_0, \delta)/\partial\delta\partial\delta^T \to \mathbf{E}\partial^2 l_t(\gamma_0, \delta)/\partial\delta\partial\delta^T$ a.s. In addition, $\partial^2 \tilde{L}_n(\gamma_0, \delta)/\partial\delta\partial\delta^T$ is continuous in Θ_δ. By Lemma 2.4.16, $\partial^2 \tilde{L}_n(\gamma_0, \delta)/\partial\delta\partial\delta^T \to \mathbf{E}\partial^2 l_t(\gamma_0, \delta)/\partial\delta\partial\delta^T$ in probability uniformly in a neighborhood of δ_0. From the proof of Lemma 2.4.8, we have that $\mathbf{E}\partial^2 l_t(\gamma_0, \delta_0)/\partial\delta\partial\delta^T$ is nonsingular. Together with Lemma 2.4.18, we verify the conditions in Theorem 1.4.2 (Newey and Mcfadden, 1994). Hence $\sqrt{n}(\delta_n - \delta_0)$ converges to a multivariate normal distribution. This completes the proof. $\boxed{\text{END}}$

2.4.6 Proofs of Lemmas

Throughout the rest of this Chapter, denote the spectral radius of a square matrix A as $\rho(A)$. Let $K > 0$, $0 < \rho < 1$ and $0 < \zeta < 1$ be generic constants. K, ρ and ζ may take different values from place to place. So we can write

$$0 < K \sum_{i \geq i_1} \rho_1^i + K \sum_{i \geq i_2} i\rho_2^i \leq K\rho^{\min(i_1, i_2)},$$

where $0 < \rho_1 < 1$, $0 < \rho_2 < 1$, $i_1 \geq 0$ and $i_2 \geq 0$.

Adopting the notation of Francq and Zakoïan (2004), let

$$\underline{\sigma}_t^2(\lambda) = \begin{pmatrix} \sigma_t^2(\lambda) \\ \sigma_{t-1}^2(\lambda) \\ \vdots \\ \sigma_{t-q+1}^2(\lambda) \end{pmatrix}_{q\times 1}, \quad \underline{c}_t(\lambda) = \begin{pmatrix} \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2(\gamma) \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{q\times 1},$$

$$B = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_q \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}_{q\times q}.$$

Then (2.2.2) can be written in vector form:

$$\underline{\sigma}_t^2(\lambda) = \underline{c}_t(\lambda) + B\underline{\sigma}_{t-1}^2(\lambda) . \tag{2.4.1}$$

Repeat (2.4.1), we obtain

$$\begin{aligned} \underline{\sigma}_t^2(\lambda) &= \underline{c}_t(\lambda) + B\underline{c}_{t-1}(\lambda) + B^2\underline{c}_{t-2}(\lambda) + \ldots + B^{t-p-1}\underline{c}_{p+1}(\lambda) \\ &\quad + B^{t-p}\underline{c}_p(\lambda) + \ldots + B^{t-1}\underline{c}_1(\lambda) + B^t\underline{\sigma}_0^2(\lambda) = \sum_{i=0}^{\infty} B^i \underline{c}_{t-i}(\lambda) . \end{aligned} \tag{2.4.2}$$

Let $\underline{\tilde{\sigma}}_t^2(\lambda)$ be the vector obtained by replacing $\sigma_{t-i}^2(\lambda)$ with $\tilde{\sigma}_{t-i}^2(\lambda)$ in $\underline{\sigma}_t^2(\lambda)$, and $\underline{\tilde{c}}_t(\lambda)$ be the vector obtained by replacing $\varepsilon_{t-i}^2(\gamma)$ with $\tilde{\varepsilon}_{t-i}^2(\gamma)$ in $\underline{c}_t(\lambda)$. Then we have the vector form of $\tilde{\sigma}_t^2(\lambda)$:

$$\begin{aligned} \underline{\tilde{\sigma}}_t^2(\lambda) &= \underline{\tilde{c}}_t(\lambda) + B\underline{\tilde{\sigma}}_{t-1}^2(\lambda) \\ &= \underline{\tilde{c}}_t(\lambda) + B\underline{\tilde{c}}_{t-1}(\lambda) + B^2\underline{\tilde{c}}_{t-2}(\lambda) + \ldots + B^{t-p-1}\underline{\tilde{c}}_{p+1}(\lambda) + B^{t-p}\underline{\tilde{c}}_p(\lambda) \\ &\quad + \ldots + B^{t-1}\underline{\tilde{c}}_1(\lambda) + B^t\underline{\tilde{\sigma}}_0^2 . \end{aligned} \tag{2.4.3}$$

Lemma 2.4.1. *Under Assumption 2, we have* $\sup_{\phi\in\Theta_\phi} |a_\phi(i)| = O(\rho^i)$, $\sup_{\varphi\in\Theta_\varphi} |a_\varphi(i)| = O(\rho^i)$, $\sup_{\gamma\in\Theta_\gamma} |a_\gamma(i)| = O(\rho^i)$, $\sup_{\varphi\in\Theta_\varphi} |\partial a_\varphi(i)/\partial\varphi_j| = O(\rho^i)$ *for* $j = 1, 2, ..., Q$. *Under Assumption 3, we have* $\sup_{\beta\in\Theta_\beta} |a_\beta(i)| = O(\rho^i)$, *and* $\sup_{\delta\in\Theta_\delta} |a_\delta(i)| = O(\rho^i)$, *further with Assumption 1, we have* $\sup_{\lambda\in\Theta} \rho(B) < 1$.

Proof: Referring to (2.3) in Ling (2005), we have $\sup_{\varphi\in\Theta_\varphi}|a_\varphi(i)| = O(\rho^i)$, and $\sup_{\gamma\in\Theta_\gamma}|a_\gamma(i)| = O(\rho^i)$. Referring to (2.5) in Ling (2005), we have $\sup_{\beta\in\Theta_\beta}|a_\beta(i)| = O(\rho^i)$, and $\sup_{\delta\in\Theta_\delta}|a_\delta(i)| = O(\rho^i)$. Referring to (2.4) in Ling (2005), we have $\sup_{\lambda\in\Theta}\rho(B) < 1$. By the definition and Assumption 2, $\mathcal{A}_\gamma^{-1}(z)$ has the same property as $\mathcal{B}_\gamma^{-1}(z)$, thus $\sup_{\phi\in\Theta_\phi}|a_\phi(i)| = O(\rho^i)$.

From the expressions of $a_\varphi(i)$ as in A.2, $a_\varphi(i) = 0$ for $i < 0$, then by (A.2.1), there exist $0 < \rho_1 < \rho_2 < 1$, such that:

$$\sup_{\varphi\in\Theta_\varphi}\left|\frac{\partial a_\varphi(i)}{\partial\varphi_j}\right| = \sup_{\varphi\in\Theta_\varphi}\left|\sum_{k=0}^{i-j} a_\varphi(k)a_\varphi(i-j-k)\right| = \sum_{k=0}^{i-j} O(\rho_1^k)O(\rho_1^{i-j-k}) = |i-j|O(\rho_1^{i-j}) = O(\rho_2^i)\,.$$

This completes the proof of Lemma 2.4.1. END

In the following proofs, we will frequently use the inequality below. Suppose $\{X_i, -\infty < t < \infty\}$ is a strictly stationary sequence, $m \geq 0$ is a integer, and $a_i = O(\rho^i)$, then we have:

$$\begin{aligned}
\mathbf{E}\left|\sum_{i=0}^{\infty} a_i X_i\right|^{m+\zeta} &\leq \mathbf{E}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\cdots\sum_{k=0}^{\infty}|a_i a_j\cdots a_k X_i X_j\cdots X_k|\left|\sum_{l=0}^{\infty} a_l X_l\right|^{\zeta} \\
&\leq \mathbf{E}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\cdots\sum_{k=0}^{\infty}|a_i a_j\cdots a_k X_i X_j\cdots X_k|\sum_{l=0}^{\infty}|a_l X_l|^{\zeta} \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\cdots\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}|a_i a_j\cdots a_k a_l^{\zeta}|\mathbf{E}|X_i X_j\cdots X_k X_l^{\zeta}|\,, \\
&\leq K\mathbf{E}|X_1|^{m+\zeta}\,,
\end{aligned} \tag{2.4.4}$$

by the fact $\mathbf{E}|X_i X_j\cdots X_k X_l^{\zeta}| \leq \mathbf{E}|X_1|^{m+\zeta}$.

Lemma 2.4.2. *Under (1.2.7), then there exists some $s > 0$, such that*

$$\mathbf{E}\sup_{\gamma\in\Theta_\gamma}|\varepsilon_t(\gamma)|^{2s} < \infty, \mathbf{E}\sup_{\lambda\in\Theta}|\sigma_t(\lambda)|^{2s} < \infty, \mathbf{E}\sup_{\gamma\in\Theta_\gamma}|\tilde{\varepsilon}_t(\gamma)|^{2s} < \infty, \ \textit{and}\ \mathbf{E}\sup_{\lambda\in\Theta}|\tilde{\sigma}_t(\lambda)|^{2s} < \infty.$$

Proof: Refer to Proposition 1 in Francq and Zakoïan (2004), under (1.2.7), we have

$$\mathbf{E}|\varepsilon_t|^{2s} < \infty, \text{and } \mathbf{E}\sigma_t^{2s} < \infty.$$

Then by (1.2.12), (2.4.4) and Lemma 2.4.1, we have

$$\mathbf{E}|Y_t - c_0|^{2s} = \mathbf{E}|\mathcal{A}_{\gamma_0}^{-1}(L)\mathcal{B}_{\gamma_0}(L)\varepsilon_t|^{2s} = \mathbf{E}\left|\mathcal{B}_{\gamma_0}(L)\sum_{i=0}^{\infty} a_{\phi_0}(i)\varepsilon_{t-i}\right|^{2s} \leq K\mathbf{E}|\varepsilon_0|^{2s} < \infty.$$

Hence $\mathbf{E}\sup_{\gamma\in\Theta_\gamma}|Y_t - c|^{2s} \leq K\mathbf{E}|Y_t - c_0|^{2s} + K\sup_{\gamma\in\Theta_\gamma}|c - c_0|^{2s} < \infty$. Again by (2.2.3), (2.2.4), (2.4.4) and Lemma 2.4.1, we have

$$\mathbf{E}\sup_{\gamma\in\Theta_\gamma}|\varepsilon_t(\gamma)|^{2s} = \mathbf{E}\sup_{\gamma\in\Theta_\gamma}\left|\sum_{i=0}^{\infty} a_\gamma(i)(Y_{t-i} - c)\right|^{2s} \leq K\mathbf{E}\sup_{\gamma\in\Theta_\gamma}|Y_1 - c|^{2s} < \infty,$$

$$\begin{aligned}
\mathbf{E}\sup_{\lambda\in\Theta}|\sigma_t(\lambda)|^{2s} &\leq K\sup_{\lambda\in\Theta}|\mathcal{B}_\delta^{-1}(1)\alpha_0|^s + K\mathbf{E}\sup_{\lambda\in\Theta}\left|\sum_{i=0}^{\infty} a_\delta(i)\varepsilon_{t-i}^2(\gamma)\right|^s \\
&\leq K\sup_{\lambda\in\Theta}|\mathcal{B}_\delta^{-1}(1)\alpha_0|^s + K\mathbf{E}\sup_{\gamma\in\Theta_\gamma}|\varepsilon_0(\gamma)|^{2s} \\
&< \infty.
\end{aligned}$$

Similarly, by (2.2.7), (2.2.8), (2.4.4), and Lemma 2.4.1, we have

$$\begin{aligned}
\mathbf{E}\sup_{\gamma\in\Theta_\gamma}|\tilde{\varepsilon}_t(\gamma)|^{2s} &\leq K\mathbf{E}\sup_{\gamma\in\Theta_\gamma}\left|\sum_{i=0}^{t-1} a_\gamma(i)(Y_{t-i} - c)\right|^{2s} + KO(\rho^{2st}) \\
&\leq K\mathbf{E}\sup_{\gamma\in\Theta_\gamma}|Y_1 - c|^{2s} + KO(\rho^{2st}) \\
&< \infty\,,
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}\sup_{\lambda\in\Theta}|\tilde{\sigma}_t(\lambda)|^{2s} &\leq K\sup_{\lambda\in\Theta}|\mathcal{B}_{\delta,t}^{-1}\alpha_0|^s + K\mathbf{E}\sup_{\lambda\in\Theta}\left|\sum_{i=0}^{t-1} a_\delta(i)\tilde{\varepsilon}_{t-i}^2(\gamma)\right|^s + KO(\rho^{st}) \\
&\leq K + K\mathbf{E}\sup_{\lambda\in\Theta}\left(\sum_{i=0}^{t-1} a_\delta(i)\sum_{j=0}^{t-i} a_\gamma^2(j)|Y_{t-i-j} - c|^2\right)^s \\
&\leq K + K\mathbf{E}\sup_{\gamma\in\Theta_\gamma}|\tilde{\varepsilon}_0(\gamma)|^{2s} \\
&< \infty\,.
\end{aligned}$$

This completes the proof of Lemma 2.4.2. END

In following proofs, we will frequently use the technique of

$$\frac{x}{b} - \frac{\tilde{x}}{\tilde{b}} = x\left(\frac{1}{b} - \frac{1}{\tilde{b}}\right) + \frac{x - \tilde{x}}{\tilde{b}}.$$

Lemma 2.4.3. *Under Assumptions 1 - 3 and Assumption 4, we have*

$$\lim_{n\to\infty} \sup_{\lambda\in\Theta} |L_n(\lambda) - \tilde{L}_n(\lambda)| = 0 \quad a.s..$$

Proof: Referring to (4.38) and (4.39) in the proof of Theorem 3.1 in Francq and Zakoïan (2004), we have

$$\sup_{\gamma\in\Theta_\gamma} |\varepsilon_k(\gamma) - \tilde{\varepsilon}_k(\gamma)| \le K\rho^k, a.s, \tag{2.4.5}$$

$$\sup_{\lambda\in\Theta} |\underline{\sigma}_t^2(\lambda) - \tilde{\underline{\sigma}}_t^2(\lambda)| \;\le\; K\rho^t \left(\sum_{k=1-p}^{t-1} \sup_{\gamma\in\Theta_\gamma} |\varepsilon_k(\gamma)| + 1\right) \quad a.s. \tag{2.4.6}$$

Based on (2.2.9), (2.2.10), by mean value theorem, there exists $\eta_t^*(\lambda)$ such that $|\eta_t(\lambda) - \eta_t^*(\lambda)| \le |\eta_t(\lambda) - \tilde{\eta}_t(\lambda)|$, we have

$$\begin{aligned}
&\sup_{\lambda\in\Theta} |L_n(\lambda) - \tilde{L}_n(\lambda)| \\
&= \sup_{\lambda\in\Theta} \left| \frac{1}{n}\sum_{t=1}^{n} \log\frac{h(\eta_t(\lambda))}{\sigma_t(\lambda)} - \frac{1}{n}\sum_{t=1}^{n} \log\frac{h(\tilde{\eta}_t(\lambda))}{\tilde{\sigma}_t(\lambda)} \right| \\
&= \sup_{\lambda\in\Theta} \left| \frac{1}{n}\sum_{t=1}^{n} \log\frac{\tilde{\sigma}_t(\lambda)}{\sigma_t(\lambda)} + \frac{1}{n}\sum_{t=1}^{n} \log\frac{h(\eta_t(\lambda))}{h(\tilde{\eta}_t(\lambda))} \right| \\
&\le \sup_{\lambda\in\Theta} \left| \frac{1}{n}\sum_{t=1}^{n} \log\frac{\tilde{\sigma}_t(\lambda)}{\sigma_t(\lambda)} \right| + \sup_{\lambda\in\Theta} \left| \frac{1}{n}\sum_{t=1}^{n} g(\eta_t^*(\lambda))\left(\eta_t(\lambda) - \tilde{\eta}_t(\lambda)\right) \right|
\end{aligned}$$

Now we will prove these two items converge to 0 almost surely respectively.

Note for any $\lambda \in \Theta$, with probability 1, and $\inf_{\lambda\in\Theta} \sigma_t(\lambda) \geq \alpha_0 > 0$, $\inf_{\lambda\in\Theta} \tilde{\sigma}_t(\lambda) \geq \alpha_0 > 0$. By (2.4.6) and the fact $\log(1+|x|) \leq |x|$, we have:

$$
\begin{aligned}
&-\frac{1}{n}\sum_{t=1}^{n}\log\frac{\tilde{\sigma}_t(\lambda)}{\sigma_t(\lambda)} \\
&= \sup_{\lambda\in\Theta}\frac{1}{2n}\sum_{t=1}^{n}\log\left(\frac{\sigma_t^2(\lambda)-\tilde{\sigma}_t^2(\lambda)}{\tilde{\sigma}_t^2(\lambda)}+1\right) \\
&\leq \sup_{\lambda\in\Theta}\frac{1}{2n}\sum_{t=1}^{n}\log\left(K\rho^t\left(\sum_{k=1-p}^{t-1}\sup_{\gamma\in\Theta_\gamma}|\varepsilon_k(\gamma)|+1\right)+1\right) \\
&\leq \sup_{\lambda\in\Theta}\frac{K}{2n}\sum_{t=1}^{n}\rho^t\left(\sum_{k=1-p}^{t-1}\sup_{\gamma\in\Theta_\gamma}|\varepsilon_k(\gamma)|+1\right).
\end{aligned}
$$

By Cesàro lemma, above expression converges to 0 almost surely if

$$
\rho^t\left(\sum_{k=1-p}^{t-1}\sup_{\gamma\in\Theta_\gamma}|\varepsilon_k(\gamma)|+1\right) \to 0,\ \textit{a.s.}\ \text{if}\ t\to\infty. \tag{2.4.7}
$$

We know $\mathbf{E}(X+Y)^\zeta \leq \mathbf{E}X^\zeta + \mathbf{E}Y^\zeta$ for all positive r.v X and Y, $0<\zeta<1$. By Markov inequality and Lemma 2.4.2, taking ζ small enough, we have

$$
\begin{aligned}
&\sum_{t=1}^{\infty}P\left\{\rho^t\left(\sum_{k=1-p}^{t-1}\sup_{\gamma\in\Theta_\gamma}|\varepsilon_k(\gamma)|+1\right)>\epsilon\right\} \\
&\leq \sum_{t=1}^{\infty}\frac{\rho^{t\zeta}}{\epsilon^\zeta}\sum_{i=1-p}^{t-1}\mathbf{E}\sup_{\lambda\in\Theta}(|\varepsilon_i(\gamma)|+1)^\zeta \\
&\leq K\sum_{t=1}^{\infty}\frac{\rho^{t\zeta}t}{\epsilon^\zeta} \\
&< \infty,
\end{aligned}
$$

Thus by Borel-Cantelli lemma, we have

$$
\sup_{\lambda\in\Theta}\frac{-1}{n}\sum_{t=1}^{n}\log\left(\frac{\tilde{\sigma}_t(\lambda)}{\sigma_t(\lambda)}\right) \longrightarrow 0\ \ \textit{a.s.}
$$

Similarly ,we can show

$$\sup_{\lambda\in\Theta}\frac{1}{n}\sum_{t=1}^{n}\log\left(\frac{\tilde\sigma_t(\lambda)}{\sigma_t(\lambda)}\right)\longrightarrow 0 \ \ a.s.$$

Then we can claim that

$$\sup_{\lambda\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\log\left(\frac{\tilde\sigma_t(\lambda)}{\sigma_t(\lambda)}\right)\right|\longrightarrow 0 \ \ a.s.$$

Now we will prove the second convergence.

$$\begin{aligned}
&\sup_{\lambda\in\Theta}\frac{1}{n}\sum_{t=1}^{n}\left|g(\eta_t^*(\lambda))\left(\eta_t(\lambda)-\tilde\eta_t(\lambda)\right)\right| \\
\le\ & \sup_{\lambda\in\Theta}\frac{1}{n}\sum_{t=1}^{n}|g(\eta_t^*(\lambda))|\frac{|\eta_t(\lambda)||\tilde\sigma_t^2(\lambda)-\sigma_t^2(\lambda)|}{(\tilde\sigma_t(\lambda)+\sigma_t(\lambda))\tilde\sigma_t(\lambda)}+\sup_{\lambda\in\Theta}\frac{1}{n}\sum_{t=1}^{n}|g(\eta_t^*(\lambda))|\frac{|\varepsilon_t(\gamma)-\tilde\varepsilon_t(\gamma)|}{\tilde\sigma_t(\lambda)} \\
\le\ & \sup_{\lambda\in\Theta}\frac{K}{n}\sum_{t=1}^{n}\rho^t\Big(\left|g(\eta_t^*(\lambda))\eta_t(\lambda)\right|+\left|g(\eta_t^*(\lambda))\right|\Big)\sum_{k=1-p}^{t-1}(|\varepsilon_k(\gamma)|+1)\ .
\end{aligned}$$

Again by Cesàro lemma, above expression converge to 0 almost surely if

$$\rho^t\sup_{\lambda\in\Theta}\left|g(\eta_t^*(\lambda))\eta_t(\lambda)\right|+\left|g(\eta_t^*(\lambda))\right|\sum_{k=1-p}^{t-1}(|\varepsilon_k(\gamma)|+1)\to 0,\ a.s. \tag{2.4.8}$$

Similarly, by Markov inequality, Cauchy inequality and Assumption 4, we have

$$\begin{aligned}
&\sum_{t=1}^{\infty}P\left(\rho^t\sup_{\lambda\in\Theta}\left|g(\eta_t^*(\lambda))\eta_t(\lambda)\right|+\left|g(\eta_t^*(\lambda))\right|\sum_{i=1-p}^{t-1}(|\varepsilon_i(\gamma)|+1)>\epsilon\right) \\
\le\ & K\sum_{t=1}^{\infty}\frac{\rho^{t\varsigma}}{\epsilon^\varsigma}\sum_{i=1-p}^{t-1}\mathbf{E}\sup_{\lambda\in\Theta}\Big(\left|g(\eta_t^*(\lambda))\eta_t(\lambda)\right|+\left|g(\eta_t^*(\lambda))\right|\Big)^\varsigma(|\varepsilon_i(\gamma)|+1)^\varsigma \\
\le\ & K\sum_{t=1}^{\infty}\frac{\rho^{t\varsigma}}{\epsilon^\varsigma}\sum_{i=1-p}^{t-1}\mathbf{E}\sup_{\lambda\in\Theta}\left|g^2(\eta_t^*(\lambda))+\eta_t^2(\lambda)\right|^\varsigma(|\varepsilon_i(\gamma)|+1)^\varsigma \\
\le\ & K\sum_{t=1}^{\infty}\frac{\rho^{t\varsigma}}{\epsilon^\varsigma}\sum_{i=1-p}^{t}\left\{\left[\mathbf{E}\left(\max(\sup_{\lambda\in\Theta}|\eta_t^*(\lambda)|,1)\right)^{4(2\iota_1-1)\varsigma}+\mathbf{E}\sup_{\lambda\in\Theta}|\eta_t^{4\varsigma}(\lambda)|\right]^{1/2}\right. \\
&\left.\left(\mathbf{E}\sup_{\gamma\in\Theta_\gamma}|\varepsilon_i(\gamma)|^{2\varsigma}+1\right)^{1/2}\right\} \\
<\ & \infty,
\end{aligned}$$

since by Lemma 2.4.2,

$$\mathbf{E}\sup_{\lambda\in\Theta}|\eta_t(\lambda)|^{\max(4(2\iota_1-1)\zeta,4\zeta)} \le K\mathbf{E}\sup_{\gamma\in\Theta_\gamma}|\varepsilon_t(\gamma)|^{\max(4(2\iota_1-1)\zeta,4\zeta)} \le K\mathbf{E}\sup_{\gamma\in\Theta_\gamma}|\varepsilon_t(\gamma)|^{2s} < \infty\,,$$

$$\mathbf{E}\sup_{\lambda\in\Theta}|\tilde{\eta}_t(\lambda)|^{\max(4(2\iota_1-1)\zeta,4\zeta)} \le K\mathbf{E}\sup_{\gamma\in\Theta_\gamma}|\tilde{\varepsilon}_t(\gamma)|^{\max(4(2\iota_1-1)\zeta,4\zeta)} \le K\mathbf{E}\sup_{\gamma\in\Theta_\gamma}|\tilde{\varepsilon}_t(\gamma)|^{2s} < \infty\,,$$

for some $\zeta > 0$ such that $\max(4(2\iota_1-1)\zeta, 4\zeta) \le 2s$. Hence $\mathbf{E}|\eta_t^*(\lambda)|^{\max(4(2\iota_1-1)\zeta,4\zeta)} < \infty$.

Thus by Borel-Cantelli lemma, (2.4.8) holds.

This completes the proof of Lemma 2.4.3. END

Lemma 2.4.4. *Under Assumptions 1 to 3, if there exists some t such that $\varepsilon_t(\gamma) = \varepsilon_t(\gamma_0)$ and $\sigma_t^2(\lambda) = \sigma_t^2(\lambda_0)$ almost surely, then $\lambda = \lambda_0$.*

Proof: We refer to (ii) in the proof of Theorem 3.1 in Francq and Zakoïan (2004). END

Denote $\mathcal{F}_{t-1}$ as the σ-algebra generated by η_{t-i}, $i \ge 1$.

Lemma 2.4.5. *Under Assumptions 1-5, if $\mathbf{E}|\eta_t|^{2\iota_1} < \infty$ (or $\mathbf{E}|\eta_t|^s < \infty$ for some $s > 0$), we have $\mathbf{E}|l_t(\lambda_0)| < \infty$. Further if $\lambda \ne \lambda_0$, then we have $\mathbf{E}l_t(\lambda_0) > \mathbf{E}l_t(\lambda)$.*

Proof: By Jensen's inequality and Lemma 2.4.2, for some small $s > 0$,

$$\mathbf{E}|\log\sigma_t| = \frac{1}{s}\mathbf{E}|\log\sigma_t^s| \le \frac{1}{s}\log^+\mathbf{E}\sigma_t^s + \frac{1}{s}\log^-\mathbf{E}\sigma_t^s < \infty$$

since $\log^-\mathbf{E}\sigma_t^s \le \max(0, -(\log\alpha_{00}^s)/2) < \infty$. Then by Assumption 4, we have

$$\mathbf{E}|l_t(\lambda_0)| \le \mathbf{E}|\log\sigma_t| + \mathbf{E}|\log h(\eta_t)| \le \mathbf{E}|\log\sigma_t| + C\mathbf{E}|\eta_t|^{2\iota_1} < \infty.$$

By (2.2.10) and the fact $\varepsilon_t = \varepsilon_t(\gamma_0)$, $\sigma_t = \sigma_t(\lambda_0)$ and $\eta_t = \varepsilon_t(\gamma_0)/\sigma_t(\lambda_0)$, we have

$$l_t(\lambda) = \log\left[\frac{\sigma_t}{\sigma_t(\lambda)}h\left(\eta_t\frac{\sigma_t}{\sigma_t(\lambda)} + \frac{\varepsilon_t(\gamma)-\varepsilon_t(\gamma_0)}{\sigma_t(\lambda)}\right)\right] - \log\sigma_t\,.$$

By the expressions of $\varepsilon_t(\gamma)$ and $\varepsilon_t(\gamma_0)$, we have $\varepsilon_t(\gamma) - \varepsilon_t(\gamma_0)$ is $\mathcal{F}_{t-1}$ adapted. Then by Remark 2.2.3 and independence of η_t with $\mathcal{F}_{t-1}$, we have for $\forall\ \lambda$,

$$\mathbf{E}\log h(\eta_t) - \mathbf{E}\left\{\log\left[\frac{\sigma_t}{\sigma_t(\lambda)}h\left(\eta_t\frac{\sigma_t}{\sigma_t(\lambda)} + \frac{\varepsilon_t(\gamma)-\varepsilon_t(\gamma_0)}{\sigma_t(\lambda)}\right)\right]\middle|\mathcal{F}_{t-1}\right\} \geq 0\,.$$

Hence $\forall\ \lambda$

$$\begin{aligned} &\mathbf{E}l_t(\lambda_0) - \mathbf{E}l_t(\lambda)| \\ &= \mathbf{E}\log h(\eta_t) - \mathbf{E}\log\left[\frac{\sigma_t}{\sigma_t(\lambda)}h\left(\eta_t\frac{\sigma_t}{\sigma_t(\lambda)} + \frac{\varepsilon_t(\gamma)-\varepsilon_t(\gamma_0)}{\sigma_t(\lambda)}\right)\right] \\ &= \mathbf{E}\left\{\mathbf{E}\log h(\eta_t) - \mathbf{E}\left\{\log\left[\frac{\sigma_t}{\sigma_t(\lambda)}h\left(\eta_t\frac{\sigma_t}{\sigma_t(\lambda)} + \frac{\varepsilon_t(\gamma)-\varepsilon_t(\gamma_0)}{\sigma_t(\lambda)}\right)\right]\middle|\mathcal{F}_{t-1}\right\}\right\} \\ &\geq 0\,. \end{aligned}$$

Thus $\mathbf{E}l_t(\lambda_0) \geq \mathbf{E}l_t(\lambda)$. If there exists λ, such that $\mathbf{E}l_t(\lambda_0) = \mathbf{E}l_t(\lambda)$. Then by the fact $\mathbf{E}X = 0$ for non-negative random variable X if and only if $X = 0$ almost surely, we have

$$\mathbf{E}\log h(\eta_t) - \mathbf{E}\left\{\log\left[\frac{\sigma_t}{\sigma_t(\lambda)}h\left(\eta_t\frac{\sigma_t}{\sigma_t(\lambda)} + \frac{\varepsilon_t(\gamma)-\varepsilon_t(\gamma_0)}{\sigma_t(\lambda)}\right)\right]\middle|\mathcal{F}_{t-1}\right\} = 0\ , a.s.$$

By Remark 2.2.3, above equation holds if and only if

$$\frac{\sigma_t}{\sigma_t(\lambda)} = 1, a.s. \quad \text{and} \quad \frac{\varepsilon_t(\gamma)-\varepsilon_t(\gamma_0)}{\sigma_t(\lambda)} = 0,\ a.s.$$

Thus by Lemma 2.4.4, we have $\lambda = \lambda_0$. This completes the proof. END

Let $\xi_{\rho,t} = 1 + \sum_{i=0}^{\infty}\rho^i|Y_{t-i} - c_{00}|$, and $\xi_{0,\rho,t} = 1 + \sum_{i=0}^{\infty}\rho^i|\varepsilon_{t-i}|$, where constant c_{00} is the true value of c_0.

Lemma 2.4.6. *Under Assumptions 1 to 3, we have for some constants $C > 0$ and $0 < \rho < 1$, such that*

(i) $\sup_{\gamma\in\Theta_\gamma}|\varepsilon_{t-1}(\gamma)|$, $\sup_{\gamma\in\Theta_\gamma}|\partial\varepsilon_t(\gamma)/\partial\gamma|$, *and* $\sup_{\gamma\in\Theta_\gamma}|\partial^2\varepsilon_t(\gamma)/\partial\gamma\partial\gamma^T|$ *are bounded by* $C\xi_{\rho,t-1}$*;*

(ii) $\sup_{\lambda\in\Theta}\sigma_t^2(\lambda)$ *is bounded by* $C\xi_{\rho,t-1}^2$*;*

(iii) *There exists a neighbor* Θ_0 *of* λ_0 *such that* $\sup_{\lambda\in\Theta_0}\sigma_t^{-2}(\lambda)\left|\partial\sigma_t^2(\lambda)/\partial\delta\right|$ *and* $\sup_{\lambda\in\Theta_0}\sigma_t^{-2}(\lambda)\left|\partial^2\sigma_t^2(\lambda)/\partial\delta\partial\delta^T\right|$ *are bounded by* $C\xi_{\rho,t-1}^s$ *for any* $s>0$*;*

(iv) $\sup_{\lambda\in\Theta}\sigma_t^{-1}(\lambda)\left|\partial\sigma_t^2(\lambda)/\partial\gamma\right|$, $\sup_{\lambda\in\Theta}\sigma_t^{-1}(\lambda)\left|\partial^2\sigma_t^2(\lambda)/\partial\gamma\partial\gamma^T\right|$, *and* $\sup_{\lambda\in\Theta}\sigma_t^{-1}(\lambda)\left|\partial^2\sigma_t^2(\lambda)/\partial\gamma\partial\delta^T\right|$ *are bounded by* $C\xi_{\rho,t-1}$*;*

(v) *There exists* $0<\rho,\varrho<1$*, such that* $\xi_{\varrho,t}\le C\xi_{0,\rho,t}$ *;*

(vi) *For any* $k>0$*, if* $\mathbf{E}|\varepsilon_0|^k<\infty$*, then* $\mathbf{E}\xi_{0,\rho,t-1}^k<\infty$.

Proof: (i) and (ii) are the same as Lemma A.1, (iii) as Lemma A.2, (iv) as Lemma A.3, and (v) as Lemma A.5 in Ling (2005), respectively.

By (2.4.4), for $0<\rho<1$, we have

$$\mathbf{E}\xi_{0,\rho,t-1}^k\le C+C\mathbf{E}|\varepsilon_0|^k<\infty\ .$$

This completes the proof of Lemma 2.4.6. END

Let $\xi_{\gamma,\rho,t}=1+\sum_{i=0}^{\infty}\rho^i|\varepsilon_{t-i}(\gamma)|$. By definition $\xi_{\gamma_0,\rho,t}=\xi_{0,\rho,t}$.

Lemma 2.4.7. *Under Assumptions 1 to 3 and 6, then there exists a neighbor* Θ_0 *of* λ_0*,* $0<\rho<1$ *and* $0\le\iota_2<1$*, such that:*

(i) *For any* $\gamma\in\Theta_\gamma$*,* $|\partial\varepsilon_t(\gamma)/\partial\gamma|$*, and* $|\partial^2\varepsilon_t(\gamma)/\partial\gamma\partial\gamma^T|$ *are bounded by* $C\xi_{\gamma,\rho,t-1}$ *;*

(ii) *For any* $\lambda\in\Theta_0$*,* $|(\partial\sigma_t^2(\lambda)/\partial\gamma)/\sigma_t(\lambda)|$ *is bounded by* $C\xi_{\gamma,\rho,t-1}$*;*

(iii) *For any* $\lambda\in\Theta_0$*,* $|\partial^2\sigma_t^2(\lambda)/\partial\gamma\partial\gamma^T|$*, and* $|\partial^2\sigma_t^2(\lambda)/\partial\gamma\partial\delta^T|$ *are bounded by* $C\xi_{\gamma,\rho,t-1}^2$*;*

(iv) *For any* $\lambda^* \neq \lambda$, λ^* *and* $\lambda \in \Theta$, *there exist* $0 < \rho < \varrho < 1$, *such that* $\sigma_t^2(\lambda^*)$ *is bounded by* $C\xi^2_{\gamma^*,\rho,t-1}$; *and* $\xi_{\gamma^*,\rho,t-1}$ *is bounded by* $C\xi_{\gamma,\varrho,t-1}$.

(v) *For any* $\lambda \in \Theta_0$, *there exist* $0 < \rho < \varrho < 1$, *such that* $\xi_{\gamma,\rho,t-1}/\sigma_t(\lambda)$ *is bounded by* $C\xi^{1-\iota_2}_{\gamma,\varrho,t-1}$; $|\eta_t(\lambda)|$ *is bounded by* $\leq C(1+|\eta_t|)\xi^{1-\iota_2}_{\gamma,\varrho,t-1}$;

(vi) *There exist* $0 < \rho < \varrho < 1$, *such that* $\sup_{\gamma\in\Theta_\gamma} \xi_{\gamma,\rho,t}$ *is bounded by* $C\xi_{\varrho,t}$.

Proof: Note that we have defined K and $0 < \rho < 1$ be generic constants taking values different from place to place. Generally we take K and ρ bigger and bigger from inequality to next inequality. In the following proof, we will frequently use the fact

$$\begin{aligned}
\sum_{i=1}^{\infty}\rho_1^i\sum_{j=0}^{\infty}\rho_2^j|x_{t-i-j}| &= \sum_{i=1}^{\infty}\sum_{j=0}^{i-1}\rho_1^{i-j}\rho_2^j|x_{t-i}| \\
&= K\sum_{i=1}^{\infty}\rho_1^i|x_{t-i}|\frac{1-(\rho_2/\rho_1)^i}{1-(\rho_2/\rho_1)} \\
&= \frac{K\rho_1}{\rho_1-\rho_2}\sum_{i=1}^{\infty}(\rho_1^i-\rho_2^i)|x_{t-i}| \\
&\leq K\sum_{i=1}^{\infty}\max(\rho_1^i,\rho_2^i)|x_{t-i}|\ , \qquad (2.4.9)
\end{aligned}$$

where $0 < \rho_1, \rho_2 < 1$, and K is some positive constant.

From (A.3.1)-(A.3.3) in Appendix A.3, by Lemma 2.4.1, we have for some $0 < \rho < 1$,

$$\left|\frac{\partial\varepsilon_t(\gamma)}{\partial\gamma}\right| = \left|\big(-\mathcal{C}_\gamma(1), -\mathcal{A}_\gamma^{-1}(L)\varepsilon_{t-i}(\gamma), -\mathcal{B}_\gamma^{-1}(L)\varepsilon_{t-j}(\gamma)\big)\right| \leq K\big(1+\sum_{i=1}^{\infty}\rho^i|\varepsilon_{t-i}(\gamma)|\big)\ .$$

From (A.3.9)-(A.3.14) in Appendix A.3, by Lemma 2.4.1 and (2.4.9), we have for some $0 < \rho < 1$,

$$
\begin{aligned}
&\left|\frac{\partial^2 \varepsilon_t(\gamma)}{\partial\gamma\partial\gamma^T}\right| \\
&= \left|\left(0, \mathcal{B}_\gamma^{-1}(1), \mathcal{A}_\gamma(1)\mathcal{B}_\gamma^{-2}(1), 0, \mathcal{A}_\gamma^{-1}(L)\mathcal{B}_\gamma^{-1}(L)\varepsilon_{t-i-j}(\gamma), \mathcal{B}_\gamma^{-2}(L)\big(\varepsilon_{t-i-j}(\gamma) + \varepsilon_{t-j}(\gamma)\big)\right)\right| \\
&\le K\big(1 + \sum_{i=1}^{\infty} \rho^i |\varepsilon_{t-i}(\gamma)|\big) .
\end{aligned}
$$

Thus part (i) holds.

To prove part (ii), we give another derivative form of $\sigma_t^2(\lambda)$. By (2.2.4),

$$
\sigma_t^2(\lambda) = \mathcal{B}_\delta{}^{-1}(1)\alpha_0 + \sum_{i=1}^{\infty} a_\delta(i)\varepsilon_{t-i}^2(\gamma).
$$

Hence

$$
\frac{\partial\sigma_t^2(\lambda)}{\partial\gamma} = 2\sum_{i=1}^{\infty} a_\delta(i)\varepsilon_{t-i}(\gamma)\frac{\partial\varepsilon_{t-i}(\gamma)}{\partial\gamma} .
$$

There exists some small positive C_1 such that for any $i \ge 1$ we have:

$$
\sigma_t(\lambda) > \left(\mathcal{B}_\delta{}^{-1}(1)\alpha_0 + a_\delta(i)\varepsilon_{t-i}^2(\gamma)\right)^{1/2} > C_1\left(1 + a_\delta(i)\varepsilon_{t-i}^2(\gamma)\right)^{1/2} \ge \frac{C_1}{\sqrt{2}}(1+\sqrt{a_\delta(i)}|\varepsilon_{t-i}(\gamma)|) .
$$

Thus with fact $|x|/(1+|x|) \le 1$, by Lemma 2.4.1, part (i) above and (2.4.9), there exist some $0 < \rho_3, \rho_4 < 1$ and positive constant K_1, such that

$$
\begin{aligned}
\left|\frac{\partial\sigma_t^2(\lambda)/\partial\gamma}{\sigma_t(\lambda)}\right| &\le \frac{K_1}{C_1}\sum_{i=1}^{\infty}\frac{a_\delta(i)|\varepsilon_{t-i}(\gamma)|}{1+\sqrt{a_\delta(i)}|\varepsilon_{t-i}(\gamma)|)}\left|\frac{\partial\varepsilon_{t-i}(\gamma)}{\partial\gamma}\right| \\
&\le K\sum_{i=1}^{\infty}\sqrt{a_\delta(i)}\left(1+\sum_{j=0}^{\infty}\rho_3^j|\varepsilon_{t-i-1-j}(\gamma)|\right) \\
&\le K + K\sum_{i=1}^{\infty}\rho_4^i\sum_{j=0}^{\infty}\rho_3^j|\varepsilon_{t-i-1-j}(\gamma)| \\
&\le K + K\sum_{i=1}^{\infty}\max(\rho_3^i, \rho_4^i)|\varepsilon_{t-i-1}(\gamma)| .
\end{aligned}
$$

This proves part (ii).

In the following proof, we will frequently use the inequality $2|xy| \le x^2 + y^2$.

From (A.3.20) and (A.3.22) in Appendix A.3, by Lemma 2.4.1, part (i) and (2.4.9), there exist $0 < \rho_5, \rho_6 < 1$, such that

$$\begin{aligned}\left|\frac{\partial^2\sigma_t^2(\lambda)}{\partial\gamma\partial\delta^T}\right| &\le K\left|\sum_{i=1}^{\infty}\rho_5^i\varepsilon_{t-i}(\gamma)\frac{\partial\varepsilon_{t-i}(\gamma)}{\partial\gamma}\right| \qquad (2.4.10)\\ &\le K\left|\sum_{i=1}^{\infty}\rho_5^i\varepsilon_{t-i}(\gamma)\left(1+\sum_{j=0}^{\infty}\rho_6^j|\varepsilon_{t-i-1-j}(\gamma)|\right)\right|\\ &\le K\sum_{i=1}^{\infty}\rho_5^i|\varepsilon_{t-i}(\gamma)| + K\sum_{i=1}^{\infty}\rho_5^i\sum_{j=0}^{\infty}\rho_6^j\left(\varepsilon_{t-i}^2(\gamma)+\varepsilon_{t-i-1-j}^2(\gamma)\right)\\ &\le K\sum_{i=1}^{\infty}\rho_5^i|\varepsilon_{t-i}(\gamma)| + K\sum_{i=1}^{\infty}\max(\rho_5^i,\rho_6^i)\varepsilon_{t-i}^2(\gamma)\ .\end{aligned}$$

Since for some $0 < \rho_7 < 1$,

$$\begin{aligned}&1+2\sum_{i=0}^{\infty}\rho_7^i|\varepsilon_{t-i}(\gamma)| + \sum_{i=0}^{\infty}\rho_7^{2i}\varepsilon_{t-i}^2(\gamma) \qquad (2.4.11)\\ &\le 1+2\sum_{i=0}^{\infty}\rho_7^i|\varepsilon_{t-i}(\gamma)| + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\rho_7^i\rho_7^j|\varepsilon_{t-i}(\gamma)||\varepsilon_{t-j}(\gamma)|\\ &= \xi_{\gamma,\rho_7,t}^2\ , \qquad (2.4.12)\end{aligned}$$

thus there exists $0 < \rho < 1$, such that $\left|\partial^2\sigma_t^2(\lambda)/\partial\gamma\partial\delta^T\right|$ is bounded by $C\xi_{\gamma,\rho,t}^2$.

By (A.3.23) in Appendix A.3 and Lemma 2.4.1, we have for any $\lambda \in \Theta$, $|\partial^2\sigma_t^2(\lambda)/\partial\gamma\partial\gamma^T|$ is bounded by

$$K\left(\sum_{i=1}^{\infty}\rho_8^i\frac{\partial\varepsilon_{t-i}(\gamma)}{\partial\gamma}\frac{\partial\varepsilon_{t-i}(\gamma)}{\partial\gamma^T}+\sum_{i=1}^{\infty}\rho_8^i\varepsilon_{t-i}(\gamma)\frac{\partial^2\varepsilon_{t-i}(\gamma)}{\partial\gamma\partial\gamma^T}\right)\ .$$

From (A.3.1)-(A.3.3) in A.3 in Appendix A.3, by (2.4.9), we have

$$\left|\sum_{i=1}^{\infty}\rho_8^i\frac{\partial\varepsilon_{t-i}(\gamma)}{\partial\gamma}\frac{\partial\varepsilon_{t-i}(\gamma)}{\partial\gamma^T}\right| \le K\sum_{i=1}^{\infty}\rho_8^i\left(1+\sum_{j=1}^{\infty}\rho^j|\varepsilon_{t-i-j}(\gamma)|\right)^2 \le C\xi_{\gamma,\rho,t}^2\ .$$

By replacing $\partial \varepsilon_{t-i}(\gamma)/\partial\gamma$ with $\partial^2 \varepsilon_{t-i}(\gamma)/\partial\gamma\partial\gamma^T$ in (2.4.10), together with (2.4.11), for some $0 < \rho_9 < \rho < 1$ and constant C, we have:

$$\left|\sum_{i=1}^{\infty} \rho_8^i \varepsilon_{t-i}(\gamma) \frac{\partial^2 \varepsilon_{t-i}(\gamma)}{\partial\gamma\partial\gamma^T}\right| \le K\sum_{i=1}^{\infty}\rho_8^i|\varepsilon_{t-i}(\gamma)| + K\sum_{i=1}^{\infty}\max(\rho_8^i,\rho_9^i)\varepsilon_{t-i}^2(\gamma) \le C\xi_{\gamma,\rho,t}^2 \,.$$

This proves part (iii).

By Lemma 2.4.1, (2.4.9), for any $\lambda^* \neq \lambda$ and $0 < \rho_{11} \le \rho_{12} \le \rho_{13} < 1$, we have:

$$\begin{aligned}
\sigma_t^2(\lambda^*) &= B_{\delta^*}^{-1}\alpha_0 + \sum_{i=1}^{\infty} a_{\delta^*}(i)\varepsilon_{t-i}^2(\gamma^*) = B_{\delta^*}^{-1}\alpha_0 + \sum_{i=1}^{\infty} a_{\delta^*}(i)\left(\sum_{j=0}^{\infty} a_{\gamma^*}(j)(Y_{t-i-j}-c^*)\right)^2 \\
&\le K + K\sum_{i=1}^{\infty} a_{\delta^*}(i)\sum_{j=0}^{\infty} a_{\gamma^*}(j)\sum_{l=0}^{\infty} a_{\gamma^*}(l)\left((Y_{t-i-j}-c^*)^2 + (Y_{t-i-l}-c^*)^2\right) \\
&\le K + K\sum_{i=1}^{\infty} a_{\delta^*}(i)\sum_{j=0}^{\infty} \rho_{11}^j (Y_{t-i-j}-c^*)^2 \\
&\le K + K\sum_{i=1}^{\infty} a_{\delta^*}(i)\sum_{j=0}^{\infty} \rho_{11}^j (c-c^*)^2 + K\sum_{i=1}^{\infty} a_{\delta^*}(i)\sum_{j=0}^{\infty} \rho_{11}^j (Y_{t-i-j}-c)^2 \\
&\le K + K\sum_{i=1}^{\infty} a_{\delta^*}(i)\sum_{j=0}^{\infty} \rho_{11}^j \left(\mathcal{A}^{-1}(L)\mathcal{B}(L)\varepsilon_{t-i-j}(\gamma)\right)^2 \\
&\le K + K\sum_{i=1}^{\infty} a_{\delta^*}(i)\sum_{j=0}^{\infty} \rho_{11}^j \sum_{l=0}^{\infty}\sum_{k=0}^{\infty} \rho_{12}^l\rho_{12}^k |\varepsilon_{t-i-j-l}(\gamma)\varepsilon_{t-i-j-k}(\gamma)| \\
&\le K + K\sum_{i=1}^{\infty} a_{\delta^*}(i)\sum_{j=0}^{\infty} \rho_{11}^j \sum_{l=0}^{\infty}\sum_{k=0}^{\infty} \rho_{12}^l\rho_{12}^k \left(\varepsilon_{t-i-j-l}^2(\gamma) + \varepsilon_{t-i-j-k}^2(\gamma)\right) \\
&\le K + K\sum_{i=1}^{\infty} a_{\delta^*}(i)\sum_{j=0}^{\infty} \rho_{11}^j \sum_{l=0}^{\infty} \rho_{12}^l \varepsilon_{t-i-j-l}^2(\gamma) \\
&\le K + K\sum_{i=1}^{\infty} \rho_{13}^i \varepsilon_{t-i}^2(\gamma) \,.
\end{aligned}$$

Thus there exist $0 < \rho_{13} < \rho < 1$,

$$\sigma_t^2(\lambda^*) \le C\xi_{\gamma,\rho,t-1}^2 \,.$$

By part (i), Lemma 2.4.1 and (2.4.9), for any $\gamma^* \neq \gamma$, there exist $0 < \rho_{14} \leq \rho_{15} \leq \rho_{16} \leq \rho_{17} < 1$, such that

$$\begin{aligned}
\xi_{\gamma^*,\rho_{14},t-1} &= 1 + \sum_{i=1}^{\infty} \rho_{14}^i |\varepsilon_{t-i}(\gamma^*)| \\
&= 1 + \sum_{i=1}^{\infty} \rho_{14}^i \left|\mathcal{A}(L)\mathcal{B}^{-1}(L)(Y_{t-i} - c^*)\right| \\
&< 1 + \sum_{i=1}^{\infty} \rho_{14}^i \sum_{j=0}^{\infty} \rho_{15}^j \left|(Y_{t-i-j} - c^*)\right| \\
&\leq 1 + \sum_{i=1}^{\infty} \rho_{14}^i \sum_{j=0}^{\infty} \rho_{15}^j |c - c^*| + \sum_{i=1}^{\infty} \rho_{14}^i \sum_{j=0}^{\infty} \rho_{15}^j |Y_{t-i-j} - c| \\
&\leq K + K \sum_{i=1}^{\infty} \rho_{14}^i \sum_{j=0}^{\infty} \rho_{15}^j \left|\mathcal{A}^{-1}(L)\mathcal{B}(L)\varepsilon_{t-i-j}(\gamma)\right| \\
&\leq K + K \sum_{i=1}^{\infty} \rho_{14}^i \sum_{j=0}^{\infty} \rho_{15}^j \sum_{l=0}^{\infty} \rho_{16}^l \left|\varepsilon_{t-i-j-l}(\gamma)\right| \\
&\leq K + K \sum_{i=1}^{\infty} \rho_{17}^i |\varepsilon_{t-i}(\gamma)| \,. \qquad (2.4.13)
\end{aligned}$$

Thus there exist $0 < \rho \leq \varrho < 1$, such that $\xi_{\gamma^*,\rho,t-1} \leq C\xi_{\gamma,\varrho,t-1}$.

This completes proof of part (iv).

By (2.4.2), we have

$$\sigma_t^2(\lambda) = \sum_{i=0}^{\infty} \underline{1}^T B^i \underline{1} \left(\alpha_0 + \sum_{j=1}^{p} \alpha_j \varepsilon_{t-i-j}^2(\gamma)\right) = K + \sum_{j=1}^{p} \sum_{i=0}^{\infty} \underline{1}^T B^i \underline{1} \alpha_j \varepsilon_{t-i-j}^2(\gamma),$$

where $\underline{1}^T = (1, 0, \cdots, 0)_{q\times 1}$. By the definition of B, $\underline{1}^T B^i \underline{1} \geq \beta_1^i$. Since λ_0 is in the interior of Θ, for any $1 \leq j \leq p$, $\inf_{\lambda\in\Theta_0} \alpha_j > 0$. Together with the fact $(1 + x^2)^{1/2} \geq (1 + x)/\sqrt{2}$ for $x \geq 0$, we have $\sigma_t(\lambda) \geq C_3(1 + \beta_1^{i/2}|\varepsilon_{t-i}(\gamma)|)$ for some small positive C_3 and any $i \geq 1$. Hence by the facts $|x|/(1 + |x|) \leq 1$ and $1/(1 + |x|) \leq 1/(1 + |x|)^{\iota_2}$

for $0 \le \iota_2 < 1$, we have:

$$\begin{aligned}
\left| \frac{\left(1+\sum_{i=1}^{\infty} \rho^i |\varepsilon_{t-i}(\gamma)|\right)}{\sigma_t(\lambda)} \right| &\le \frac{1}{C_3}\left(1+\sum_{i=1}^{\infty} \frac{\rho^i |\varepsilon_{t-i}(\gamma)|}{1+\beta_1^{i/2}|\varepsilon_{t-i}(\gamma)|}\right) \\
&\le \frac{1}{C_3} + \frac{1}{C_3}\sum_{i=1}^{\infty}\left(\frac{\rho}{\beta_1^{\iota_2/2}}\right)^i |\varepsilon_{t-i}(\gamma)|^{1-\iota_2}\left(\frac{\beta_1^{i/2}|\varepsilon_{t-i}(\gamma)|}{1+\beta_1^{i/2}|\varepsilon_{t-i}(\gamma)|}\right)^{\iota_2} \\
&\le \frac{1}{C_3} + \frac{1}{C_3}\sum_{i=1}^{\infty}\left(\frac{\rho}{\beta_1^{\iota_2/2}}\right)^i |\varepsilon_{t-i}(\gamma)|^{1-\iota_2} ,
\end{aligned}$$

and

$$\begin{aligned}
\sum_{i=1}^{\infty} \varrho^i |\varepsilon_{t-i}(\gamma)|^{1-\iota_2} &= \left(1+\sum_{j=1}^{\infty}\varrho^j|\varepsilon_{t-j}(\gamma)|\right)^{1-\iota_2} \sum_{i=1}^{\infty} \varrho^i \frac{|\varepsilon_{t-i}(\gamma)|^{1-\iota_2}}{\left(1+\sum_{j=1}^{\infty}\varrho^j|\varepsilon_{t-j}(\gamma)|\right)^{1-\iota_2}} \\
&\le \left(1+\sum_{j=1}^{\infty}\varrho^j|\varepsilon_{t-j}(\gamma)|\right)^{1-\iota_2} \sum_{i=1}^{\infty} \varrho^i \frac{|\varepsilon_{t-i}(\gamma)|^{1-\iota_2}}{\left(\varrho^i|\varepsilon_{t-i}(\gamma)|\right)^{1-\iota_2}} \\
&\le \left(1+\sum_{j=1}^{\infty}\varrho^j|\varepsilon_{t-j}(\gamma)|\right)^{1-\iota_2} \sum_{i=1}^{\infty} \varrho^{i\iota_2} .
\end{aligned}$$

Then by taking ι_2 small enough, such that $0 < \varrho = \rho/\beta_1^{\iota_2/2} < 1$, we have for some constant C:

$$\left|\frac{\xi_{\gamma,\rho,t-1}}{\sigma_t(\lambda)}\right| \le C\xi_{\gamma,\varrho,t-1}^{1-\iota_2} .$$

Now by part (i) and (iv), there exists $0 < \rho < 1$, such that σ_t and $\partial\varepsilon_t(\gamma^*)/\partial\gamma^*$ are bounded by $C\xi_{\gamma,\rho,t-1}$. Then for any $\lambda \in \Theta$ and $0 < \rho < \rho^* < \varrho < 1$, we have

$$\begin{aligned}
|\eta_t(\lambda)| &= \left|\frac{\varepsilon_t(\gamma)}{\sigma_t(\lambda)}\right| = \left|\frac{\varepsilon_t(\gamma_0) + (\gamma - \gamma_0)\partial\varepsilon_t(\gamma^*)/\partial\gamma}{\sigma_t(\lambda)}\right| \\
&\leq |\eta_t|\frac{\sigma_t}{\sigma_t(\lambda)} + \left|(\gamma - \gamma_0)\frac{\partial\varepsilon_t(\gamma^*)/\partial\gamma}{\sigma_t(\lambda)}\right| \\
&\leq K(1 + |\eta_t|)\frac{\xi_{\gamma^*,\rho,t-1}}{\sigma_t(\lambda)} \\
&\leq K(1 + |\eta_t|)\frac{\xi_{\gamma,\rho^*,t-1}}{\sigma_t(\lambda)} \\
&\leq K(1 + |\eta_t|)\xi^{1-\iota_2}_{\gamma,\varrho,t-1}\ .
\end{aligned}$$

This proves part (v).

By Lemma 2.4.1 and (2.4.9), there exist $0 < \rho_{18} \leq \rho_{19} \leq \rho_{20} < 1$, such that

$$\begin{aligned}
\sup_{\gamma\in\Theta_\gamma}\sum_{i=0}^{\infty}\rho_{18}^i|\varepsilon_{t-i}(\gamma)| &\leq K\sup_{\gamma\in\Theta_\gamma}\sum_{i=0}^{\infty}\rho_{18}^i\left|\sum_{j=0}^{\infty}\rho_{19}^j(Y_{t-i-j} - c)\right| \\
&\leq K\sup_{\gamma\in\Theta_\gamma}\sum_{i=0}^{\infty}\rho_{18}^i\sum_{j=0}^{\infty}\rho_{19}^j|c_{00} - c| + K\sum_{i=0}^{\infty}\rho_{18}^i\sum_{j=0}^{\infty}\rho_{19}^j\,|Y_{t-i-j} - c_{00}| \\
&\leq K\left(1 + \sum_{i=0}^{\infty}\rho_{20}^i\,|Y_{t-i} - c_{00}|\right)
\end{aligned}$$

Thus there is $0 < \rho \leq \varrho < 1$, such that $\xi_{\gamma,\rho,t} < C\xi_{\varrho,t}$, which proves part (vi).

This completes the proof of Lemma 2.4.7. END

Lemma 2.4.8. *Under Assumptions 1 to 5, if* $\mathbf{E}|\varepsilon_t|^{2(1-\iota_2)} < \infty$, *and* $\mathbf{E}|\eta_0|^{4\iota_1} < \infty$, *then*

$$\mathbf{E}\left|\frac{\partial l_t(\lambda_0)}{\partial\lambda}\frac{\partial l_t(\lambda_0)}{\partial\lambda^T}\right| < \infty \textit{ and } \mathbf{E}\left|\frac{\partial^2 l_t(\lambda_0)}{\partial\lambda\partial\lambda^T}\right| < \infty.$$

Proof: By (2.2.12), Assumption 4, Lemmas 2.4.6 and 2.4.7, and the independence of η_t with $\xi_{\varrho,t-1}$, we have

$$
\begin{aligned}
&\mathbf{E}\left|\frac{\partial l_t(\lambda_0)}{\partial\lambda}\frac{\partial l_t(\lambda_0)}{\partial\lambda^T}\right| \\
&\leq \mathbf{E}\Bigg\{\frac{1}{4}\Big(1+g(\eta_t)\eta_t\Big)^2\left|\frac{\partial\sigma_t^2(\lambda_0)/\partial\lambda}{\sigma_t^2}\frac{\partial\sigma_t^2(\lambda_0)/\partial\lambda^T}{\sigma_t^2}\right| \\
&\quad +\frac{1}{2}\Big|g(\eta_t)+g^2(\eta_t)\eta_t\Big|\left|\frac{\partial\sigma_t^2(\lambda_0)/\partial\lambda}{\sigma_t^2}\frac{\partial\varepsilon_t(\gamma_0)/\partial\lambda^T}{\sigma_t}\right|+g^2(\eta_t)\left|\frac{\partial\varepsilon_t(\gamma_0)/\partial\lambda}{\sigma_t}\frac{\partial\varepsilon_t(\gamma_0)/\partial\lambda^T}{\sigma_t}\right|\Bigg\} \\
&\leq K\mathbf{E}\left(1+|\eta_t|^{4\iota_1}+|\eta_t|^{2\iota_1}+\max(|\eta_t|,1)^{4\iota_1-1}+\max(|\eta_t|,1)^{4\iota_1-2}\right)\mathbf{E}\xi_{0,\rho,t-1}^{2(1-\iota_2)} \\
&< \infty\,.
\end{aligned}
$$

By (2.2.13), Assumption 4, Lemmas 2.4.6 and 2.4.7, and the independence of η_t with $\xi_{\varrho,t-1}$, we have

$$
\begin{aligned}
\mathbf{E}\left|\frac{\partial^2 l_t(\lambda_0)}{\partial\lambda\partial\lambda^T}\right| &\leq K\mathbf{E}\left(1+|\eta_t|^{2\iota_1}+\max(|\eta_t|,1)^{2\iota_1-1}+\max(|\eta_t|,1)^{2\iota_1-2}\right)\mathbf{E}\xi_{0,\rho,t-1}^{2(1-\iota_2)} \\
&< \infty\,.
\end{aligned}
$$

This completes the proof of Lemma 2.4.8. END

Lemma 2.4.9. *Under Assumptions 1-7, if* $\mathbf{E}|\varepsilon_t|^{2(1-\iota_2)}<\infty$, *and* $\mathbf{E}|\eta_t|^{4\iota_1}<\infty$, *then* $\mathcal{I}$ *is nonsingular. Further if* $\mathbf{E}g'(\eta_0)\leq 0$, $\mathbf{E}g'(\eta_0)\eta_0^2\leq 1$ *and* $\mathbf{E}g'(\eta_0)\eta_0=0$, *then* $\mathcal{J}$ *is nonsingular.*

Proof: Referring to the proof of Theorem 3.2 in Francq and Zakoïan (2004), for any vector $r_\delta\subset\mathbb{R}^{p+q+1}$ and $r_\gamma\subset\mathbb{R}^{P+Q+1}$, we have:

$$r_\delta^T\frac{\partial\sigma_t^2(\lambda_0)}{\partial\delta}=0\quad a.s$$

if and only if $r_\delta=0$; and

$$r_\gamma^T\frac{\partial\varepsilon_t(\lambda_0)}{\partial\gamma}=0,\quad a.s$$

if and only if $r_\gamma = 0$. This implies

$$\mathbf{E}\left(\frac{\partial\sigma_t^2(\lambda_0)/\partial\lambda}{\sigma_t^2}\frac{\partial\sigma_t^2(\lambda_0)/\partial\lambda^T}{\sigma_t^2}\right) \text{ and } \mathbf{E}\left(\frac{\partial\varepsilon_t(\gamma_0)/\partial\lambda}{\sigma_t}\frac{\partial\varepsilon_t(\gamma_0)/\partial\lambda^T}{\sigma_t}\right)$$

are nonsingular.

The existence of $\mathcal{I}$ and $\mathcal{J}$ are established in the Lemma 2.4.8. We prove the nonsingularity of $\mathcal{I}$ by a contradiction method similar to that in the proof of Theorem 3.2 in Francq and Zakoïan (2004). Assume $\mathcal{I}$ is singular, then there exists a vector $r = (r_\gamma, r_\delta) \neq 0$, such that $r^T\partial l_t(\lambda_0)/\partial\lambda = 0$ a.s. From (2.2.12), we have

$$-\frac{1}{2}\Big(1+g(\eta_t)\eta_t\Big)r^T\frac{\partial\sigma_t^2(\lambda_0)/\partial\lambda}{\sigma_t^2} + g(\eta_t)r^T\frac{\partial\varepsilon_t(\gamma_0)/\partial\lambda}{\sigma_t} = 0, \quad a.s. \tag{2.4.14}$$

Taking the variance of the left-hand side of (2.4.14) conditional on $\mathcal{F}_{t-1}$, and using the fact that η_t is independent with $(\partial\sigma_t^2/\partial\lambda)/\sigma_t^2$ and $(\partial\varepsilon_t/\partial\lambda)/\sigma_t$, and $\mathbf{E}g(\eta_t) = 0$ (see Remark 2.2.3), we have almost surely:

$$\begin{aligned}
0 &= \frac{1}{4}\mathbf{E}\left(g(\eta_t)\eta_t+1\right)^2\left(r^T\frac{\partial\sigma_t^2(\lambda_0)/\partial\lambda}{\sigma_t^2}\right)^2 \\
&\quad -\mathbf{E}g^2(\eta_t)\eta_t\left(r^T\frac{\partial\sigma_t^2(\lambda_0)/\partial\lambda}{\sigma_t^2}\right)\left(r^T\frac{\partial\varepsilon_t(\gamma_0)/\partial\lambda}{\sigma_t}\right) + \mathbf{E}g^2(\eta_t)\left(r^T\frac{\partial\varepsilon_t(\gamma)/\partial\lambda}{\sigma_t}\right)^2 \\
&= K_1a_t^2 - K_2a_tb_t + K_3b_t^2 \\
&= \frac{4K_1K_3-K_2^2}{4K_3}a_t^2 + K_3\left(\frac{K_2}{2K_3}a_t - b_t\right)^2 .
\end{aligned}$$

By Assumption 7, $(4K_1K_3 - K_2^2)/4K_3$ is negative. By stationarity, we have either $b_t = \frac{K_2+(K_2^2-4K_1K_3)^{1/2}}{2K_3}a_t$ or $b_t = \frac{K_2-(K_2^2-4K_1K_3)^{1/2}}{2K_3}a_t$ almost surely for all t. Without loss of generality, take the later case and substitute it into (2.4.14), we obtain for all t,

$$\left\{-\frac{1}{2}\Big(1+g(\eta_t)\eta_t\Big) + \frac{K_2-(K_2^2-4K_1K_3)^{1/2}}{2K_3}g(\eta_t)\right\}r^T\frac{\partial\sigma_t^2(\lambda_0)/\partial\lambda}{\sigma_t^2} = 0, \quad a.s.$$

The term in the parenthesis can not be 0 almost surely if $P(g(\eta_t)\eta_t+1 = Cg(\eta_t)) < 1$ for any constant C. While by Remark 2.2.6, Assumption 7 implies $P[g(\eta_t)\eta_t + 1 = Cg(\eta_t)] < 1$ for any constant C. Hence $a_t = 0$ $a.s.$ and thus $b_t = 0$ almost surely, that is

$$r^T \frac{\partial \sigma_t^2(\lambda_0)}{\partial \lambda} = 0, \text{ and } r^T \frac{\partial \varepsilon_t(\lambda_0)}{\partial \lambda} = 0, \quad a.s.$$

Since $\partial \varepsilon_t(\lambda_0)/\partial \delta = 0$, $r^T \partial \varepsilon_t(\lambda_0)/\partial \lambda = 0$ $a.s$ if and only if $r_\gamma = 0$. With $r_\gamma = 0$, $r^T \partial \sigma_t^2(\lambda_0)/\partial \lambda = 0$ $a.s$ means $r_\delta^T \partial \sigma_t^2(\lambda_0)/\partial \delta = 0$ $a.s$, which holds if and only if $r_\delta = 0$. Thus we have $r = 0$, which contradicts with the assumption. This proves the non-singularity of $\mathcal{I}$.

Since $\mathbf{E}(g(\eta_t)\eta_t + 1) = 0$ and $\mathbf{E}g(\eta_t) = 0$ (see Remark 2.2.3), further with $\mathbf{E}g'(\eta_t)\eta_t = 0$, by (2.2.13) and the independence of η_t with $(\partial \sigma_t^2/\partial \lambda)/\sigma_t^2$ and $(\partial \varepsilon_t/\partial \lambda)/\sigma_t$, we have

$$\mathcal{J} = \frac{1}{4}(1-\mathbf{E}g'(\eta_t)\eta_t^2)\mathbf{E}\left(\frac{\partial \sigma_t^2(\lambda_0)/\partial \lambda}{\sigma_t^2}\frac{\partial \sigma_t^2(\lambda_0)/\partial \lambda^T}{\sigma_t^2}\right)-\mathbf{E}g'(\eta_t)\mathbf{E}\left(\frac{\partial \varepsilon_t(\gamma_0)/\partial \lambda}{\sigma_t}\frac{\partial \varepsilon_t(\gamma_0)/\partial \lambda^T}{\sigma_t}\right). \tag{2.4.15}$$

By the given conditions, we have $\mathbf{E}g'(\eta_t)\eta_t^2 < 1$, $\mathbf{E}g'(\eta_t) \leq 0$. Hence $\mathcal{J}$ is a sum of a positive definite matrix and a positive semi-definite matrix. Thus $\mathcal{J}$ is nonsingular.

Lemma 2.4.10. *Under Assumptions 1-7, if* η_0 *is symmetrically distributed,* $\mathbf{E}|\varepsilon_t|^{2(1-\iota_2)} < \infty$ *and* $\mathbf{E}|\eta_t|^{\max(1,4\iota_1)} < \infty$*, then we have:*

(i) *if* $\mathbf{E}g^2(\eta_0)\eta_0 = 0$*, then* $\mathcal{I}$ *is block-diagonal;*

(ii) *if* $\mathbf{E}g'(\eta_0)\eta_0 = 0$*, then* $\mathcal{J}$ *is block-diagonal.*

Proof:

$$\mathcal{I} = \begin{pmatrix} \mathbf{E}\left(\frac{\partial l_0(\lambda_0)}{\partial\gamma}\frac{\partial l_0(\lambda_0)}{\partial\gamma^T}\right) & \mathbf{E}\left(\frac{\partial l_0(\lambda_0)}{\partial\delta}\frac{\partial l_0(\lambda_0)}{\partial\gamma^T}\right) \\ \mathbf{E}\left(\frac{\partial l_0(\lambda_0)}{\partial\gamma}\frac{\partial l_0(\lambda_0)}{\partial\delta^T}\right) & \mathbf{E}\left(\frac{\partial l_0(\lambda_0)}{\partial\delta}\frac{\partial l_0(\lambda_0)}{\partial\delta^T}\right) \end{pmatrix} = \begin{pmatrix} \mathcal{I}_1 & \mathcal{I}_4 \\ \mathcal{I}_3 & \mathcal{I}_2 \end{pmatrix},$$

where $\mathcal{I}_1$ and $\mathcal{I}_2$ have the expressions as those in (2.2.14) and (2.2.15).

Since $\partial\varepsilon_t(\gamma)/\partial\delta = 0$, $\mathbf{E}g(\eta_t) = \mathbf{E}g^2(\eta_t)\eta_t = 0$, by (2.2.12), we have:

$$\begin{aligned} \mathcal{I}_4 &= \frac{1}{4}\mathbf{E}(1+g(\eta_t)\eta_t)^2\mathbf{E}\frac{\partial\sigma_t^2(\lambda_0)/\partial\delta}{\sigma_t^2}\frac{\partial\sigma_t^2(\lambda_0)/\partial\gamma^T}{\sigma_t^2} \\ &\quad -\frac{1}{2}\mathbf{E}(g(\eta_t)+g^2(\eta_t)\eta_t)\mathbf{E}\frac{\partial\sigma_t^2(\lambda_0)/\partial\delta}{\sigma_t^2}\frac{\partial\varepsilon_t(\gamma_0)/\partial\gamma^T}{\sigma_t} \\ &= \frac{1}{4}(\mathbf{E}(g(\eta_t)\eta_t)^2-1)\mathbf{E}\frac{\partial\sigma_t^2(\lambda_0)/\partial\delta}{\sigma_t^2}\frac{\partial\sigma_t^2(\lambda_0)/\partial\gamma^T}{\sigma_t^2}\ . \end{aligned}$$

Refer to (ii) in the proof of theorem 3.2 in Francq and Zakoïan (2004), for $1 \le i \le p+q+1$, $1 \le l \le q$ and $1 \le j \le P+Q+1$, we have

$$\mathbf{E}\frac{\partial\sigma_t^2(\lambda_0)/\partial\delta}{\sigma_t^2}\frac{\partial\sigma_t^2(\lambda_0)/\partial\gamma^T}{\sigma_t^2} = 0\ . \tag{2.4.16}$$

Thus $\mathcal{I}_4 = 0$. $\mathcal{I}_3$ is the transpose of $\mathcal{I}_4$, so $\mathcal{I}_3^T = \mathcal{I}_4 = 0$.

Since $\partial\varepsilon_t(\gamma)/\partial\delta = 0$, $\mathbf{E}g(\eta_t) = 0$, $\mathbf{E}g(\eta_t)\eta_t + 1 = 0$ (see Remark 2.2.3), based on (2.4.15), one non-diagonal of $\mathcal{J}$ is:

$$\frac{1}{4}\mathbf{E}(1-\mathbf{E}g'(\eta_t)\eta_t^2)\mathbf{E}\frac{\partial\sigma_t^2(\lambda_0)/\partial\delta}{\sigma_t^2}\frac{\partial\sigma_t^2(\lambda_0)/\partial\gamma^T}{\sigma_t^2}\ .$$

Thus with (2.4.16), the non-diagonals of $\mathcal{J}$ are null.

This completes the proof of Lemma 2.4.9. $\boxed{\text{END}}$

Lemma 2.4.11. *Under Assumptions 1-5, then*

$$\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{\partial l_t(\lambda_0)}{\partial\lambda}-\frac{\partial\tilde{l}_t(\lambda_0)}{\partial\lambda}\right)\right| \longrightarrow 0,$$

$$\sup_{\lambda\in\Theta_0}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{\partial^2 l_t(\lambda)}{\partial\lambda\partial\lambda^T}-\frac{\partial^2 \tilde{l}_t(\lambda)}{\partial\lambda\partial\lambda^T}\right)\right|\longrightarrow 0,$$

in probability as $n\longrightarrow\infty$, where Θ_0 is a neighborhood of λ_0.

Proof: Analogous to (2.2.12) we obtain:

$$\frac{\partial\tilde{l}_t(\lambda)}{\partial\lambda}=-\frac{1}{2}\Big\{1+g(\tilde{\eta}_t(\lambda))\tilde{\eta}_t(\lambda)\Big\}\frac{\partial\tilde{\sigma}_t^2(\lambda)/\partial\lambda}{\tilde{\sigma}_t^2(\lambda)}+g(\tilde{\eta}_t(\lambda))\frac{\partial\tilde{\varepsilon}_t(\gamma)/\partial\lambda}{\tilde{\sigma}_t(\lambda)}\ .$$

The first and second derivatives of $\tilde{\sigma}_t^2(\lambda)$ and $\tilde{\varepsilon}_t(\gamma)$ are similar with those of $\sigma_t^2(\lambda)$ and $\varepsilon_t(\gamma)$ and are given in Appendix A.6. With Assumption 4, inequalities (A.7.1) to (A.7.5) and (A.7.10) in Appendix A.7, by the mean value theorem, there exists η_t^* such that $|\eta_t^*-\eta_t|\leq|\eta_t-\tilde{\eta}_t|$, and

$$\begin{aligned}
&\left|\frac{\partial l_t(\lambda_0)}{\partial\lambda}-\frac{\partial\tilde{l}_t(\lambda_0)}{\partial\lambda}\right|\\
=&\ \Bigg|-\frac{1}{2}\Big(1+g(\eta_t)\eta_t\Big)\left(\frac{\partial\sigma_t^2(\lambda_0)/\partial\lambda}{\sigma_t^2}-\frac{\partial\tilde{\sigma}_t^2(\lambda_0)/\partial\lambda}{\tilde{\sigma}_t^2}\right)\\
&-\frac{1}{2}\Big(g(\eta_t)\eta_t-g(\tilde{\eta}_t)\tilde{\eta}_t\Big)\frac{\partial\tilde{\sigma}_t^2(\lambda_0)/\partial\lambda}{\tilde{\sigma}_t^2}\\
&+g(\eta_t)\left(\frac{\partial\varepsilon_t(\lambda_0)/\partial\lambda}{\sigma_t}-\frac{\partial\tilde{\varepsilon}_t(\gamma_0)/\partial\lambda}{\tilde{\sigma}_t}\right)+\Big(g(\eta_t)-g(\tilde{\eta}_t)\Big)\frac{\partial\tilde{\varepsilon}_t(\lambda_0)/\partial\lambda}{\tilde{\sigma}_t}\Bigg|\\
\leq&\ K\rho^t\,|1+g(\eta_t)\eta_t|\,S_{t-1}(\gamma_0)\xi_{\rho,t-1}+\frac{1}{2}\left|(g'(\eta_t^*)\eta_t^*+g(\eta_t^*))\,(\eta_t-\tilde{\eta}_t)\frac{\partial\tilde{\sigma}_t^2(\lambda_0)/\partial\lambda}{\tilde{\sigma}_t^2}\right|\\
&+K\rho^t|g(\eta_t)|S_{t-1}(\gamma_0)\xi_{\rho,t-1}+\left|g'(\eta_t^*)(\eta_t-\tilde{\eta}_t)\frac{\partial\tilde{\varepsilon}_t(\lambda_0)}{\tilde{\sigma}_t}\right|\\
\leq&\ K\rho^t\Big(1+\eta_t^{2\iota_1}+\max(\eta_t,1)^{2\iota_1-1}\Big)S_{t-1}(\gamma_0)\xi_{\rho,t-1}\\
&+K\rho^tS_{t-1}(\gamma_0)(1+|\eta_t|)\Big(\max(\eta_t^*,1)^{2\iota_1-1}+\max(\eta_t^*,1)^{2\iota_1-2}\Big)(1+K\rho^tS_{t-1}(\gamma_0))\xi_{\rho,t-1}\\
\leq&\ K\rho^tS_{t-1}^2(\gamma_0)\xi_{\rho,t-1}\Big(1+\eta_t^{2\iota_1}+\max(\eta_t,1)^{2\iota_1-1}+(1+|\eta_t|)\big(\max(\eta_t^*,1)^{2\iota_1-1}\\
&+\max(\eta_t^*,1)^{2\iota_1-2}\big)\Big)\ ,
\end{aligned}$$

where

$$S_{t-1}(\gamma)=\sum_{i=1-p}^{t-1}\left(|\varepsilon_i(\gamma)|+1\right) .$$

Then by the Markov inequality and the Cauchy Schwarz inequality, Lemma 2.4.2, for some $0<s<1$ small enough, we have

$$\begin{aligned}
&P\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left|\frac{\partial l_t(\lambda_0)}{\partial\lambda}-\frac{\partial\tilde{l}_t(\lambda_0)}{\partial\lambda}\right|>\epsilon\right)\\
\leq\ &\frac{K^s n^{-s/2}}{\epsilon^s}\sum_{t=1}^{n}\rho^{ts}\left(\mathbf{E}S_{t-1}^{4s}(\gamma_0)\mathbf{E}\xi_{\rho,t-1}^{2s}\right)^{1/2}\Big\{1+\mathbf{E}|\eta_t|^{4s\iota_1}+\mathbf{E}\max(\eta_t,1)^{4s\iota_1-2s}\\
&\quad+\left[\left(1+\mathbf{E}|\eta_t|^{4s}\right)\left(\mathbf{E}\max(\eta_t^*,1)^{8s\iota_1-4s}+\mathbf{E}\max(\eta_t^*,1)^{8s\iota_1-8s}\right)\right]^{1/2}\Big\}^{1/2}\\
\leq\ &\frac{Kn^{-s/2}}{\epsilon^s}\sum_{t=1}^{n}\rho^{ts}t\longrightarrow 0.
\end{aligned}$$

The second convergence in the lemma can be proved by similar arguments. Based on the expression of $\partial^2\tilde{l}_t(\lambda)/\partial\lambda$, analogous to (2.2.13) we can obtain the expression of $\partial^2\tilde{l}_t(\lambda)/\partial\lambda$, and then

$$\sup_{\lambda\in\Theta_0}\left|\frac{\partial^2 l_t(\lambda)}{\partial\lambda\partial\lambda^T}-\frac{\partial^2\tilde{l}_t(\lambda)}{\partial\lambda\partial\lambda^T}\right| \;=\; I_{11}+I_{12}+I_{13}\ ,$$

where

$$\begin{aligned}
I_{11} &= \sup_{\lambda\in V(\lambda_0)} \Bigg| -\frac{1}{2}\Big(1+g\big(\eta_t(\lambda)\big)\eta_t(\lambda)\Big)\left(\frac{\partial^2\sigma_t^2(\lambda)/\partial\lambda}{\sigma_t^2}-\frac{\partial^2\tilde{\sigma}_t^2(\lambda)/\partial\lambda}{\tilde{\sigma}_t^2}\right) \\
&\quad -\frac{1}{2}\Big(g\big(\eta_t(\lambda)\big)\eta_t(\lambda)-g(\tilde{\eta}_t(\lambda))\tilde{\eta}_t(\lambda)\Big)\frac{\partial^2\tilde{\sigma}_t^2(\lambda)/\partial\lambda}{\tilde{\sigma}_t^2} \\
&\quad +\frac{1}{4}\Big(2+3g\big(\eta_t(\lambda)\big)\eta_t(\lambda)+g'\big(\eta_t(\lambda)\big)\eta_t^2(\lambda)\Big) \\
&\quad \times\left\{\frac{\partial\sigma_t^2(\lambda)/\partial\lambda}{\sigma_t^2}\frac{\partial\sigma_t^2(\lambda)/\partial\lambda^T}{\sigma_t^2}-\frac{\partial\tilde{\sigma}_t^2(\lambda)/\partial\lambda}{\tilde{\sigma}_t^2}\frac{\partial\tilde{\sigma}_t^2(\lambda)/\partial\lambda^T}{\tilde{\sigma}_t^2}\right\} \\
&\quad +\frac{1}{4}\Big(3g\big(\eta_t(\lambda)\big)\eta_t(\lambda)+g'\big(\eta_t(\lambda)\big)\eta_t^2(\lambda)-3g\big(\tilde{\eta}_t(\lambda)\big)\tilde{\eta}_t(\lambda)-g'\big(\tilde{\eta}_t(\lambda)\big)\tilde{\eta}_t^2(\lambda)\Big) \\
&\quad \times\left(\frac{\partial\tilde{\sigma}_t^2(\lambda)/\partial\lambda}{\tilde{\sigma}_t^2}\frac{\partial\tilde{\sigma}_t^2(\lambda)/\partial\lambda^T}{\tilde{\sigma}_t^2}\right)
\end{aligned}$$

$$\begin{aligned}
I_{12} &= \frac{1}{2}\Big(g'\big(\eta_t(\lambda)\big)\eta_t(\lambda)-g\big(\eta_t(\lambda)\big)\Big) \\
&\quad \times\left\{\frac{\partial\sigma_t^2(\lambda)/\partial\lambda}{\sigma_t^2}\frac{\partial\varepsilon_t(\lambda)/\partial\lambda^T}{\sigma_t}-\frac{\partial\tilde{\sigma}_t^2(\lambda)/\partial\lambda}{\tilde{\sigma}_t^2}\frac{\partial\tilde{\varepsilon}_t(\lambda)/\partial\lambda^T}{\tilde{\sigma}_t}\right\} \\
&\quad +\frac{1}{2}\left\{\Big(g'\big(\eta_t(\lambda)\big)\eta_t(\lambda)-g\big(\eta_t(\lambda)\big)\Big)-\Big(g'\big(\tilde{\eta}_t(\lambda)\big)\tilde{\eta}_t(\lambda)-g(\tilde{\eta}_t(\lambda))\Big)\right\} \\
&\quad \times\frac{\partial\tilde{\sigma}_t^2(\lambda)/\partial\lambda}{\tilde{\sigma}_t^2}\frac{\partial\tilde{\varepsilon}_t(\lambda)/\partial\lambda^T}{\tilde{\sigma}_t} \\
&\quad +\frac{1}{2}\Big(g'\big(\eta_t(\lambda)\big)\eta_t(\lambda)-g\big(\eta_t(\lambda)\big)\Big) \\
&\quad \times\left\{\frac{\partial\varepsilon_t(\lambda)/\partial\lambda}{\sigma_t}\frac{\partial\sigma_t^2(\lambda)/\partial\lambda^T}{\sigma_t^2}-\frac{\partial\tilde{\varepsilon}_t(\lambda)/\partial\lambda}{\tilde{\sigma}_t}\frac{\partial\tilde{\sigma}_t^2(\lambda)/\partial\lambda^T}{\tilde{\sigma}_t^2}\right\} \\
&\quad +\frac{1}{2}\left\{\Big(g'\big(\eta_t(\lambda)\big)\eta_t(\lambda)-g\big(\eta_t(\lambda)\big)\Big)-\Big(g'\big(\tilde{\eta}_t(\lambda)\big)\tilde{\eta}_t(\lambda)-g(\tilde{\eta}_t(\lambda))\Big)\right\} \\
&\quad \times\frac{\partial\tilde{\varepsilon}_t(\lambda)/\partial\lambda}{\tilde{\sigma}_t}\frac{\partial\tilde{\sigma}_t^2(\lambda)/\partial\lambda^T}{\tilde{\sigma}_t^2},
\end{aligned}$$

$$
\begin{aligned}
I_{13} &= g'(\eta_t(\lambda))\left\{\frac{\partial\varepsilon_t(\lambda)/\partial\lambda}{\sigma_t}\frac{\partial\varepsilon_t(\lambda)/\partial\lambda^T}{\sigma_t} - \frac{\partial\tilde{\varepsilon}_t(\lambda)/\partial\lambda}{\tilde{\sigma}_t}\frac{\partial\tilde{\varepsilon}_t(\lambda)/\partial\lambda^T}{\tilde{\sigma}_t}\right\} \\
&+\Big(g'(\eta_t(\lambda)) - g'(\eta_t(\lambda))\Big)\frac{\partial\tilde{\varepsilon}_t(\lambda)/\partial\lambda}{\tilde{\sigma}_t}\frac{\partial\tilde{\varepsilon}_t(\lambda)/\partial\lambda^T}{\tilde{\sigma}_t} \\
&+g(\eta_t(\lambda))\left\{\frac{\partial^2\varepsilon_t(\lambda)/\partial\lambda\partial\lambda^T}{\sigma_t} - \frac{\partial^2\tilde{\varepsilon}_t(\lambda)/\partial\lambda\partial\lambda^T}{\tilde{\sigma}_t}\right\} \\
&+\Big(g(\eta_t(\lambda)) - g(\tilde{\eta}_t(\lambda))\Big)\frac{\partial^2\tilde{\varepsilon}_t(\lambda)/\partial\lambda\partial\lambda^T}{\tilde{\sigma}_t}\Bigg| .
\end{aligned}
$$

By Assumption 4, inequalities (A.7.1) to (A.7.11) in Appendix A.7, and by the mean value theorem, there exists $\eta_t^*(\lambda)$ such that $|\eta_t^*(\lambda)-\eta_t(\lambda)| \le |\eta_t(\lambda)-\tilde{\eta}_t(\lambda)|$, and

$$
\begin{aligned}
&\sup_{\lambda\in\Theta_0}\left|\frac{\partial^2 l_t(\lambda)}{\partial\lambda\partial\lambda^T} - \frac{\partial^2\tilde{l}_t(\lambda)}{\partial\lambda\partial\lambda^T}\right| \\
\le\ & K\rho^t S_{t-1}^2\xi_{\rho,t-1}^2\left(1+\eta_t^{2\iota_1}(\lambda)+\max(\eta_t(\lambda),1)^{2\iota_1-1}+\max(\eta_t(\lambda),1)^{2\iota_1-2}\right) \\
&+K\rho^t S_{t-1}(1+|\eta_t(\lambda)|)(1+K\rho^t S_{t-1}^2)\xi_{\rho,t-1}^2 \\
&\times\Big(\eta_t^*(\lambda)^{2\iota_1}+\max(\eta_t^*(\lambda),1)^{2\iota_1-1}+\max(\eta_t^*(\lambda),1)^{2\iota_1-2}\Big) .
\end{aligned}
$$

Then by Lemma 2.4.2, the Markov inequality and the Cauchy-Schwarz inequality, similarly we can show:

$$
\sup_{\lambda\in\Theta_0}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial^2 l_t(\lambda)}{\partial\lambda\partial\lambda^T} - \frac{\partial^2\tilde{l}_t(\lambda)}{\partial\lambda\partial\lambda^T}\right)\right| \longrightarrow 0, \text{ in probability as } n\to\infty.
$$

This completes the proof of Lemma 2.4.11. END

Lemma 2.4.12. *Under Assumptions 1 to 7,* $\mathbf{E}|\varepsilon_t|^{2(\iota_1+1)(1-\iota_2)}$*, if* $\mathbf{E}|\eta_t|^{\max(1,2\iota_1)} < \infty$*, then*

$$
\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\lambda^*)}{\partial\lambda_i\partial\lambda_j} \longrightarrow \mathcal{J}(i,j),
$$

almost surely for any λ^* *between* $\tilde{\lambda}_n$ *and* λ_0.

Proof: By Lemma 2.4.8, $\mathcal{J}$ exists. By ergodic theorem,

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\lambda_0)}{\partial\lambda\partial\lambda^T} = \mathcal{J}, \ a.s.$$

Since $\tilde{\lambda}_n \longrightarrow \lambda_0$ almost surely, it suffices to prove that for any $\epsilon > 0$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sup_{\lambda\in\Theta_0}\left|\frac{\partial^2 l_t(\lambda)}{\partial\lambda\partial\lambda^T} - \frac{\partial^2 l_t(\lambda_0)}{\partial\lambda\partial\lambda^T}\right| \le \epsilon, \ a.s.$$

Furthermore since $\partial^2 l_t(\lambda)/\partial\lambda\partial\lambda^T$ is stationary and ergodic, thus it reduces to prove:

$$\sup_{\lambda\in\Theta_0}\mathbf{E}\left|\frac{\partial^2 l_t(\lambda)}{\partial\lambda\partial\lambda^T}\right| < \infty\,.$$

By (2.2.13), Assumption 4, Lemma 2.4.6 and 2.4.7, the independence of η_t with ξ_{t-1}, we have:

$$\begin{aligned}
&\sup_{\lambda\in\Theta_0}\mathbf{E}\left|\frac{\partial^2 l_t(\lambda)}{\partial\lambda\partial\lambda^T}\right| \\
\le\ & K\sup_{\lambda\in\Theta_0}\mathbf{E}\left\{\left(1+\eta_t^{2\iota_1}(\lambda)+\max(\eta_t(\lambda),1)^{2\iota_1-1}+\max(\eta_t(\lambda),1)^{2\iota_1-2}\right)\xi_{\rho,t-1}^{2(1-\iota_2)}\right\} \\
\le\ & K\sup_{\lambda\in\Theta_0}\left\{\mathbf{E}\xi_{t-1}^{2(1-\iota_2)}+\mathbf{E}(1+|\eta_t|)^{2\iota_1}\mathbf{E}\xi_{t-1}^{2(\iota_1+1)(1-\iota_2)}+\mathbf{E}(1+|\eta_t|)^{2\iota_1-1}\mathbf{E}\xi_{t-1}^{(2\iota_1+1)(1-\iota_2)}\right. \\
&\left.+\mathbf{E}(1+|\eta_t|)^{2\iota_1-2}\mathbf{E}\xi_{t-1}^{2\iota_1(1-\iota_2)}\right\} \\
<\ & \infty\,.
\end{aligned}$$

Since $\tilde{\lambda}_n \to \lambda_0$ almost surely, as Θ_0 decreases to the singleton λ_0, $\mathbf{E}\partial^2 l_t(\lambda^*)/\partial\lambda_i\partial\lambda_j \to \mathcal{J}(i,j)$ almost surely. This completes the proof of Lemma 2.4.12. END

Lemma 2.4.13. *Under Assumptions 1-5, if* $\mathbf{E}|\varepsilon_t|^{2(1-\iota_2)} < \infty$, *and* $\mathbf{E}|\eta_t|^{\max(1,4\iota_1)} < \infty$, *then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial l_t(\lambda_0)}{\partial\lambda} \Rightarrow N(0,\mathcal{I}).$$

Proof: This Lemma can be proved by a central limit theorem for martingale differences. It is clear that $\partial l_t(\lambda_0)/\partial\lambda$ is stationary and ergodic. By Remark 2.2.3, we have

$$\mathbf{E}\left(\frac{\partial l_t(\lambda_0)}{\partial\lambda}\bigg|\mathcal{F}_{t-1}\right) = 0 \ .$$

Lemma 2.4.8 shows that $Var(\partial l_t(\lambda_0)/\partial\lambda)$ exists. By Lemma 2.4.9, we have $Var(\partial l_t(\lambda_0)/\partial\lambda)$ is non-degenerate. So for any $r \in \mathbb{R}^{(1+P+Q+1+p+q)}$, the sequence $\{r^T\partial l_t(\lambda_0)/\partial\lambda, \mathcal{F}_{t-1}\}$ is a square-integrable stationary martingale difference. Then by the central limit theorem of Billingsley (1961) and the Wold-Cramér device, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial l_t(\lambda_0)}{\partial\lambda} \Rightarrow N(0,\mathcal{I}) \ .$$

This completes the proof of Lemma 2.4.13. END

Let $\sigma_t^2(\hat{\gamma}_n,\delta) = \alpha_0+\sum_{i=1}^{p}\alpha_j\varepsilon_{t-i}(\hat{\gamma}_n)+\sum_{j=1}^{q}\beta_j\sigma_{t-j}^2(\hat{\gamma}_n,\delta)$ and $\eta_t(\hat{\gamma}_n,\delta) = \varepsilon_t(\hat{\gamma}_n)/\sigma_t(\hat{\gamma}_n,\delta)$. Also define

$$L_n(\hat{\gamma}_n,\delta) = \frac{1}{n}\sum_{t=1}^{n} l_t(\hat{\gamma}_n,\delta) = \frac{1}{n}\sum_{t=1}^{n}\log\frac{h\big(\eta_t(\hat{\gamma}_n,\delta)\big)}{\sigma_t(\hat{\gamma}_n,\delta)}.$$

Similarly define

$$\tilde{L}_n(\hat{\gamma}_n,\delta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\hat{\gamma}_n,\delta) = \frac{1}{n}\sum_{t=1}^{n}\log\frac{h\big(\tilde{\eta}_t(\hat{\gamma}_n,\delta)\big)}{\tilde{\sigma}_t(\hat{\gamma}_n,\delta)}.$$

Lemma 2.4.14. *Under Assumptions 1, 3 and 4, if* $\mathbf{E}|\varepsilon_t|^{2\iota_1(1-\iota_2)+\zeta} < \infty$ *for some* $0 < \zeta < 1$ *(or* $\mathbf{E}|\varepsilon_t|^s < \infty$ *for some* $s > 0$ *if* $\iota_1 = 0$*), and* $\mathbf{E}|\eta_t|^{\max(1,2\iota_1)} < \infty$*, then as n goes to infinity, we have:*

$$\sup_{\delta\in\Theta_\delta}|L_n(\hat{\gamma}_n,\delta) - L_n(\gamma_0,\delta)| = o_p(1) \ ,$$

$$\sup_{\delta\in\Theta_\delta}\left|\tilde{L}_n(\hat{\gamma}_n,\delta) - L_n(\hat{\gamma}_n,\delta)\right| = 0, \ \ a.s.$$

Proof: By the mean value theorem, we have:

$$\begin{aligned}
&\sup_{\delta\in\Theta_\delta}|L_n(\hat{\gamma}_n,\delta)-L_n(\gamma_0,\delta)| \\
\leq\ & \sup_{\delta\in\Theta_\delta}\left|\frac{1}{n}\sum_{t=1}^{n}\log\left(\frac{\sigma_t(\gamma_0,\delta)}{\sigma_t(\hat{\gamma}_n,\delta)}\right)\right| + \sup_{\delta\in\Theta_\delta}\frac{1}{n}\left|\sum_{t=1}^{n}\log h\left(\frac{\varepsilon_t}{\sigma_t(\gamma_0,\delta)}\right)-\log h\left(\frac{\varepsilon_t(\hat{\gamma}_n)}{\sigma_t(\hat{\gamma}_n,\delta)}\right)\right| \\
\leq\ & \sup_{\delta\in\Theta_\delta}\left|\frac{1}{n}\sum_{t=1}^{n}\log\left(\frac{\sigma_t(\gamma_0,\delta)}{\sigma_t(\hat{\gamma}_n,\delta)}\right)\right| + \sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}|g(\eta^*_{\lambda t})|\left|\frac{\varepsilon_t}{\sigma_t(\gamma_0,\delta)}-\frac{\varepsilon_t(\hat{\gamma}_n)}{\sigma_t(\hat{\gamma}_n,\delta)}\right| \\
=\ & L_1+L_2,
\end{aligned}$$

where $|\gamma^*-\gamma_0|<|\hat{\gamma}_n-\gamma_0|$, $|\eta^*_{\lambda t}-\varepsilon_t/\sigma_t(\gamma_0,\delta)|\leq|\varepsilon_t/\sigma_t(\gamma_0,\delta)-\varepsilon_t(\hat{\gamma}_n)/\sigma_t(\hat{\gamma}_n,\delta)|$.

We will show following both L_1 and L_2 are $o_p(1)$.

Ling (2005) has showed that L_1 is $o_p(1)$. For completeness, we give his proof below.

With probability 1, $\sigma_t(\lambda)>\alpha_0>0$ for any λ, by mean value theorem, there exists γ^* between γ_0 and $\hat{\gamma}_n$, such that

$$\begin{aligned}
&\sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}\log\left(\frac{\sigma_t(\gamma_0,\delta)}{\sigma_t(\hat{\gamma}_n,\delta)}\right) \\
&= \sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}\log\left(1+(\gamma_0-\hat{\gamma}_n)\frac{\partial\sigma_t(\gamma^*,\delta)/\partial\gamma}{\sigma_t(\hat{\gamma}_n,\delta)}\right) \\
&\leq \sup_{\lambda\in\Theta}\frac{1}{n\zeta}\sum_{t=1}^{n}\log\left(1+\alpha_0^{-1}|\gamma_0-\hat{\gamma}_n|\left|\frac{\partial\sigma_t(\gamma^*,\delta)}{\partial\gamma}\right|\right)^{\zeta}.
\end{aligned}$$

There exists an $\zeta>0$ such that $\mathbf{E}\sup_{\lambda\in\Theta}|\partial\sigma_t(\lambda^*,\delta)/\partial\gamma|^{\zeta}<\infty$. For any $\epsilon>0$, first taking π small enough such that $\log(1+\pi^{\zeta}\alpha_0^{-\zeta}\mathbf{E}\sup_{\lambda\in\Theta}|\partial\sigma_t(\lambda)/\partial\gamma|^{\zeta})<\epsilon^2\zeta$ and then n large enough such that $P(|\gamma_0 - \hat{\gamma}_n| \geq \pi) \leq \epsilon$, it follows that

$$
\begin{aligned}
& P\left\{\frac{1}{n\zeta}\sum_{t=1}^{n}\log\left(1+\alpha_0^{-1}|\gamma_0-\hat{\gamma}_n|\sup_{\lambda\in\Theta}\left|\frac{\partial\sigma_t(\lambda)}{\partial\gamma}\right|\right)^{\zeta}\geq\epsilon\right\} \\
\leq\ & P\left\{\frac{1}{n\zeta}\sum_{t=1}^{n}\log\left(1+\alpha_0^{-1}|\gamma_0-\hat{\gamma}_n|\sup_{\lambda\in\Theta}\left|\frac{\partial\sigma_t(\lambda)}{\partial\gamma}\right|\right)^{\zeta}\geq\epsilon,\ |\gamma_0-\hat{\gamma}_n|\leq\pi\right\}+\epsilon \\
\leq\ & \frac{1}{n\zeta\epsilon}\sum_{t=1}^{n}\mathbf{E}\log\left(1+\alpha_0^{-1}\pi\sup_{\lambda\in\Theta}\left|\frac{\partial\sigma_t(\lambda)}{\partial\gamma}\right|\right)^{\zeta}+\epsilon \\
=\ & \frac{1}{\zeta\epsilon}\mathbf{E}\log\left(1+\alpha_0^{-1}\pi\sup_{\lambda\in\Theta}\left|\frac{\partial\sigma_t(\lambda)}{\partial\gamma}\right|\right)^{\zeta}+\epsilon \\
\leq\ & \frac{1}{\zeta\epsilon}\log\left(1+\alpha_0^{-\zeta}\pi^{\zeta}\mathbf{E}\sup_{\lambda\in\Theta}\left|\frac{\partial\sigma_t(\lambda)}{\partial\gamma}\right|^{\zeta}\right)+\epsilon \\
\leq\ & 2\epsilon\ ,
\end{aligned}
$$

where the last second inequality holds by Jenson's inequality. Thus

$$P\left\{\sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}\log\left(\frac{\sigma_t(\gamma_0,\delta)}{\sigma_t(\hat{\gamma}_n,\delta)}\right)\geq\epsilon\right\}\leq 2\epsilon.$$

Similarly we can show that

$$P\left\{\sup_{\delta\in\Theta_\delta}\frac{-1}{n}\sum_{t=1}^{n}\log\left(\frac{\sigma_t(\gamma_0,\delta)}{\sigma_t(\hat{\gamma}_n,\delta)}\right)\geq\epsilon\right\}\leq 2\epsilon.$$

Thus we can claim that

$$\left|\sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}\log\left(\frac{\sigma_t(\gamma_0,\delta)}{\sigma_t(\hat{\gamma}_n,\delta)}\right)\right| = o_p(1)\ .$$

Next we show L_2 is $o_p(1)$.

$$
\begin{aligned}
&\sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}|g(\pi^*_{\lambda t})|\left|\frac{\varepsilon_t}{\sigma_t(\gamma_0,\delta)}-\frac{\varepsilon_t(\hat\gamma_n)}{\sigma_t(\hat\gamma_n,\delta)}\right| \\
\le\ & \sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}|g(\pi^*_{\lambda t})\varepsilon_t|\left|\frac{1}{\sigma_t(\gamma_0,\delta)}-\frac{1}{\sigma_t(\hat\gamma_n,\delta)}\right|+\sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}|g(\pi^*_{\lambda t})|\left|\frac{\varepsilon_t-\varepsilon_t(\hat\gamma_n)}{\sigma_t(\hat\gamma_n,\delta)}\right| \\
\le\ & \sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}|g(\pi^*_{\lambda t})||\varepsilon_t|\left|\frac{(\gamma_0-\hat\gamma_n)\partial\sigma_t^2(\gamma^*,\delta)/\partial\gamma}{\sigma_t(\gamma_0,\delta)\sigma_t(\hat\gamma_n,\delta)(\sigma_t(\gamma_0,\delta)+\sigma_t(\hat\gamma_n,\delta))}\right|^{\zeta} \\
&\times\left|\frac{1}{\sigma_t(\gamma_0,\delta)}-\frac{1}{\sigma_t(\hat\gamma_n,\delta)}\right|^{1-\zeta}+\sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}|g(\pi^*_{\lambda t})|\left|\frac{(\gamma_0-\hat\gamma_n)\partial\varepsilon_t(\gamma^*)/\partial\gamma}{\sigma_t(\hat\gamma_n,\delta)}\right| \\
\le\ & \sup_{\lambda\in\Theta}\frac{K}{n}\sum_{t=1}^{n}|g(\pi^*_{\lambda t})|\frac{|\varepsilon_t|}{\sigma_t(\lambda)}|\gamma_0-\hat\gamma_n|^{\zeta}\left|\frac{\partial\sigma_t^2(\lambda)}{\partial\gamma}\right|^{\zeta} \\
&+\sup_{\lambda\in\Theta}\frac{1}{n}\sum_{t=1}^{n}|g(\pi^*_{\lambda t})||\gamma_0-\hat\gamma_n|\left|\frac{\partial\varepsilon_t(\gamma^*)/\partial\gamma}{\sigma_t(\lambda)}\right| .
\end{aligned}
$$

From (2.4.13) in Lemma 2.4.7, we know that for any $\gamma^*\neq\gamma$, there exists ρ and constant C such that $\partial\varepsilon_t(\gamma^*)/\partial\gamma$ are bounded by $C\xi_{\gamma,\rho,t-1}$. Thus by Lemmas 2.4.6 and 2.4.7, we have:

$$
\begin{aligned}
\sup_{\delta\in\Theta_\delta}\left|\frac{\varepsilon_t}{\sigma_t(\gamma_0,\delta)}\right| &= |\eta_t|\sup_{\delta\in\Theta_\delta}\left|\frac{\sigma_t}{\sigma_t(\gamma_0,\delta)}\right| \\
&\le |\eta_t|\sup_{\lambda\in\Theta}\left|\frac{\xi_{\rho_1,t-1}}{\sigma_t(\lambda)}\right|\le|\eta_t|\sup_{\lambda\in\Theta}\left|\frac{\xi_{\gamma_0,\rho_2,t-1}}{\sigma_t(\lambda)}\right| \\
&\le |\eta_t|\xi^{1-\iota_2}_{\gamma,\rho_3,t-1}\le K|\eta_t|\xi^{1-\iota_2}_{\rho,t-1}
\end{aligned}
$$

$$
\begin{aligned}
\sup_{\delta\in\Theta_\delta}\left|\frac{\varepsilon_t(\hat\gamma_n)}{\sigma_t(\hat\gamma_n,\delta)}\right| &= \sup_{\delta\in\Theta_\delta}\left|\frac{\varepsilon_t+(\hat\gamma_n-\gamma_0)\partial\varepsilon_t(\gamma^*)/\partial\gamma}{\sigma_t(\hat\gamma_n,\delta)}\right| \qquad (2.4.17)\\
&\le |\eta_t|\sup_{\lambda\in\Theta}\left|\frac{\sigma_t}{\sigma_t(\lambda)}\right|+|\hat\gamma_n-\gamma_0|\sup_{\lambda\in\Theta}\left|\frac{\xi_{\gamma^*,\rho_1,t-1}}{\sigma_t(\lambda)}\right| \\
&\le K|\eta_t|\xi^{1-\iota_2}_{\rho,t-1}+K|\hat\gamma_n-\gamma_0|\xi^{1-\iota_2}_{\rho,t-1} .
\end{aligned}
$$

Thus by Assumption 4, if $2\iota_1 - 1 > 0$, we have:

$$|g(\eta^*_{\lambda t})| \le K|\eta_t|^{2\iota_1-1}\xi_{\rho,t-1}^{(2\iota_1-1)(1-\iota_2)} + K|\hat{\gamma}_n - \gamma_0|^{2\iota_1-1}\xi_{\rho,t-1}^{(2\iota_1-1)(1-\iota_2)} . \qquad (2.4.18)$$

If $2\iota_1 - 1 \le 0$, we have $|g(\eta^*_{\lambda t})| \le K$. In the following proof, we give the case of $2\iota_1 - 1 > 0$. The proof for case of $2\iota_1 - 1 \le 0$ is similar.

Since $\hat{\gamma}_n \to \gamma_0$ in probability, $\mathbf{E}|\eta_t|^{\max(1,2\iota_1)} < \infty$ and $\mathbf{E}\xi_{\rho,t-1}^{2\iota_1(1-\iota_2)+\zeta} < \infty$ due to $\mathbf{E}|\varepsilon_t|^{2\iota_1(1-\iota_2)+\zeta} < \infty$, and η_t being independent with $\xi_{\rho,t-1}$. With (2.4.18) and Lemma 2.4.7, by the weak law of large numbers and ergodic theorem, we have:

$$\begin{aligned}
&\sup_{\lambda\in\Theta}\frac{1}{n}\sum_{t=1}^{n}|g(\eta^*_{\lambda t})|\frac{|\varepsilon_t|}{\sigma_t(\lambda)}\left|\frac{(\gamma_0-\hat{\gamma}_n)\partial\sigma_t^2(\lambda)}{\partial\gamma}\right|^{\zeta}\\
\le\ & K|\gamma_0-\hat{\gamma}_n|^{\zeta}\frac{1}{n}\sum_{t=1}^{n}|\eta_t|^{2\iota_1}\xi_{\rho,t-1}^{2\iota_1(1-\iota_2)+\zeta}\\
&+K|\gamma_0-\hat{\gamma}_n|^{2\iota_1-1+\zeta}\frac{1}{n}\sum_{t=1}^{n}|\eta_t|\xi_{\rho,t-1}^{2\iota_1(1-\iota_2)+\zeta}\\
=\ & o_p(1) .
\end{aligned}$$

Similarly by (2.4.18) and Lemma 2.4.7, the weak law of large numbers and ergodic theorem, we have:

$$\begin{aligned}
&\sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}|g(\eta^*_{\lambda t})|\left|\frac{(\gamma_0-\hat{\gamma}_n)\partial\varepsilon_t(\gamma^*)/\partial\gamma}{\sigma_t(\hat{\gamma}_n,\delta)}\right|\\
\le\ & |\gamma_0-\hat{\gamma}_n|\frac{1}{n}\sum_{t=1}^{n}|\eta_t|^{2\iota_1-1}\xi_{\rho,t-1}^{(2\iota_1-1)(1-\iota_2)}\sup_{\lambda\in\Theta}\left|\frac{\xi_{\gamma^*,\rho_1,t-1}}{\sigma_t(\lambda)}\right|\\
&+K|\hat{\gamma}_n-\gamma_0|^{2\iota_1}\frac{1}{n}\sum_{t=1}^{n}\xi_{\rho,t-1}^{(2\iota_1-1)(1-\iota_2)}\sup_{\lambda\in\Theta}\left|\frac{\xi_{\gamma^*,\rho_1,t-1}}{\sigma_t(\lambda)}\right|\\
\le\ & |\gamma_0-\hat{\gamma}_n|\frac{1}{n}\sum_{t=1}^{n}|\eta_t|^{2\iota_1-1}\xi_{\rho,t-1}^{2\iota_1(1-\iota_2)}+K|\hat{\gamma}_n-\gamma_0|^{2\iota_1}\frac{1}{n}\sum_{t=1}^{n}\xi_{\rho,t-1}^{2\iota_1(1-\iota_2)} = o_p(1) .
\end{aligned}$$

Thus

$$\sup_{\delta\in\Theta_\delta} |L_n(\hat{\gamma}_n,\delta) - L_n(\gamma_0,\delta)| = o_p(1).$$

By Lemma 2.4.3, we have

$$\sup_{\delta\in\Theta_\delta} \left|\tilde{L}_n(\hat{\gamma}_n,\delta) - L_n(\hat{\gamma}_n,\delta)\right| = 0, \ a.s.$$

This completes the proof of Lemma 2.4.14. END

Lemma 2.4.15. *Under Assumptions 1, 3 and 4, Assumption 5 with $w = 1$, then $\mathbf{E}|l_t(\gamma_0,\delta_0)| < \infty$, furthermore if $\delta \neq \delta_0$, then $\mathbf{E}l_t(\gamma_0,\delta_0) > \mathbf{E}l_t(\gamma_0,\delta)$.*

Proof: We have shown $\mathbf{E}|l_t(\gamma_0,\delta_0)| < \infty$ in Lemma 2.4.5. With the modified Assumption 5 and Lemma 2.4.4, it is straightforward that $\mathbf{E}l_t(\gamma_0,\delta)$ has unique maximum at $\delta = \delta_0$. END

$$\begin{aligned}\frac{\partial^2 l_t(\gamma_0,\delta)}{\partial\delta\partial\delta^T} &= -\frac{1}{2}\Big(1+g\big(\eta_t(\gamma_0,\delta)\big)\eta_t(\gamma_0,\delta)\Big)\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial\delta^T}{\sigma_t^2(\gamma_0,\delta)} \\ &\quad +\frac{1}{4}\Big(2+3g\big(\eta_t(\gamma_0,\delta)\big)\eta_t(\gamma_0,\delta)+g'\big(\eta_t(\gamma_0,\delta)\big)\eta_t^2(\gamma_0,\delta)\Big) \\ &\quad \times\frac{\partial\sigma_t^2(\gamma_0,\delta)/\partial\delta}{\sigma_t^2(\gamma_0,\delta)}\frac{\partial\sigma_t^2(\gamma_0,\delta)/\partial\delta^T}{\sigma_t^2(\gamma_0,\delta)}\ . \end{aligned} \tag{2.4.19}$$

Similarly we can write down $\partial^2\tilde{l}_t(\gamma_0,\delta)/\partial\delta\partial\delta^T$.

Lemma 2.4.16. *Under Assumptions 1, 3 and 4, Assumption 5 with $w = 1$, if $\mathbf{E}|\varepsilon_t|^{2\iota_1(1-\iota_2)+2\zeta} < \infty$ (or $\mathbf{E}|\varepsilon_t|^s < \infty$ for some $s > 0$ if $\iota_1 = 0$) and $\mathbf{E}|\eta_t|^{\max(1,2\iota_1)} \leq \infty$, then*

$$\sup_{\delta\in\Theta_\delta}\left|\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial^2 l_t(\hat{\gamma}_n,\delta)}{\partial\delta\partial\delta^T} - \frac{\partial^2 l_t(\gamma_0,\delta)}{\partial\delta\partial\delta^T}\right)\right| = o_p(1),$$

$$\sup_{\delta\in\Theta_\delta}\left|\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial^2 \tilde{l}_t(\hat{\gamma}_n,\delta)}{\partial\delta\partial\delta^T} - \frac{\partial^2 l_t(\hat{\gamma}_n,\delta)}{\partial\delta\partial\delta^T}\right)\right| = o_p(1),$$

Proof: First we without loss of generality, we assume $2\iota_1 - 1 > 0$, by Assumption 4, Lemmas 2.4.6 and 2.4.7, and the mean value theorem, we have:

$$
\begin{aligned}
&\sup_{\delta\in\Theta_\delta} \frac{1}{n}\sum_{t=1}^{n}\Bigg| g\left(\frac{\varepsilon_t(\hat\gamma_n)}{\sigma_t(\hat\gamma_n,\delta)}\right)\frac{\varepsilon_t(\hat\gamma_n)}{\sigma_t(\hat\gamma_n,\delta)}\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)} \\
&\quad - g\left(\frac{\varepsilon_t}{\sigma_t(\gamma_0,\delta)}\right)\frac{\varepsilon_t}{\sigma_t(\gamma_0,\delta)}\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\Bigg| \\
\le\ & \sup_{\delta\in\Theta_\delta} \frac{1}{n}\sum_{t=1}^{n} |g'(\eta^*_{\lambda t})\eta^*_{\lambda t} + g(\eta^*_{\lambda t})| \left|\frac{\varepsilon_t(\hat\gamma_n)}{\sigma_t(\hat\gamma_n,\delta)} - \frac{\varepsilon_t}{\sigma_t(\gamma_0,\delta)}\right| \left|\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\right| \\
\le\ & \sup_{\delta\in\Theta_\delta} \frac{1}{n}\sum_{t=1}^{n} |\eta^*_{\lambda t}|^{2\iota_1-1}|\varepsilon_t| \left|\frac{1}{\sigma_t(\gamma_0,\delta)} - \frac{1}{\sigma_t(\hat\gamma_n,\delta)}\right|^{\varsigma} \left|\frac{1}{\sigma_t(\gamma_0,\delta)} - \frac{1}{\sigma_t(\hat\gamma_n,\delta)}\right|^{1-\varsigma} \\
&\times \left|\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\right| + \sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}|\eta^*_{\lambda t}|^{2\iota_1-1}\left|\frac{\varepsilon_t-\varepsilon_t(\hat\gamma_n)}{\sigma_t(\hat\gamma_n,\delta)}\right|\left|\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\right| \\
\le\ & \sup_{\lambda\in\Theta}\frac{K|\hat\gamma_n-\gamma_0|^{\varsigma}}{n}\sum_{t=1}^{n}|\eta^*_{\lambda t}|^{2\iota_1-1}\left|\frac{\varepsilon_t}{\sigma_t(\lambda)}\right|\left|\frac{\partial\sigma_t^2(\gamma,\delta)}{\partial\gamma}\right|^{\varsigma}\left|\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\right| \\
&+ \sup_{\delta\in\Theta_\delta}\frac{K|\hat\gamma_n-\gamma_0|}{n}\sum_{t=1}^{n}|\eta^*_{\lambda t}|^{2\iota_1-1}\left|\frac{\partial\varepsilon_t(\gamma^*)/\partial\gamma}{\sigma_t(\hat\gamma_n,\delta)}\right|\left|\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\right| ,
\end{aligned}
$$

where $|\varepsilon_t/\sigma_t(\gamma_0,\delta) - \eta^*_{\lambda t}| \le |\varepsilon_t/\sigma_t(\gamma_0,\delta) - \varepsilon_t(\hat\gamma_n)/\sigma_t(\hat\gamma_n,\delta)|$, γ^* between $\hat\gamma_n$ and γ_0.

By (2.4.18), Lemmas 2.4.6 and 2.4.7, we have

$$
\begin{aligned}
&|\eta^*_{\lambda t}|^{2\iota_1-1}\left|\frac{\varepsilon_t}{\sigma_t(\lambda)}\right|\left|\frac{\partial\sigma_t^2(\gamma^*,\delta)}{\partial\gamma}\right|^{\varsigma}\left|\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\right| \\
&\le |\eta_t|^{2\iota_1}\xi_{\rho,t-1}^{2\iota_1(1-\iota_2)+2\varsigma} + |\hat\gamma_n-\gamma_0|^{2\iota_1-1}|\eta_t|\xi_{\rho,t-1}^{2\iota_1(1-\iota_2)+2\varsigma} ,
\end{aligned}
$$

$$
\begin{aligned}
&|\eta^*_{\lambda t}|^{2\iota_1-1}\left|\frac{\partial\varepsilon_t(\gamma^*)/\partial\gamma}{\sigma_t(\hat\gamma_n,\delta)}\right|\left|\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\right| \\
&\le |\eta_t|^{2\iota_1-1}\xi_{\rho,t-1}^{2\iota_1(1-\iota_2)+\varsigma} + |\hat\gamma_n-\gamma_0|^{2\iota_1-1}\xi_{\rho,t-1}^{2\iota_1(1-\iota_2)+\varsigma} .
\end{aligned}
$$

Since $(\hat\gamma_n - \gamma_0) = o_p(1)$, by law of large number and ergodic Theorem, we have:

$$
\sup_{\lambda\in\Theta}\frac{K|\hat\gamma_n-\gamma_0|^{\varsigma}}{n}\sum_{t=1}^{n}|\eta^*_{\lambda t}|^{2\iota_1-1}\left|\frac{\varepsilon_t}{\sigma_t(\lambda)}\right|\left|\frac{\partial\sigma_t^2(\gamma^*,\delta)}{\partial\gamma}\right|^{\varsigma}\left|\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\right| = o_p(1) ,
$$

$$\sup_{\delta\in\Theta_\delta} \frac{K|\hat{\gamma}_n-\gamma_0|}{n}\sum_{t=1}^{n}|\eta^*_{\lambda t}|^{2\iota_1-1}\left|\frac{\partial\varepsilon_t(\gamma^*)/\partial\gamma}{\sigma_t(\hat{\gamma}_n,\delta)}\right|\left|\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\right| = o_p(1)\ .$$

Thus

$$\sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}\left|g\left(\frac{\varepsilon_t(\hat{\gamma}_n)}{\sigma_t(\hat{\gamma}_n,\delta)}\right)\frac{\varepsilon_t(\hat{\gamma}_n)}{\sigma_t(\hat{\gamma}_n,\delta)}\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\right.$$
$$\left.-g\left(\frac{\varepsilon_t}{\sigma_t(\gamma_0,\delta)}\right)\frac{\varepsilon_t}{\sigma_t(\gamma_0,\delta)}\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\right| = o_p(1)\ .$$

By Assumption 4, (2.4.17), Lemmas 2.4.6 and 2.4.7, and the mean value theorem, for some γ^* between $\hat{\gamma}_n$ and γ_0, we have

$$\begin{aligned}
&\sup_{\delta\in\Theta_\delta}\left|\frac{1}{n}\sum_{t=1}^{n}g\left(\frac{\varepsilon_t(\hat{\gamma}_n)}{\sigma_t(\hat{\gamma}_n,\delta)}\right)\frac{\varepsilon_t(\hat{\gamma}_n)}{\sigma_t(\hat{\gamma}_n,\delta)}\frac{\partial^2\sigma_t^2(\hat{\gamma}_n,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\hat{\gamma}_n,\delta)}\right.\\
&\quad\left.-g\left(\frac{\varepsilon_t(\hat{\gamma}_n)}{\sigma_t(\hat{\gamma}_n,\delta)}\right)\frac{\varepsilon_t(\hat{\gamma}_n)}{\sigma_t(\hat{\gamma}_n,\delta)}\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\right|\\
\le\ &\sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}\left|g\left(\frac{\varepsilon_t(\hat{\gamma}_n)}{\sigma_t(\hat{\gamma}_n,\delta)}\right)\frac{\varepsilon_t(\hat{\gamma}_n)}{\sigma_t(\hat{\gamma}_n,\delta)}\right|\left|\frac{\partial^2\sigma_t^2(\hat{\gamma}_n,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\hat{\gamma}_n,\delta)}-\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\right|^{\zeta}\\
&\times\left|\frac{\partial^2\sigma_t^2(\hat{\gamma}_n,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\hat{\gamma}_n,\delta)}-\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\right|^{1-\zeta}\\
\le\ &\sup_{\delta\in\Theta_\delta}\frac{2}{n}\sum_{t=1}^{n}\left|\frac{\varepsilon_t(\hat{\gamma}_n)}{\sigma_t(\hat{\gamma}_n,\delta)}\right|^{2\iota_1}\xi_{\rho,t-1}^{(1-\zeta)\zeta}\,|\hat{\gamma}_n-\gamma_0|^{\zeta}\left|\frac{\partial^3\sigma_t^2(\gamma^*,\delta)/\partial\delta\partial^T\delta\partial\gamma}{\sigma_t^2(\gamma^*,\delta)}\right|^{\zeta}\\
\le\ &\sup_{\delta\in\Theta_\delta}K\,|\hat{\gamma}_n-\gamma_0|^{\zeta}\,\frac{1}{n}\sum_{t=1}^{n}|\eta_t|^{2\iota_1}\xi_{\rho,t-1}^{2\iota_1(1-\iota_2)+2\zeta}+\sup_{\delta\in\Theta_\delta}K\,|\hat{\gamma}_n-\gamma_0|^{2\iota_1+\zeta}\,\frac{1}{n}\sum_{t=1}^{n}\xi_{\rho,t-1}^{2\iota_1(1-\iota_2)+2\zeta}\\
=\ &o_p(1)\ .
\end{aligned}$$

Since by a trivial extension of (iii) in Lemma 2.4.7, we have $|\partial^3\sigma_t^2(\gamma^*,\delta)/\partial\delta\partial^T\delta\partial\gamma| < \xi^2_{\gamma^*,\rho,t-1}$. The expression of $|\partial^3\sigma_t^2(\gamma^*,\delta)/\partial\delta\partial^T\delta\partial\gamma|$ can be obtained from (A.3.16)-(A.3.19) and (A.3.21) in Appendix A.3.

Thus we have:

$$\sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}\left|g\left(\frac{\varepsilon_t(\hat\gamma_n)}{\sigma_t(\hat\gamma_n,\delta)}\right)\frac{\varepsilon_t(\hat\gamma_n)}{\sigma_t(\hat\gamma_n,\delta)}\frac{\partial^2\sigma_t^2(\hat\gamma_n,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\hat\gamma_n,\delta)}\right.$$
$$\left.-g\left(\frac{\varepsilon_t}{\sigma_t(\gamma_0,\delta)}\right)\frac{\varepsilon_t}{\sigma_t(\gamma_0,\delta)}\frac{\partial^2\sigma_t^2(\gamma_0,\delta)/\partial\delta\partial^T\delta}{\sigma_t^2(\gamma_0,\delta)}\right| = o_p(1).$$

Then by the same arguments, we can prove

$$\sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}\left\{\left[2+3g\left(\frac{\varepsilon_t(\hat\gamma_n)}{\sigma_t(\hat\gamma_n,\delta)}\right)\frac{\varepsilon_t(\hat\gamma_n)}{\sigma_t(\hat\gamma_n,\delta)}+g'\left(\frac{\varepsilon_t(\hat\gamma_n)}{\sigma_t(\hat\gamma_n,\delta)}\right)\frac{\varepsilon_t^2(\hat\gamma_n)}{\sigma_t^2(\hat\gamma_n,\delta)}\right]\right.$$
$$\times\frac{\partial\sigma_t^2(\hat\gamma_n,\delta)/\partial\delta}{\sigma_t^2(\hat\gamma_n,\delta)}\frac{\partial\sigma_t^2(\hat\gamma_n,\delta)/\partial\delta^T}{\sigma_t^2(\hat\gamma_n,\delta)}$$
$$-\left[2+3g\left(\frac{\varepsilon_t}{\sigma_t(\gamma_0,\delta)}\right)\frac{\varepsilon_t}{\sigma_t(\gamma_0,\delta)}+g'\left(\frac{\varepsilon_t}{\sigma_t(\gamma_0,\delta)}\right)\frac{\varepsilon_t^2}{\sigma_t^2(\gamma_0,\delta)}\right]$$
$$\left.\times\frac{\partial\sigma_t^2(\gamma_0,\delta)/\partial\delta}{\sigma_t^2(\gamma_0,\delta)}\frac{\partial\sigma_t^2(\gamma_0,\delta)/\partial\delta^T}{\sigma_t^2(\gamma_0,\delta)}\right\} = o_p(1).$$

This proves the first convergence in the lemma.

By (A.7.2)-(A.7.11) in Appendix A.7, with the same argument as that in the proof of Lemma 2.4.11, we can prove the second convergence in this lemma.

This completes the proof of Lemma 2.4.16. END

$$\begin{aligned}\frac{\partial^2 l_t(\lambda)}{\partial\delta\partial\gamma^T} &= -\frac{1}{2}\Big(1+g(\eta_t(\lambda))\eta_t(\lambda)\Big)\frac{\partial^2\sigma_t^2(\lambda)/\partial\delta\partial\gamma^T}{\sigma_t^2(\lambda)}\\ &\quad+\frac{1}{4}\Big(2+3g(\eta_t(\lambda))\eta_t(\lambda)+g'(\eta_t(\lambda))\eta_t^2(\lambda)\Big)\frac{\partial\sigma_t^2(\lambda)/\partial\delta}{\sigma_t^2(\lambda)}\frac{\partial\sigma_t^2(\lambda)/\partial\gamma^T}{\sigma_t^2(\lambda)}\\ &\quad-\frac{1}{2}\Big(g'(\eta_t(\lambda))\eta_t(\lambda)+g(\eta_t(\lambda))\Big)\frac{\partial\sigma_t^2(\lambda)/\partial\delta}{\sigma_t^2(\lambda)}\frac{\partial\varepsilon_t(\gamma)/\partial\gamma^T}{\sigma_t(\lambda)}\,. \end{aligned}\tag{2.4.20}$$

Lemma 2.4.17. *Under Assumptions 1, 3 and 5, if* $\mathbf{E}|\varepsilon_t|^{2(1-\iota_2)}<\infty$ *(or* $\mathbf{E}|\varepsilon_t|^s<\infty$ *for some* $s>0$ *if* $\iota_1=0$*) and* $\mathbf{E}|\eta_t|^{\max(1,2\iota_1)}\leq\infty$ *, then*

$$\mathbf{E}\sup_{\delta\in\Theta_\delta}\left|\frac{\partial^2 l_t(\gamma_0,\delta)}{\partial\delta\partial\delta^T}\right|<\infty.$$

if η_t is symmetric about zero and $\mathbf{E}g'(\eta_t)\eta_t = 0$ *then*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left|\frac{\partial l_t(\gamma_0,\delta_0)}{\partial\delta} - \frac{\partial l_t(\hat{\gamma}_n,\delta_0)}{\partial\delta}\right| = o_p(1),$$

Proof: From (2.4.19), by Lemmas 2.4.6 and 2.4.7, we have:

$$\mathbf{E}\sup_{\delta\in\Theta_\delta}\left|\frac{\partial^2 l_t(\gamma_0,\delta)}{\partial\delta\partial\delta^T}\right| \le \mathbf{E}(1+|\eta_t|^{2\iota_1}\xi_{\rho,t-1}^{2\iota_1(1-\iota_2)})\xi_{\rho,t-1}^{\zeta} \le \mathbf{E}\xi_{\rho,t-1}^{\zeta}+\mathbf{E}|\eta_t|^{2\iota_1}\mathbf{E}\xi_{\rho,t-1}^{2\iota_1(1-\iota_2)+\zeta} < \infty.$$

Since η_t is symmetric. From (2.4.20), by Remark 2.2.3 and (2.4.16), we have

$$\mathbf{E}\frac{\partial^2 l_t(\lambda_0)}{\partial\delta\partial\gamma^T} = 0\ .$$

Note that $\hat{\gamma}_n$ is $\sqrt{n}$ consistent, by mean value theorem and Lemma 2.4.12, we have:

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\frac{\partial l_t(\gamma_0,\delta_0)}{\partial\delta} - \frac{\partial l_t(\hat{\gamma}_n,\delta_0)}{\partial\delta}\right\} &= \frac{\hat{\gamma}_n-\gamma_0}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial^2 l_t(\gamma^*,\delta_0)}{\partial\delta\partial\gamma^T} \\ &= \sqrt{n}(\hat{\gamma}_n-\gamma_0)\left(\mathbf{E}\frac{\partial^2 l_t(\lambda_0)}{\partial\delta\partial\gamma}+o(1)\right) = o_p(1)\end{aligned}$$

This completes the proof. END

Lemma 2.4.18. *Under Assumptions 1, 3, 4 and 6, Assumption 5 with $w = 1$, if* $\mathbf{E}\varepsilon_t^{2(1-\iota_2)} < \infty$ *and* $\mathbf{E}|\eta_t|^{2\iota_1} \le \infty$ *(or* $\mathbf{E}|\varepsilon_t|^s < \infty$ *and* $\mathbf{E}|\eta_t|^s < \infty$ *for some $s > 0$ if $\iota_1 = 0$), then* $n^{-1/2}\sum_{t=1}^{n}\partial l_t(\delta_0)/\partial\delta \to N(0,\Sigma_3)$*, where Σ_3 has the form in Theorem 2.2.5 .*

Proof: This lemma can be proved with the same arguments as that in the proof of Lemma 2.4.13. END

2.5 APPENDIX

A.1 Proof of Proposition 2.2.1

To show $\mathbf{E}\log[wh(w\eta_t + v)] < \mathbf{E}\log h(\eta_t)$ for any $w \neq 1$ and $v \neq 0$, it is enough to show the partial derivatives of $\mathbf{E}\log[wh(w\eta_t + v)]$ with respect to w and v

$$\frac{1}{w} + \mathbf{E}\frac{\partial h(w\eta_t + v)/\partial w}{h(w\eta_t + v)} = \frac{1}{w} + \mathbf{E}\big(\eta_t g(w\eta_t + v)\big) = 0, \qquad \text{(A.1.1)}$$

$$\frac{\partial h(w\eta_t + v)/\partial v}{h(w\eta_t + v)} = \mathbf{E}g(w\eta_t + v) = 0, \qquad \text{(A.1.2)}$$

have a unique solution at $w = 1$ and $v = 0$.

Since g is odd and η_t is symmetric about zero, it is clear $\mathbf{E}g(\eta_t) = 0$. Obviously $w = 1, v = 0$ is a solution of (A.1.1) and (A.1.2). We first show that for any given $w > 0$, (A.1.2) has a unique solution at $v = 0$. This is equivalent to show that for any $v \neq 0$, $\mathbf{E}g(w\eta_t + v) \neq 0$. Let $f(x)$ be the probability density function of η_t.

Since g is odd, we have

$$\begin{aligned}
\mathbf{E}g(w\eta_t + v) &= \int_{-\infty}^{\infty} g(wx + v) f(x) dx \\
&= \frac{1}{w}\int_{-\infty}^{\infty} g(x) f\left(\frac{x - v}{w}\right) dx \\
&= \frac{1}{w}\int_{-\infty}^{0} g(x) f\left(\frac{x - v}{w}\right) dx + \frac{1}{w}\int_{0}^{\infty} g(x) f\left(\frac{x - v}{w}\right) dx \\
&= \frac{-1}{w}\int_{\infty}^{0} g(-x) f\left(\frac{-x - v}{w}\right) dx + \frac{1}{w}\int_{0}^{\infty} g(x) f\left(\frac{x - v}{w}\right) dx \\
&= \frac{1}{w}\int_{0}^{\infty} g(x) \left\{ f\left(\frac{x - v}{w}\right) - f\left(\frac{-x - v}{w}\right) \right\} dx\,.
\end{aligned}$$

For part (i), since $\mathbf{E}\eta_t^2 < \infty$, then $(1 - F(x))x^2 \to 0$ as $x \to \infty$. Together with the given conditions $g(0) = 0$ and $|g(x)x| \leq Kx^2$, we have

$$\begin{aligned}
\mathbf{E}g(w\eta_t + v) &= g(x)\left\{F\left(\frac{x-v}{w}\right) - F\left(\frac{x+v}{w}\right)\right\}\Bigg|_0^\infty \\
&\quad -\frac{1}{w}\int_0^\infty g'(x)\left\{F\left(\frac{x-v}{w}\right) - F\left(\frac{x+v}{w}\right)\right\}dx \\
&= \lim_{x\to\infty} g(x)\left\{1 - F\left(\frac{x+v}{w}\right)\right\} - \lim_{x\to\infty} g(x)\left\{1 - F\left(\frac{x-v}{w}\right)\right\} - 0 \\
&\quad -\frac{1}{w}\int_0^\infty g'(x)\left\{F\left(\frac{x-v}{w}\right) - F\left(\frac{x+v}{w}\right)\right\}dx \\
&= -\frac{1}{w}\int_0^\infty g'(x)\left\{F\left(\frac{x-v}{w}\right) - F\left(\frac{x+v}{w}\right)\right\}dx\ ,
\end{aligned}$$

which is negative if $v > 0$ or positive if $v < 0$. Hence for part (i), it is left to prove $\mathbf{E}\big(w\eta_t g(w\eta_t)\big) = -1$ if and only if $w = 1$. By using integration by parts again and since $(1 - F(x))x^2 \to 0$, we have

$$\begin{aligned}
&\mathbf{E}\big(w\eta_t g(w\eta_t)\big) - \mathbf{E}\big(\eta_t g(\eta_t)\big) \\
&= \int_{-\infty}^\infty \frac{1}{w} x g(x) f\left(\frac{x}{w}\right)dx - \int_{-\infty}^\infty x g(x) f(x) dx \\
&= 2\int_0^\infty x g(x) d\left(F\left(\frac{x}{w}\right) - F(x)\right) \\
&= 2xg(x)\left(F\left(\frac{x}{w}\right) - F(x)\right)\Bigg|_0^\infty - 2\int_0^\infty (g(x) + xg'(x))\left(F\left(\frac{x}{w}\right) - F(x)\right)dx \\
&= -2\int_0^\infty (g(x) + xg'(x))\left(F\left(\frac{x}{w}\right) - F(x)\right)dx,
\end{aligned}$$

which is zero if and only if $w = 1$, since $g'(x) \leq 0$ but not always equals 0 for $x \geq 0$. Thus $\mathbf{E}\big(w\eta_t g(w\eta_t)\big) = -1$ if and only $w = 1$.

Furthermore, it is easy to check that (A.1.1) and (A.1.2) cannot be satisfied when $w \to \infty$ or (and) $v \to \infty$. This completes the proof of part (i).

For part (ii), since $f(x)$ is even and decreasing for $x > 0$, $g(x) \leq 0$ but not always

0 for $x > 0$, we have

$$\mathbf{E}g(wX+v) = \frac{1}{w}\int_0^\infty g(x)\left\{f\left(\frac{x-v}{w}\right) - f\left(\frac{x+v}{w}\right)\right\}dx,$$

which is negative if $v > 0$, since $f((x-v)/w) - f((x+v)/w) \geq 0$ for any $x > 0$, or is positive if $v < 0$, since $f((x-v)/w) - f((x+v)/w) \leq 0$ for any $x > 0$. Thus $\mathbf{E}g(w\eta_t + v) = 0$ if and only if $v = 0$.

Since $g(x)x$ is a strictly monotone function,

$$\mathbf{E}\big(w\eta_t g(w\eta_t)\big) - \mathbf{E}\big(\eta_t g(\eta_t)\big) = 2\int_0^\infty (wxg(wx) - xg(x))f(x)dx,$$

which is 0 if and only if $w = 1$.

It is easy to check that (A.1.1) and (A.1.2) can not be satisfied when $w \to \infty$ or (and) $v \to \infty$. This completes the proof part (i) and (ii).

The proofs of part (iii) and part (iv) are the same as the proof of part (ii). END

A.2 Expressions of $a_\gamma(i)$, $a_\delta(i)$, $\tilde{\varepsilon}_t(\gamma)$, $\tilde{\sigma}_t(\lambda)$

By comparing the coefficients of z^i on both sides of $\mathcal{A}_\phi(z) = \mathcal{B}_\varphi(z)\sum_{i=0}^{\infty} a_\gamma(i)z^i$, we have if $P < Q$:

$$
\begin{aligned}
a_\gamma(0) &= 1, \\
a_\gamma(1) &= -\phi_1 - \varphi_1 a_\gamma(0), \\
a_\gamma(2) &= -\phi_2 - \varphi_1 a_\gamma(1) - \varphi_2, \\
&\vdots \\
a_\gamma(P) &= -\phi_P - \varphi_1 a_\gamma(P-1) - \cdots - \varphi_{P-1} a_\gamma(1) - \varphi_P, \\
a_\gamma(P+1) &= -\varphi_1 a_\gamma(P) - \cdots - \varphi_P a_\gamma(1) - \varphi_{P+1}, \\
&\vdots \\
a_\gamma(Q) &= -\varphi_1 a_\gamma(Q-1) - \cdots - \varphi_{Q-1} a_\gamma(1) - \varphi_Q, \\
a_\gamma(Q+1) &= -\varphi_1 a_\gamma(Q) - \cdots - \varphi_Q a_\gamma(1),
\end{aligned}
$$

if $P \geq Q$,

$$
\begin{aligned}
a_\gamma(0) &= 1, \\
a_\gamma(1) &= -\phi_1 - \varphi_1 a_\gamma(0), \\
a_\gamma(2) &= -\phi_2 - \varphi_1 a_\gamma(1) - \varphi_2, \vdots
\end{aligned}
$$

$$
\begin{aligned}
a_\gamma(Q) &= -\phi_Q - \varphi_1 a_\gamma(Q-1) - \cdots - \varphi_{Q-1} a_\gamma(1) - \varphi_Q, \\
&\vdots \\
a_\gamma(P) &= -\phi_P - \varphi_1 a_\gamma(P-1) - \cdots - \varphi_{P-1} a_\gamma(1), \\
a_\gamma(P+1) &= -\varphi_1 a_\gamma(P) - \cdots - \varphi_Q a_\gamma(P+1-Q),
\end{aligned}
$$

for $i > \max(P, Q)$,

$$a_\gamma(i) = -\varphi_1 a_\gamma(i-1) - \cdots - \varphi_Q a_\gamma(i-Q).$$

Similar with the calculation of $a_\gamma(i)$, by comparing the coefficients of z^i on both sides of $1 = \mathcal{B}_\varphi(z) \sum_{i=0}^{\infty} a_\varphi(i) z^i$, we can obtain the expression of $a_\varphi(i)$:

$$
\begin{aligned}
a_\varphi(0) &= 1, \\
a_\varphi(1) &= -\varphi_1, \\
a_\varphi(2) &= -\varphi_1 a_\varphi(1) - \varphi_2, \\
&\vdots \\
a_\varphi(Q) &= -\varphi_1 a_\varphi(Q-1) - \varphi_2 a_\varphi(Q-2) - \cdots - \varphi_{Q-1} a_\varphi(1) - \varphi_Q,
\end{aligned}
$$

for $i > Q$,

$$a_\varphi(i) = -\varphi_1 a_\varphi(i-1) - \varphi_2 a_\varphi(Q-2) - \cdots - \varphi_Q a_\varphi(i-Q).$$

Let $a_\varphi(i) = 0$ for $i < 0$. Then based on the expressions of $a_\varphi(i)$, for $i > 0$ and $1 \leq j \leq Q$, we have:

$$\frac{\partial a_\varphi(i)}{\partial \varphi_j} = -\sum_{k=0}^{\infty} a_\varphi(k) a_\varphi(i-j-k) \; . \tag{A.2.1}$$

Similarly, by comparing the coefficients of z^i on both sides of

$$\mathcal{A}_\alpha(z) = \mathcal{B}_\beta(z) \sum_{i=0}^{\infty} a_\delta(i) z^i,$$

we can write down the expressions of $a_\delta(i)$ in terms of $\alpha_1, \cdots, \alpha_p$ and $\beta_1, \cdots, \beta_q$. If $p < q$, we have:

$$\begin{aligned}
a_\delta(0) &= 0, \\
a_\delta(1) &= \alpha_1, \\
a_\delta(2) &= \alpha_2 + \beta_1 a_\delta(1), \\
&\vdots \\
a_\delta(p) &= \alpha_p + \beta_1 a_\delta(p-1) + \cdots + \beta_{p-1} a_\delta(1), \\
a_\delta(p+1) &= \beta_1 a_\delta(p) + \beta_2 a_\delta(p-1) + \cdots + \beta_p a_\delta(1), \\
&\vdots \\
a_\delta(q) &= \beta_1 a_\delta(q-1) + \beta_2 a_\delta(q-2) + \cdots + \beta_{q-1} a_\delta(1), \\
a_\delta(q+1) &= \beta_1 a_\delta(q) + \beta_2 a_\delta(q-1) + \cdots + \beta_q a_\delta(1),
\end{aligned}$$

if $p \geq q$,

$$\begin{aligned} a_\delta(0) &= 0, \\ a_\delta(1) &= \alpha_1, \\ a_\delta(2) &= \alpha_2 + \beta_1 a_\delta(1), \\ &\vdots \end{aligned}$$

$$\begin{aligned} a_\delta(q) &= \alpha_q + \beta_1 a_\delta(q-1) + \cdots + \beta_{q-1} a_\delta(1), \\ &\vdots \\ a_\delta(p) &= \alpha_p + \beta_1 a_\delta(p-1) + \beta_2 a_\delta(p-2) + \cdots + \beta_q a_\delta(p-q), \\ a_\delta(p+1) &= \beta_1 a_\delta(p) + \beta_2 a_\delta(p-1) + \cdots + \beta_q a_\delta(p+1-q), \end{aligned}$$

for $i > \max(p, q)$,

$$a_\delta(i) = \beta_1 a_\delta(i-1) + \beta_2 a_\delta(i-2) + \cdots + \beta_q a_\delta(i-q).$$

Since α_i for $1 \leq i \leq p$ and β_j for $1 \leq j \leq q$ are non-negative, by the expressions of $a_\delta(l)$, we have $a_\delta(l) \geq 0$ for $l \geq 0$.

By the expressions of $a_\gamma(i)$, we have, if $P < Q$,

$$
\begin{aligned}
\tilde{\varepsilon}_1(\gamma) &= (Y_1 - c) - \sum_{i=1}^{P} \phi_i (Y_{1-i} - c) - \sum_{i=1}^{Q} \varphi_{i-1}\tilde{\varepsilon}_{1-i} = a_\gamma(0)(Y_1 - c) + R_{\varepsilon,\gamma,1} \ , \\
\tilde{\varepsilon}_2(\gamma) &= (Y_2 - c) - \phi_1(Y_1 - c) - \sum_{i=2}^{P} \phi_i (Y_{2-i} - c) - \varphi_1 \tilde{\varepsilon}_1(\gamma) - \sum_{i=2}^{Q} \varphi_i \tilde{\varepsilon}_{2-i} \\
&= (Y_2 - c) + (-\phi_1 - \varphi_1)(Y_1 - c) + R_{\varepsilon,\gamma,2} = \sum_{i=0}^{1} a_\gamma(i)(Y_{2-i} - c) + R_{\varepsilon,\gamma,2} \ , \\
&\vdots \\
\tilde{\varepsilon}_{P+1}(\gamma) &= (Y_{P+1} - c) - \sum_{i=1}^{P} \phi_i (Y_{P+1-i} - c) - \sum_{i=1}^{P} \varphi_i \tilde{\varepsilon}_{P+1-i}(\gamma) - \sum_{i=P+1}^{Q} \varphi_i \tilde{\varepsilon}_{P+1-i} \\
&= (Y_{P+1} - c) + (-\phi_1 - \varphi_1)(Y_P - c) + \cdots \\
&\quad + (-\phi_P - \varphi_1 a_\gamma(P-1) - \cdots - \varphi_{P-1} a_\gamma(1) - \varphi_P)(Y_1 - c) + R_{\varepsilon,\gamma,P+1} \\
&= \sum_{i=0}^{P} a_\gamma(i)(Y_{P+1-i} - c) + R_{\varepsilon,\gamma,P+1} \ , \\
&\vdots \\
\tilde{\varepsilon}_Q(\gamma) &= (Y_{P+1} - c) - \sum_{i=1}^{P} \phi_i (Y_{P+1-i} - c) - \sum_{i=1}^{Q-1} \varphi_i \tilde{\varepsilon}_{Q-i}(\gamma) - \varphi_Q \tilde{\varepsilon}_0 \\
&= \sum_{i=0}^{Q-1} a_\gamma(i)(Y_{Q-i} - c) + R_{\varepsilon,\gamma,Q} \ ,
\end{aligned}
$$

If $P \geq Q$, we have

$$\begin{aligned}
\tilde{\varepsilon}_1(\gamma) &= (Y_1 - c) - \sum_{i=1}^{P} \phi_i (Y_{1-i} - c) - \sum_{i=1}^{Q} \varphi_{i-1} \tilde{\varepsilon}_{1-i} = a_\gamma(0)(Y_1 - c) + R_{\varepsilon,\gamma,1} \ , \\
\tilde{\varepsilon}_2(\gamma) &= (Y_2 - c) - \phi_1(Y_1 - c) - \sum_{i=2}^{P} \phi_i (Y_{2-i} - c) - \varphi_1 \tilde{\varepsilon}_1(\gamma) - \sum_{i=2}^{Q} \varphi_i \tilde{\varepsilon}_{2-i} \\
&= (Y_2 - c) + (-\phi_1 - \varphi_1)(Y_1 - c) + R_{\varepsilon,\gamma,2} = \sum_{i=0}^{1} a_\gamma(i)(Y_{2-i} - c) + R_{\varepsilon,\gamma,2} \ , \\
&\vdots \\
\tilde{\varepsilon}_{Q+1}(\gamma) &= (Y_{Q+1} - c) - \sum_{i=1}^{Q} \phi_i (Y_{Q+1-i} - c) - \sum_{i=Q+1}^{P} \phi_i (Y_{Q+1-i} - c) - \sum_{i=1}^{Q} \varphi_i \tilde{\varepsilon}_{Q+1-i}(\gamma) \\
&= (Y_{Q+1} - c) + (-\phi_1 - \varphi_1)(Y_Q - c) + \cdots \\
&\quad + (-\phi_Q - \varphi_1 a_\gamma(Q-1) - \cdots - \varphi_{Q-1} a_\gamma(1) - \varphi_Q)(Y_1 - c) + R_{\varepsilon,\gamma,Q+1} \\
&= \sum_{i=0}^{Q} a_\gamma(i)(Y_{Q+1-i} - c) + R_{\varepsilon,\gamma,Q+1} \ , \\
&\vdots \\
\tilde{\varepsilon}_P(\gamma) &= (Y_P - c) - \sum_{i=1}^{P-1} \phi_i (Y_{P-i} - c) - \phi_P(Y_0 - c) - \sum_{i=1}^{Q} \varphi_i \tilde{\varepsilon}_{P-i}(\gamma) \\
&= \sum_{i=0}^{P-1} a_\gamma(i)(Y_{P-i} - c) + R_{\varepsilon,\gamma,P} \quad ,
\end{aligned}$$

for $j > \max(P, Q)$,

$$\tilde{\varepsilon}_j(\gamma) = (Y_j - c) - \sum_{i=1}^{P} \phi_i (Y_{j-i} - c) - \sum_{i=1}^{Q} \varphi_i \tilde{\varepsilon}_{j-i}(\gamma) = \sum_{i=0}^{j-1} a_\gamma(i)(Y_{j-i} - c) + R_{\varepsilon,\gamma,j} \ ,$$

where

$$R_{\varepsilon,\gamma,j} = -\sum_{i=1}^{Q} \varphi_i R_{\varepsilon,\gamma,j-i}, \text{ for } \ j \geq Q \ ,$$

which is a recurrence function of φ_i for $1 \le i \le q$. By Assumption 2, and the property of recurrence sequence, we have $\sup_{\gamma\in\Theta_\gamma} |R_{\varepsilon,\gamma,j}| = O(\rho^j)$ a.s. for $0 < \rho < 1$, $j \ge 1$.

Thus

$$\tilde{\varepsilon}_t(\gamma) = (Y_t - c) - \sum_{i=1}^{P} \phi_i (Y_{t-i} - c) - \sum_{i=1}^{Q} \varphi_i \tilde{\varepsilon}_{t-i}(\gamma) = \sum_{i=0}^{t-1} a_\gamma(i)(Y_{t-i} - c) + O(\rho^t)\,. \quad (A.2.2)$$

By the expressions of $a_\delta(i)$, we have, if $p \le q$,

$$\begin{aligned}
\tilde{\sigma}_1^2(\lambda) &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \tilde{\varepsilon}_{1-i}^2 + \sum_{j=1}^{q} \beta_j \tilde{\sigma}_{1-j}^2 = R_{\sigma,\delta,1} \\
\tilde{\sigma}_2^2(\lambda) &= \alpha_0 + \alpha_1 \tilde{\varepsilon}_1^2(\gamma) + \sum_{i=2}^{p} \alpha_i \tilde{\varepsilon}_{2-i}^2 + \beta_1 \tilde{\sigma}_1^2(\lambda) + \sum_{j=2}^{q} \beta_j \tilde{\sigma}_{2-j}^2 = \alpha_1 \tilde{\varepsilon}_1^2(\gamma) + R_{\sigma,\delta,2} \\
&= a_\delta(1)\tilde{\varepsilon}_1^2(\gamma) + R_{\sigma,\delta,2} = \sum_{i=0}^{1} a_\delta(i)\tilde{\varepsilon}_{2-i}^2(\gamma) + R_{\sigma,\delta,2} \\
\tilde{\sigma}_3^2(\lambda) &= \alpha_0 + \alpha_1 \tilde{\varepsilon}_2^2(\gamma) + \alpha_2 \tilde{\varepsilon}_1^2(\gamma) + \sum_{i=3}^{p} \alpha_i \tilde{\varepsilon}_{3-i}^2 + \beta_1 \tilde{\sigma}_2^2(\lambda) + \beta_2 \tilde{\sigma}_1^2(\lambda) + \sum_{j=3}^{q} \beta_j \tilde{\sigma}_{3-j}^2 \\
&= \alpha_1 \tilde{\varepsilon}_2^2(\gamma) + (\alpha_2 + \beta_1 a_\delta(1))\tilde{\varepsilon}_1^2(\gamma) + R_{\sigma,\delta,2} \\
&= \sum_{i=0}^{2} a_\delta(i)\tilde{\varepsilon}_{3-i}^2(\gamma) + R_{\sigma,\delta,3} \\
&\vdots \\
\tilde{\sigma}_{p+1}^2(\lambda) &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \tilde{\varepsilon}_{p+1-i}^2(\gamma) + \sum_{j=1}^{p} \beta_j \tilde{\sigma}_{p+1-j}^2(\lambda) + \sum_{j=p+1}^{q} \beta_j \tilde{\sigma}_{p+1-j}^2 \\
&= \alpha_1 \tilde{\varepsilon}_p^2(\gamma) + (\alpha_2 + \beta_1 a_\delta(1))\tilde{\varepsilon}_{p-1}^2(\gamma) + \cdots + (\alpha_p + \beta_1 a_\delta(p-1) + \cdots \\
&\quad + \beta_{p-1} a_\delta(1))\tilde{\varepsilon}_1^2(\gamma) + R_{\sigma,\delta,p+1} \\
&= \sum_{i=0}^{p} a_\delta(i)\tilde{\varepsilon}_{p+1-i}^2(\gamma) + R_{\sigma,\delta,p+1} \\
&\vdots
\end{aligned}$$

$$\begin{aligned}
\tilde{\sigma}_q^2(\lambda) &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \tilde{\varepsilon}_{q-i}^2(\gamma) + \sum_{j=1}^{q-1} \beta_j \tilde{\sigma}_{q-j}^2(\lambda) + \beta_q \tilde{\sigma}_0^2 \\
&= \alpha_1 \tilde{\varepsilon}_p^2(\gamma) + (\alpha_2 + \beta_1 a_\delta(1)) \tilde{\varepsilon}_{p-1}^2(\gamma) + \cdots \\
&\quad + (\beta_1 a_\delta(q-1) + \beta_2 a_\delta(q-2) + \cdots + \beta_{q-1} a_\delta(1)) \tilde{\varepsilon}_1^2(\gamma) + R_{\sigma,\delta,p+1} \\
&= \sum_{i=0}^{q-1} a_\delta(i) \tilde{\varepsilon}_{q-i}^2(\gamma) + R_{\sigma,\delta,q}
\end{aligned}$$

if $p > q$,

$$\begin{aligned}
\tilde{\sigma}_1^2(\lambda) &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \tilde{\varepsilon}_{1-i}^2 + \sum_{j=1}^{q} \beta_j \tilde{\sigma}_{1-j}^2 = R_{\sigma,\delta,1} \\
\tilde{\sigma}_2^2(\lambda) &= \alpha_0 + \alpha_1 \tilde{\varepsilon}_1^2(\gamma) + \sum_{i=2}^{p} \alpha_i \tilde{\varepsilon}_{2-i}^2 + \beta_1 \tilde{\sigma}_1^2(\lambda) + \sum_{j=2}^{q} \beta_j \tilde{\sigma}_{2-j}^2 = \alpha_1 \tilde{\varepsilon}_1^2(\gamma) + R_{\sigma,\delta,2} \\
&= a_\delta(1) \tilde{\varepsilon}_1^2(\gamma) + R_{\sigma,\delta,2} = \sum_{i=0}^{1} a_\delta(i) \tilde{\varepsilon}_{2-i}^2(\gamma) + R_{\sigma,\delta,2} \\
\tilde{\sigma}_3^2(\lambda) &= \alpha_0 + \alpha_1 \tilde{\varepsilon}_2^2(\gamma) + \alpha_2 \tilde{\varepsilon}_1^2(\gamma) + \sum_{i=3}^{p} \alpha_i \tilde{\varepsilon}_{3-i}^2 + \beta_1 \tilde{\sigma}_2^2(\lambda) + \beta_2 \tilde{\sigma}_1^2(\lambda) + \sum_{j=3}^{q} \beta_j \tilde{\sigma}_{3-j}^2 \\
&= \alpha_1 \tilde{\varepsilon}_2^2(\gamma) + (\alpha_2 + \beta_1 a_\delta(1)) \tilde{\varepsilon}_1^2(\gamma) + R_{\sigma,\delta,3} \\
&= \sum_{i=0}^{2} a_\delta(i) \tilde{\varepsilon}_{3-i}^2(\gamma) + R_{\sigma,\delta,3}
\end{aligned}$$

$$\vdots$$

$$\begin{aligned}
\tilde{\sigma}_{q+1}^2(\lambda) &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \tilde{\varepsilon}_{q+1-i}^2(\gamma) + \sum_{i=q+1}^{p} \alpha_i \tilde{\varepsilon}_{q+1-i}^2 + \sum_{j=1}^{q} \beta_j \tilde{\sigma}_{q+1-j}^2(\lambda) \\
&= \alpha_1 \tilde{\varepsilon}_q^2(\gamma) + (\alpha_2 + \beta_1 a_\delta(1)) \tilde{\varepsilon}_{q-1}^2(\gamma) + \cdots + (\alpha_q + \beta_1 a_\delta(q-1) + \cdots \\
&\quad + \beta_{q-1} a_\delta(1)) \tilde{\varepsilon}_1^2(\gamma) + R_{\sigma,\delta,q+1} \\
&= \sum_{i=0}^{q} a_\delta(i) \tilde{\varepsilon}_{q+1-i}^2(\gamma) + R_{\sigma,\delta,q+1}
\end{aligned}$$

$$\vdots$$

$$\begin{aligned}
\tilde{\sigma}_p^2(\lambda) &= \alpha_0 + \sum_{i=1}^{p-1} \alpha_i \tilde{\varepsilon}_{p-i}^2(\gamma) + \alpha_p \tilde{\varepsilon}_0^2 + \sum_{j=1}^{q} \beta_j \tilde{\sigma}_{p-j}^2(\lambda) \\
&= \alpha_1 \tilde{\varepsilon}_{p-1}^2(\gamma) + (\alpha_2 + \beta_1 a_\delta(1))\tilde{\varepsilon}_{p-2}^2(\gamma) + \cdots \\
&\quad + (\alpha_{p-1} + \beta_1 a_\delta(p-2) + \beta_2 a_\delta(p-3) + \cdots + \beta_q a_\delta(p-q-1))\tilde{\varepsilon}_1^2(\gamma) + R_{\sigma,\delta,p} \\
&= \sum_{i=0}^{p-1} a_\delta(i)\tilde{\varepsilon}_{p-i}^2(\gamma) + R_{\sigma,\delta,p} \ ,
\end{aligned}$$

for $i > \max(p,q)$,

$$\begin{aligned}
\tilde{\sigma}_i^2(\lambda) &= \alpha_0 + \sum_{j=1}^{p} \alpha_j \tilde{\varepsilon}_{i-j}^2(\gamma) + \sum_{j=1}^{q} \beta_j \tilde{\sigma}_{i-j}^2(\lambda) \\
&= \alpha_1 \tilde{\varepsilon}_{i-1}^2(\gamma) + (\alpha_2 + \beta_1 a_\delta(1))\tilde{\varepsilon}_{i-2}^2(\gamma) + \cdots \\
&\quad + (\beta_1 a_\delta(i-2) + \beta_2 a_\delta(i-3) + \cdots + \beta_q a_\delta(i-q-1))\tilde{\varepsilon}_1^2(\gamma) + R_{\sigma,\delta,i} \\
&= \sum_{j=0}^{i-1} a_\delta(j)\tilde{\varepsilon}_{i-j}^2(\gamma) + R_{\sigma,\delta,i} \ ,
\end{aligned}$$

where

$$R_{\sigma,\delta,i} = \sum_{j=1}^{q} \varphi_j R_{\sigma,\delta,i-j}, \text{ for } \ i \geq q,$$

which is a recurrence function of β_j for $1 \leq j \leq q$. By Assumption 3, and property of recurrence sequence, we have $\sup_{\delta\in\Theta_\delta} |R_{\sigma,\delta,i}| = O(\rho^i)$ a.s. for $0 < \rho < 1$, $i \geq 1$.

Thus

$$\tilde{\sigma}_t^2(\lambda) = \sum_{i=0}^{t-1} a_\delta(i)\tilde{\varepsilon}_{t-i}^2(\gamma) + O(\rho^t) \ . \tag{A.2.3}$$

A.3 Expressions of first and second derivatives of $\varepsilon_t(\gamma)$ and $\sigma_t^2(\lambda)$

By (2.2.1), we have the first derivatives of $\varepsilon_t(\gamma)$:

$$
\begin{aligned}
\frac{\partial \varepsilon_t(\gamma)}{\partial c} &= -\mathcal{C}_\gamma(1) = -\left(1-\sum_{j=1}^{p}\phi_j\right)\sum_{i=0}^{\infty} a_\varphi(i), & \text{(A.3.1)}\\
\frac{\partial \varepsilon_t(\gamma)}{\partial \phi_i} &= -\mathcal{B}_\gamma^{-1}(L)(Y_{t-i}-c) = -\sum_{j=0}^{\infty} a_\varphi(j)(Y_{t-i-j}-c) & \text{(A.3.2)}\\
&= -\mathcal{A}_\gamma^{-1}(L)\varepsilon_{t-i}(\gamma) = -\sum_{j=0}^{\infty} a_\phi(j)\varepsilon_{t-i-j}(\gamma), \quad 1\le i\le P, &\\
\frac{\partial \varepsilon_t(\gamma)}{\partial \varphi_j} &= -\mathcal{B}_\gamma^{-1}(L)\varepsilon_{t-j}(\gamma) = -\sum_{i=0}^{\infty} a_\varphi(i)\varepsilon_{t-i-j}(\gamma), \quad 1\le j\le Q, & \text{(A.3.3)}\\
\frac{\partial \varepsilon_t(\gamma)}{\partial \delta} &= 0\,. & \text{(A.3.4)}
\end{aligned}
$$

By (2.4.2), we have the first derivatives of $\sigma_t^2(\lambda)$:

$$
\begin{aligned}
\frac{\partial \underline{\sigma}_t^2(\lambda)}{\partial \alpha_0} &= \sum_{k=0}^{\infty} B^k \underline{1}, & \text{(A.3.5)}\\
\frac{\partial \underline{\sigma}_t^2(\lambda)}{\partial \alpha_i} &= \sum_{k=0}^{\infty} B^k \underline{\varepsilon}_{t-i-k}^2(\gamma), \quad 1\le i\le p, & \text{(A.3.6)}\\
\frac{\partial \underline{\sigma}_t^2(\lambda)}{\partial \beta_i} &= \sum_{k=1}^{\infty}\left(\sum_{l=1}^{k} B^{l-1}B^{(i)}B^{k-l}\right)\underline{c}_{t-k}(\lambda), \quad 1\le i\le q, & \text{(A.3.7)}\\
\frac{\partial \sigma_t^2(\lambda)}{\partial \gamma_j} &= \sum_{k=0}^{\infty} B^k(1,1)\sum_{i=1}^{p} 2\alpha_i\varepsilon_{t-k-i}(\gamma)\frac{\partial \varepsilon_{t-k-i}(\gamma)}{\partial \gamma_j}, \quad 1\le j\le P+Q+1 & \text{(A.3.8)}
\end{aligned}
$$

where $\underline{1} = (1,0,...,0)^T_{q\times 1}$, $\underline{\varepsilon}_t(\gamma) = (\varepsilon_t(\gamma),0,...,0)^T_{q\times 1}$, $B^{(i)}$ is a $q\times q$ matrix with $(1,i)th$ element 1 and all other elements 0.

Based on (A.3.1)-(A.3.4),we have the second derivatives of $\varepsilon_t(\gamma)$:

$$\frac{\partial^2 \varepsilon_t(\gamma)}{\partial c \partial c} = 0, \tag{A.3.9}$$

$$\frac{\partial^2 \varepsilon_t(\gamma)}{\partial c \partial \phi_i} = \mathcal{B}_\gamma^{-1}(1) = \sum_{k=0}^{\infty} a_\varphi(k), \tag{A.3.10}$$

$$\frac{\partial^2 \varepsilon_t(\gamma)}{\partial c \partial \varphi_i} = \mathcal{A}_\gamma(1)\mathcal{B}_\gamma^{-2}(1) = -\left(1 - \sum_{j=1}^{p} \phi_j\right) \sum_{k=0}^{\infty} \frac{\partial a_\varphi(k)}{\partial \varphi_i}, \tag{A.3.11}$$

$$\frac{\partial^2 \varepsilon_t(\gamma)}{\partial \phi_i \partial \phi_j} = 0, \quad 1 \le i, j \le P, \tag{A.3.12}$$

$$\begin{aligned} \frac{\partial^2 \varepsilon_t(\gamma)}{\partial \phi_i \partial \varphi_j} &= \mathcal{A}_\gamma^{-1}(L)\mathcal{B}_\gamma^{-1}(L)\varepsilon_{t-i-j}(\gamma) \\ &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} a_\phi(k) a_\varphi(l) \varepsilon_{t-i-j-l-k}(\gamma), \quad 1 \le i \le P, 1 \le j \le Q, \end{aligned} \tag{A.3.13}$$

$$\begin{aligned} \frac{\partial^2 \varepsilon_t(\gamma)}{\partial \varphi_i \partial \varphi_j} &= -\mathcal{B}_\gamma^{-2}(L)\big(\varepsilon_{t-i-j}(\gamma) + \varepsilon_{t-j}(\gamma)\big) \\ &= -\sum_{k=0}^{\infty} \frac{\partial a_\varphi(k)}{\partial \varphi_j} \varepsilon_{t-i-k}(\gamma) - \sum_{k=0}^{\infty} a_\varphi(k) \frac{\partial \varepsilon_{t-i-k}(\gamma)}{\partial \varphi_j}, \quad 1 \le i, j \le Q, \end{aligned} \tag{A.3.14}$$

$$\frac{\partial^2 \varepsilon_t(\gamma)}{\partial \delta \partial \lambda} = 0\,. \tag{A.3.15}$$

Based on (A.3.5)-(A.3.8),we have the second derivatives of $\sigma_t^2(\lambda)$:

$$\frac{\partial^2 \underline{\sigma}_t^2(\lambda)}{\partial \alpha_0 \partial \alpha_0} = \frac{\partial^2 \underline{\sigma}_t^2(\lambda)}{\partial \alpha_0 \partial \alpha_i} = \frac{\partial^2 \underline{\sigma}_t^2(\lambda)}{\partial \alpha_0 \partial \gamma_j} = 0, \quad 1 \le i \le p, \le j \le 1 + P + Q, \tag{A.3.16}$$

$$\frac{\partial^2 \underline{\sigma}_t^2(\lambda)}{\partial \alpha_0 \partial \beta_j} = \sum_{k=1}^{\infty} \left\{ \sum_{l=1}^{k} B^{l-1} B^{(j)} B^{k-l} \right\} \underline{1}, \quad 1 \le j \le q, \tag{A.3.17}$$

$$\frac{\partial^2 \underline{\sigma}_t^2(\lambda)}{\partial \alpha_i \partial \alpha_j} = 0, \quad 1 \le i, j \le p, \tag{A.3.18}$$

$$\begin{aligned} \frac{\partial \underline{\sigma}_t^2(\lambda)}{\partial \alpha_i \partial \beta_j} &= \sum_{k=1}^{\infty} \left(\sum_{l=1}^{k} B^{l-1} B^{(j)} B^{k-l} \right) \underline{\varepsilon}_{t-i-k}^2(\gamma), \\ & \quad 1 \le i \le p, 1 \le j \le q, \end{aligned} \tag{A.3.19}$$

$$\frac{\partial^2 \underline{\sigma}_t^2(\lambda)}{\partial\alpha_i\partial\gamma_j} = 2\sum_{k=0}^{\infty} B^k(1,1)\varepsilon_{t-i-k}(\gamma)\frac{\partial\varepsilon_{t-i-k}(\gamma)}{\partial\gamma_j}, \tag{A.3.20}$$

$$1 \le i \le p, 1 \le j \le P+Q+1,$$

$$\begin{aligned}\frac{\partial^2 \underline{\sigma}_t^2(\lambda)}{\partial\beta_j\partial\beta_{j^*}} &= \sum_{k=2}^{\infty}\left\{\sum_{i=2}^{k}\left[\left(\sum_{l=1}^{i-1} B^{l-1}B^{(j^*)}B^{i-1-l}\right)B^{(j)}B^{k-i}\right]\right. \\ &\left.+\sum_{i=1}^{k-1}\left[B^{i-1}B^{(j)}\left(\sum_{l=1}^{k-i} B^{l-1}B^{(j^*)}B^{k-i-l}\right)\right]\right\}\underline{c}_{t-k}(\lambda), \quad 1 \le j, j^* \le q,\end{aligned} \tag{A.3.21}$$

$$\frac{\partial^2 \sigma_t^2(\lambda)}{\partial\beta_i\partial\gamma_j} = 2\sum_{k=1}^{\infty}\left(\sum_{l=1}^{k} B^{l-1}B^{(i)}B^{k-l}(1,1)\sum_{m=1}^{p}\alpha_m\varepsilon_{t-k-m}(\gamma)\frac{\partial\varepsilon_{t-k-m}(\gamma)}{\partial\gamma_j}\right) \tag{A.3.22}$$

$$1 \le i \le q, 1 \le j \le P+Q+1\ ,$$

$$\begin{aligned}\frac{\partial^2 \sigma_t^2(\lambda)}{\partial\gamma_j\partial\gamma_{j^*}} &= 2\sum_{k=0}^{\infty} B^k(1,1)\sum_{i=1}^{p}\alpha_i\left(\frac{\partial\varepsilon_{t-k-i}(\gamma)}{\partial\gamma_j}\frac{\partial\varepsilon_{t-k-i}(\gamma)}{\partial\gamma_{j^*}}\right. \\ &\left.+\varepsilon_{t-k-i}(\gamma)\frac{\partial^2\varepsilon_{t-k-i}(\gamma)}{\partial\gamma_j\partial\gamma_{j^*}}\right), \quad 1 \le j, j^* \le P+Q+1.\end{aligned} \tag{A.3.23}$$

A.4 Modification for pure GARCH with non-zero constant mean

Some modifications for pure GARCH with non-zero constant mean are listed below:

Modifications 1:

When $P = Q = 0$, $c_0 \neq 0$, model (1.2.12)-1.2.13 reduces to pure GARCH with nonzero constant conditional mean. The parameter space reduce to (c, δ) and $\{\varepsilon_t\}$ becomes the observations. Initial values of $\varepsilon_0, \cdots, \varepsilon_{1-p}$, $\tilde{\sigma}_0^2, \cdots, \tilde{\sigma}_{1-q}^2$ are required in calculating $\tilde{\sigma}_t^2$ and (2.4.3) is replaced by:

$$\tilde{\underline{\sigma}}_t^2(\lambda) = \underline{c}_t(c) + B\underline{c}_{t-1}(c) + B^2\underline{c}_{t-2}(c) + \dots + B^{t-p-1}\underline{c}_{p+1}(c) \qquad \text{(A.4.1)}$$
$$+B^{t-p}\tilde{\underline{c}}_p(c) + \dots + B^{t-1}\tilde{\underline{c}}_1(c) + B^t\tilde{\underline{\sigma}}_0^2(\lambda)$$

Based on (A.4.1), we can adjust the derivatives of $\tilde{\sigma}_t^2$.

Modifications 2:

The first and second derivatives of $l_t(\delta, c)$ still have the form of (2.2.12) and (2.2.13). Some derivatives of $\varepsilon_t(c)$ and $\sigma_t(\lambda)$ are simplified as:

$$\frac{\partial \varepsilon_t(c)}{\partial \lambda} = (1, 0, \cdots, 0), \qquad \frac{\partial^2 \varepsilon_t(c)}{\partial \lambda \partial \lambda^T} = \underline{0},$$

$$\frac{\partial \sigma_t^2(\lambda)}{\partial c} = -2\sum_{i=0}^{\infty} B^i(1,1)\sum_{j=1}^{p}\alpha_j(\varepsilon_{t-i-j} - c), \qquad \frac{\partial^2 \sigma_t^2(\lambda)}{\partial c \partial c} = 2\sum_{i=0}^{\infty} B^i(1,1)\sum_{j=1}^{p}\alpha_j,$$

$$\frac{\partial^2 \sigma_t^2(\lambda)}{\partial c \partial \alpha_j} = -2\sum_{i=0}^{\infty} B^i(1,1)(\varepsilon_{t-i-j} - c),$$

$$\frac{\partial^2 \sigma_t^2(\lambda)}{\partial c \partial \beta_k} = -2\sum_{i=0}^{\infty}\left(\sum_{l=1}^{i} B^{l-1}B^{(k)}B^{i-l}(1,1)\sum_{j=1}^{p}\alpha_j(\varepsilon_{t-i-j} - c)\right).$$

Modifications 3:

Based on Modifications 2, we have for any $\lambda \in \Theta_0$:

$$\begin{aligned}
\sup_{\lambda\in\Theta}\left|\frac{\partial^2\sigma_t^2(\lambda)/\partial c\partial\alpha_j}{\sigma_t^2(\lambda)}\right| &= \frac{2}{\alpha_j}\left|\frac{\sum_{i=0}^{\infty} B^i(1,1)\alpha_j(\varepsilon_{t-i-j} - c)}{K + \sum_{i=0}^{\infty} B^i(1,1)\sum_{j=1}^{p}\alpha_j(\varepsilon_{t-i-j} - c)^2}\right| \\
&\leq \sup_{\lambda\in\Theta}\frac{2}{\alpha_j}\left|\frac{\sum_{i=0}^{\infty} B^i(1,1)\alpha_j(\varepsilon_{t-i-j} - c)^2 I|\varepsilon_{t-i-j} - c| \geq 1}{K + \sum_{i=0}^{\infty} B^i(1,1)\alpha_j(\varepsilon_{t-i-j} - c)^2}\right| \\
&\quad + \sup_{\lambda\in\Theta}\frac{2}{\alpha_j}\left|\frac{\sum_{i=0}^{\infty} B^i(1,1)\alpha_j I|\varepsilon_{t-i-j} - c| < 1}{K + \sum_{i=0}^{\infty} B^i(1,1)\alpha_j(\varepsilon_{t-i-j} - c)^2}\right| \\
&\leq K.
\end{aligned}$$

Similarly we can show

$$\sup_{\lambda\in\Theta}\left|\frac{\partial\sigma_t^2(\lambda)/\partial c}{\sigma_t^2(\lambda)}\right|, \quad \sup_{\lambda\in\Theta}\left|\frac{\partial^2\sigma_t^2(\lambda)/\partial c\partial\beta_j}{\sigma_t^2(\lambda)}\right|$$

are bounded by some constants. Then together with Lemma 2.4.6, we have that

$$\sup_{\lambda\in\Theta}\left|\frac{\partial\varepsilon_t(c)/\partial\lambda}{\sigma_t(\lambda)}\right|,\ \sup_{\lambda\in\Theta}\left|\frac{\partial^2\varepsilon_t(c)/\partial\lambda\partial^T\lambda}{\sigma_t(\lambda)}\right|,\ \sup_{\lambda\in\Theta}\left|\frac{\partial\sigma_t^2(\lambda)/\partial\lambda}{\sigma_t^2(\lambda)}\right|,\ \sup_{\lambda\in\Theta}\left|\frac{\partial^2\sigma_t^2(\lambda)/\partial\lambda\partial^T\lambda}{\sigma_t^2(\lambda)}\right|$$

are bounded by $C\xi_{\rho,t-1}^{\zeta}$, which has any moments.

By Lemma 2.4.7, we have $\eta_t(\lambda) = \eta_t\sigma_t(\lambda_0)/\sigma_t(\lambda) + (c_0 - c)/\sigma_t(\lambda)$ is bounded uniformly by $C|1+\eta_t|\xi_{\rho,t-1}^{1-\zeta}$ for any $\lambda \in \Theta_0$. Thus in the proof of Lemma 2.4.12, for pure GARCH with $c \neq 0$, we can relax the moment condition of ε_0 to $2\iota_1(1-\iota_2)$ (or some $s > 0$ if $\iota_1 = 0$).

A.5 Modification for pure ARMA

When $p = q = 0$, ARMA(P,Q)-GARCH(p,q) model reduces to pure ARMA(P,Q) model (1.2.12). Modified assumption 5 implies $g(\varepsilon_t/\sqrt{\alpha_0}) = 0$.

In this reduced model, $\{\varepsilon_t\}$ is a sequence of IID random variables with mean 0 and variance α_0. With α_0 being nuisance parameter, the parameters are reduced to γ. Initial values $Y_0, ..., Y_{1-P}$ are required.

For pure ARMA, $l_t = \log[h(\varepsilon_t(\gamma)/\sqrt{\alpha_0})/\sqrt{\alpha_0}]$. The first and second derivatives of $l_t(\gamma)$ are simplified as:

$$\frac{\partial l_t(\gamma)}{\partial\gamma} = g\left(\frac{\varepsilon_t(\gamma)}{\sqrt{\alpha_0}}\right)\frac{\partial\varepsilon_t(\gamma)/\partial\gamma}{\sqrt{\alpha_0}},$$

$$\frac{\partial^2 l_t(\gamma)}{\partial\gamma\partial\gamma^T} = g'\left(\frac{\varepsilon_t(\gamma)}{\sqrt{\alpha_0}}\right)\frac{\partial\varepsilon_t(\gamma)/\partial\gamma}{\sqrt{\alpha_0}}\frac{\partial\varepsilon_t(\gamma)/\partial\gamma^T}{\sqrt{\alpha_0}} + g\left(\frac{\varepsilon_t(\gamma)}{\sqrt{\alpha_0}}\right)\frac{\partial^2\varepsilon_t(\gamma)/\partial\gamma\partial\gamma^T}{\sqrt{\alpha_0}}.$$

Lemmas 2.4.3 to 2.4.5 and 2.4.8 to 2.4.12 still hold with modified conditions. These Lemmas may have more simple forms for this special case. For instance:

Lemma 2.4.4 is adjusted as: Under Assumption 2, if there exists some t such that $\varepsilon_t(\gamma) = \varepsilon_t(\gamma_0)$ almost surely, then $\gamma = \gamma_0$.

Lemma 2.4.9 is modified as: Under Assumptions 1, 2, 5 to 6, if $\mathbf{E}|\varepsilon_t|^2 < \infty$, then $\mathcal{I}$ is nonsingular.

Thus by modified Assumption 4 and Lemma 2.4.6, for pure ARMA, if $\mathbf{E}|\varepsilon_0|^2 < \infty$, then $\mathbf{E}\partial^2 l_t(\gamma)/\partial\gamma\partial\gamma^T < \infty$ as in Lemma 2.4.12 .

A.6 Expressions of first and second derivatives of $\tilde{\varepsilon}_t(\gamma)$ and $\tilde{\sigma}_t(\lambda)$

Since the initial values are fixed, analogous to (A.3.1)-(A.3.4) and (A.3.9)-(A.3.15), by method of induction, differentiate (2.2.5), we have:

$$\frac{\partial\tilde{\varepsilon}_t(\gamma)}{\partial c} = -\left(1-\sum_{j=1}^{p}\phi_j\right)\sum_{i=0}^{t-1} a_\varphi(i), \tag{A.6.1}$$

$$\begin{aligned}\frac{\partial\tilde{\varepsilon}_t(\gamma)}{\partial \phi_j} &= -\sum_{i=0}^{t-1} a_\varphi(i)(Y_{t-i-j}-c) &\text{(A.6.2)}\\ &= -\sum_{i=0}^{t-1} a_\varphi(i)\sum_{k=0}^{\infty} a_\gamma(k)\varepsilon_{t-i-j-k}(\gamma), \quad 1\le j\le P, \text{if } \ t-i-j>0,\end{aligned}$$

$$\frac{\partial\tilde{\varepsilon}_t(\gamma)}{\partial \varphi_j} = -\sum_{i=0}^{t-1} a_\varphi(i)\tilde{\varepsilon}_{t-i-j}(\gamma), \quad 1\le j\le Q, \tag{A.6.3}$$

$$\frac{\partial\tilde{\varepsilon}_t(\gamma)}{\partial \delta} = 0, \tag{A.6.4}$$

$$\frac{\partial^2 \tilde{\varepsilon}_t(\gamma)}{\partial c \partial c} = 0, \tag{A.6.5}$$

$$\frac{\partial^2 \tilde{\varepsilon}_t(\gamma)}{\partial c \partial \phi_i} = \sum_{k=0}^{t-1} a_\varphi(k), \quad 1 \le i \le P, \tag{A.6.6}$$

$$\frac{\partial^2 \tilde{\varepsilon}_t(\gamma)}{\partial c \partial \varphi_i} = -\left(1 - \sum_{j=1}^{p} \phi_j\right) \sum_{k=0}^{t-1} \frac{\partial a_\varphi(k)}{\partial \varphi_i}, \quad 1 \le i \le Q. \tag{A.6.7}$$

$$\frac{\partial^2 \tilde{\varepsilon}_t(\gamma)}{\partial \phi_i \partial \phi_j} = 0, \quad 1 \le i, j \le P, \tag{A.6.8}$$

$$\begin{aligned} \frac{\partial^2 \tilde{\varepsilon}_t(\gamma)}{\partial \phi_i \partial \varphi_j} &= -\sum_{k=0}^{t-1} \frac{\partial a_\varphi(k)}{\partial \varphi_j}(Y_{t-i-k} - c) \\ &= -\sum_{k=0}^{t-1} \frac{\partial a_\varphi(k)}{\partial \varphi_j} \sum_{l=0}^{\infty} a_\gamma(l) \varepsilon_{t-i-k-l}(\gamma), \text{ if } t-i-k>0, \\ & \quad 1 \le i \le P, 1 \le j \le Q, \end{aligned} \tag{A.6.9}$$

$$\begin{aligned} \frac{\partial^2 \tilde{\varepsilon}_t(\gamma)}{\partial \varphi_i \partial \varphi_j} &= -\sum_{k=0}^{t-1} \frac{\partial a_\varphi(k)}{\partial \varphi_j} \tilde{\varepsilon}_{t-k-i}(\gamma) - \sum_{k=0}^{t-1} a_\varphi(k) \frac{\partial \tilde{\varepsilon}_{t-k-i}(\gamma)}{\partial \varphi_j}, \\ & \quad 1 \le i, j \le Q, \end{aligned} \tag{A.6.10}$$

$$\frac{\partial^2 \tilde{\varepsilon}_t(\gamma)}{\partial \delta \partial \gamma} = 0. \tag{A.6.11}$$

Analogous to (A.3.9)-(A.3.23), differentiate (2.4.3), we have:

$$\frac{\partial \underline{\tilde{\sigma}}_t^2(\lambda)}{\partial \alpha_0} = \sum_{k=0}^{t-1} B^k \underline{1}, \tag{A.6.12}$$

$$\frac{\partial \underline{\tilde{\sigma}}_t^2(\lambda)}{\partial \alpha_i} = \sum_{k=0}^{t-1} B^k \underline{\tilde{\varepsilon}}_{t-i-k}^2(\gamma), \quad 1 \le i \le p, \tag{A.6.13}$$

$$\frac{\partial \underline{\tilde{\sigma}}_t^2(\lambda)}{\partial \beta_i} = \sum_{k=1}^{t-1} \left(\sum_{l=1}^{k} B^{l-1} B^{(i)} B^{k-l} \right) \underline{\tilde{c}}_{t-i-k}(\gamma), \quad 1 \le i \le q, \tag{A.6.14}$$

$$\frac{\partial \tilde{\sigma}_t^2(\lambda)}{\partial \gamma_j} = \sum_{k=0}^{t-1} B^k(1,1) \sum_{i=1}^{p} 2\alpha_i \tilde{\varepsilon}_{t-k-i}(\gamma) \frac{\partial \tilde{\varepsilon}_{t-k-i}(\gamma)}{\partial \gamma_j}, \tag{A.6.15}$$

$$1 \le j \le P+Q+1.$$

$$\frac{\partial^2 \tilde{\sigma}_t^2(\lambda)}{\partial \alpha_0 \partial \alpha_0} = \frac{\partial^2 \tilde{\sigma}_t^2(\lambda)}{\partial \alpha_0 \partial \alpha_i} = \frac{\partial^2 \tilde{\sigma}_t^2(\lambda)}{\partial \alpha_0 \partial \gamma_j} = 0, \quad 1 \le i \le p, \le j \le 1+P+Q, \tag{A.6.16}$$

$$\frac{\partial^2 \underline{\tilde{\sigma}}_t^2(\lambda)}{\partial \alpha_0 \partial \beta_j} = \sum_{k=1}^{t-1} \left\{ \sum_{l=1}^{k} B^{l-1} B^{(j)} B^{k-l} \right\} \underline{1}, \quad 1 \le j \le q, \tag{A.6.17}$$

$$\frac{\partial^2 \tilde{\sigma}_t^2(\lambda)}{\partial \alpha_i \partial \alpha_j} = 0, \quad 1 \le i, j \le p, \tag{A.6.18}$$

$$\frac{\partial \underline{\tilde{\sigma}}_t^2(\lambda)}{\partial \alpha_i \partial \beta_j} = \sum_{k=1}^{t-1} \left(\sum_{l=1}^{k} B^{l-1} B^{(j)} B^{k-l} \right) \underline{\tilde{\varepsilon}}_{t-i-k}^2(\gamma), \tag{A.6.19}$$

$$1 \le i \le p, 1 \le j \le q,$$

$$\frac{\partial^2 \tilde{\sigma}_t^2(\lambda)}{\partial \alpha_i \partial \gamma_j} = 2 \sum_{k=0}^{t-1} B^k(1,1) \tilde{\varepsilon}_{t-i-k}(\gamma) \frac{\partial \tilde{\varepsilon}_{t-i-k}(\gamma)}{\partial \gamma_j}, \tag{A.6.20}$$

$$1 \le i \le p, 1 \le j \le P+Q+1,$$

$$\frac{\partial^2 \underline{\tilde{\sigma}}_t^2(\lambda)}{\partial \beta_j \partial \beta_{j'}} = \sum_{k=2}^{t-1} \left\{ \sum_{i=2}^{k} \left[\left(\sum_{l=1}^{i-1} B^{l-1} B^{(j')} B^{i-1-l} \right) B^{(j)} B^{k-i} \right] \right. \tag{A.6.21}$$

$$\left. + \sum_{i=1}^{k-1} \left[B^{i-1} B^{(j)} \left(\sum_{l=1}^{k-i} B^{l-1} B^{(j')} B^{k-i-l} \right) \right] \right\} \underline{\tilde{c}}_{t-k}, \quad 1 \le j, j' \le q,$$

$$\frac{\partial^2\tilde{\underline{\sigma}}_t^2(\lambda)}{\partial\beta_i\partial\gamma_j} = 2\sum_{k=1}^{t-1}\left(\sum_{l=1}^{k}B^{l-1}B^{(i)}B^{k-l}\sum_{m=1}^{p}\tilde{\varepsilon}_{t-i-k-m}(\gamma)\frac{\partial\tilde{\varepsilon}_{t-i-k-m}(\gamma)}{\partial\gamma_j}\right) \quad (A.6.22)$$

$$1 \le i \le q, 1 \le j \le P+Q+1.$$

$$\frac{\partial^2\tilde{\sigma}_t^2(\lambda)}{\partial\gamma_j\partial\gamma_j'} = 2\sum_{k=0}^{t-1}B^k(1,1)\sum_{i=1}^{p}\alpha_i\left(\frac{\partial\tilde{\varepsilon}_{t-k-i}(\gamma)}{\partial\gamma_j}\frac{\partial\tilde{\varepsilon}_{t-k-i}(\gamma)}{\partial\gamma_j'} + \tilde{\varepsilon}_{t-k-i}(\gamma)\frac{\partial^2\tilde{\varepsilon}_{t-k-i}(\gamma)}{\partial\gamma_j\partial\gamma_j'}\right), \quad 1 \le i,j \le P+Q+1. \quad (A.6.23)$$

A.7 Difference between $\varepsilon_t(\gamma)$ and $\tilde{\varepsilon}_t(\gamma)$, $\sigma_t(\lambda)$ and $\tilde{\sigma}_t(\lambda)$, as well as between their derivatives

Refer to (4.55) in Francq and Zakoïan (2004), we have

$$\sup_{\gamma\in\Theta_\gamma}\max\left\{|\varepsilon_k(\gamma)-\tilde{\varepsilon}_k(\gamma)|, \left|\frac{\partial\varepsilon_k(\gamma)}{\partial\gamma}-\frac{\partial\tilde{\varepsilon}_k(\gamma)}{\partial\gamma}\right|\right\} \le K\rho^k, \quad a.s. \quad (A.7.1)$$

Based on the expressions of $\partial^2\varepsilon_k(\gamma)/\partial\gamma\partial\gamma^T$, and $\partial^2\tilde{\varepsilon}_k(\gamma)/\partial\gamma\partial\gamma^T$, by Lemma 2.4.1 and a trivial extension of (A.7.1), we have:

$$\sup_{\gamma\in\Theta_\gamma}\left|\frac{\partial^2\varepsilon_k(\gamma)}{\partial\gamma\partial\gamma^T}-\frac{\partial^2\tilde{\varepsilon}_k(\gamma)}{\partial\gamma\partial\gamma^T}\right| \le K\rho^k, \quad a.s. \quad (A.7.2)$$

Refer to (iii) in the proof of Theorem 3.2 in Francq and Zakoïan (2004), we have

$$\sup_{\lambda\in\Theta}\left|\frac{\partial\sigma_t^2(\lambda)}{\partial\lambda}-\frac{\partial\tilde{\sigma}_t^2(\lambda)}{\partial\lambda}\right| \le K\rho^t, \left|\frac{1}{\sigma_t^2(\lambda)}-\frac{1}{\tilde{\sigma}_t^2(\lambda)}\right| \le K\rho^t\frac{S_{t-1}(\gamma)}{\sigma_t^2(\lambda)}, \quad (A.7.3)$$

note that

$$S_{t-1}(\gamma) = \sum_{i=1-p}^{t-1}\left(|\varepsilon_i(\gamma)|+1\right).$$

With similar proof as that for first item in (A.7.3) in (iii) of the proof of Theorem 3.2 in Francq and Zakoïan (2004), we have

$$\sup_{\lambda\in\Theta}\left|\frac{\partial^2\sigma_t^2(\lambda)}{\partial\lambda_i\partial\lambda_j}-\frac{\partial^2\tilde{\sigma}_t^2(\lambda)}{\partial\lambda_i\partial\lambda_j}\right| \le K\rho^t.$$

Based on (2.4.6), (A.7.1) to (A.7.3), by Lemma 2.4.6, for any $\lambda \in \Theta$ we have:

$$\begin{aligned}
|\eta_t(\lambda) - \tilde{\eta}_t(\lambda)| &\leq |\varepsilon_t(\gamma)| \left| \frac{1}{\sigma_t(\lambda)} - \frac{1}{\tilde{\sigma}_t(\lambda)} \right| + \frac{1}{\tilde{\sigma}_t(\lambda)} |\varepsilon_t(\gamma) - \tilde{\varepsilon}_t(\gamma)| \\
&\leq |\eta_t(\lambda)| \left| \frac{\sigma_t^2(\lambda) - \tilde{\sigma}_t^2(\lambda)}{\tilde{\sigma}_t(\lambda)(\sigma_t(\lambda) + \tilde{\sigma}_t(\lambda))} \right| + K\rho^t \\
&\leq K\rho^t S_{t-1}(\gamma)(1 + |\eta_t|(\lambda)),
\end{aligned} \tag{A.7.4}$$

$$\begin{aligned}
&\left| \frac{\partial \sigma_t^2(\lambda)/\partial \lambda}{\sigma_t^2(\lambda)} - \frac{\partial \tilde{\sigma}_t^2(\lambda)/\partial \lambda}{\tilde{\sigma}_t^2(\lambda)} \right| \\
\leq\ & \frac{\partial \sigma_t^2(\lambda)}{\partial \lambda} \left| \frac{1}{\sigma_t^2(\lambda)} - \frac{1}{\tilde{\sigma}_t^2(\lambda)} \right| + \frac{1}{\tilde{\sigma}_t^2(\lambda)} \left| \frac{\partial \sigma_t^2(\lambda)}{\partial \lambda} - \frac{\partial \tilde{\sigma}_t^2(\lambda)}{\partial \lambda} \right| \\
\leq\ & K\rho^t \left(1 + S_{t-1}(\gamma) \left| \frac{\partial \sigma_t^2(\lambda)/\partial \lambda}{\sigma_t^2(\lambda)} \right| \right), \\
\leq\ & K\rho^t S_{t-1}(\gamma) \xi_{\rho,t-1},
\end{aligned} \tag{A.7.5}$$

$$\begin{aligned}
&\left| \frac{\partial \sigma_t^2(\lambda)/\partial \lambda}{\sigma_t^2(\lambda)} \frac{\partial \sigma_t^2(\lambda)/\partial \lambda^T}{\sigma_t^2(\lambda)} - \frac{\partial \tilde{\sigma}_t^2(\lambda)/\partial \lambda}{\tilde{\sigma}_t^2(\lambda)} \frac{\partial \tilde{\sigma}_t^2(\lambda)/\partial \lambda^T}{\tilde{\sigma}_t^2(\lambda)} \right| \\
\leq\ & \left| \frac{\partial \sigma_t^2(\lambda)/\partial \lambda}{\sigma_t^2(\lambda)} \right| \left| \frac{\partial \sigma_t^2(\lambda)/\partial \lambda^T}{\sigma_t^2(\lambda)} - \frac{\partial \tilde{\sigma}_t^2(\lambda)/\partial \lambda^T}{\tilde{\sigma}_t^2(\lambda)} \right| \\
&+ \left| \frac{\partial \sigma_t^2(\lambda)/\partial \lambda}{\sigma_t^2(\lambda)} - \frac{\partial \tilde{\sigma}_t^2(\lambda)/\partial \lambda}{\tilde{\sigma}_t^2(\lambda)} \right| \left| \frac{\partial \tilde{\sigma}_t^2(\lambda)/\partial \lambda^T}{\tilde{\sigma}_t^2(\lambda)} \right| \\
\leq\ & K\rho^t S_{t-1}(\gamma) \xi_{\rho,t-1}^2 + K\rho^t S_{t-1}(\gamma) \xi_{\rho,t-1}^2 (1 + K\rho^t S_{t-1}(\gamma)) \\
\leq\ & K\rho^t S_{t-1}^2(\gamma) \xi_{\rho,t-1}^2 .
\end{aligned} \tag{A.7.6}$$

Similarly, for any $\lambda \in \Theta$, we can show:

$$\left| \frac{\partial^2 \sigma_t^2(\lambda)/\partial \lambda \partial \lambda^T}{\sigma_t^2(\lambda)} - \frac{\partial^2 \tilde{\sigma}_t^2(\lambda)/\partial \lambda \partial \lambda^T}{\tilde{\sigma}_t^2(\lambda)} \right| \leq K\rho^t S_{t-1}(\gamma) \xi_{\rho,t-1}, \tag{A.7.7}$$

$$\left| \frac{\partial \sigma_t^2(\lambda)/\partial \lambda}{\sigma_t^2(\lambda)} \frac{\partial \varepsilon_t/\partial \lambda^T}{\sigma_t(\lambda)} - \frac{\partial \tilde{\sigma}_t^2(\lambda)/\partial \lambda}{\tilde{\sigma}_t^2(\lambda)} \frac{\partial \tilde{\varepsilon}_t/\partial \lambda^T}{\tilde{\sigma}_t(\lambda)} \right| \leq K\rho^t S_{t-1}^2(\gamma) \xi_{\rho,t-1}^2, \tag{A.7.8}$$

$$\left| \frac{\partial \varepsilon_t(\gamma)/\partial \lambda}{\sigma_t(\lambda)} \frac{\partial \varepsilon_t(\gamma)/\partial \lambda^T}{\sigma_t(\lambda)} - \frac{\partial \tilde{\varepsilon}_t(\gamma)/\partial \lambda}{\tilde{\sigma}_t(\lambda)} \frac{\partial \tilde{\varepsilon}_t(\gamma)/\partial \lambda^T}{\tilde{\sigma}_t(\lambda)} \right| \leq K\rho^t S_{t-1}^2(\gamma) \xi_{\rho,t-1}^2, \tag{A.7.9}$$

$$\left|\frac{\partial \varepsilon_t(\gamma)/\partial \lambda}{\sigma_t(\lambda)} - \frac{\partial \tilde{\varepsilon}_t(\gamma)/\partial \lambda}{\tilde{\sigma}_t(\lambda)}\right| \leq K\rho^t S_{t-1}(\gamma)\xi_{\rho,t-1}, \tag{A.7.10}$$

$$\left|\frac{\partial^2 \varepsilon_t(\gamma)/\partial \lambda \partial \lambda^T}{\sigma_t(\lambda)} - \frac{\partial^2 \tilde{\varepsilon}_t(\gamma)/\partial \lambda \partial \lambda^T}{\tilde{\sigma}_t(\lambda)}\right| \leq K\rho^t S_{t-1}(\gamma)\xi_{\rho,t-1}. \tag{A.7.11}$$

Chapter 3

High Moment Partial Sum Processes of Residuals

In this Chapter we study some high moment partial sum processes based on residuals from an ARMA-GARCH/IGARCH model, originally proposed by Kulperger and Yu (2005) for a pure GARCH model. We show that the k-th power partial sum process of residuals converges to a Brownian process plus two correction terms, where the correction terms always depend on ARMA-GARCH parameters. We also consider the CUSUM and the self-normalized processes (standardized by the residual sample mean and variance), which behave as if the residuals were asymptotically IID distributed.

This Chapter is organized as follows. Section 3.1 exhibits some existing results of empirical processes and high moment partial sum processes based on (G)ARCH models. Section 3.2 presents the assumptions and our results. The proofs are postponed to Section 3.3.

3.1 Introduction

Several authors have studied the residuals from non-linear time series models. They showed that the residuals from non-linear time series models behavior different with those from linear time series models.

Boldin (1998) first studied the empirical process of an ARCH(1) and showed that the limiting distribution depends on the parameters of the model. Horváth, Kokoszka and Teyssiére (2001) extended the result to ARCH(p) model. Kawczak, Kulperger and Yu (2002) showed that the limiting distributions of the empirical process and the partial sum process based on residuals from a stationary ARCH-M model are no longer distribution free and hence the residuals cannot be treated as asymptotically IID. They showed that the limiting Gaussian process for the empirical process is a standard Brownian bridge plus an additional term, while the one for partial sum process is a standard Brownian motion plus an additional term. They showed that Kolmogorov-Smirnov test for goodness-of-fit based on residuals differs from the one based on IID sample. The Kolmogorov-Smirnov test produced smaller size and poorer power and is not applicable for ARCH-M models.

Kulperger and Yu (2005) studied some processes based on residuals of pure GARCH (IGARCH) models. These processes are partial sum processes of k-th powers of residuals, CUSUM processes and self-normalized partial sum processes. They showed that the k-th power partial sum process converges to a Brownian motion process plus a correction term that depends on the k-th moment of the innovation sequence. If k-th moment of the innovation is 0, then the correction term is gone and the partial sum moment process converges weakly to the same Gaussian process as if the residuals were IID with same distribution as the innovation. Further, they showed that CUSUM processes and self-normalized partial sum processes converge to Gaussian processes as if the residuals were asymptotically IID. They applied those results for following applications: CUSUM statistics for testing structure change, Jarque-Bera omnibus statistic for testing normality of the unobservable innovation distribution and kernel density estimation of the innovation.

Based on the theories in Chapter 2, we can extend Kulperger and Yu's (2005) results to ARMA-GARCH/IGARCH models. Similar to the extension of QMLE theory from a pure GARCH to an ARMA-GARCH model (see Remark 3.5 in Francq and Zakoïab, 2004), this extension also leads to non-trivial problems and additional assumptions are required for the approximation of hight moment partial sum processes for the ARMA-GARCH model. Basically we require some additional moments on GARCH errors that are not needed for a pure GARCH model. The applications of these results will be introduced in Chapter 4 with numerical samples.

3.2 Assumptions and Results

In model (1.2.12)-(1.2.13), $\mathbf{E}\eta_0^2 = 1$ and $\mathbf{E}\eta_0 = 0$ are assumed. When η_0 has a finite k-th moment, denote $\mu_k = \mathbf{E}(\eta_0^k)$. Thus $\mu_1 = 0$ and $\mu_2 = 1$.

Throughout this chapter we assume that $\hat{\lambda}_n = (\hat{\gamma}_n^T, \hat{\delta}_n^T)^T$ is an estimator of λ based on a sample $Y_1, \ldots, Y_n$, and that it is $\sqrt{n}$ consistent as defined in Assumption 8. To show the results, we assume:

Assumption 8. $\sqrt{n}|\hat{\lambda}_n - \lambda_0| = O_P(1)$,

Assumption 9. $E|\varepsilon_0|^{2\iota} < \infty$ for some $\iota > 0$.

Lemma 2.4.6 (v) and (vi) implies that the moments of $\xi_{\varrho,t}$ and $\xi_{0,\rho,t}$ are determined by the moment of ε_t. For example, $\xi_{0,\rho,t}$ has 2ι moment if $\mathbf{E}|\varepsilon_t|^{2\iota} < \infty$. Furthermore, $\xi_{0,\rho,t-1}^{1-\iota_3}$ has $2\iota/(1-\iota_3)$ moment, where $0 \leq \iota_3 < 1$. Thus there exists some ι so that $1 - \iota_3 < \iota \leq 1$ and hence $2\iota/(1-\iota_3) > 2$, which is crucial for the following proofs. This means that we can choose either $\iota = 1$ or some $0 < \iota < 1$. Notice that $\iota = 1$ corresponds to the ARMA model with finite variance GARCH errors (the ARMA-GARCH model) and $0 < \iota < 1$ to the ARMA model with infinite variance GARCH errors which includes the ARMA-IGARCH model. Throughout the rest of this chapter, we assume that Assumption 9 holds for such a $0 < \iota \leq 1$.

In Chapter 2, we denote $\tilde{\varepsilon}_t = \tilde{\varepsilon}_t(\hat{\gamma}_n)$. Correspondingly denote $\tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\hat{\lambda}_n)$. Then the residual at time t is

$$\tilde{\eta}_t = \tilde{\eta}_t(\hat{\lambda}_n) = \frac{\tilde{\varepsilon}_t(\hat{\gamma}_n)}{\tilde{\sigma}_t(\hat{\lambda}_n)} = \frac{\tilde{\varepsilon}_t}{\tilde{\sigma}_t}.$$

The k-th ($k = 1, 2, 3, 4, \ldots$) order high moment partial sum process of residuals is defined as

$$\tilde{S}_n^{(k)}(u) = \sum_{t=1}^{[nu]} \tilde{\eta}_t^k, \ 0 \leq u \leq 1, \tag{3.2.1}$$

where $[nu]$ is the ceiling integer of nu.

Its counterpart based on the IID innovations is defined as

$$S_n^{(k)}(u) = \sum_{t=1}^{[nu]} \eta_t^k, \ 0 \leq u \leq 1. \tag{3.2.2}$$

Theorem 3.2.1. *Suppose that Assumptions 1 to 3, 6, 8 and 9 hold and let $k \geq 1$ be an integer. If $\mathbf{E}|\eta_0|^k < \infty$, then*

$$\sup_{0\leq u\leq 1}\left|\frac{1}{\sqrt{n}}\left(\tilde{S}_n^{(k)}(u) - S_n^{(k)}(u)\right) + \frac{ku\mu_k}{2}\left\langle \Lambda, \sqrt{n}(\hat{\lambda}_n - \lambda_0)\right\rangle\right.$$
$$\left. - ku\mu_{k-1}\left\langle \Gamma, \sqrt{n}(\hat{\gamma}_n - \gamma_0)\right\rangle\right| = o_P(1),$$

where $\Lambda = \mathbf{E}\left(\partial \log \sigma_0^2(\lambda_0)/\partial\lambda\right)$, $\Gamma = \mathbf{E}\sigma_0^{-1}(\lambda_0)\left(\partial\varepsilon_0(\gamma_0)/\partial\gamma\right)$, and $\langle \mathbf{x}, \mathbf{y}\rangle$ is the inner product of the vectors $\mathbf{x}$ and $\mathbf{y}$.

Remark 3.2.1. *Theorem 3.2.1 shows that the asymptotic properties of the high moment partial sum process $\{\tilde{S}_n^{(k)}(u), 0 \leq u \leq 1\}$ always depend on the parameters of the model for any integer k. This is different from the pure GARCH case where there is no such a term as $ku\mu_{k-1}\langle \Gamma, \sqrt{n}(\hat{\gamma}_n - \gamma_0)\rangle$ and the approximation for $\{\tilde{S}_n^{(k)}(u), 0 \leq u \leq 1\}$ can be parameter free if $\mu_k = 0$ for some odd k. See Remark 1.1 in Kulperger and Yu (2005).*

Remark 3.2.2. *The discussion after Theorem 4.1 in Ling (2005) indicates that $\beta_{01} \neq 0$ is critical when $0 < \iota < 1$ in Assumption 9. With the same reason, Theorem 3.2.1 cannot be applied to the ARMA-IARCH model though it holds for the ARMA-ARCH model after removing the redundant parameters. Further, Theorem 3.2.1 holds also for a pure GARCH model after dropping Assumptions 2 and 9 and letting vector $\Gamma \equiv 0$. Thus, for the pure GARCH model, Theorem 3.2.1 imposes weaker conditions than those given in Kulperger and Yu (2005). Mainly, the parameter space Θ_δ is wider and the condition*

$$\lim_{x\to 0} x^{-\tau} P\{|\eta_0| \leq x\} = 0 \text{ for some } \tau > 0$$

is dropped.

Remark 3.2.3. *Lemma 2.4.6 and Assumption 9 imply the existence of Λ and Γ.*

By Theorem 3.2.1, we immediately obtain the following CUSUM result, Corollary 3.2.1. It implies that the CUSUM normalized high moment partial sum process $\{\tilde{S}_n^{(k)}(u) - u\tilde{S}_n^{(k)}(1), 0 \le u \le 1\}$ behaves as if the residuals $\{\tilde{\eta}_t, 1 \le t \le n\}$ were asymptotically the same as the unobservable innovations $\{\eta_t, 1 \le t \le n\}$.

Corollary 3.2.1. *Suppose that Assumptions 1 to 3, 6, 8 and 9 hold and let $k \ge 1$ be an integer. If $\mathbf{E}|\eta_0|^k < \infty$, then*

$$\sup_{0\le u\le 1} \frac{1}{\sqrt{n}} \left| \left(\tilde{S}_n^{(k)}(u) - u\tilde{S}_n^{(k)}(1)\right) - \left(S_n^{(k)}(u) - uS_n^{(k)}(1)\right) \right| = o_P(1) \ .$$

The next result follows immediately from Corollary 3.2.1 based on the invariance principle for partial sums for an IID sequence $\{\eta_t^k\}$ (see for example Billingsley, 1999).

Corollary 3.2.2. *Suppose that Assumptions 1 to 3, 6, 8 and 9 hold . Let $k \ge 1$ be an integer and $\zeta_k^2 = \mathbf{E}(\eta_0^k - \mu_k)^2 < \infty$. If $E|\eta_0|^{2k} < \infty$ for some integer $k \ge 1$ then*

$$\left\{ \frac{\tilde{S}_n^{(k)}(u) - u\tilde{S}_n^{(k)}(1)}{\zeta_k \sqrt{n}}, \ 0 \le u \le 1 \right\}$$

converges weakly in the Skorokhod space $D[0,\ 1]$ with J_1 topology to a Brownian bridge $\{B_0(u), 0 \le u \le 1\}$.

Before formulating the next result, we need to modify the high moment partial sum processes of (3.2.1) and (3.2.2). The k-th order moment residual centered partial sum process is defined as

$$\tilde{T}_n^{(k)}(u) = \sum_{t=1}^{[nu]} \left(\tilde{\eta}_t - \bar{\tilde{\eta}}\right)^k, \ 0 \le u \le 1 \ ,$$

where $\bar{\tilde{\eta}}$ is the sample mean of the residuals. Its counterpart based on the IID innovations is

$$T_n^{(k)}(u) = \sum_{t=1}^{[nu]} \left(\eta_t - \bar{\eta}\right)^k, \ 0 \le u \le 1 \ ,$$

where $\bar{\eta}$ is the sample mean of innovations.

Theorem 3.2.2. *Suppose that Assumptions 1 to 3, 6, 8 and 9 hold and let $k \geq 1$ be an integer. If $\mathbf{E}|\eta_0|^k < \infty$, then*

$$\sup_{0 \leq u \leq 1} \left| \frac{1}{\sqrt{n}} \left(\tilde{T}_n^{(k)}(u) - T_n^{(k)}(u) \right) + \frac{ku\mu_k}{2} \left\langle \Lambda, \sqrt{n}(\hat{\lambda}_n - \lambda_0) \right\rangle \right| = o_P(1).$$

Remark 3.2.4. *Theorem 3.2.2 shows that the sample mean centering is able to remove the parameter term $ku\mu_{k-1} \langle \Gamma, \sqrt{n}(\hat{\gamma}_n - \gamma_0) \rangle$ in Theorem 3.2.1. This matches with a result for high moment partial sum processes for a stationary ARMA model in Yu (2005).*

Let $\tilde{\sigma}^2_{(n)} = \tilde{T}_n^{(2)}(1)/n$ be the sample moment estimator of μ_2. Just like in the pure GARCH case, $\tilde{\sigma}^2_{(n)}$ will play an important role when it is used to self-normalized $\tilde{T}_n^{(k)}(u)$. Denote $\sigma^2_{(n)} = T_n^{(2)}(1)/n$, and note that it is the sample variance of the true innovations, except with divisor n instead of $n-1$, which does not matter as long as large sample properties are concerned.

Theorem 3.2.3. *Suppose that Assumptions 1 to 3, 6, 8 and 9 hold and let $k \geq 1$ be an integer. If $\mathbf{E}|\eta_0|^{\max\{k,2\}} < \infty$, then*

$$\sup_{0 \leq u \leq 1} \frac{1}{\sqrt{n}} \left| \frac{\tilde{T}_n^{(k)}(u)}{\tilde{\sigma}^k_{(n)}} - \frac{T_n^{(k)}(u)}{\sigma^k_{(n)}} \right| = o_P(1).$$

Let $\nu_k = \mu_k / \mu_2^{k/2}$ for $k \geq 1$ and define $\nu_0 = 1$. For each $k \geq 1$, let $\{B^{(k)}(u), 0 \leq u \leq 1\}$ be a zero mean Gaussian process with covariance

$$\begin{aligned} EB^{(k)}(u)B^{(k)}(v) &= (\nu_{2k} - \nu_k^2)(u \wedge v) + k\nu_{k-1}(k\nu_{k-1} + k\nu_k\nu_3 - 2\nu_{k+1})uv \\ &\quad + k\nu_k \left((1 - k/4)\nu_k + k\nu_k\nu_4/4 - \nu_{k+2} \right) uv \qquad (3.2.3) \end{aligned}$$

for any $0 \leq u, v \leq 1$, where $u \wedge v = \min(u, v)$.

If $\mu_{2k} < \infty$, then Lemma 3.8 in Kulperger and Yu (2005) implies

$$\left\{ \frac{1}{\sqrt{n}} \left(\frac{T_n^{(k)}(u)}{\sigma_{(n)}^k} - nu\nu_k \right), \ 0 \le u \le 1 \right\}$$

converges weakly to the Gaussian process $\{B^{(k)}(u),\ 0 \le u \le 1\}$. By Theorem 3.2.3, we immediately obtain the following corollary.

Corollary 3.2.3. *If Assumptions 1 to 3, 6, 8 and 9 hold , then* $\mathbf{E}|\eta_0|^{2k} < \infty$ *for some integer* $k \ge 1$ *implies that*

$$\left\{ \frac{1}{\sqrt{n}} \left(\frac{\tilde{T}_n^{(k)}(u)}{\tilde{\sigma}_{(n)}^k} - nu\nu_k \right), \ 0 \le u \le 1 \right\}$$

converges weakly to the Gaussian process $\{B^{(k)}(u),\ 0 \le u \le 1\}$.

When $k = 1$, (3.2.3) becomes $EB^{(1)}(u)B^{(1)}(v) = u \wedge v - uv$ for any $0 \le u, v \le 1$, that is, $\{B^{(1)}(u),\ 0 \le u \le 1\}$ is a Brownian bridge. For $k = 2$, (3.2.3) implies that $EB^{(2)}(u)B^{(2)}(v) = (\nu_4 - 1)(u \wedge v - uv)$ for any $0 \le u, v \le 1$. This means $\{B^{(2)}(u)/\sqrt{\nu_4 - 1},\ 0 \le u \le 1\}$ is also a Brownian bridge. In general, the Gaussian process $\{B^{(k)}(u),\ 0 \le u \le 1\}$ for $k \ge 3$ depends on the moments of the innovation distribution and cannot be identified to be a specific process known in the literature, such as a Brownian motion or a Brownian bridge. More details can be found in Kulperger and Yu (2005). Here we just give the following corollary that will be used to construct the Jarque-Bera test statistic given in the next Chapter.

Corollary 3.2.4. *Suppose that Assumptions 1 to 3, 6, 8 and 9 hold. Assume also that* $k \ge 1$ *is an odd number and* $\mu_3 = \mu_k = \mu_{k+2} = \mu_{2k+1} = 0$. *Then* $E|\eta_0|^{2(k+1)} < \infty$ *implies that*

$$\left\{ \frac{1}{\sqrt{n}} \left(\frac{\tilde{T}_n^{(k)}(x)}{\tilde{\sigma}_{(n)}^k} - nx\nu_k,\ \frac{\tilde{T}_n^{(k+1)}(y)}{\tilde{\sigma}_{(n)}^{k+1}} - ny\nu_{k+1} \right), \ 0 \le x, y \le 1 \right\}.$$

converges weakly in the Skorokhod space $D^2[0,\ 1]$ *to a two dimensional Gaussian process* $\{(B^{(k)}(x),\ B^{(k+1)}(y))\ 0 \le x,y \le 1\}$, *where* $\{B^{(k)}(x),\ 0 \le x \le 1\}$ *and* $\{B^{(k+1)}(y),\ 0 \le y \le 1\}$ *are two independent zero mean Gaussian processes defined by*

$$\begin{aligned} EB^{(i)}(x)B^{(i)}(y) &= (\nu_{2i} - \nu_i^2)(x \wedge y) + i\nu_{i-1}(i\nu_{i-1} + i\nu_i\nu_3 - 2\nu_{i+1})xy \\ &\quad + i\nu_i\left((1 - i/4)\nu_i + i\nu_i\nu_4/4 - \nu_{i+2}\right)xy,\ i = k, k+1, \end{aligned}$$

for any $0 \le x, y \le 1$.

3.3 Proofs

This section begins with a proof of Theorem 3.2.1. It is given in a sketch or overview form with the details given in a series of lemmas, which are placed in the later part of this section. The proofs of Theorems 3.2.2 and 3.2.3 rely on the proof of Theorem 3.2.1.

3.3.1 Proof of Theorem 3.2.1

Proof of Theorem 3.2.1: Let $\hat{\varepsilon}_t = \varepsilon_t(\hat{\gamma}_n)$, $\hat{\sigma}_t^2 = \sigma_t^2(\hat{\lambda}_n)$,

$$\hat{\eta}_t = \eta_t(\hat{\lambda}_n) = \frac{\varepsilon_t(\hat{\gamma}_n)}{\sigma_t(\hat{\lambda}_n)} = \frac{\hat{\varepsilon}_t}{\hat{\sigma}_t},\ 1 \le t \le n, \quad \text{and} \quad \hat{S}_n^{(k)}(u) = \sum_{t=1}^{[nu]} \hat{\eta}_t^k,\ 0 \le u \le 1.$$

Then Theorem 3.2.1 follows if we can show that

$$\begin{aligned} \sup_{0\le u\le 1} \Bigg| \frac{1}{\sqrt{n}}\left(\hat{S}_n^{(k)}(u) - S_n^{(k)}(u)\right) + \frac{ku\mu_k}{2}\left\langle \Lambda, \sqrt{n}(\hat{\lambda}_n - \lambda_0)\right\rangle \\ -ku\mu_{k-1}\left\langle \Gamma, \sqrt{n}(\hat{\gamma}_n - \gamma_0)\right\rangle \Bigg| = o_P(1) \end{aligned} \tag{3.3.1}$$

and

$$\sup_{0\le u\le 1} \frac{1}{\sqrt{n}}\left|\tilde{S}_n^{(k)}(u) - \hat{S}_n^{(k)}(u)\right| = o_P(1). \tag{3.3.2}$$

The proof of (3.3.2) is left in Lemma 3.3.1. Let

$$g_t(\lambda) = \frac{\sqrt{n}\Big(\sigma_t^2(\lambda_0 + n^{-1/2}\lambda) - \sigma_t^2(\lambda_0)\Big)}{\sigma_t^2(\lambda_0)}, \; \lambda \in \mathbb{R}^{P+Q+p+q+2} \tag{3.3.3}$$

and

$$Z_t(\gamma) = \frac{\sqrt{n}\Big(\varepsilon_t(\gamma_0 + n^{-1/2}\gamma) - \varepsilon_t(\gamma_0)\Big)}{\sigma_t(\lambda_0)}. \tag{3.3.4}$$

Note the we adopt the same notation in Kulperger and Yu (2005). And there is no relationship between the function $g_t(\lambda)$ defined here and the function $g(x)$ defined in Chapter 2.

Though $g_t(\lambda)$ and $Z_t(\gamma)$ depend on n, we omit it for convenience of notation. By the definition of $\hat{\eta}_t$, it is not difficult to see

$$\hat{\eta}_t = \frac{\eta_t + n^{-1/2} Z_t(\sqrt{n}(\hat{\gamma}_n - \gamma_0))}{\sqrt{1 + n^{-1/2} g_t(\sqrt{n}(\hat{\lambda}_n - \lambda_0))}}. \tag{3.3.5}$$

Thus, by Assumption 8, to prove (3.3.1), we need to prove for any $b > 0$ that

$$\sup_{0\le u\le 1} \sup_{|\lambda|\le b} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{[nu]} \left(\frac{\eta_t + n^{-1/2} Z_t(\gamma)}{\sqrt{1 + n^{-1/2} g_t(\lambda)}} \right)^k \right.$$
$$\left. - \frac{1}{\sqrt{n}} S_n^{(k)}(u) + \frac{k u \mu_k}{2} \langle \Lambda, \lambda \rangle - k u \mu_{k-1} \langle \Gamma, \gamma \rangle \right| = o_P(1).$$

This last part follows by

$$\sup_{0\le u\le 1} \sup_{|\lambda|\le b} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{[nu]} \left(\frac{\eta_t + n^{-1/2} Z_t(\gamma)}{\sqrt{1 + n^{-1/2} g_t(\lambda)}} \right)^k \right. \tag{3.3.6}$$
$$\left. - \frac{1}{\sqrt{n}} \sum_{t=1}^{[nu]} \left(\frac{\eta_t}{\sqrt{1 + n^{-1/2} g_t(\lambda)}} \right)^k - \frac{k}{n} \sum_{t=1}^{[nu]} \left(\frac{1}{\sqrt{1 + n^{-1/2} g_t(\lambda)}} \right)^k \eta_t^{k-1} Z_t(\gamma) \right| = o_P(1),$$

$$\sup_{0\le u\le 1} \sup_{|\lambda|\le b} \left| \frac{1}{n} \sum_{t=1}^{[nu]} \left(\frac{1}{\sqrt{1 + n^{-1/2} g_t(\lambda)}} \right)^k \eta_t^{k-1} Z_t(\gamma) - u \mu_{k-1} \langle \Gamma, \gamma \rangle \right| = o_P(1), \tag{3.3.7}$$

and

$$\sup_{0\le u\le 1}\sup_{|\lambda|\le b}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{[nu]}\left(\frac{\eta_t}{\sqrt{1+n^{-1/2}g_t(\lambda)}}\right)^k-\frac{1}{\sqrt{n}}S_n^{(k)}(u)+\frac{ku\mu_k}{2}\langle\Lambda,\lambda\rangle\right|=o_P(1). \tag{3.3.8}$$

To prove (3.3.6) to (3.3.8), we need to find how fast the following terms

$$\max_{1\le t\le n}\sup_{|\lambda|\le b}\frac{|g_t(\lambda)|}{\sqrt{n}} \quad \text{and} \quad \max_{1\le t\le n}\sup_{|\lambda|\le b}|g_t(\lambda)-\langle\partial\log\sigma_t^2(\lambda_0)/\partial\lambda,\lambda\rangle| \tag{3.3.9}$$

converge in probability to zero. In pure GARCH case, Lemma 3.3 in Kulperger and Yu (2005) show that both $\sup_{|\lambda|\le b}|g_t(\lambda)|$ and $\sup_{|\lambda|\le b}\sqrt{n}|g_t(\lambda)-\langle\partial\log\sigma_t^2(\lambda_0)/\partial\lambda,\lambda\rangle|$ have any finite moments and hence one can use a well-known result that if $\{X_n, n\ge 0\}$ is a sequence of identically distributed r.v.'s. with $E|X_0|^{\kappa^*}<\infty$ for some $\kappa^*>0$, then

$$\max_{1\le t\le n}|X_t|=o_P(n^{1/\kappa^*}). \tag{3.3.10}$$

In addition the following important approximation

$$\max_{1\le t\le n}\sup_{|\lambda|\le b}\left|\frac{1}{\sqrt{1+n^{-1/2}g_t(\lambda)}}-\left(1-\frac{\langle\partial\log\sigma_t^2(\lambda_0)/\partial\lambda,\lambda\rangle}{2\sqrt{n}}\right)\right|=o_P\left(\frac{1}{\sqrt{n}}\right) \tag{3.3.11}$$

can be established as well.

Unfortunately, if we adapted the same approach used in Kulperger and Yu (2005), we would need at least $E|\varepsilon_0|^{16}<\infty$ due to variation contributed from ARMA components. In fact (3.3.11) may be not feasible for ARMA-GARCH models under a minimum moment condition on GARCH errors. Notice by (3.3.3) that we only need to work with the function $\sigma_t^2(\lambda)$ in the neighborhood $|\lambda-\lambda_0|\le b/\sqrt{n}$ of λ_0. Indeed we are able to take such an advantage in Lemma 3.3.3 and find proper convergence rates of (3.3.9) in Lemma 3.3.4 without requiring higher moments on GARCH errors.

(3.3.6) and (3.3.7) are proven in Lemmas 3.3.6 and 3.3.7, respectively. We divide the proof of (3.3.8) into three parts. Let $w(x) = 1/\sqrt{1+x}$ and $w^{(i)}(x)$ denote the i-th derivative of $w(x)$. Let M be a positive integer which will be determined in Lemma 3.3.5. Then equation (3.3.8) follows by

$$\sup_{|\lambda|\le b} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} |\eta_t|^k \left| \left(\frac{1}{\sqrt{1+n^{-1/2}g_t(\lambda)}}\right)^k - \left(1+\sum_{i=1}^{M} \frac{w^{(i)}(0)}{i!}\frac{\langle \partial \log \sigma_t^2(\lambda_0)/\partial\lambda, \lambda\rangle^i}{n^{i/2}}\right)^k \right| = o_P(1), \tag{3.3.12}$$

$$\sup_{|\lambda|\le b} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} |\eta_t|^k \left| \left(1+\sum_{i=1}^{M} \frac{w^{(i)}(0)}{i!}\frac{\langle \partial \log \sigma_t^2(\lambda_0)/\partial\lambda, \lambda\rangle^i}{n^{i/2}}\right)^k - \left(1-\frac{\langle \partial \log \sigma_t^2(\lambda_0)/\partial\lambda, \lambda\rangle}{2\sqrt{n}}\right)^k \right| = o_P(1), \tag{3.3.13}$$

and

$$\sup_{0\le u\le 1}\sup_{|\lambda|\le b} \left| \frac{1}{\sqrt{n}}\sum_{t=1}^{[nu]} \eta_t^k \left(1-\frac{\langle \partial \log \sigma_t^2(\lambda_0)/\partial\lambda, \lambda\rangle}{2\sqrt{n}}\right)^k - \frac{1}{\sqrt{n}} S_n^{(k)}(u) + \frac{ku\mu_k}{2}\langle \Lambda, \lambda\rangle \right| = o_P(1). \tag{3.3.14}$$

They are proven in Lemma 3.3.8 to 3.3.10, respectively. Now we completely finish the proof of Theorem 3.2.1.

As a consequence of Theorem 3.2.1, we can obtain for any $1 \le i \le k$

$$\frac{1}{n}\sum_{t=1}^{n} \bar{\eta}_t^i = \frac{1}{n}\sum_{t=1}^{n} \eta_t^i - \frac{i\mu_i}{2\sqrt{n}}\left\langle \Lambda, \sqrt{n}(\hat{\lambda}_n - \lambda_0)\right\rangle + \frac{i\mu_{i-1}}{\sqrt{n}}\langle \Gamma, \sqrt{n}(\hat{\gamma}_n - \gamma_0)\rangle + o_P\left(\frac{1}{\sqrt{n}}\right). \tag{3.3.15}$$

In particular, since $\mu_0 = 1$ and $\mu_1 = 0$, we have

$$\bar{\bar{\eta}} = \bar{\eta} + \frac{1}{\sqrt{n}}\langle \Gamma, \sqrt{n}(\hat{\gamma}_n - \gamma_0)\rangle + o_P\left(\frac{1}{\sqrt{n}}\right). \tag{3.3.16}$$

3.3.2 Proof of Theorem 3.2.2

Proof of Theorem 3.2.2: When $k = 1$, Theorem 3.2.2 follows by

$$\sup_{0\le u\le 1}\left|\tilde{T}_n^{(1)}(u) - T_n^{(1)}(u)\right| \le \sup_{0\le u\le 1}\left|\left(\tilde{S}_n^{(1)}(u) - u\tilde{S}_n^{(1)}(1)\right) - \left(S_n^{(1)}(u) - uS_n^{(1)}(1)\right)\right| + |\bar{\bar{\eta}} - \bar{\eta}|$$

and Corollary 3.2.1 and (3.3.16).

Next we consider the case $k \ge 2$. Theorem 3.2.1 and weak law of large number implies that

$$\sup_{0\le u\le 1}\frac{1}{n}\left|\tilde{S}_n^{(i)}(u)\right| \le \frac{1}{n}\sum_{t=1}^{n}|\eta_t^i| + O_P\left(\frac{1}{\sqrt{n}}\right) = O_P(1)$$

for $1 \le i \le k$, while (3.3.16) and CLT imply

$$\bar{\bar{\eta}} = O_P\left(\frac{1}{\sqrt{n}}\right) \quad \text{and} \quad \bar{\eta} = O_P\left(\frac{1}{\sqrt{n}}\right).$$

In addition,

$$\sup_{0\le u\le 1}\left|\frac{1}{n}S_n^{(k-1)}(u) - u\mu_{k-1}\right| = o_P(1)$$

follows by Lemma 3.3.2. Thus, by Theorem 3.2.1 and (3.3.16), we have uniformly in

u that

$$\begin{aligned}
\frac{1}{\sqrt{n}}\tilde{T}_n^{(k)}(u) &= \frac{1}{\sqrt{n}}\tilde{S}_n^{(k)}(u) - \frac{k\bar{\tilde{\eta}}}{\sqrt{n}}\tilde{S}_n^{(k-1)}(u) + \frac{1}{\sqrt{n}}\sum_{i=0}^{k-2}\binom{k}{i}(-1)^{k-i}\bar{\tilde{\eta}}^{k-i}\tilde{S}_n^{(i)}(u) \\
&= \frac{1}{\sqrt{n}}S_n^{(k)}(u) - \frac{ku\mu_k}{2}\left\langle \Lambda, \sqrt{n}(\hat{\lambda}_n - \lambda_0)\right\rangle + ku\mu_{k-1}\left\langle \Gamma, \sqrt{n}(\hat{\gamma}_n - \gamma_0)\right\rangle \\
&\quad -k\left(\bar{\eta} + \frac{1}{\sqrt{n}}\left\langle \Gamma, \sqrt{n}(\hat{\gamma}_n - \gamma_0)\right\rangle + o_P\left(\frac{1}{\sqrt{n}}\right)\right)\times \\
&\quad \left(\frac{1}{\sqrt{n}}S_n^{(k-1)}(u) - \frac{(k-1)u\mu_{k-1}}{2}\left\langle \Lambda, \sqrt{n}(\hat{\lambda}_n - \lambda_0)\right\rangle\right. \\
&\quad \left.+(k-1)u\mu_{k-2}\left\langle \Gamma, \sqrt{n}(\hat{\gamma}_n - \gamma_0)\right\rangle + o_P\left(\frac{1}{\sqrt{n}}\right)\right) + o_P(1) \\
&= \frac{1}{\sqrt{n}}S_n^{(k)}(u) - \frac{k\bar{\eta}}{\sqrt{n}}S_n^{(k-1)}(u) - \frac{ku\mu_k}{2}\left\langle \Lambda, \sqrt{n}(\hat{\lambda}_n - \lambda_0)\right\rangle + o_P(1) \\
&= \frac{1}{\sqrt{n}}T_n^{(k)}(u) - \frac{ku\mu_k}{2}\left\langle \Lambda, \sqrt{n}(\hat{\lambda}_n - \lambda_0)\right\rangle + o_P(1).
\end{aligned}$$

This completes the proof of Theorem 3.2.2.

3.3.3 Proof of Theorem 3.2.3

Proof of Theorem 3.2.3: First Theorem 3.2.2 implies for any $1 \le i \le k$

$$\frac{1}{n}\sum_{t=1}^{n}\left(\tilde{\eta}_t - \bar{\tilde{\eta}}\right)^i = \frac{1}{n}\sum_{t=1}^{n}\left(\eta_t - \bar{\eta}\right)^i - \frac{i\mu_i}{2\sqrt{n}}\left\langle \Lambda, \sqrt{n}(\hat{\lambda}_n - \lambda_0)\right\rangle + o_P\left(\frac{1}{\sqrt{n}}\right). \quad (3.3.17)$$

Since we assume $\mathbf{E}\eta_0^2 < \infty$ for any $k \ge 1$, (3.3.17) implies

$$\tilde{\sigma}_{(n)}^2 = \sigma_{(n)}^2 - \frac{\mu_2}{\sqrt{n}}\left\langle \Lambda, \sqrt{n}(\hat{\lambda}_n - \lambda_0)\right\rangle + o_P\left(\frac{1}{\sqrt{n}}\right). \quad (3.3.18)$$

Let

$$\tilde{L}_n(u) = \sqrt{n}\left(\frac{1}{n}\tilde{T}_n^{(k)}(u) - u\nu_k\tilde{\sigma}_{(n)}^k\right) \text{ and } L_n(u) = \sqrt{n}\left(\frac{1}{n}T_n^{(k)}(u) - u\nu_k\sigma_{(n)}^k\right).$$

Then

$$\sup_{0\le u\le 1}\frac{1}{\sqrt{n}}\left|\frac{\tilde{T}_n^{(k)}(u)}{\tilde{\sigma}_{(n)}^k}-\frac{T_n^{(k)}(u)}{\sigma_{(n)}^k}\right| \le \frac{\sup_{0\le u\le 1}|\tilde{L}_n(u)-L_n(u)|}{\tilde{\sigma}_{(n)}^k}+\sup_{0\le u\le 1}|L_n(u)|\left|\frac{1}{\tilde{\sigma}_{(n)}^k}-\frac{1}{\sigma_{(n)}^k}\right|.$$

Notice that (3.3.18) implies

$$\left|\frac{1}{\tilde{\sigma}_{(n)}^k}-\frac{1}{\sigma_{(n)}^k}\right| = O_P\left(\frac{1}{\sqrt{n}}\right).$$

Therefore we can prove Theorem 3.2.3 if

$$\sup_{0\le u\le 1}|\tilde{L}_n(u)-L_n(u)| = o_P(1) \tag{3.3.19}$$

and

$$\sup_{0\le u\le 1}\frac{|L_n(u)|}{\sqrt{n}} = o_P(1). \tag{3.3.20}$$

By the facts that $\bar{\eta}=O_P(1/\sqrt{n})$ and $\sigma_{(n)}^2=\mu_2+o_P(1)$,

$$\begin{aligned}\frac{L_n(u)}{\sqrt{n}} &= \frac{1}{n}\sum_{t=1}^{[nu]}(\eta_t-\bar{\eta})^k - u\frac{\mu_k}{\mu_2^{k/2}}(\mu_2+o_P(1))^{k/2}\\ &= \frac{1}{n}\sum_{t=1}^{[nu]}\eta_t^k - u\mu_k + o_P(1)\end{aligned}$$

thus (3.3.20) holds by Lemma 3.3.2.

When $k=1$, (3.3.19) follows directly from Theorem 3.2.2 since $\mu_1=0$. Let us consider the case $k\ge 2$. By (3.3.18) and a first order Taylor's approximation with remainder we have

$$\begin{aligned}\sqrt{n}\tilde{\sigma}_{(n)}^k &= \sqrt{n}\left(\sigma_{(n)}^2-\frac{\mu_2}{\sqrt{n}}\left\langle\Lambda,\sqrt{n}(\hat{\lambda}_n-\lambda_0)\right\rangle+o_P\left(\frac{1}{\sqrt{n}}\right)\right)^{k/2}\\ &= \sqrt{n}\left(\sigma_{(n)}^2-\frac{\mu_2}{\sqrt{n}}\left\langle\Lambda,\sqrt{n}(\hat{\lambda}_n-\lambda_0)\right\rangle\right)^{k/2}+o_P(1)\\ &= \sqrt{n}\sigma_{(n)}^k-\frac{k\sigma_{(n)}^{k-2}\mu_2}{2}\left\langle\Lambda,\sqrt{n}(\hat{\lambda}_n-\lambda_0)\right\rangle+o_P(1).\end{aligned}$$

Putting the above into $\tilde{L}_n(u)$, together with Theorem 3.2.2 and the fact that $\sigma^2_{(n)} = \mu_2 + o_P(1)$, we obtain

$$\begin{aligned}\tilde{L}_n(u) &= \frac{1}{\sqrt{n}} T_n^{(k)}(u) - \frac{ku\mu_k}{2}\left\langle \Lambda, \sqrt{n}(\hat{\lambda}_n - \lambda_0)\right\rangle \\ &\quad -u\nu_k\left(\sqrt{n}\sigma^k_{(n)} - \frac{k\mu_2^{k/2}}{2}\left\langle \Lambda, \sqrt{n}(\hat{\lambda}_n - \lambda_0)\right\rangle\right) + o_P(1) \\ &= L_n(u) + o_P(1)\end{aligned}$$

uniformly in $0 \le u \le 1$. This concludes the proof of (3.3.19) and hence Theorem 3.2.3.

3.3.4 Proof of preliminary results

The remainder of this section gives the various Lemmas needed in the proofs above. Throughout the rest of proofs, C denote a finite positive constant which may change values from place to place but does not depend on t. We also use the following inequality

$$\left|(x+\Delta)^k - x^k\right| \le k2^{k-1}|\Delta|\left(|x|^{k-1} + |\Delta|^{k-1}\right) \qquad (3.3.21)$$

on several occasions.

Lemma 3.3.1. *Under Assumptions 1 to 3, 6, 8 and 9, we have for* $k \ge 1$

$$\sup_{0\le u\le 1} \frac{1}{\sqrt{n}}\left|\tilde{S}_n^{(k)}(u) - \hat{S}_n^{(k)}(u)\right| = o_P(1).$$

Proof: By definitions of $\tilde{\eta}_t$ and $\hat{\eta}_t$, we can rewrite

$$\tilde{\eta}_t = \hat{\eta}_t + \frac{\hat{\eta}_t(\hat{\sigma}_t - \tilde{\sigma}_t) + \tilde{\varepsilon}_t - \hat{\varepsilon}_t}{\tilde{\sigma}_t}$$

and hence

$$\tilde{S}_n^{(k)}(u) = \hat{S}_n^{(k)}(u) + \sum_{i=1}^{k} \binom{k}{i} \sum_{t=1}^{[nu]} \hat{\eta}_t^{k-i} \left(\frac{\hat{\eta}_t(\hat{\sigma}_t - \tilde{\sigma}_t) + \tilde{\varepsilon}_t - \hat{\varepsilon}_t}{\tilde{\sigma}_t} \right)^i .$$

Since we assume that $\alpha_0 > 0$ and Θ_δ is compact, by (2.2.2) and (2.2.6), there exists a constant $C > 0$ such that almost surely

$$\tilde{\sigma}_t \geq C \ \text{ and } \hat{\sigma}_t \geq C \ \text{ for all } t \geq 1.$$

Thus, by using the inequality $|a+b|^i \leq 2^{i-1}(|a|^i + |b|^i)$ for any real a, b and integer $i \geq 1$, Lemma 3.3.1 is proven if we can show for all $i = 1, \ldots, k$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} |\hat{\varepsilon}_t|^k |\hat{\sigma}_t^2 - \tilde{\sigma}_t^2|^i = o_P(1) \tag{3.3.22}$$

and

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} |\hat{\varepsilon}_t|^{k-i} |\hat{\varepsilon}_t - \tilde{\varepsilon}_t|^i = o_P(1). \tag{3.3.23}$$

Now, by Lemma 2.4.6 (i) and (2.4.6),

$$|\hat{\varepsilon}_t|^k |\hat{\sigma}_t^2 - \tilde{\sigma}_t^2|^i \leq C \xi_{\rho,t}^k \rho^{it} S_t^i(\gamma_0).$$

By Assumption 8, Lemma 2.4.6 (v), and Hölder's inequality, taking $\tau * = \iota/(4k)$,

$$E^2 |\xi_{\rho,t}^k S_t^i(\gamma_0)|^{\tau *} \leq E \xi_{\rho,t}^{2k\tau *} E S_t^{2i\tau *}(\gamma_0) \leq C(1+t) \ \text{ for all } 1 \leq t \leq n.$$

Hence we have

$$E\left(\sum_{t=1}^{n} |\hat{\varepsilon}_t|^k |\hat{\sigma}_t^2 - \tilde{\sigma}_t^2|^i \right)^{\tau *} \leq C \sum_{t=1}^{\infty} \sqrt{1+t} \rho^{it\tau *} < \infty.$$

This proves (3.3.22). One can prove (3.3.23) similarly. This completes the proof of Lemma 3.3.1.

Lemma 3.3.2. *Let $X_t = h(\eta_t, \eta_{t-1}, \ldots)$ and suppose that $E|X_0| < \infty$. Then*

$$\sup_{0\le u\le 1}\left|\frac{1}{n}\sum_{t=1}^{[nu]} X_t - uEX_0\right| = o_P(1) \ .$$

Proof: We refer to Lemma 3.6 of Kulperger and Yu (2005).

The following lemma is a key result that provides proper convergence rates used in proving (3.3.9) under minimum moment conditions on GARCH errors. Notice that λ is in the neighborhood $|\lambda - \lambda_0| \le b/\sqrt{n}$ of λ_0. Denote for any $b > 0$,

$$I_n = \max_{1\le t\le n}\sup_{|\lambda-\lambda_0|\le b/\sqrt{n}}\frac{|\partial\sigma_t^2(\lambda)/\partial\lambda|}{\sigma_t^2(\lambda_0)} \quad \text{and} \quad J_n = \max_{1\le t\le n}\sup_{|\lambda-\lambda_0|\le b/\sqrt{n}}\frac{|\partial^2\sigma_t^2(\lambda)/\partial\lambda\partial\lambda^T|}{\sigma_t^2(\lambda_0)}.$$

Lemma 3.3.3. *Under Assumptions 1 to 3, 6, 8 and 9, we have*

$$I_n = o_P\left(n^{1/\kappa}\right) \quad \text{and} \quad J_n = o_P\left(n^{1/\kappa}\right),$$

where $\kappa = 2\iota/(1-\iota_3) > 2$.

Proof: By Lemma 2.4.6(iii),

$$\begin{aligned}
\sup_{|\lambda-\lambda_0|\le b/\sqrt{n}}\frac{|\partial\sigma_t^2(\lambda)/\partial\delta|}{\sigma_t^2(\lambda_0)} &= \sup_{|\lambda-\lambda_0|\le b/\sqrt{n}}\frac{|\partial\sigma_t^2(\lambda)/\partial\delta|}{\sigma_t^2(\lambda)}\left|1+\frac{\sigma_t^2(\lambda)-\sigma_t^2(\lambda_0)}{\sigma_t^2(\lambda_0)}\right| \\
&\le C\xi_{\rho,t-1}^{s}\left(1+\sup_{|\lambda-\lambda_0|\le b/\sqrt{n}}|\lambda-\lambda_0|\frac{|\partial\sigma_t^2(\lambda)/\partial\lambda|}{\sigma_t^2(\lambda_0)}\right) \\
&\le C\xi_{\rho,t-1}^{s}\left(1+\frac{bI_n}{\sqrt{n}}\right).
\end{aligned}$$

By Lemma 2.4.6 (iv) and (v), 2.4.7 (v), we have

$$\begin{aligned}
\sup_{|\lambda-\lambda_0|\le b/\sqrt{n}}\frac{|\partial\sigma_t^2(\lambda)/\partial\gamma|}{\sigma_t^2(\lambda_0)} &= \frac{1}{\sigma_t(\lambda_0)}\sup_{|\lambda-\lambda_0|\le b/\sqrt{n}}\frac{|\partial\sigma_t^2(\lambda)/\partial\gamma|}{\sigma_t(\lambda)}\left|1+\frac{\sigma_t(\lambda)-\sigma_t(\lambda_0)}{\sigma_t(\lambda_0)}\right| \\
&\le \frac{C\xi_{\rho,t-1}}{\sigma_t(\lambda_0)}\left(1+\sup_{|\lambda-\lambda_0|\le b/\sqrt{n}}|\lambda-\lambda_0|\frac{|\partial\sigma_t^2(\lambda)/\partial\lambda|}{\sigma_t^2(\lambda_0)}\right) \\
&\le C\xi_{0,\rho_1,t-1}^{1-\iota_3}\left(1+\frac{bI_n}{\sqrt{n}}\right).
\end{aligned}$$

Putting the above together after choosing $s < 1 - \iota_3$, we obtain

$$I_n \leq C \max_{1\leq t\leq n} \xi_{0,\rho_1,t-1}^{1-\iota_3}\left(1+\frac{bI_n}{\sqrt{n}}\right).$$

Though I_n appears on the right hand side in the above, the extra $\sqrt{n}$ term will make it small so we can move it to the left hand side as long as

$$\max_{1\leq t\leq n} \xi_{0,\rho_1,t-1}^{1-\iota_3} = o_P\left(n^{1/2}\right)$$

which follows immediately by Assumption 9 and (3.3.10). In fact, Assumption 9 implies $E\xi_{0,\rho_1,t}^{(1-\iota_3)\kappa} < \infty$ and hence by (3.3.10)

$$\max_{1\leq t\leq n} \xi_{0,\rho_1,t-1}^{1-\iota_3} = o_P\left(n^{1/\kappa}\right).$$

Therefore we prove the first half of Lemma 3.3.3.

To prove the second half, we adapt the same approach as we use in the first half. We have

$$J_n \leq C \max_{1\leq t\leq n} \xi_{0,\rho_1,t-1}^{1-\iota_3}\left(1+\frac{bI_n}{\sqrt{n}}\right).$$

This proves the second half. Thus Lemma 3.3.3 is proved.

Lemma 3.3.4. *Under Assumptions 1 to 3, 6, 8 and 9, we have for any* $b > 0$

$$\max_{1\leq t\leq n} \sup_{|\lambda|\leq b} \frac{|g_t(\lambda)|}{\sqrt{n}} = o_P\left(n^{1/\kappa-1/2}\right)$$

and

$$\max_{1\leq t\leq n} \sup_{|\lambda|\leq b} |g_t(\lambda) - \langle \partial \log \sigma_t^2(\lambda_0)/\partial\lambda, \lambda\rangle| = o_P\left(n^{1/\kappa-1/2}\right).$$

Proof: The proof follows easily from Lemma 3.3.3 and one or two terms Taylor expansion of $\sigma_t^2(\lambda)$. The details are omitted.

Lemma 3.3.5. *Under Assumptions 1 to 3, 6, 8 and 9, there exists an integer M such that for any $b > 0$*

$$\max_{1\le t\le n}\sup_{|\lambda|\le b}\left|\frac{1}{\sqrt{1+n^{-1/2}g_t(\lambda)}}-\left(1+\sum_{i=1}^{M}\frac{w^{(i)}(0)}{i!}\frac{\langle\partial\log\sigma_t^2(\lambda_0)/\partial\lambda,\lambda\rangle^i}{n^{i/2}}\right)\right| = o_P\left(n^{-1/2}\right).$$

Proof: By M-term Taylor expansion of $w(x)$, when x is small,

$$w(x) = 1+\sum_{i=1}^{M}\frac{w^{(i)}(0)}{i!}x^i + O\left(x^{M+1}\right).$$

Thus Lemma 3.3.4 implies

$$\begin{aligned}\max_{1\le t\le n}\sup_{|\lambda|\le b}\left|\frac{1}{\sqrt{1+n^{-1/2}g_t(\lambda)}}-\left(1+\sum_{i=1}^{M}\frac{w^{(i)}(0)}{i!}\frac{g_t^i(\lambda)}{n^{i/2}}\right)\right| &= O_P\left(\max_{1\le t\le n}\sup_{|\lambda|\le b}\left|\frac{g_t(\lambda)}{\sqrt{n}}\right|^{M+1}\right)\\ &= o_P\left(n^{(M+1)(1/\kappa-1/2)}\right)\\ &= o_P\left(n^{-1/2}\right)\end{aligned}$$

if $M \ge \kappa/(\kappa-2)-1$. Now we can finish the proof of Lemma 3.3.5 if we show for each $i = 1, 2, \ldots, M$ that

$$\max_{1\le t\le n}\sup_{|\lambda|\le b}\left|g_t^i(\lambda)-\langle\partial\log\sigma_t^2(\lambda_0)/\partial\lambda,\lambda\rangle^i\right| = o_P\left(n^{(i-1)/2}\right). \qquad (3.3.24)$$

When $i = 1$, (3.3.24) follows directly from Lemma 3.3.4. Let us consider the cases $i = 2, 3, \ldots, M$. From Lemma 3.3.3 we have for each $i = 2, 3, \ldots, M$ that

$$\max_{1\le t\le n}\sup_{|\lambda|\le b}\left|\langle\partial\log\sigma_t^2(\lambda_0)/\partial\lambda,\lambda\rangle\right|^{i-1} = o_P\left(n^{(i-1)/\kappa}\right)$$

which, together with (3.3.21) and Lemma 3.3.4, implies

$$\begin{aligned}
&\max_{1\le t\le n}\sup_{|\lambda|\le b}\left|g_t^i(\lambda)-\langle\partial\log\sigma_t^2(\lambda_0)/\partial\lambda,\lambda\rangle^i\right| \\
&= o_P\left(n^{1/\kappa-1/2}\right)\left(\max_{1\le t\le n}\sup_{|\lambda|\le b}\left|\langle\partial\log\sigma_t^2(\lambda_0)/\partial\lambda,\lambda\rangle\right|^{i-1}+o_P\left(n^{(1/\kappa-1/2)(i-1)}\right)\right) \\
&= o_P\left(n^{1/\kappa-1/2}\right)\left(o_P\left(n^{(i-1)/\kappa}\right)+o_P\left(n^{(1/\kappa-1/2)(i-1)}\right)\right) \\
&= o_P\left(n^{(i-1)/2}\right)
\end{aligned}$$

since $\kappa > 2$. This proves (3.3.24) and Lemma 3.3.5.

Lemma 3.3.6. *Under Assumptions 1 to 3, 6, 8 and 9, for any $b > 0$ and $k \ge 1$, $E|\eta_0|^k < \infty$ implies that*

$$\sup_{0\le u\le 1}\sup_{|\lambda|\le b}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{[nu]}\left(\frac{\eta_t+n^{-1/2}Z_t(\gamma)}{\sqrt{1+n^{-1/2}g_t(\lambda)}}\right)^k\right.$$
$$\left.-\frac{1}{\sqrt{n}}\sum_{t=1}^{[nu]}\left(\frac{\eta_t}{\sqrt{1+n^{-1/2}g_t(\lambda)}}\right)^k-\frac{k}{n}\sum_{t=1}^{[nu]}\left(\frac{1}{\sqrt{1+n^{-1/2}g_t(\lambda)}}\right)^k\eta_t^{k-1}Z_t(\gamma)\right|=o_P(1).$$

Proof: The case $k = 1$ is trivial. We consider the case $k \ge 2$. By Newton's binomial formula, Lemma 3.3.6 follows by

$$\sup_{|\lambda|\le b}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{1}{\sqrt{1+n^{-1/2}g_t(\lambda)}}\right)^k\frac{|\eta_t|^{k-i}|Z_t(\gamma)|^i}{n^{i/2}}=o_P(1)$$

for each $i = 2,\ldots,k$. By Lemma 3.3.4, we can reduce the above to

$$\sum_{t=1}^{n}\frac{|\eta_t|^{k-i}\sup_{|\gamma|\le b}|Z_t(\gamma)|^i}{n^{(i+1)/2}}=o_P(1)$$

which can be proven if for each $i = 2,\ldots,k$

$$E\left(\sum_{t=1}^{n}\frac{|\eta_t|^{k-i}\sup_{|\gamma|\le b}|Z_t(\gamma)|^i}{n^{i/2}}\right)^{2/i}=O(1). \tag{3.3.25}$$

By Lemma 2.4.6 (i) and (v)

$$\sup_{|\gamma|\le b}|Z_t(\gamma)| \le \frac{b}{\sigma_t(\lambda_0)}\sup_{|\gamma-\gamma_0|\le b/\sqrt{n}}\left|\frac{\partial\varepsilon_t(\gamma)}{\partial\gamma}\right| \le \frac{bC\xi_{\rho,t-1}}{\sigma_t(\lambda_0)} \le C\xi_{0,\rho_1,t-1}^{1-\iota_3}. \tag{3.3.26}$$

Now we can easily prove (3.3.25) since η_t and $\xi_{0,\rho_1,t-1}$ are independent, $E|\eta_0|^k < \infty$ and $E\xi_{0,\rho_1,t-1}^{(1-\iota_3)\kappa} < \infty$. The proof of Lemma 3.3.6 is finished.

Lemma 3.3.7. *Under Assumptions 1 to 3, 6, 8 and 9, for any $b > 0$ and $k \ge 1$, $E|\eta_0|^k < \infty$ implies that*

$$\sup_{0\le u\le 1}\sup_{|\lambda|\le b}\left|\frac{1}{n}\sum_{t=1}^{[nu]}\left(\frac{1}{\sqrt{1+n^{-1/2}g_t(\lambda)}}\right)^k\eta_t^{k-1}Z_t(\gamma) - u\mu_{k-1}\langle\Gamma,\gamma\rangle\right| = o_P(1).$$

Proof: First we get rid of the term $g_t(\lambda)$ by using Lemma 3.3.4 and the same argument in proving (3.3.25). Mainly, by $|1/(1+x)^{k/2}-1| = O(x)$ for small x and (3.3.26), we have

$$\begin{aligned}
&\sup_{|\lambda|\le b}\frac{1}{n}\sum_{t=1}^{n}\left|\left(\frac{1}{\sqrt{1+n^{-1/2}g_t(\lambda)}}\right)^k\eta_t^{k-1}Z_t(\gamma) - \eta_t^{k-1}Z_t(\gamma)\right| \\
&= O_P\left(\max_{1\le t\le n}\sup_{|\lambda|\le b}\frac{|g_t(\lambda)|}{\sqrt{n}}\right)\frac{1}{n}\sum_{t=1}^{n}|\eta_t|^{k-1}\xi_{0,\rho_1,t-1}^{1-\iota_3} \\
&= o_P(1).
\end{aligned}$$

Next we need to prove that

$$\sup_{|\gamma|\le b}\frac{1}{n}\sum_{t=1}^{n}|\eta_t|^{k-1}\left|Z_t(\gamma) - \left\langle\frac{\partial\varepsilon_t(\gamma_0)/\partial\gamma}{\sigma_t(\lambda_0)},\gamma\right\rangle\right| = o_P(1). \tag{3.3.27}$$

To this end, we get by using two terms Taylor expansion and Lemma 2.4.6 (i) and (v),

$$\sup_{|\gamma|\le b}\left|Z_t(\gamma) - \left\langle\frac{\partial\varepsilon_t(\gamma_0)/\partial\gamma}{\sigma_t(\lambda_0)},\gamma\right\rangle\right| \le \frac{b}{\sqrt{n}}\frac{\sup_{|\gamma|\le b}|\partial^2\varepsilon_t(\gamma)/\partial\gamma\partial\gamma^T|}{\sigma_t(\lambda_0)} \le \frac{C\xi_{0,\rho_1,t-1}^{1-\iota_3}}{\sqrt{n}}.$$

This proves (3.3.27).

Finally we can prove Lemma 3.3.7 if

$$\sup_{0\le u\le 1}\sup_{|\gamma|\le b}\frac{1}{n}\left|\sum_{t=1}^{[nu]}\eta_t^{k-1}\left\langle\frac{\partial\varepsilon_t(\gamma_0)/\partial\gamma}{\sigma_t(\lambda_0)},\gamma\right\rangle - u\mu_{k-1}\left\langle\Gamma,\gamma\right\rangle\right| \tag{3.3.28}$$

The proof of (3.3.28) follows by taking $\sup_{|\gamma|\le b}$ into the inner product first, then applying Lemma 3.3.2 and noting that $\left\langle\frac{\partial\varepsilon_t(\gamma_0)/\partial\gamma}{\sigma_t(\lambda_0)},\gamma\right\rangle = X(\eta_{t-1},\eta_{t-2},\ldots)$ for an appropriate function X. This completes the proof of Lemma 3.3.7.

Lemma 3.3.8. *Under Assumptions 1 to 3, 6, 8 and 9, for any $b>0$ and $k\ge 1$, $E|\eta_0|^k<\infty$ implies that*

$$\sup_{|\lambda|\le b}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|\eta_t|^k\left|\left(\frac{1}{\sqrt{1+n^{-1/2}g_t(\lambda)}}\right)^k - \left(1+\sum_{i=1}^{M}\frac{w^{(i)}(0)}{i!}\frac{\langle\partial\log\sigma_t^2(\lambda_0)/\partial\lambda,\lambda\rangle^i}{n^{i/2}}\right)^k\right| = o_P(1).$$

Proof: Lemma 3.3.8 follows easily from (3.3.21), LLN, and Lemmas 3.3.4 and 3.3.5. The detail is omitted.

Lemma 3.3.9. *Under Assumptions 1 to 3, 6, 8 and 9, for any $b>0$ and $k\ge 1$, $E|\eta_0|^k<\infty$ implies that*

$$\sup_{|\lambda|\le b}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|\eta_t|^k\left|\left(1+\sum_{i=1}^{M}\frac{w^{(i)}(0)}{i!}\frac{\langle\partial\log\sigma_t^2(\lambda_0)/\partial\lambda,\lambda\rangle^i}{n^{i/2}}\right)^k - \left(1-\frac{\langle\partial\log\sigma_t^2(\lambda_0)/\partial\lambda,\lambda\rangle}{2\sqrt{n}}\right)^k\right| = o_P(1).$$

Proof: By (3.3.21) and then by taking $\sup_{|\gamma|\le b}$ into the inner products, we find the dominate term in the above is

$$\sum_{i=2}^{M}\sum_{t=1}^{n}|\eta_t|^k\frac{|\partial\log\sigma_t^2(\lambda_0)/\partial\lambda|^i}{n^{(i+1)/2}}$$

which is $o_P(1)$ by following the same way as we prove (3.3.25).

Lemma 3.3.10. *Under Assumptions 1 to 3, 6, 8 and 9, for any $b > 0$ and $k \geq 1$, $E|\eta_0|^k < \infty$ implies that*

$$\sup_{0\leq u\leq 1}\sup_{|\lambda|\leq b}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{[nu]}\eta_t^k\left(1-\frac{\langle\partial\log\sigma_t^2(\lambda_0)/\partial\lambda,\lambda\rangle^i}{2\sqrt{n}}\right)^k -\frac{1}{\sqrt{n}}S_n^{(k)}(u)+\frac{ku\mu_k}{2}\langle\Lambda,\lambda\rangle\right| = o_P(1).$$

Proof: By Newton's binomial formula and by using similar way in proving Lemma 3.3.9, the dominate term left in the above is

$$\sup_{0\leq u\leq 1}\sup_{|\lambda|\leq b}\left|\frac{1}{n}\sum_{t=1}^{[nu]}\eta_t^k\langle\partial\log\sigma_t^2(\lambda_0)/\partial\lambda,\lambda\rangle - u\mu_k\langle\Lambda,\lambda\rangle\right|$$

which follows easily from Lemma 3.3.2. Now we complete the proof of Lemma 3.3.10.

Chapter 4

Diagnostic Test of ARMA-GARCH Models

With the results in Chapter 2 and 3, we can investigate some properties of QMLE and conduct model diagnostic tests based on residuals with numerical examples.

In fitting ARMA-GARCH, we notice that Splus module S+FinMetrics version 1 and 2 does not scale the QMLE. One scaling approach based on outputs from the Splus S+FinMetrics is introduced. Then we verify numerically the relative efficiency of QMLE based on different likelihood kernels. Fitting ARMA-IGARCH model is also considered.

We also study by Monte Carlo simulation the residual-based diagnostic tests like: CUSUM test for model structural change and Jarque-Bera test for normality of innovation.

Finally, some open problems are presented as the future work.

4.1 Fitting ARMA-GARCH Models

4.1.1 Fitting ARMA-GARCH by Splus module S+FinMetrics

Usually $\mathbf{E}\eta_0^2 = 1$ is assumed to identify model (1.2.13). As mentioned in Remark 2.2.4 in Chapter 2, when we fit data by a likelihood kernel other than the standard normal density, we may have to scale η_t such that $\eta_t^{**} = a\eta_t$. As showed in Section 2.3.5, it results in a scaling of $\alpha_0, \alpha_1, \cdots, \alpha_p$ only. If a is known, after estimating λ^{**}, we can scale the estimators $\tilde{\alpha}_{0n}^{**}, \tilde{\alpha}_{1n}^{**}, \cdots, \tilde{\alpha}_{pn}^{**}$ by multiplying a^2 back to obtain estimators of $\alpha_0, \alpha_1, \cdots, \alpha_p$ in the original model.

The algorithm of fitting ARMA-GARCH model in Splus module S+FinMetrics version 1 and 2 does not scale the estimates after the parameters are estimated. In addition, standardized residuals, the estimates of conditional variances and asymptotic variances are not scaled either. This could lead to wrong inference and poor prediction.

In the following, we use an example of ARMA(1,1)-GARCH(1,1) to show the scaling problem and give an approach to amend it based on the Splus outputs.

```
> module(finmetrics)
> data <- sim.arma.garch(n = 10000, n0 = 500, arch = c(0.0002, 0.2),
   garch = c( 0.5), dist.par = 0, mu = 0, ar = c(0.4), ma = c(0.6))
```

Instead using the command of "simulate.garch" in S+FinMetrics, we write our own command "sim.arma.garch" (the code is appended) to generate the data. This command produces an output including an ARMA-GARCH series $\{y_t, 1 \le t \le n\}$, GARCH errors $\{\varepsilon_t, 1 \le t \le n\}$, GARCH innovations $\{\eta_t, 1 \le t \le n\}$ and conditional variances $\{\sigma_t^2, 1 \le t \le n\}$. The model parameters are given in the command. *dist.par* $= 0$ means the innovation is generated from the standard normal distribution. n is sample size and n_0 is the starting value. Figure (4.1) display the simulated data.

```
> par(mfrow = c(2, 2))
> tsplot(data$series, main = "Series")
> tsplot(data$error, main = "Error")
> tsplot(data$sigma.sq, main = "Sigma")
> tsplot(data$innov, main = "Innovation")
```

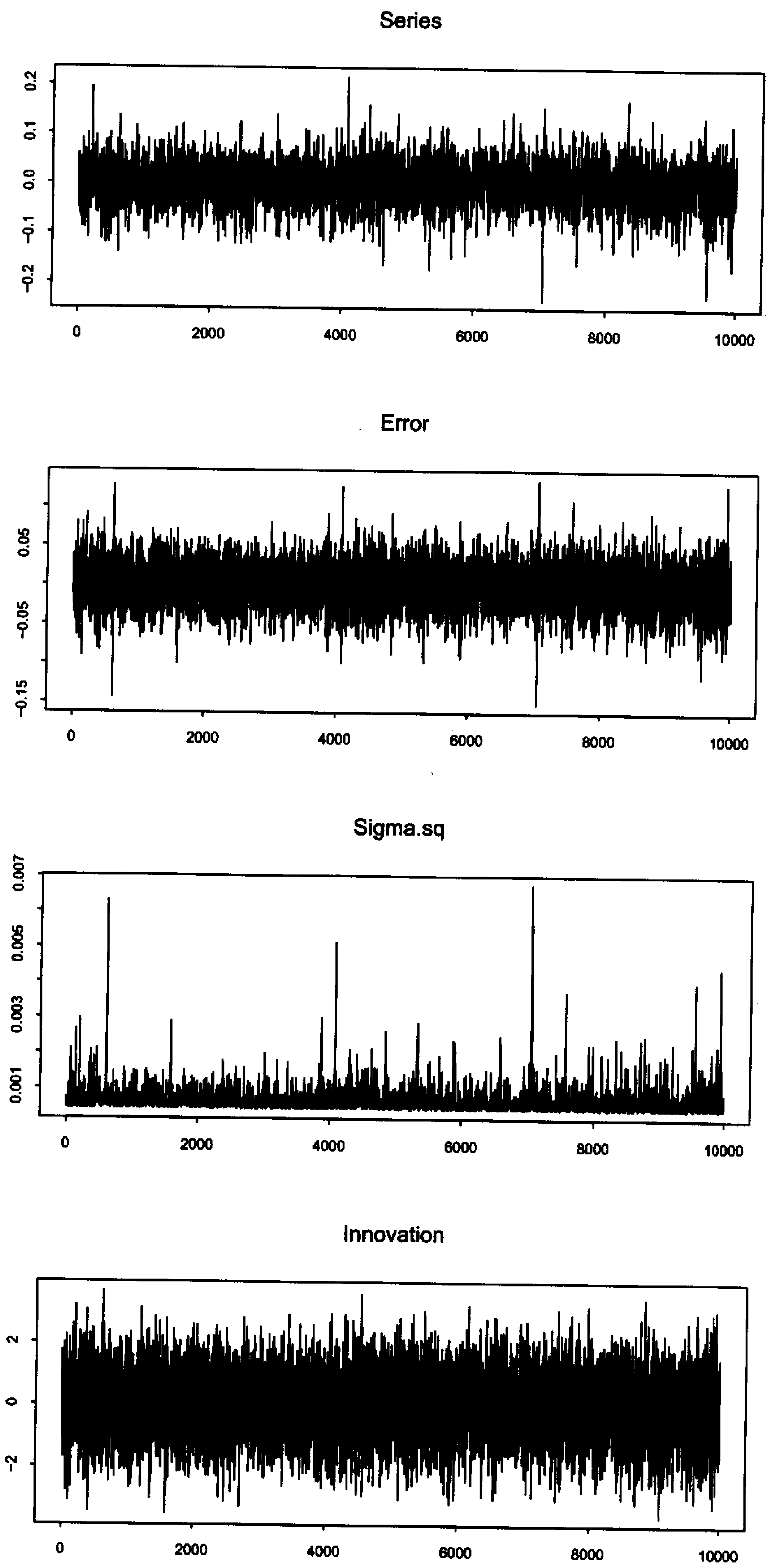

Figure 4.1: Plots of y_t, ε_t , σ_t^2, η_t.

We fit the series y_t with an ARMA(1,1)-GARCH(1,1) model based on the standard normal kernel and the student t(3) kernel respectively by the command "garch" built in S+FinMetrics. Since $\eta_t \sim N(0,1)$, we have $\mathbf{E}\eta_0 = 0$ and $\mathbf{E}\eta_0^2 = 1$. By Proposition 2.2.1, Assumption 5 is satisfied, so we do not need to scale the model when fit the data with standard normal kernel. The estimators actually are MLE. But we have to scale η_t when applying the t(3) kernel. The scale parameter a is chosen such that $\mathbf{E}a\eta_0/(3+1) = 0$ and $\mathbf{E}(3+(a\eta_0)^2)^{-1} = 1/(3+1)$. Solve the equations, we have $a \approx 1.26$.

```
> series <- data$series
> fit.nm <- garch(series = series, formula.mean =  ~ -1 + arma(1, 1),
    formula.var =  ~ garch(1, 1), cond.dist = "gaussian", trace = FALSE)
> fit.t3 <- garch(series = series, formula.mean =  ~ -1 + arma(1, 1),
    formula.var =  ~ garch(1, 1), cond.dist = "t", dist.par = 3, dist.est
     = F, trace = FALSE)
```

We have shown in Chapter 2, both of these (Q)MLEs are asymptotically consistent and normally distributed even the likelihood kernels are different. After scaling, both of these estimators should be very close to the true values.

```
> coef.nm.splus <- fit.nm$coef
> coef.t3.splus <- fit.t3$coef
> list(coef.nm.splus, coef.t3.splus)
[[1]]:
                COEF
   AR(1) 0.3850685255
   MA(1) 0.6283150831
```

```
       A 0.0001963835
 ARCH(1) 0.1994200015
GARCH(1) 0.5051757511

[[2]]:
                 COEF
   AR(1) 0.3872451994
   MA(1) 0.6222308782
       A 0.0003761551
 ARCH(1) 0.3720943546
GARCH(1) 0.5062610707
```

It can be seen that estimators of ϕ_1, φ_1, β_1 are very close to true values in both two fittings. But the estimates of α_0 and α_1 from the two fittings are quite different. Estimates of α_0 and α_1 based on the normal kernel fit are closer to the true value than the estimates based on the t(3) kernel.

Since $\mathbf{E}\eta_0^2 = 1$, we would expect the variances of standardized residuals from the two fittings to be close to 1.

```
> res.nm.splus <- residuals(fit.nm, st = T)
> res.t3.splus <- residuals(fit.t3, st = T)
> var(res.nm.splus)
[1] 1.002256
> var(res.t3.splus)
[1] 0.5267372
> var(data$innov)
[1] 1.000595
```

It can be seen that the sample variance of standardized residuals based on the t(3) fitting is far away from 1.

Not only are the parameter estimators and the residuals of fit based on the t(3) kernel not scaled, neither are the estimated conditional standard deviation sequence $\tilde{\sigma}_t$ and the estimated asymptotic variance. One approach of solving this scaling issue is to apply a correction parameter a_{t3}. Denote $\tilde{\eta}_{t3s_i}$, $i = 1, \cdots, n$ be the standardized residuals given by S+finMetrics based on the t(3) fitting. Since $\mathbf{E}\eta_0^2 = 1$, we expect the sample variance of the standardized residuals to be close to 1. So a_{t3} is set to $n/\sum_{i=1}^{n}(\tilde{\eta}_{t3s_i})^2$.

```
> a.t3 = 1/mean(res.t3.splus^2)
> a.t3
[1] 1.898638
```

This correction parameter is just used to scale the Splus estimation. Now we can use it to correct the problems. For example:

(i) Rescale the QMLES:

Denote $\tilde{\phi}_{1t3s}$, $\tilde{\varphi}_{1t3s}$, $\tilde{\beta}_{1t3s}$ $\tilde{\alpha}_{0t3s}$, $\tilde{\alpha}_{1t3s}$ as the estimator based on the t(3) kernel fitting given by S+FinMetrics. Let $\tilde{\phi}_{1t3c}$, $\tilde{\varphi}_{1t3c}$, $\tilde{\beta}_{1t3c}$ $\tilde{\alpha}_{0t3c}$, $\tilde{\alpha}_{1t3c}$ be the properly rescaled estimators based on the t(3) kernel. To correct the estimators, let $\tilde{\phi}_{1t3c} = \tilde{\phi}_{1t3s}$, $\tilde{\varphi}_{1t3c} = \tilde{\varphi}_{1t3s}$, $\tilde{\alpha}_{0t3c} = \tilde{\alpha}_{0t3s}/a_{t3}$, $\tilde{\alpha}_{1t3c} = \tilde{\alpha}_{1t3s}/a_{t3}$, $\tilde{\beta}_{1t3c} = \tilde{\beta}_{1t3s}$.

```
> coef.t3.correct <- coef.t3.splus/c(1, 1, a.t3, a.t3, 1)
> coef.t3.correct
                 COEF
   AR(1) 0.3872451994
```

```
   MA(1) 0.6222308782
       A 0.0001981184
 ARCH(1) 0.1959796218
GARCH(1) 0.5062610707
```

The modified QMLE of the t(3) kernel fitting now are very close to the true parameters.

(ii) Correct the standardized residuals:

We begin with a look of the density of the standardized residuals (in Figure 4.2) of the two fittings given by Splus Finmatrics.

```
> plot(density(res.t3.splus), type = "p", col = 1, pch = 2, main =
    "Density Plots of Splus Residuals of two fittings", xlab = "", ylab =
    "")
> lines(density(res.nm.splus), type = "p", col = 5, pch = 0)
> lines(density(data$innov), type = "p", col = 6, pch = 3)
> legend(1.6, 0.5, legend = c("t(3)", "normal", "original"), marks = c(2, 0,
    3), col = c(1, 5, 6,  ), bty = "n")
```

The density of residuals based on the normal kernel fitting overlaps with that of the true innovation, which implies the fitting is good. There is a big difference between the density of residuals based on the t(3) kernel and that of the true innovation.

Denote correct standardized residuals as $\tilde{\eta}_{t3c_i}$. Then we correct the residuals by $\tilde{\eta}_{t3c_i} = \sqrt{a_{t3}}\tilde{\eta}_{t3s_i}$.

```
> res.t3.correct <- res.t3.splus * sqrt(a.t3)
```

Plot again the density (in Figure 4.2) of the modified residuals.

```
> plot(density(res.t3.correct), pch = 2, col = 1, main =
    "Density Plot of Modified Residuals", xlab = "", ylab = "")
> lines(density(res.nm.splus), type = "p", pch = 0, col = 2)
> lines(density(data$innov), type = "p", pch = 3, col = 3)
> legend(2, 0.38, legend = c("t(3)", "normal", "original"), marks = c(2, 0,
    3), col = c(1, 5, 6,  ), bty = "n")
```

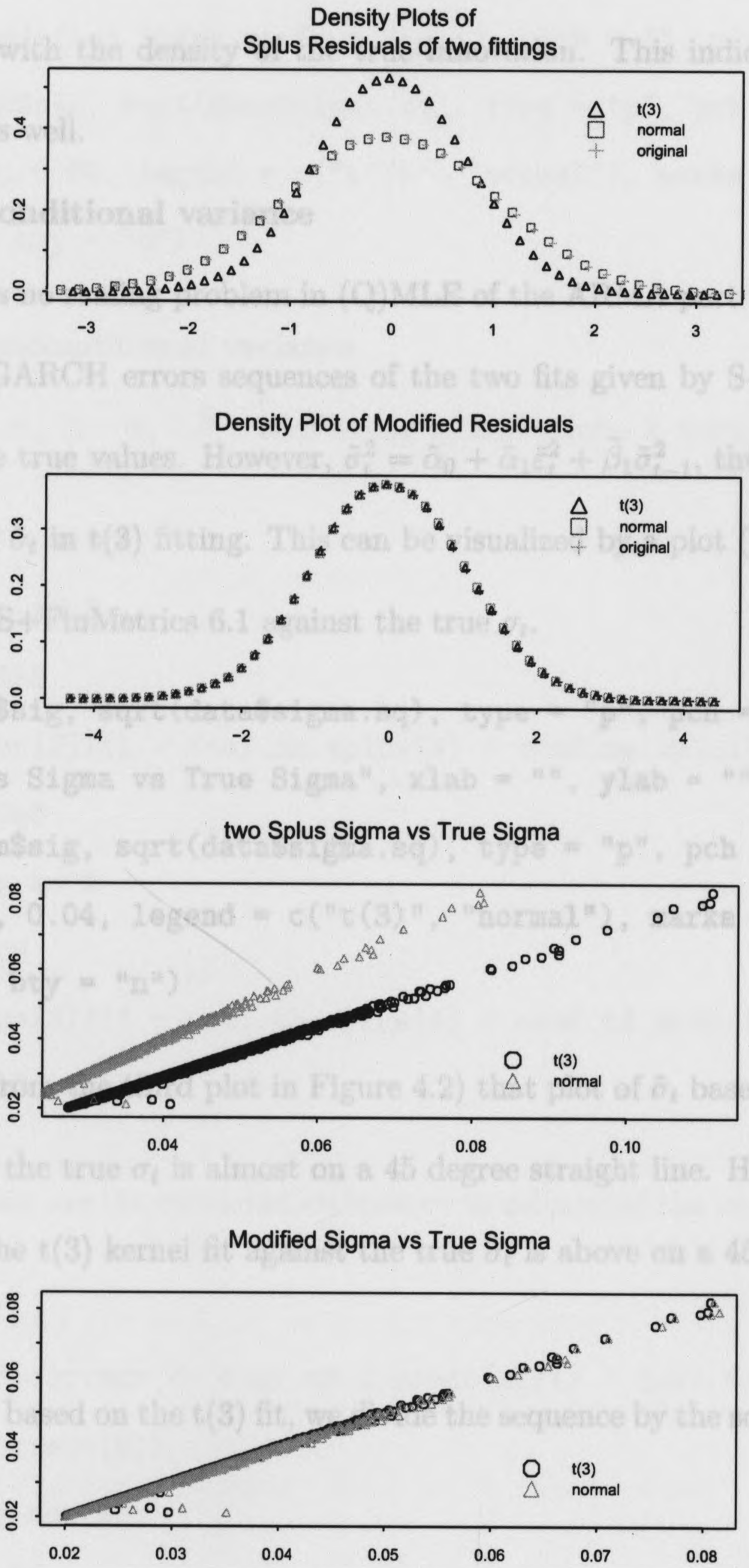

Figure 4.2: Plots of QMLEs density and conditional standard deviation, scaled vs not scaled

The plots shows, after scaling, both the densities of residuals from the two fits almost overlap with the density of the true innovation. This indicates the scaling parameter works well.

(iii) Correct conditional variance

Since there is no scaling problem in (Q)MLE of the ARMA part for both fits, the two estimated GARCH errors sequences of the two fits given by S+FinMetrics are very close to the true values. However, $\tilde{\sigma}_t^2 = \tilde{\alpha}_0 + \tilde{\alpha}_1\tilde{\varepsilon}_t^2 + \tilde{\beta}_1\tilde{\sigma}_{t-1}^2$, thus wrong QMLE results in wrong $\tilde{\sigma}_t$ in t(3) fitting. This can be visualized by a plot (in Figure 4.2) of the $\tilde{\sigma}_t$ given by S+FinMetrics 6.1 against the true σ_t.

```
> plot(fit.t3$sig, sqrt(data$sigma.sq), type = "p", pch = 1, col = 1, main =
    "two Splus Sigma vs True Sigma", xlab = "", ylab = "")
> lines(fit.nm$sig, sqrt(data$sigma.sq), type = "p", pch = 2, col = 2)
> legend(0.08, 0.04, legend = c("t(3)", "normal"), marks = c(1, 2), col = c(
    1, 2,  ), bty = "n")
```

It can be seen (from the third plot in Figure 4.2) that plot of $\tilde{\sigma}_t$ based on the normal kernel fit versus the true σ_t is almost on a 45 degree straight line. However, the plot of $\tilde{\sigma}_t$ based on the t(3) kernel fit against the true σ_t is above on a 45 degree straight line.

To correct $\tilde{\sigma}_t$ based on the t(3) fit, we divide the sequence by the scaling parameter as:

```
> sigma.t3.correct = fit.t3$sig/a.t3^0.5
```

It can be seen (from the forth plot in Figure 4.2) that plots of the corrected sequences versus the true σ_t overlap and almost lie on a 45 degree straight line.

```
> plot(sigma.t3.correct, sqrt(data$sigma.sq), type = "p", pch = 1, col = 1,
    main = "Modified Sigma vs True Sigma", xlab = "", ylab = "")
> lines(fit.nm$sig, sqrt(data$sigma.sq), type = "p", pch = 2, col = 2)
> legend(0.06, 0.04, legend = c("t(3)", "normal"), marks = c(1, 2), col = c(
    1, 2,  ), bty = "n")
```

(iv) Correct unconditional variance

Since $\mathbf{E}\tilde{\sigma}_t^2 = \tilde{\alpha}_0/(1-\tilde{\alpha}_1-\tilde{\beta}_1)$, thus wrong QMLE results in wrong $\mathbf{E}\tilde{\sigma}_t^2$ in the t(3) kernel fit.

```
> fit.nm$asymp.sd^2
[1] 0.0006647958
> coef.nm.splus[3]/(1 - coef.nm.splus[4] - coef.nm.splus[5])
[1] 0.0006647958
> fit.t3$asymp.sd^2
[1] 0.003092248
> coef.t3.splus[3]/(1 - coef.t3.splus[4] - coef.t3.splus[5])
[1] 0.003092248
```

To correct, we can use the corrected estimators to calculated the unconditional variance.

```
> asymp.var.t3.correct <- coef.t3.correct[3]/(1 - coef.t3.correct[4] -
    coef.t3.correct[5])
> asymp.var.t3.correct
[1] 0.0006653643
```

Now it is very close to the unconditional standard deviation based on the normal kernel fitting.

(v) Check our assumption

Assumption 5 can be verified by checking if $\mathbf{E}\{1/(3+(a\eta_t)^2)\}$ is 1/4.

With residuals based on t(3) kernel given by Splus, we have:

```
> mean(1/(3 + res.t3.splus^2))
[1] 0.2924377
```

which is a little bit away from 0.25.

With the rescaled residuals, we have:

```
> mean(1/(3 + (1.26 * res.t3.correct)^2))
[1] 0.2497764
```

which is very close to 0.25.

This implies that the scale parameter correctly modified the fit.

4.1.2 Efficiency of QMLE

(i) Pure ARMA

As shown in Theorem 2.2.4 in Chapter 2, with α_0 being nuisance parameter, the asymptotic variance of $\tilde{\gamma}_n$ for pure ARMA is $\tau^2_{arma}\mathcal{D}_\varepsilon^{-1}$. For a given distribution of ε_t, $\mathcal{D}_\varepsilon$ is not determined by the choice of kernel. Thus efficiency of the QMLE depends on τ^2_{arma}, which is determined by the likelihood kernel h. Table 4.1 lists some τ^2_{arma} based on several different distribution of ε_t and likelihood kernels.

It can be seen from Table 4.1 that for a given distribution of ε_t, the MLE is the most efficient. The closer is the likelihood kernel to the density of η_t, the smaller is τ^2_{arma}. It also can be seen that when η_t is heavy-tailed, fitting with the normal kernel

Table 4.1: τ^2_{arma} for different h and distributions of η_t

	$\eta_t \sim t(3)$	$\eta_t \sim t(6)$	$\eta_t \sim SDE$	$\eta_t \sim N(0,1)$
$h \sim t(3)$	1.499	1.307	1.278	1.110
$h \sim t(6)$	1.532	1.286	1.375	1.049
$h \sim SDE$	1.830	1.687	1.004	1.552
$h \sim N(0,1)$	2.993	1.5	2	1

Table 4.2: τ^{**2} for different h and distributions of η_t

	$\eta_t \sim$ GED(.5)	$\eta_t \sim$ GED(1)	$\eta_t \sim$ GED(1.5)	$\eta_t \sim$ N(0,1)	$\eta_t \sim$ t(6)
$h \sim$ SGED(.5)	2.000	1.090	0.837	0.720	0.850
$h \sim$ SGED(.8)	2.122	1.012	0.735	0.614	0.784
$h \sim$ SGED(1)	2.330	1.000	0.698	0.570	0.778
$h \sim$ SGED(1.5)	3.444	1.065	0.667	0.514	0.879
$h \sim$ N(0,1)	6.050	1.250	0.690	0.500	1.250
$h \sim$ SGED(3)	27.127	2.110	0.867	0.543	17.128
$h \sim$t(6)	2.223	1.032	0.700	0.554	0.750
$h \sim$t(12)	2.413	1.035	0.674	0.520	0.774

is less efficient than fitting with the student's t kernel. However, for a fixed kernel h, it is hard to compare the efficiency among different ε_t, since the variance depends on both τ^2_{arma} and $\mathcal{D}_\varepsilon^{-1}$. $\mathcal{D}_\varepsilon^{-1}$ is decided by distribution of ε_t and model parameters.

(ii) Pure GARCH

As to pure GARCH, it can been seen from Section 2.3.5, that for a given distribution of η_t, asymptotic variance of $\tilde{\delta}_n$ is determined only by τ^{**2}, which depends on the choice of h. Table 4.2 presents some τ^{**2} for several distributions of η_t and likelihood kernels.

In Table 4.2, GED(v) is the generalized error distribution (Nelson, 1991). GED(1) is equivalent to double exponential distribution or Laplace distribution. GED(2) is equivalent to the standard normal distribution.

Table 4.2 shows that for a fixed distribution of η_t, the closer is the likelihood kernel to the density of η_t, the smaller is the τ^{**2}. Similar to the findings in the Table 4.1, Table 4.2 shows that when η_t is heavy-tailed, fitting by the normal kernel is less efficient than fitting by a student t kernel. Similarly for a fixed kernel h, it is hard to compare the efficiency among different distribution of η_t, since the variance depends on both τ^{**2} and $\mathcal{D}^{-1}$. $\mathcal{D}^{-1}$ is decided by the distribution of η_t and model parameters.

(iii) Pure GARCH with Nonzero Mean

Theorem 2.2.5 in Chapter 2 implies that the common practice of using the sample mean to center financial data is workable when $\mathbf{E}|\varepsilon_t|^2$ is finite. While Corollary 2.2.1 shows that estimation of the mean term by the sample average is less efficient.

Lu (2001) showed by a simulation study that the asymptotic efficiency ratio of sample average estimator and QMLE of the mean term c depends on other parameters. When the model is close to IGARCH, QMLE is much more efficient than the sample average estimator. Table 4.3 presents some simulation results of GARCH(1,1) with $\eta_0 \sim N(0,1)$ and samples size 10000 for different parameters, repeated 2000 times. When $\alpha_1 + \beta_1$ is fixed, the efficiency ratio seems to depend more on α_1.

Table 4.3: Ratio of $Var(\bar{\varepsilon}_n)$ and σ_c^2

α_0	α_1	β_1	$Var(\bar{\varepsilon}_n)/\sigma_c^2$
0.05	0.9	0.05	13.7
0.05	0.85	0.1	12.1
0.05	0.8	0.15	11.1
0.05	0.75	0.2	9.7
0.05	0.7	0.25	8.6
0.05	0.65	0.3	7.6
0.05	0.5	0.45	5.0
0.05	0.3	0.65	2.7
0.05	0.1	0.85	1.2
0.05	0.75	0.05	3.6
0.05	0.7	0.1	3.3
0.05	0.6	0.2	2.6
0.05	0.5	0.3	2.2
0.05	0.4	0.4	1.8
0.05	0.45	0.05	1.6
0.05	0.4	0.1	1.4
0.05	0.3	0.2	1.3

(iv) For ARMA-GARCH

Theorem 2.2.2 in Chapter 2 shows the variance of QMLE depends on the choice of the likelihood kernel and the distribution of the innovation η_t. Under some conditions, QMLE for the ARMA part and the GARCH part are asymptotically independent and the efficiency of the QMLE for the GARCH part is decided by τ^{**2} for a fixed distribution of η_t. In general, it is hard to compare the efficiency of QMLE for the ARMA part.

We verify the result by a simulation study. First we generate ARMA(1,1)-GARCH(1,1) data with sample size 10000, $\phi_1 = 0.4$, $\varphi_1 = 0.6$, $\alpha_0 = 0.0002$, $\alpha_1 = 0.2$, $\beta_1 = 0.5$, and η_t from the standard normal distribution. Then we fit the data with the standard normal kernel and the student t(3) kernel respectively. The estimates from the t(3) fit have been adjusted as discussed in Section 4.1.1. We repeat this procedure 5000 times. Thus we have 5000 estimates for each true parameters. The sample means and variances of these 5000 replications should be very close to the true mean parameters and asymptotic variances of the (Q)MLEs.

Let's first have a look of the average across all 5000 replicates of the two fits for each parameter.

```
> mean.n1 = mean(coef.n[, 1])
> mean.t1 = mean(coef.t3[, 1])
> mean.n2 = mean(coef.n[, 2])
> mean.t2 = mean(coef.t3[, 2])
> mean.n3 = mean(coef.n[, 3])
> mean.t3 = mean(coef.t3[, 3])
> mean.n4 = mean(coef.n[, 4])
> mean.t4 = mean(coef.t3[, 4])
> mean.n5 = mean(coef.n[, 5])
> mean.t5 = mean(coef.t3[, 5])
> c(mean.n1, mean.n2, mean.n3, mean.n4, mean.n5)
[1] 0.3984329078 0.6013272417 0.0002019203 0.1992662602 0.4965470515
> c(mean.t1, mean.t2, mean.t3, mean.t4, mean.t5)
[1] 0.3988031599 0.6003895511 0.0002039537 0.1975628389 0.4957456241
```

It shows that the means of these (Q)MLEs for the two fits are close to the true values.

Next we plot the density of (Q)MLEs of the two fits (in Figure 4.3).

It can be seen from Figure 4.3 that the density of QMLEs based on the t(3) kernel fit is more spread, which implies the t(3) kernel fit is less efficient than the normal kernel fit.

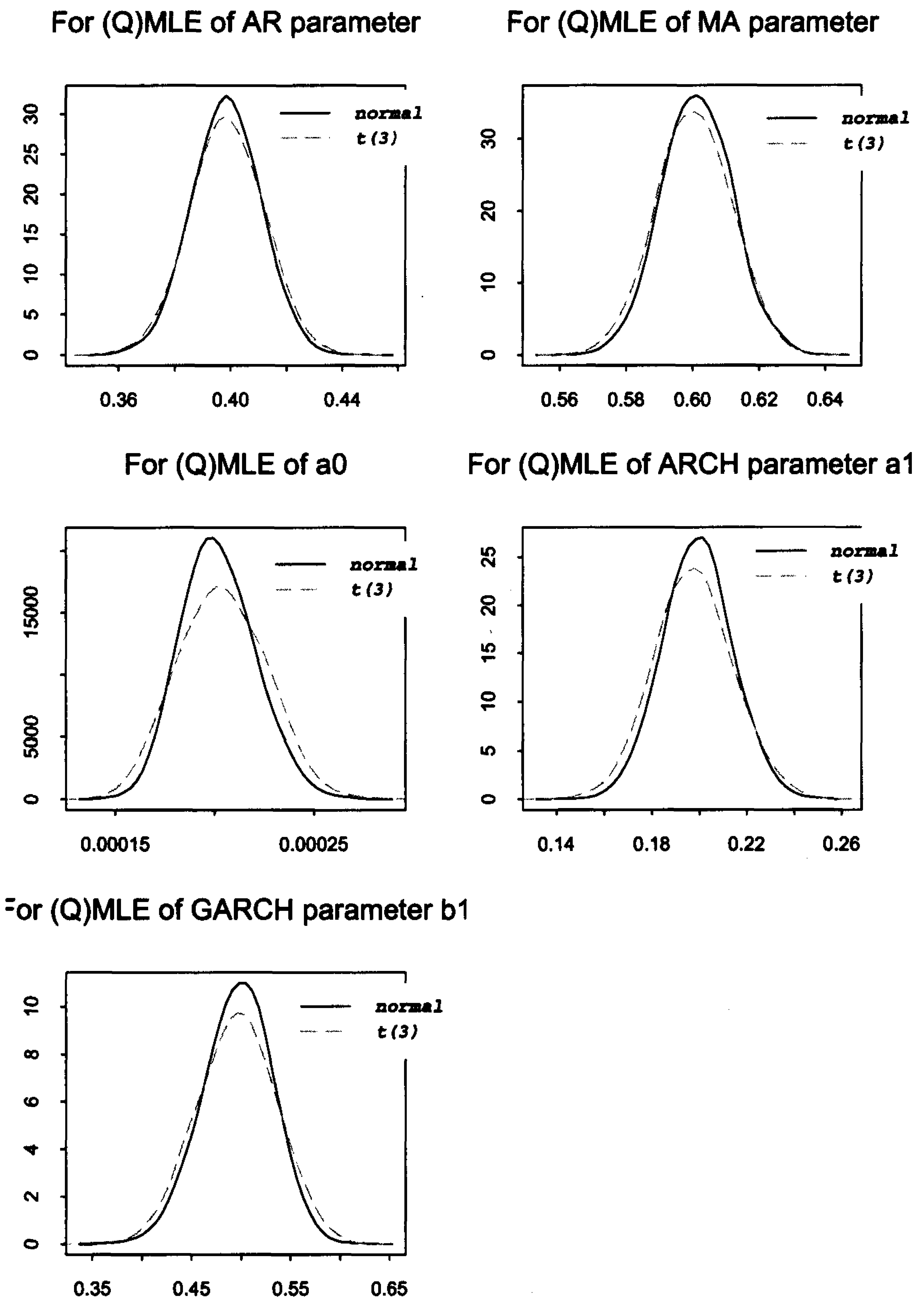

Figure 4.3: Density Plots of (Q)MLEs for the two fits

A direct calculation of the variance, we have

```
> var.n1 = var(coef.n[, 1])
> var.t1 = var(coef.t3[, 1])
> var.n2 = var(coef.n[, 2])
> var.t2 = var(coef.t3[, 2])
> var.n3 = var(coef.n[, 3])
> var.t3 = var(coef.t3[, 3])
> var.n4 = var(coef.n[, 4])
> var.t4 = var(coef.t3[, 4])
> var.n5 = var(coef.n[, 5])
> var.t5 = var(coef.t3[, 5])
> c(var.n1/var.t1, var.n2/var.t2, var.n3/var.t3,
   var.n4/var.t4, var.n5/var.t5)
[1] 0.8529636 0.8701201 0.7046332 0.7781677 0.7672930
```

It can be seen that the efficiency of the ARMA part and the GARCH part is different. It also confirms that the MLE from the normal kernel fit is more efficient than QMLE from the t(3) kernel fit.

We also fit the data by the student t distributions with other degrees of freedom. It shows when degree of freedom is big enough, there are no big difference between the efficiency of the QMLEs based on the normal kernel and the student t kernel.

4.1.3 Two Step Estimation of ARMA-GARCH

We show this by a simulation example. First we generate ARMA(1,1)-GARCH(1,1) data with sample size 2000 and $c = 1$, $\phi_1 = 0.4$, $\varphi_1 = 0.5$, $\alpha_0 = 0.005$, $\alpha_1 = 0.3$, $\beta_1 = 0.6$, η_t from the standard normal distribution. For two step estimation, since ε_t

is heavy-tailed, by Theorem 2.2.4, QMLE based on a student t kernel will be more efficient than that based on the normal kernel. We first fit the data with ARMA model by the standard normal kernel and the student t(3) respectively. Estimates of c, ϕ_1, φ_1 and residual sequences e_t are obtained correspondingly. Then using $\{e_t, 1 \leq t \leq n\}$ as observations, we obtain the estimates of $\alpha_0, \alpha_1, \beta_1$ by fitting these two sequences of $\{e_t, 1 \leq t \leq n\}$, respectively, with GARCH model based on the normal kernel. For the ARMA parameters, by using the estimators from the t(3) kernel as initial values, we obtained the local QMLEs by one step replication based on normal kernel and the t(3) kernel respectively as given by (3.9) in Ling (2005). As to one step estimation, we fit the data with ARMA-GARCH based on the normal kernel.

Since $\eta_t \sim N(0,1)$, we do not need to scale the estimators for the GARCH part. Note that the c in our model is different from what is in the Splus. Since $\phi_1 = 0.4$, $c = 1$ in our model implies $c = .6$ in Splus.

We repeat this procedure 2000 times. Thus we have 2000 estimations for both one step and two step estimations.

Since ε_t is not independent and presents ARCH effect, fitting ARMA-GARCH data by ARMA model may lead to bad estimation of ARMA parameters.

Also since $\eta_t \sim N(0,1)$ is symmetric about 0 and h is the normal kernel. We have $\mathbf{E}(g^2(\eta_0)\eta_0) = 0$ and $\mathbf{E}(g'(\eta_0)\eta_0) = 0$. Thus by Theorem 2.2.5, the variance of the estimators of the ARMA part does not affect the variance of the estimator of the GARCH parameters in the two step estimation. By Remark 2.2.10, estimators of the ARMA parameters and the GARCH parameters are asymptotically independent in one step estimation.

Figure 4.4 plots the densities (the curve of histogram of the 2000 estimations) of the QMLEs for the ARMA parameters. It can be seen that the density of the MLE from two step estimation are heavier than those from one step estimation. Also it can be seen that density of QMLEs based on the normal kernel are heavier than that based on the t(3) kernel, which is confirmed from Table 4.1. With one step replication, the efficiency of local QMLE is somewhat improved. But they are still less accurate then one step estimation.

Figure 4.5 plots the densities the QMLEs for the GARCH parameter. The variances of one step and two step estimations are almost the same. It can be seen that the densities of the estimators in both one step and two step estimations almost overlap.

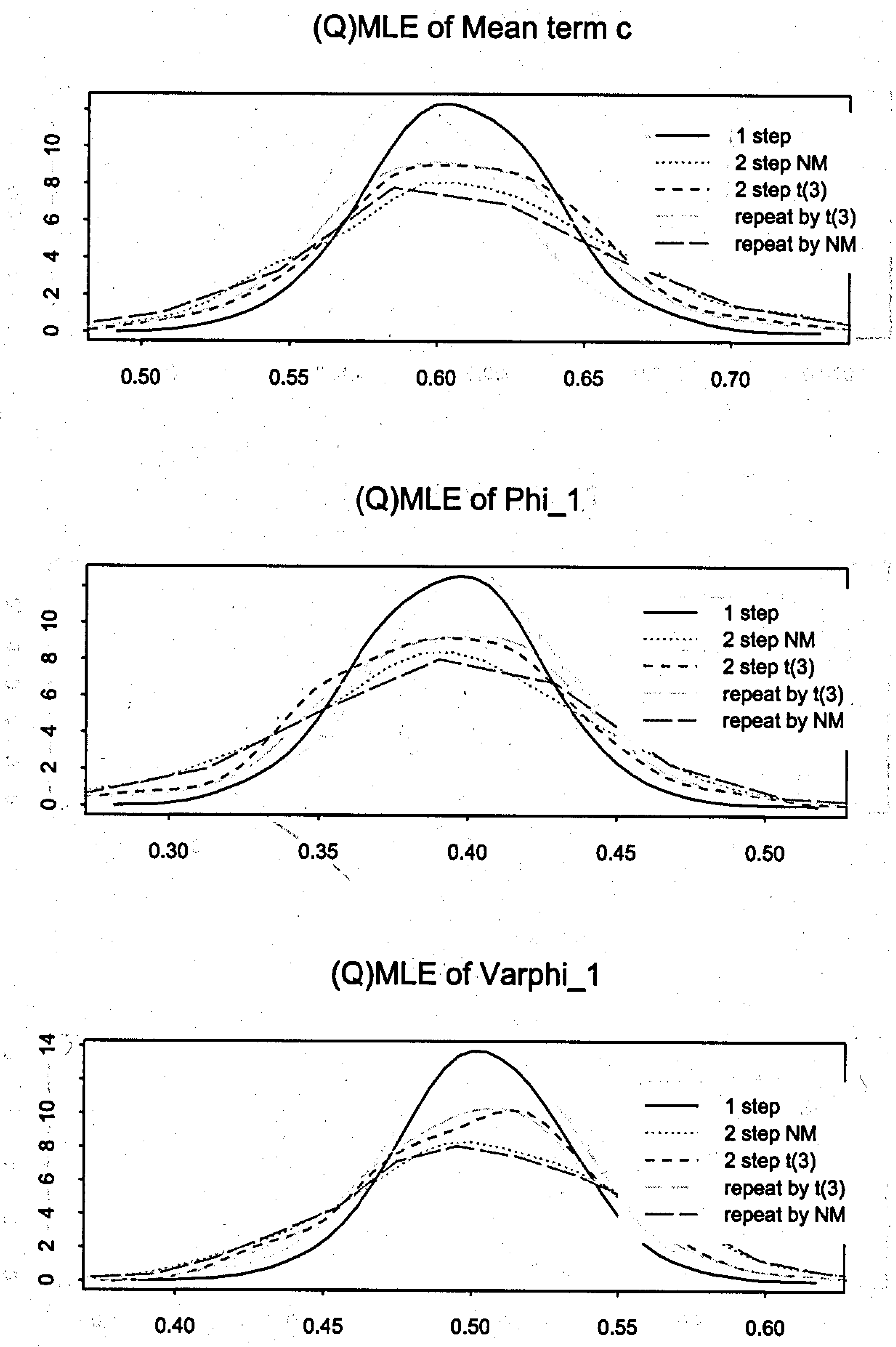

Figure 4.4: Density Plot of (Q)MLEs for ARMA parameters from One Step and Two Step Estimations

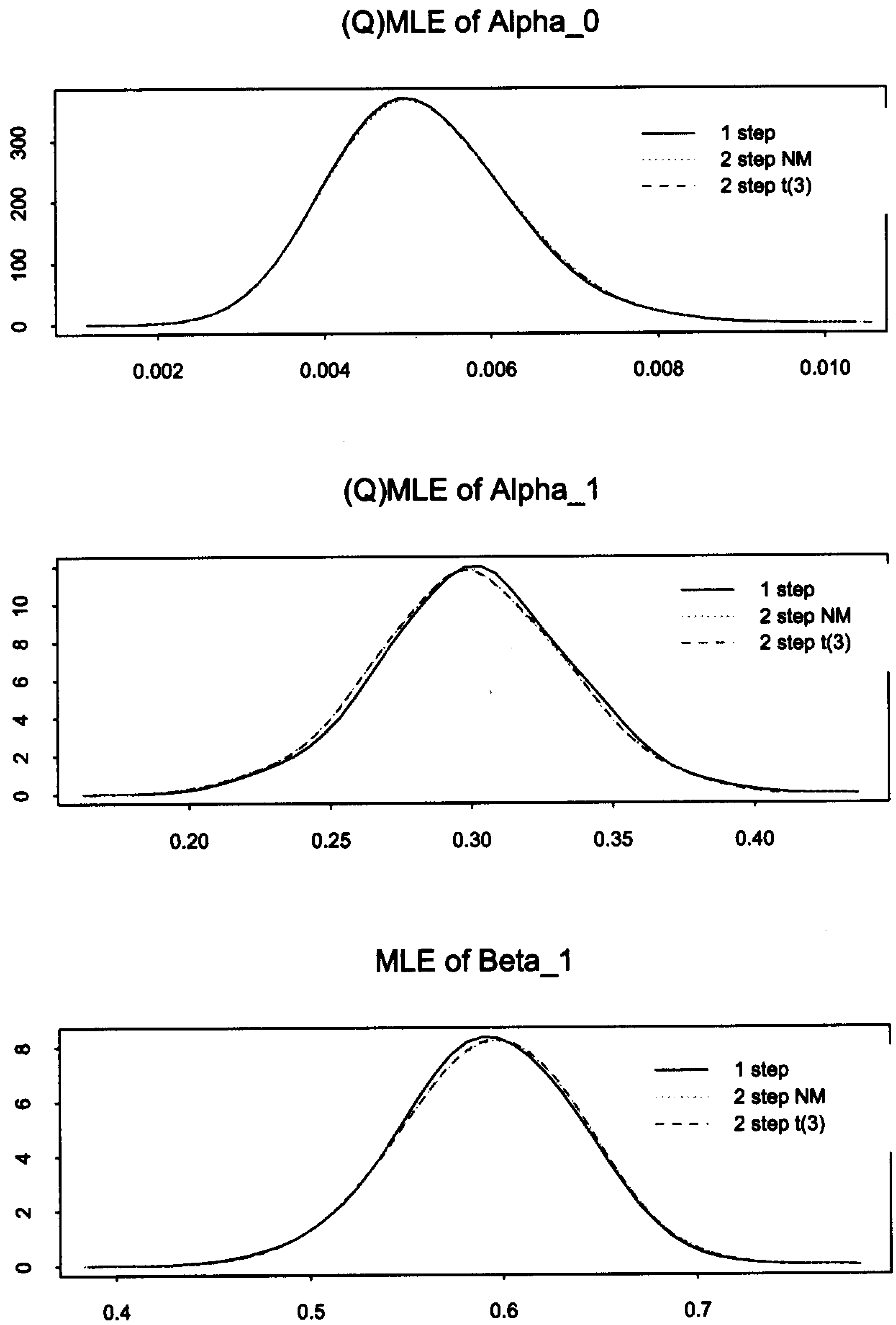

Figure 4.5: Density Plot of (Q)MLEs for GARCH parameters from One Step and Two Step Estimations

4.2 Model Diagnosis

As introduced in Chapter 1, after fitting a model, it is a good practice to test the model assumptions like: randomness of residuals, remainder ARCH effect, structural change, and distribution of residuals and etc.. In this section, we only present some numeric examples based on results from Chapters 2 and 3, in particular the CUSUM test for the change point problem, and Jarque-Bera test for distribution of residuals.

4.2.1 Change Point Problems

In the modelling of financial time series analysis, usually the sequences are assumed to be stationary or the model parameters is assumed to be constant over the time period. However financial time series often suffer from structural changes due to changes in political and social events. Ignoring this can lead to a poor estimation and false conclusions. Thus detecting possible changes in the stochastic structure of a time series has become an important area of research in the last two decades and has drawn much attention from many researchers. Recently, there is a growing interest in testing for and estimating changes in parameters of econometric models. So far, a large number of articles have been published in various journals. See, for instance, Brown, Durbin and Evans (1975), Wichern, Miller and Hsu (1976), Zacks (1983), Krishnaiah and Miao (1988) and Csörgő and Horváth (1997) among the others.

Kokoszka and Leipus (2000) studied a change point for an ARCH process based on the original observations. Kim, Cho, and Lee (2000) constructed a CUSUM test based on the squares of the original data of a GARCH(1,1) model. Lee, Tokutsu and

Maekawa (2003) improved the test of Kim, Cho, and Lee by constructing a test based on the standardized residuals.

Berkes, Horváth, Kokoszka (2004) proposed a test for change in the parameters of a GARCH(p q) model. The test is based on approximate likelihood scores and does not require the observations to have finite variance. They show that the test has asymptotical correct size under some weak assumptions on the model errors.

Kulperger and Yu (2005) showed the CUSUM processes based on residuals from a GARCH(p, q) process behaves as if they were asymptotically IID as the unobservable innovations. And they applied this result to detect change-point in a GARCH(p, q) model. In particular, Yu (2004) demonstrated with numerical examples that the CUSUM test based on standardized residuals of GARCH(p, q) has reasonable size and nice power with large sample sizes. There are substantial power gains when the innovation distribution is t(8) comparing to standard normal. This test can be used to perform near-integrated GARCH(1,1) with a comparison to Kim, Cho, and Lee's test, which could not perform at all for the near-integrated GARCH(1,1).

Based on the results in Chapter 3, we can extend the results of Kulperger and Yu (2005) to ARMA-GARCH processes.

(i) a Structural Change in the Conditional Mean

First we consider a structural change in the conditional mean for ARMA-GARCH model. Due to the masking effect of ARMA, we consider only a structural change on constant term c_0 in the pure GARCH model with nonzero c_0, AR-GARCH model and MA-GARCH model.

(1) Pure GARCH model with nonzero c_0

The null hypothesis is "no-change in the conditional mean"

$$H_0 : \left\{ \begin{array}{l} Y_t = \sigma_t \eta_t + c_0 \\ \sigma_t^2 = \alpha_{00} + \sum_{i=1}^{p} \alpha_{0i}(Y_{t-i} - c_0)^2 + \sum_{j=1}^{q} \beta_{0j}\sigma_{t-j}^2 \end{array} \right\}, \ t = 1, \ldots, n$$

and the alternative is "one change in the conditional mean"

$$H_a : \left\{ \begin{array}{l} \left. \begin{array}{l} Y_t = \sigma_t \eta_t + c_0 \\ \sigma_t^2 = \alpha_{00} + \sum_{i=1}^{p} \alpha_{0i}(Y_{t-i} - c_0)^2 + \sum_{j=1}^{q} \beta_{0j}\sigma_{t-j}^2 \end{array} \right\}, \ t = 1, \ldots, [nu^*] \\ \left. \begin{array}{l} Y_t = \sigma_t \eta_t + c_0' \\ \sigma_t^2 = \alpha_{00} + \sum_{i=1}^{p} \alpha_{0i}(Y_{t-i} - c_0')^2 + \sum_{j=1}^{q} \beta_{0j}\sigma_{t-j}^2 \end{array} \right\}, \ t = [nu^*] + 1, \ldots, n, \end{array} \right.$$

where $c_0 \neq c_0'$ and $0 < u^* < 1$.

(2) MA-GARCH model

The null hypothesis is "no-change of c_0 in the conditional mean"

$$H_0 : \left\{ \begin{array}{l} Y_t - c_0 = \varepsilon_t + \sum_{j=1}^{Q} \varphi_j \varepsilon_{t-j} \\ \varepsilon_t = \sigma_t \eta_t \ \text{ and } \ \sigma_t^2 = \alpha_{00} + \sum_{i=1}^{p} \alpha_{0i}(Y_{t-i} - c_0)^2 + \sum_{j=1}^{q} \beta_{0j}\sigma_{t-j}^2 \end{array} \right\}, \ t = 1, \ldots, n$$

and the alternative is "one change of c_0 in the conditional mean"

$$H_a : \left\{ \begin{array}{l} \left\{ \begin{array}{l} Y_t - c_0 = \varepsilon_t + \sum_{j=1}^{Q} \varphi_j \varepsilon_{t-j} \ , \ \text{ if } \ t = 0, \ldots, [nu^*] \\ Y_t - c_0' = \varepsilon_t + \sum_{j=1}^{Q} \varphi_j \varepsilon_{t-j} \ , \ \text{ if } \ t = [nu^*] + 1, \ldots, n \end{array} \right. \\ \varepsilon_t = \sigma_t \eta_t \ \text{ and } \ \sigma_t^2 = \alpha_{00} + \sum_{i=1}^{p} \alpha_{0i}\varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j}\sigma_{t-j}^2 \ , \ t = 1, \ldots, n \end{array} \right.$$

where $c_0 \neq c_0'$ and $0 < u^* < 1$.

(3) AR-GARCH model

The null hypothesis is "no-change of c_0 in the conditional mean"

$$H_0 : \left\{ \begin{array}{l} Y_t - c_0 = \sum_{i=1}^{P} \phi_i (Y_{t-j} - c_0) + \varepsilon_t \\ \varepsilon_t = \sigma_t \eta_t \ \text{ and } \ \sigma_t^2 = \alpha_{00} + \sum_{i=1}^{p} \alpha_{0i}(Y_{t-i} - c_0)^2 + \sum_{j=1}^{q} \beta_{0j}\sigma_{t-j}^2 \end{array} \right\}, \ t = 1, \ldots, n$$

and the alternative is "one change of c_0 in the conditional mean"

$$H_a : \begin{cases} \begin{cases} Y_t - c_0 = \sum_{i=1}^{P} \phi_i (Y_{t-j} - c_0) + \varepsilon_t \ , & \text{if} \quad t = 0, \ldots, [nu^*] \\ Y_t - c_0' = \sum_{i=1}^{P} \phi_i (Y_{t-j} - c_0') + \varepsilon_t \ , & \text{if} \quad t = [nu^*] + 1, \ldots, n \end{cases} \\ \varepsilon_t = \sigma_t \eta_t \ \text{ and } \ \sigma_t^2 = \alpha_{00} + \sum_{i=1}^{p} \alpha_{0i} \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j} \sigma_{t-j}^2 \ , \ t = 1, \ldots, n \end{cases}$$

where $c_0 \neq c_0'$ and $0 < u^* < 1$.

To test above hypothesis, we use the standard CUSUM test constructed from residuals as

$$CUSUM^{(1)} = \max_{1 \leq i < n} \frac{\left| \sum_{t=1}^{i} \tilde{\eta}_t - i \bar{\tilde{\eta}} \right|}{\tilde{\sigma}_{(n)} \sqrt{n}} \ .$$

By Corollary 3.2.2, under H_0,

$$CUSUM^{(1)} \xrightarrow{\mathcal{D}} \sup_{0 \leq u \leq 1} |B_0(u)| \ ,$$

where $\{B_0(u), \ 0 \leq u \leq 1\}$ is a Brownian bridge. Hence we can reject H_0 in favor of H_a if $CUSUM^{(1)}$ is large.

Table 4.4 lists the simulation results of GARCH(1,1) with nonzero mean. The parameter in H_0 is $(c_0, \alpha_{00}, \alpha_{01}, \beta_{01}) = (0, 0.5, 0.1, 0.8)$. Table 4.5 lists the simulation results of AR(1)-GARCH(1,1) with nonzero mean. The parameter in H_0 is $(c_0, \phi_{01}, \alpha_{00}, \alpha_{01}, \beta_{01}) = (0, 0.4, 0.5, 0.1, 0.8)$. Table 4.6 lists the simulation results of MA(1)-GARCH(1,1) with nonzero mean. The parameter in H_0 is $(c_0, \varphi_{01}, \alpha_{00}, \alpha_{01}, \beta_{01}) = (0, 0.6, 0.5, 0.1, 0.8)$. In These 3 tables, all the break points in H_a are $u^* = 0.5$ and c_0 changes from 0 to 0.5 after $[u^* n]$. Critical values 1.358 and 1.2239 are chosen for significance levels $\alpha = 5\%$ and $\alpha = 10\%$ respectively. Five thousand replications are used.

Table 4.4: Size and Power of $CUSUM^{(1)}$ Statistic for GARCH(1,1) with Nonzero Mean

$\eta_0 \sim N(0,1)$	n=300	n=600	n=1000	n=3000
Size				
$\alpha = .05$	0.038	0.038	0.043	.050
$\alpha = .1$	0.075	0.079	0.094	.099
Power				
$\alpha = .05$	0.387	0.689	0.902	1
$\alpha = .1$	0.509	0.792	0.944	1

Table 4.5: Size and Power of $CUSUM^{(1)}$ Statistic for AR(1)-GARCH(1,1) with Nonzero Mean

$\eta_0 \sim N(0,1)$	n=300	n=600	n=1000	n=3000
Size				
$\alpha = .05$	.037	.043	.049	.050
$\alpha = .1$	.077	.086	.091	.102
Power				
$\alpha = .05$	0.142	0.288	0.480	.930
$\alpha = .1$	0.234	0.403	0.595	.961

Table 4.6: Size and Power of $CUSUM^{(1)}$ Statistic for MA(1)-GARCH(1,1) with Nonzero Mean

$\eta_0 \sim N(0,1)$	n=300	n=600	n=1000	n=3000
Size				
$\alpha = .05$	0.044	0.035	0.045	.052
$\alpha = .1$	0.087	0.081	0.098	.105
Power				
$\alpha = .05$	0.164	0.315	0.521	0.947
$\alpha = .1$	0.255	0.436	0.632	0.97

From Table 4.4 - Table 4.6, it can be seen that sizes are somewhat conservative when sample size n is small. When sample size increases, the sizes are very close to the nominal significance level. The powers increase with the sample size and are bigger than 93% as sample size is bigger than 3000.

(ii) a Structural Change in Conditional Variance

Next we consider a change in the conditional variance of an ARMA-GARCH model with null hypothesis as "no-change in the conditional variance"

$$H_{0'}: \left\{ \begin{array}{l} Y_t - c_0 = \sum_{i=1}^{P} \phi_{0i}(Y_{t-i} - c_0) + \varepsilon_t + \sum_{j=1}^{Q} \psi_{0j}\varepsilon_{t-j} \\ \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \alpha_{00} + \sum_{i=1}^{p} \alpha_{0i}\varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j}\sigma_{t-j}^2 \end{array} \right\}, \ t = 1, \ldots, n$$

against the "one change in the conditional variance" alternative

$$H_{a'}: \left\{ \begin{array}{l} Y_t - c_0 = \sum_{i=1}^{P} \phi_{0i}(Y_{t-i} - c_0) + \varepsilon_t + \sum_{j=1}^{Q} \psi_{0j}\varepsilon_{t-j} \\ \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \left\{ \begin{array}{ll} \alpha_{00} + \sum_{i=1}^{p} \alpha_{0i}\varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j}\sigma_{t-j}^2 & \text{if} \ \ t = 1, \ldots, [nu^*] \\ \alpha'_{00} + \sum_{i=1}^{p} \alpha'_{0i}\varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta'_{0j}\sigma_{t-j}^2 & \text{if} \ \ t = [nu^*]+1, \ldots, n, \end{array} \right. \end{array} \right.$$

where $(\alpha_{00}, \alpha_{01}, \ldots, \alpha_{0p}, \beta_{01}, \ldots, \beta_{0q}) \neq (\alpha'_{00}, \alpha'_{01}, \ldots, \alpha'_{0p}, \beta'_{01}, \ldots, \beta'_{0q})$ and $0 < u^* < 1$. The statistic is defined as

$$CUSUM^{(2)} = \max_{1 \le i < n} \frac{\left| \sum_{t=1}^{i} (\tilde{\eta}_t - \bar{\tilde{\eta}})^2 - i\tilde{\sigma}_{(n)}^2 \right|}{\tilde{\zeta}_2 \sqrt{n}},$$

where

$$\tilde{\zeta}_2^2 = \frac{1}{n} \sum_{t=1}^{n} \left((\tilde{\eta}_t - \bar{\tilde{\eta}})^2 - \tilde{\sigma}_{(n)}^2 \right)^2$$

is an estimator of $\zeta_2^2 = E(\eta_0^2 - \mu_2)^2 = \mu_2^2(\nu_4 - 1)$. Therefore, by Corollaries 3.2.3,

Table 4.7: Size and Power of $CUSUM^{(2)}$ statistic for ARMA(1,1)-GARCH(1,1)

$\eta_0 \sim N(0,1)$	n=500	n=1000	n=1500	n=3000
Null	0.036	0.036	0.037	0.044
$\alpha'_{00} = .0003$	0.236	0.754	0.929	0.999
$\alpha'_{01} = .167$	0.334	0.416	0.621	0.914
$\beta'_{01} = .767$	0.306	0.655	0.851	0.994
$\eta_0 \sim t(9)$	n=500	n=1000	n=1500	n=3000
Null	0.039	0.046	0.048	0.043
$\alpha'_{00} = .0003$	0.191	0.442	0.645	0.930
$\alpha'_{01} = .167$	0.148	0.367	0.551	0.868
$\beta'_{01} = .767$	0.182	0.425	0.636	0.938

under H'_0,

$$CUSUM^{(2)} \xrightarrow{\mathcal{D}} \sup_{0 \le u \le 1} |B_0(u)|,$$

where $\{B_0(u),\ 0 \le u \le 1\}$ is a Brownian bridge. Hence we can reject $H_{0'}$ in favor of $H_{a'}$ whenever $CUSUM^{(2)}$ is large.

Table 4.7 presents the simulation results of ARMA(1,1)-GARCH(1,1) with $\eta_t \sim N(0,1)$ and $\eta_t \sim t(9)$ respectively. The parameter in $H_{0'}$

$$(c_0, \phi_{01}, \varphi_{01}, \alpha_{00}, \alpha_{01}, \beta_{01}) = (0, 0.4, 0.6, 0.0002, 0.1, 0.8).$$

Break point 0.5 in $H_{a'}$ is used. Replication is 5000 times. Each time there is only one change in the conditional variance. Critical values 1.358 is chosen for significance level 5%.

Table 4.7 shows some similar conclusion of ARMA(1,1)-GARCH(1,1) to that of GARCH(1,1) in Yu (2004). There is size distortion, which is less serious when the sample size is bigger. Also there are power losses when the innovation distribution changes from normal to student t.

4.2.2 Jarque-Bera Test for Normality

Although normality of the innovations in ARMA-(I)GARCH model is not necessary for the estimation, the efficiency of QMLE is related to the density presumed. The closer is the likelihood kernel to the density of innovation, the more efficient is the QMLE. Empirically the innovation density is leptokurtic and not normally distributed. Thus normality test is quite important in diagnosis of goodness-of-fit, efficiency test and inference.

A popular graphical method for examining normality is the normal quantile-quantile plot (QQ-plot). QQ-plot is a scatter plot of the standardized empirical quantiles of the residuals against the quantiles of the standard normal distribution. If the data is normally distributed, then the quantiles will lie approximately on a 45 degree line.

In econometrics a normality test is customarily performed by Jarque-Bera (JB) test for its straightforward interpretation and implementation.

The JB statistic is defined as:

$$JB = \frac{nb_1^2}{6} + \frac{n(b_2-3)^2}{24} \xrightarrow{\mathcal{D}} \chi^2_{(2)}, \tag{4.2.1}$$

where $b_1 = m_3/m_2^{3/2}$, $b_2 = m_4/m_2^2$ respectively, and m_i is the i^{th} central moment of the sample with size n.

The JB test was formally derived by Jarque and Bera (1987) as a Lagrange

Multiplier test of normality of the regression residuals versus the alternative that the error distribution belongs to the Pearson family, which includes the beta, gamma and student's t distribution and others. They showed JB is asymptotically equivalent to the likelihood ratio test, implying it has the same asymptotic power characteristics including maximum local asymptotic power (Cox and Hinkley (1974)). Hence a test based on JB is asymptotically locally most powerful. They also showed that JB is asymptotically distributed as $\chi^2(2)$.

There are some reasons which limit the application of JB test. One of the reasons is that the asymptotic validity of the JB test has been only proved for limit stationary models. It is unclear if this test can be extended. Recently, Kulperger and Yu (2005) extended it to GARCH(p,q) models.

A second limitation is that JB statistic does not take the serial correlation into account. In time series modelling, due to mis-specifying the model or other reasons, the residuals may not be identically and independently distributed. For example, Kawczak, Kulperger and Yu (2005) has shown that residuals of ARCH models cannot be treated as IID in general.

Another limitation is that the asymptotic distribution of the JB statistic may provide a poor approximation in finite samples.

Urzúa (1996) adjusted the JB statistic by using the exact means and variance of b_1, b_2 in (4.2.1) instead of the asymptotic means and variance. He showed that the adjusted statistic behaves better for small and medium size samples by simulation.

Kilian and Demiroglu (2000) studied the Jarque-Bera test for vector error-correction (VEC) models and level vector autoregressions (VAR) containing possibly integrated or cointegrated variables. They also proposed to use bootstrap critical values to improve the small-sample performance of the test in stationary VAR models and in VEC models and compared the accuracy of the asymptotic and the bootstrap version of the Jarque-Bera test by simulation.

Lu (2001) extended the JB statistic to test normal and student t distribution for residuals from ARCH models. In testing he used adjusted critical values for finite sample instead of applying directly the χ_2 critical value. With this correction, the size and power of the test are improved. In the work, he obtained equations of the JB statistic critical value (size .10 and .05) for both the normal distribution and student t distribution with different sample size n. To do this, he used Monte Carlo simulation to obtain critical values critical value for different sample sizes n (and degree freedom d for student t distribution) and then regressed on sample size n (and degree freedom d for student t distribution). The critical value formulas are listed here:

Critical value for normality test with size 0.10 is:

$$JB_{nm.1} = 4.60517 - \frac{11.438}{n^{1/2}} + \frac{290.146}{n} - \frac{5767.467}{n^{3/2}} + \frac{30798.127}{n^2}, \quad n \geq 100 \quad (4.2.2)$$

Critical value for normality test with size 0.05 is:

$$JB_{nm.05} = 5.991645 - \frac{16.912}{n^{1/2}} + \frac{519.764}{n} - \frac{7754.753}{n^{3/2}} + \frac{36092.983}{n^2}, \quad n \geq 100 \quad (4.2.3)$$

Critical value for the student $t(d)$ test with size 0.10 is:

$$\begin{aligned} JB_{t.1} = 4.60517 &- \frac{4.44}{d^{1/2}} + \frac{53.75}{d} - \frac{149.1}{d^{3/2}} - \frac{23050}{n^2} + \frac{1810300}{n^3} \\ &- \frac{38.15}{(dn)^{1/2}} + \frac{3294}{dn} - \frac{358.4}{dn^{1/2}} + \frac{10070}{d^2 n} - \frac{120200}{dn^2}, \quad n \geq 100 \quad (4.2.4) \end{aligned}$$

Critical value for student $t(d)$ test with size 0.05 is:

$$JB_{t.05} = 5.991645 - \frac{40.88}{d} + \frac{361.44}{d^{3/2}} - \frac{816.43}{d^2} - \frac{10.27}{n^{1/2}} - \frac{142.28}{n^{3/2}}$$

$$+\frac{139.68}{(dn)^{1/2}} - \frac{2013}{dn} - \frac{608.86}{dn^{1/2}} + \frac{24640}{d^2 n}, \; n \geq 100 \qquad (4.2.5)$$

Kulperger and Yu (2005) extended the JB statistic to test normal distribution of GARCH(p,q) innovations. They defined the test as:

$$JB = \frac{n}{\sigma_s^2}\left(\hat{s}_n - \kappa_3\right)^2 + \frac{n}{\sigma_\kappa^2}\left(\hat{\kappa}_n - \kappa_4\right)^2, \qquad (4.2.6)$$

where, η_k is the GARCH(p,q) innovation, $\mu_k = \mathbf{E}(\eta_k)$, $\kappa_k = \mu_k/\mu_2^{k/2}$, and

$$\hat{s}_n = \frac{\sum_{t=1}^n (\hat{\eta}_t - \bar{\hat{\eta}})^3}{(\sum_{t=1}^n (\hat{\eta}_t - \bar{\hat{\eta}})^2)^{3/2}}$$

$$\hat{\kappa}_n = \frac{\sum_{t=1}^n (\hat{\eta}_t - \bar{\hat{\eta}})^4}{(\sum_{t=1}^n (\hat{\eta}_t - \bar{\hat{\eta}})^2)^2}$$

$$\sigma_s^2 = (\kappa_6 - \kappa_3^2) + 3(3 + 3\kappa_3^2 - 2\kappa_4) + 3\kappa_3(\kappa_3/4 + 3\kappa_3\kappa_4/4 - \kappa_5)$$

and

$$\sigma_\kappa^2 = (\kappa_8 - \kappa_4^2) + 4\kappa_3(4\kappa_3 + 4\kappa_3\kappa_4 - 2\kappa_5) + 4\kappa_4(\kappa_4^2 - \kappa_6)\ .$$

Based on results in Chapter 3, we can further extended it to test distribution of ARMA-(I)GARCH innovations. The statistic is defined same as in (4.2.6).

Koul and Ling (2005) proposed a test based on a vector of certain weighted residual empirical processes and used it to test the normality of the GARCH innovation distribution (as in Table 4.8). To compare the performance of Koul and Ling's (2005)

Table 4.8: Size and Power of test, Koul and Ling (2005)

		n=200			n=400	
	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
Null	.089	.041	.006	.102	.053	.008
H_{a1}	.171	.086	.021	.348	.226	.058
H_{a2}	.309	.180	.056	.590	.453	.200
H_{a3}	.570	.434	.201	.909	.882	.581
H_{a4}	.407	.247	.060	.793	.640	.283
H_{a5}	1	1	1	1	1	1

test with the JB test, we use the same model AR(1)-GARCH(1,1), same parameters of $(\phi_{01}, \alpha_{00}, \alpha_{01}, \beta_{01}) = (0.5, 0.025, 0.25, 0.5)$ and with same sample size 200 and 400. The procedure is replicated 1000 times. The Null distribution of η_t is N(0,1) and its alternatives are set as:

$$H_{a1} : \eta_t \sim \sqrt{3/5}\, t(5); \quad H_{a2} : \eta_t \sim \sqrt{1/2}\, t(4); \quad H_{a3} : \eta_t \sim \sqrt{1/3}\, t(3);$$

$$H_{a4} : \eta_t \sim \text{double exponential}\ ; \quad H_{a5} : \eta_t \sim [0.5N(-3,1) + 0.5N(3,1)]/\sqrt{10}\ .$$

Table 4.9 lists the JB test results based on χ_2 critical values. Table 4.10 lists the JB test results based on corrected critical values by Lu (2001). It can be seen that JB test based on both corrected critical values and χ_2 critical values are much more powerful than Koul and Ling's (2005) test. In particular, JB test has substantial power gains under H_{a1} and H_{a2}. The sizes are conservative when χ_2 critical values are used. While applying corrected critical values, the sizes are very close to the nominal significance level for $n = 400$, though the sizes are still somewhat conservative for $n = 200$.

Table 4.9: Size and Power of JB statistic for AR(1)-GARCH(1,1) based on χ_2 critical value

		n=200			n=400	
	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
Null	.064	.032	.015	.064	.034	.008
H_{a1}	.831	.794	.715	.991	.984	.959
H_{a2}	.932	.912	.854	1	.997	.991
H_{a3}	.985	.981	.969	1	1	1
H_{a4}	.954	.933	.875	.998	.998	.995
H_{a5}	1	1	1	1	1	1

Table 4.10: Size and Power of JB statistic for AR(1)-GARCH(1,1) based on corrected critical value

	n=200		n=400	
	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$
Null	.088	.043	.096	.052
H_{a1}	.845	.810	.986	.975
H_{a2}	.944	.915	.997	.993
H_{a3}	.989	.980	1	1
H_{a4}	.965	.933	.998	.998
H_{a5}	1	1	1	1

4.3 Conclusions and Discussions

In Section 4.1.1, with a being unknown, due to $\mathbf{E}\eta_0^2 = 1$ and $\mathbf{E}\eta_t^{**2} = a^2\mathbf{E}\eta_t^2 = a^2$, we can estimate a^2 by $n^{-1}\sum_{i=1}^n \hat{\eta}_t^{**2}$. One shortcoming of this estimation of a is that it brings extra variation to the estimator of $\alpha_0, \alpha_1, \cdots, \alpha_p$. And the normality of rescaled estimators requires 4th moment of η_t. Another shortcoming is that the sample variance of $\hat{\eta}_t^2$ is constant 1 for any sampling. To relax the moment requirements of η_t and constant sample variance of $\hat{\eta}_t^2$, we can assume other conditions, i.e. $\mathbf{E}|\eta_0| = 1$ or $\mathbf{E}\big(g(\eta_0)\eta_0\big) = -1$ to identify model (1.2.13).

In the proof of normality theorem in Chapter 2, we require the distribution of η_t to be symmetric about 0. If the innovation distribution is not symmetric, we can employ other models which consider the leverage effect as introduced in Chapter 1. The problem is how to test the symmetry, which will be considered in my future work.

In my future work, I will consider if the results in Chapter 2 and 3 can be extended to multivariate ARMA-GARCH.

4.4 APPENDIX

A.1 Splus code of simulating ARMA-GARCH

```
#E eta^2 =1

#Normal distribution: dist.par=0; Double Exponential distribution:
dist.part=1; #Student distribution: dist.par > 2.
```

```
sim.garch <- function(n, n0, arch, garch, dist.par) {
    module(finmetrics)
    if(dist.par == 0)
        innov <- rnorm(n + n0)
    else if(dist.par == 1)
        innov <- rdexp(n + n0, rate = sqrt(2))
    else innov <- rt(n+n0, df=dist.par)*sqrt((dist.par - 2)/dist.par)
    x <- innov
    h <- rep(arch[1], (n + n0))
    p <- length(arch)
    q <- length(garch)
    m <- max(c(p - 1, q))
    if(m == 0)
        return(innov)
    if(p - 1 < 1)
        return("error: NO arch")
    x[1] <- sqrt(h[1]) * x[1]

    #ARCH
    if(q < 1) {
        for(i in 2:p) {
            for(j in 2:i)
                h[i] <- h[i] + (arch[j] * (x[i - j + 1])^   2)
            x[i] <- sqrt(h[i]) * x[i]
        }
        for(i in (m + 2):(n0 + n)) {
            for(j in 2:p)
```

```
            h[i] <- h[i] + arch[j] * (x[i - j + 1])^2
        x[i] <- sqrt(h[i]) * x[i]
    }
    error <- x
    sigma.sq <- h
    return(error, sigma.sq, innov)
}

#GARCH initial
if((p - 1) == q) {
    for(i in 2:p) {
        for(j in 2:i)
            h[i] <- h[i] + (arch[j] * (x[i - j + 1])^   2)
            + garch[j - 1] * h[i - j + 1]
        x[i] <- sqrt(h[i]) * x[i]
    }
}
else if((p - 1) < q) {
    for(i in 2:p) {
        for(j in 2:i)
            h[i] <- h[i] + (arch[j] * (x[i - j + 1])^2)
            + garch[j - 1] * h[i - j + 1]
        x[i] <- sqrt(h[i]) * x[i]
    }
    for(i in (p + 1):(q + 1)) {
        for(j in 2:p)
            h[i] <- h[i] + arch[j] * (x[i - j + 1])^2
```

```
        for(j in 2:i)
            h[i] <- h[i] + garch[j - 1] * h[i - j + 1]
        x[i] <- sqrt(h[i]) * x[i]
    }
}
else {
    for(i in 2:(q + 1)) {
        for(j in 2:i)
            h[i] <- h[i] + arch[j] * (x[i - j + 1])^2
            + garch[j - 1] * h[i - j + 1]
        x[i] <- sqrt(h[i]) * x[i]
    }
    for(i in (q + 2):p) {
        for(j in 2:(q + 1))
            h[i] <- h[i] + garch[j - 1] * h[i - j + 1]
        for(j in 2:i)
            h[i] <- h[i] + arch[j] * (x[i - j + 1])^2
        x[i] <- sqrt(h[i]) * x[i]
    }
}

# GARCH
for(i in (m + 2):(n0 + n)) {
    for(j in 1:q)
        h[i] <- h[i] + garch[j] * h[i - q]
    for(j in 2:p)
        h[i] <- h[i] + arch[j] * (x[i - j + 1])^2
```

```
        x[i] <- sqrt(h[i]) * x[i]
    }
    list(error = x, sigma.sq = h, innov = innov)
}

sim.arma.garch <- function(n, n0, arch, garch,
dist.par, mu, ar, ma)
{
    e.sig.eta <- sim.garch(n = n, n0 = n0, arch = arch,
    garch = garch, dist.par = dist.par)
    e <- e.sig.eta$error
    x <- e
    p <- length(ar)
    q <- length(ma)
    m <- max(c(p, q))
    if(m == 0) {
        series <- x[ - (1:n0)] + mu
        error <- e[ - (1:n0)]
        sigma.sq <- e.sig.eta$sigma.sq[ - (1:n0)]
        innov <- e.sig.eta$innov[ - (1:n0)]
        return(series, error, sigma.sq, innov)
    }
    x[1] = e[1] + mu

    #AR-GARCH
    if(q < 1) {
        for(i in 2:(p + 1))
```

```
    x[i] <- mu + sum(ar[1:(i - 1)] * (x[(i - 1):1] - mu)) + e[i]
    for(i in (p + 2):(n + n0))
    x[i] <- mu + sum(ar[1:p] * (x[(i - 1):(i - p)] - mu)) + e[i]
    series <- x[ - (1:n0)]
    error <- e[ - (1:n0)]
    sigma.sq <- e.sig.eta$sig[ - (1:n0)]
    innov <- e.sig.eta$innov[ - (1:n0)]
    return(series, error, sigma.sq, innov)
}

#MA-GARCH

if(p < 1) {
    for(i in 2:(q + 1))
        x[i] <- mu + e[i] + sum(ma[1:(i - 1)] * e[(i - 1):
            1])
    for(i in (q + 2):(n + n0))
        x[i] <- mu + e[i] + sum(ma[1:q] * e[(i - 1):(i - q)])
    series <- x[ - (1:n0)]
    error <- e[ - (1:n0)]
    sigma.sq <- e.sig.eta$sig[ - (1:n0)]
    innov <- e.sig.eta$innov[ - (1:n0)]
    return(series, error, sigma.sq, innov)
}

#ARMA-GARCH, initial
```

```
if(p == q) {
    for(i in 2:(p + 1))
    x[i] <- mu + sum(ar[1:(i - 1)] * (x[(i - 1):1] - mu)) + e[i]
        + sum(ma[1:(i - 1)] * e[(i - 1):1])
    }

else if(p < q) {
    for(i in 2:(p + 1))
    x[i] <- mu + sum(ar[1:(i - 1)] * (x[(i - 1):1] - mu)) + e[i]
        + sum(ma[1:(i - 1)] * e[(i - 1):1])
    for(i in (p + 2):(q + 1))
    x[i] <- mu + sum(ar[1:p] * (x[(i - 1):(i - p)] - mu)) + e[i]
        + sum(ma[1:(i - 1)] * e[(i - 1):1])
    }

else {
    for(i in 2:(q + 1))
    x[i] <- mu + sum(ar[1:(i - 1)] * (x[(i - 1):1] - mu)) + e[i]
        + sum(ma[1:(i - 1)] * e[(i - 1):1])
    for(i in (q + 2):(p + 1))
    x[i] <- mu + sum(ar[1:(i - 1)] * (x[(i - 1):1] - mu)) + e[i]
        + sum(ma[1:q] * e[(i - 1):(i - q)])
    }

#ARMA-GARCH, initial

for(i in (m + 2):(n0 + n))
```

```
    x[i] <- mu + sum(ar[1:p] * (x[(i - 1):(i - p)] - mu)) + e[  i]
        + sum(ma[1:q] * e[(i - 1):(i - q)])

    list(series = x[ - (1:n0)], error = e[ - (1:n0)],
    sigma.sq = e.sig.eta$ sigma.sq[ - (1:n0)],
    innov = e.sig.eta$innov[ - (1:n0)])
}
```

References

[1] Baillie, R.T. and R.J. Myers, (1991), Bivariate GARCH Estimation of the Optimal Commodity Futures Hedge, *Journal of Applied Econometrics*, **6**, 109 - 124.

[2] Baillie, R.T. Bollerslev, T. and H.O. Mikkelsen (1996). Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, **74**, 3 - 30.

[3] Baillie, R.T. (1996) Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, **73**, 5-59.

[4] Berkes, I. and Horváth, L. (2003). The rate of consistency of the quasi-maximum likelihood estimator. *Statistics and Probability Letters* **61** 133-143.

[5] Berkes, I. , Horváth, L., and Kokoszka, P. (2003). GARCH processes: structure and estimation. *Bernoulli* **9 2003** 201-207

[6] Berkes, I. and Horváth, L. (2004) . *The efficiency of the estimatiors of the parameters in GARCH processes. Ann. Statist.* **32 (2004) no. 2,** 633-655

[7] I. Berkes, L. Horvath and P. Kokoszka, Testing for parameter constancy in GARCH(p,q) models. *Statistics and Probability Letters*, **70**, 263-273, 2004.

[8] B. Mandelbrot, (1963), *The variation of certain speculative prices.* The Journal of Business. **36(4)**,394-419.

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

[10] Billingsley, P. (1995). Probability and Measure. New York: Wiley.

[11] Nelson D.B. (1991). Conditional Heteroscedasticity in Asset Returns: A New Approach, *Econometrica* **59**, 347370.

[12] Billingsley, P. (1999). *Convergence of probability measures.* 2nd Edition. Wiley, New York.

[13] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327.

[14] Bollerslev, T., (1990), Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model, *Review of Economics and Statistics*, **72**, 498 - 505.

[15] Bougerol, P. and Picard, N. (1992a). Strict stationary of generalized autoregressive processes. *Ann. Probab.* **20** 1714-1730.

[16] Bougerol, P. and Picard, N. (1992b). Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics* **52** 115-127.

[17] Brock, W.A., Dechert, W., and Scheinkman J. (1996), A Test for Independence based on the Correlation Dimension, *Econometrics Reviews*, **15**, 197-235.

[18] Brown, R.L., Durbin, J. and Evans, J.M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society* **B37**, 149-163.

[19] Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets.* Princeton University Press.

[20] Csörgó, M. and L. Horváth, (1997), *Limit Theorems in Change-Point Analysis,* NewYork: John Wiley.

[21] Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). A Long Memory Property of Stock Market Returns and a New Model, *Journal of Empirical Finance*,**1**, 83-106.

[22] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50** 987–1008.

[23] Engle, R. F., Lilien, D.M., and Robins, R. P. (1987). Estimating Time Varying Risk Premia in the Term-Structure: the ARCH-M Model, *Econometrica*, **55** (2), 391-407.

[24] Engle, R. F., and Ng, V. (1993). Measuring and Testing the Impact of News on Volatility, *Journal of Finance*, **48**(5), 1749-1778.

[25] Francq, C. and Zakoïan J. M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10 (2004), no. 4**, 605-637

[26] Fuller, W. (1996). Introduction to Statistical Time Series, Second Edition. John Wiley, New York.

[27] Hall, P. and Q. Yao (2003). Inference in ARCH and GARCH Models with Heavy-tailed Errors. *Econometrica* **71**, 285-317.

[28] Hinich, M. and Patterson, D.M. (1995), Detecting Epochs of Transient Dependence in White Noise, unpublished manuscript, University of Texas at Austin.

[29] Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance*, **48** (5), 1779-1801.

[30] Gourieroux, Christian (2001). *Financial econometrics : problems, models, and methods.* Princeton University Press, Oxford.

[31] Jarque, C., Bera, A. (1987) A test for normality of observations and regression residuals. *International Statistical Review*, **55**, 163-172.

[32] Karolyi A. (1995) A Multivariate GARCH Model of International Transmissions od Stock Returns and Volatility: The Case of the United States and Canada, *Journal of Business and Economic Statistics*, **13**,n.1.

[33] Kawczak, J., Kulperger, R. J. and Yu, H. (2003). The empirical distribution and partial sum process of residuals from a stationary ARCH-M process. Submitted, under revision.

[34] Kim, S., Cho, S. and Lee, S. (2000). On the cusum test for parameter changes in GARCH(1,1) models. *Commun. in Statist. Theory & Meth.* **29**, 445-462.

[35] Kokoszka P., Leipus R. (2000). Change-point estimation in ARCH models. *Bernoulli.* **6**. 513-539.

[36] Krishnaiah, P.R. and B.Q. Miao, (1988), *Review about Estimation of Change Points*, in Handbook of Statistics, Vol. **7**, ed. by P.R. Krishnaiah and C.R. Rao, New York: Elsevier.

[37] Kulperger, R. J. and Yu, H. (2005). High moment partial sum processes of residuals in garch models and their applications. *Annals of Statistics.* **33**, no. 5 (2005), 23952422.

[38] Kullback, S. (1959). *Information theory and statistics.* John Wiley and Sons, New York, NY.

[39] Kullback, S., and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79–86.

[40] Lee, S., Ha, J., Na, O. and Na, S. (2003). The Cusum Test for Parameter Change in Time Series Models. *Scand. J. Statist.* **30**, 781-796.

[41] Lee, S.W. and B.E. Hansen, (1994). Asymptotic Theory for the GARCH(1, 1) quasimaximum likelihood estimator, *Econometric Theory* **10** 29–52.

[42] Li, W. K., (2004). *Diagnostic Checks in Time Seires.* Chapman & Hall/CRC, Boca Raton.

[43] Ling, S. (2005). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. To appear in *Journal of Econometrics*

[44] Ling, S. and Li, W. K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Amer. Statist. Assoc.* **92** 1184–1194.

[45] Ling, S. and Li, W. K. (1998). Limiting distributions of maximum likelihood estimators for unstalbe ARMA models with GARCH errors. *Ann. Statist.* **26** 84–125.

[46] Ling, S. and Li, W. K. (2003). Asymptotic inference for unit root processes with GARCH(1,1) errors. *Econometric Theory* **19** 541–564.

[47] Ling, S. and McAleer (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory.* **19** 280–310.

[48] Ling, S. and Li, W.K. (1997) Diagnostic checking of nonlinear multivariate time series with multivariate arch errors. *Journal of Time Series Analysis.* **18**, 447–464.

[49] Linton, O. (1997). An Asymptotic Expansion in the GARCH(1,1) Model. *Econometric Theory* **13**, 558–581.

[50] Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models, *Biometrika* **65**, 297–303.

[51] Lumsdaine, R.L. (1996). Consistency and asymptotic normality of the quasi-

maximum likelihood estimator in IGARCH(1, 1) and covariance stationary GARCH(1, 1) models. *Econometrica* **64**, 575-596.

[52] Lu, P. G. (2001). Identification of distribution for ARCH innovation. Master's Project. University of Western Ontario.

[53] Luukkonen,R., Saikkonen P., and Terasvirta, T. (1988) Testing linearity against smooth transition autoregessive models. *Biometrika*, **75**, 491-499.

[54] McLeod, A. I. and W. K. Li (1983). Diagnostic checking ARMA time series models using squaredresidual autocorrelations, *Journal of Time Series Analysis*, **4**, 269-273.

[55] Nelson D.B. (1991). Conditional Heteroscedasticity in Asset Returns: A New Approach, *Econometrica* **59**, 347-370.

[56] Newey, W. K. and D. McFadden (1994), *Large Sample Estimation and Hypothesis Testing*, Handbook of Econometrics, Vol **4**.

[57] Osborne, M.F.M. (1959) *Brownian motion in the stock market.* Operations Research, 7, 145-173; 7, 807-811.

[58] Ron C. Mittelhammer, George G. Judge, Douglas J. Miller. (2000) *Econometric foundations.* Cambridge University Press, New York.

[59] Rossi, P. (1996). Modelling Stock Market Volatility: bridging the gap to continuous time. Academic Press, Toronto.

[60] Wichern, D. W., R. B. Miller, and D. A. Hsu, (1976), changes of variance in first-order autoregressive time series modelswith an application, *Journal of the Royal Statistical Society,* Series C (Applied Statistics), **25**, 248-56.

[61] White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1-25.

[62] White, H. (1994). *Estimation, Inference, and Specification Analysis.* Cambridge University Press.

[63] Stout, W. F. (1974). *Almost Sure Covnergence.* Academic Press, New York.

[64] Terence C. Mills. (1999) *The econometric modelling of financial time series* . Cambridge University Press, New York.

[65] Tsay, R.S. (1986), Nonlinearity tests for Time Series, *Biometrica,* **73**, 461-466.

[66] Urzúa. C. M. (1996). On the correct use of omnibus tests for normality. *Economic Letters* **53**. 247–251.

[67] Yu, H. (2004). Analyzing residual processes of (G)ARCH time series models. *Fields Institute Communications.* **44** 469–486.

[68] Yu, H. (2005). High moment partial sum processes of residuals in ARMA models and their applications. To appear in *Journal of Time Series Analysis.*

[69] Zacks, S. (1983). Survey of classical and Bayesian approaches to the change-point problem : fixed sample and sequential procedures of testing and estimation. In

Recent Advances in Statistics. M. H. Rivzi et al. eds. 245-269, Academic Press, New York. 11

[70] Zakoian, J. (1991). Threshold Heteroskedasticity Model, unpublished manuscript, INSEE.