August 2011

# Asymptotic Theory of General Multivariate GARCH Models

Weibin Jiang
*The University of Western Ontario*

Supervisor
Dr. Reg J. Kulperger
*The University of Western Ontario*

Joint Supervisor
Dr. Hao Yu
*The University of Western Ontario*

Graduate Program in Statistics and Actuarial Sciences

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

# ASYMPTOTIC THEORY OF GENERAL MULTIVARIATE GARCH MODELS

(Thesis format: Monograph)

by

Weibin Jiang

Graduate Program in Statistical & Actuarial Sciences

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

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
School of Graduate and Postdoctoral Studies

**CERTIFICATE OF EXAMINATION**

Supervisor:

________________

Dr. R. J. Kulperger

Co-Supervisor:

________________

Dr. H. Yu

Examiners:

________________

Dr. Rafal Kulik

________________

Dr. Adam Metzler

________________

Dr. A. Ian McLeod

________________

Dr. Rogemar Mamon

The thesis by

**Weibin Jiang**

entitled:

**Asymptotic Theory of General Multivariate 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

Generalized autoregressive conditional heteroscedasticity (GARCH) models are widely used in financial markets. Parameters of GARCH models are usually estimated by the quasi-maximum likelihood estimator (QMLE). In recent years, economic theory often implies equilibrium between the levels of time series, which makes the application of multivariate models a necessity. Unfortunately the asymptotic theory of the multivariate GARCH models is far from coherent since many algorithms on the univariate case do not extend to multivariate models naturally. This thesis studies the asymptotic theory of the QMLE under mild conditions. We give some counterexamples for the parameter identifiability result in Jeantheau [1998] and provide a better necessary and sufficient condition. We prove the ergodicity of the conditional variance process on an application of theorems by Meyn and Tweedie [2009]. Under those conditions, the consistency and asymptotic normality of the QMLE can be proved by the standard compactness argument and Taylor expansion of the score function.

We also give numeric examples on verifying the assumptions and the scaling issue when estimating GARCH parameters in S+ FinMetrics.

**Keywords:** General multivariate GARCH, asymptotic theory, ergodicity, stationarity, consistency, asymptotic normality, VEC, BEKK.

To my parents and Ruoqiao.

Acknowledgements

First of all I would like to express my sincere gratitude to my Ph.D. supervisors Dr. Reg J. Kulperger and Dr. Hao Yu, who are also my Master supervisors. This thesis could not have been completed without the their guidance, support and help. Their patience and kindness, as well as academic experiences, have been invaluable to me.

I greatly appreciate my course instructors during my graduate study, including but not limited to Dr. A. Ian McLeod, Dr. Rogemar Mamon, Dr. Serge Provost, Dr. W. John Braun, Dr. Duncan Murdoch and Dr. Ricardas Zitikis. I have benefited and learned a lot from their teaching.

I am grateful to my colleagues and friends Jonathan Lee, Raymond Zhang, Yanyan Zang, Weiwei Liu, Na Lei, Mark Wolters and Zhe Sheng, for being the surrogate family during the many years I stayed in London, Ontario. The support I received from the department coordinators Ms. Jennifer Dungavell and Ms. Jane Bai has also been most helpful.

Finally, I wish to express special thanks to my parents and Ruoqiao for their understanding, endless patience and encouragement when it was most required. They helped me concentrate on completing this dissertation and supported me mentally during the course of this work.

# Contents

# List of Figures

# List of Tables

# Notations

$\rho(A)$ The spectral radius of the square matrix $A$, i.e., $\rho(A) = \max\{|\lambda_i| : \lambda_i$ is an eigenvalue of $A\}$.

$\odot$ Hadamard or elementwise product of matrices.

$\otimes$ Kronecker product.

$\|v\|$ The Euclidean norm of vector $v$.

$\|A\|$ The spectral norm of matrix $A$, i.e., $\|A\| = \sqrt{\rho(A^T A)}$.

$\|A\|_2$ The Euclidean/Frobenius norm of matrix $A$.

$A^T$ The transpose of matrix $A$ (or a vector).

$\mathbf{tr}(A)$ Trace of matrix $A$.

$\xrightarrow{\mathbf{a.s.}}$ Almost surely convergence.

$\xrightarrow{\mathcal{D}}$ Converge in distribution.

$C_1, C_2, \cdots$ Generic constants taking different values from time to time.

**vec**$(\cdot)$ The operator that stacks a $d \times d$ matrix column by column as a $d^2 \times 1$ vector.

**mat**$(\cdot)$ The inverse operator of vec$(\cdot)$.

**vech**$(\cdot)$ The operator that stacks the lower triangular portion of a $d \times d$ matrix as a $d(d+1)/2 \times 1$ vector column by column.

**math**$(\cdot)$ The inverse operator of vech$(\cdot)$.

$I_m$ $m \times m$ identity matrix.

$D_m$ $m^2 \times \frac{m(m+1)}{2}$ duplication matrix such that for a symmetric matrix $A$, $\text{vec}(A) = D_m \text{vech}(A)$.

$D_m^+$ The generalized inverse of $D_m$. $D_m^+$ is such that for a symmetric matrix $A$, $\text{vech}(A) = D_m^+ \text{vec}(A)$ and $D_m^+ D_m = I_{m(m+1)/2}$.

$K_{mn}$ **or** $K_{m,n}$ $mn \times mn$ commutation matrix such that for $A(m \times n)$, $\text{vec}(A^T) = K_{mn}\text{vec}(A)$.

$\mathcal{F}_t$ The information filter generated by the observable data at times less than or equal $t$.

$P(x, A)$ The one-step transition probability for a Markov chain (denoted by $\mathbf{\Phi}$), i.e., $P(\Phi_1 \in A | \Phi_0 = x)$.

$P^n(x, A)$ The $n$-step transition probability for a Markov chain (denoted by $\mathbf{\Phi}$), i.e., $P(\Phi_n \in A | \Phi_0 = x)$.

# Chapter 1

# Introduction

## 1.1 The Univariate GARCH Model

In financial markets, estimating volatilities is essential in derivative pricing and risk management. For example, in order to evaluate stock option prices in the future, forecast of volatilities are usually required. Let $y_t$ be the continuously compounded return or the proportional change of a market variable during day $t$, i.e.,

$$y_t = \log \frac{S_t}{S_{t-1}} \quad \text{or} \quad y_t = \frac{S_t - S_{t-1}}{S_{t-1}}.$$

The difference between these two expressions are tiny when the time increment is small, since the proportional change is the first order Taylor expansion of the continuously compounded return. In contrast to the original asset prices, the continuously compounded return or the proportional change do not depend on monetary units. The MLE of the variance (square of the volatility) using the most recent $q$ observations is

$$\sigma_t^2 = \frac{1}{q}\sum_{i=1}^{q}(y_{t-i} - \bar{y})^2,$$

where $\bar{y} = \sum_{i=1}^{q} y_{t-i}$. Since in this thesis we are only interested in the volatility part, $\bar{y}$ is assumed to be zero and the formula for variance becomes

$$\sigma_t^2 = \frac{1}{q}\sum_{i=1}^{q} y_{t-i}^2. \tag{1.1}$$

In (1.1), every observation has equal effect on the volatility. It is more appropriate to assign more weight on recent data. The model becomes

$$\sigma_t^2 = \sum_{i=1}^{q} \alpha_i y_{t-i}^2,$$

where $\sum_{i=1}^{q} \alpha_i = 1$. A further extension of the model is to add a long-run average volatility term, which leads to that

$$\sigma_t^2 = \gamma V + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 = c + \sum_{i=1}^{q} \alpha_i y_{t-i}^2,$$

where $\gamma + \sum_{i=1}^{q} \alpha_i = 1$. This is known as an autoregressive conditional heteroscedasticity (ARCH) model if we assign $c = \gamma V$. The univariate ARCH($q$)

model, which was first introduced by Engle [1982], is defined as

$$\begin{aligned} y_t &= \sigma_t \xi_t, \\ \sigma_t^2 &= c + \sum_{i=1}^{q} \alpha_i y_{t-i}^2, \end{aligned} \tag{1.2}$$

where $\{y_t\}$ is the observed process, $\xi_t \overset{\text{i.i.d.}}{\sim} (0,1)$[1] and is independent of $\mathcal{F}_{t-1}$, $c \geq 0$, $\alpha_i \geq 0$ for $1 \leq i \leq q-1$, $\alpha_q > 0$. However, in practice, people usually find that a large number of lags $q$ is needed, which results in a large amount of model parameters to be estimated. It is also well known that in financial markets, large changes tend to be followed by large changes, and small changes tend to be followed by small changes. This volatile behavior in financial markets is usually referred to as "volatility clustering". In the past several decades, the generalized autoregressive conditional heteroscedasticity (GARCH) models are commonly used to describe volatilities. Bollerslev [1986] presented the GARCH$(p,q)$ model, where (1.2) was generalized as

$$\sigma_t^2 = c + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \tag{1.3}$$

where $\beta_j \geq 0$ for $1 \leq j \leq p-1$, $\beta_p > 0$.

[1]Note that we do not assume any distributional property on $\xi_t$ except the mean and variance. It may or may not be normally distributed.

**An Example: Modeling Stock Price Proportional Change Using GARCH Model** Suppose a stock price $S_t$ follows the model

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_t,$$

where $W_t$ is a standard Brownian motion. Note that for a given $\Delta t$, $W_t - W_{t-\Delta t} \sim N(0, \Delta t)$. Discretizing the stock price model gives

$$\frac{S_t - S_{t-\Delta t}}{\sqrt{\Delta t} S_{t-\Delta t}} = \mu_t \sqrt{\Delta t} + \sigma_t z_t,$$

where $z_t$ is a standard normal random number. The left hand side can be treated as the observed sequence. Using the GARCH setting, the conditional volatility can be modeled by

$$\sigma_t^2 = c + \sum_{i=1}^{q} \alpha_i \left( \frac{S_{t-i\Delta t} - S_{t-(i+1)\Delta t}}{\sqrt{\Delta t} S_{t-(i+1)\Delta t}} \right)^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j\Delta t}^2.$$

To obtain the one-step prediction for $S_t$, the procedure is as follows:

1. Estimate the model parameters $c$, $\alpha_i$'s and $\beta_j$'s using the observed data.

2. Compute the estimated conditional variance sequence $\hat{\sigma}_t^2, \hat{\sigma}_{t-\Delta t}^2, \hat{\sigma}_{t-2\Delta t}^2, \cdots$.

3. Predict the future conditional variance as

$$\hat{\sigma}_{t+\Delta t}^2 = \hat{c} + \sum_{i=1}^{q} \hat{\alpha}_i \left( \frac{S_{t-(i-1)\Delta t} - S_{t-i\Delta t}}{\sqrt{\Delta t} S_{t-i\Delta t}} \right)^2 + \sum_{j=1}^{p} \hat{\beta}_j \hat{\sigma}_{t-(j-1)\Delta t}^2.$$

4. Simulate a standard normal random number $z_{t+\Delta t}$.

5. $\mu_{t+\Delta t}$ can be predicted using its own model, e.g., the ARMA model.

6. The future stock price can be calculated as

$$S_{t+\Delta t} = S_t(1 + \hat{\mu}_{t+\Delta t}\Delta t + \hat{\sigma}_{t+\Delta t}\sqrt{\Delta t} z_{t+\Delta t}).$$

The asymptotic theory of GARCH models involves strong consistency and asymptotic normality of the quasi-maximum likelihood estimator (QMLE). The asymptotic theory of the univariate model was first established by Weiss [1986] for ARCH models. The GARCH results were first demonstrated in Lee and Hansen [1994] and Lumsdaine [1996], both for the GARCH$(1,1)$ model. Berkes and Horvàth [2004], Berkes and Horvàth [2003] and Berkes et al. [2003] extended the theory into the GARCH$(p,q)$ case. By far the weakest assumptions were given by Francq and Zakoïan [2004], in which they assume the finite fourth moment of the innovations.

Strong stationarity and ergodicity are required to achieve the asymptotic result. Nelson [1990] gave necessary and sufficient conditions for stationarity and ergodicity for the GARCH$(1,1)$ model. Bougerol and Picard [1992] proved that the GARCH$(p,q)$ process is strictly stationary and ergodic if and only if its top Lyapunov exponent is strictly negative.

## 1.2 Multivariate GARCH Models

Economic theory often implies equilibrium between the levels of time series. For each model we developed to capture variances, there is a corresponding model which can be used to track covariances. For example, a similar estimate for the covariance between two time series $\{x_t\}$ and $\{y_t\}$ using the GARCH setting is

$$\mathrm{Cov}(x_t, y_t) = c + \sum_{i=1}^{q} \alpha_i x_{t-i} y_{t-i} + \sum_{j=1}^{p} \beta_j \mathrm{Cov}(x_{t-j}, y_{t-j}).$$

This fact makes the application of multivariate models a necessity. In this thesis, we are interested in general multivariate GARCH models. A general $d$-dimensional GARCH$(p,q)$ model, usually called the VEC model (see Bollerslev et al. [1998]), is given by

$$\begin{aligned} y_t &= H_t^{1/2} \xi_t, \\ h_t &= c + \sum_{i=1}^{q} A_i \eta_{t-i} + \sum_{j=1}^{p} B_j h_{t-j}, \end{aligned} \tag{1.4}$$

where

$$\begin{aligned} h_t &= \mathrm{vech}(H_t), \\ \eta_t &= \mathrm{vech}(y_t y_t^T), \\ \xi_t &\overset{\text{i.i.d.}}{\sim} (0, I_d), \end{aligned}$$

$A_i$'s and $B_j$'s are square parameter matrices of order $N = d(d+1)/2$ and $c$ is an $N \times 1$ parameter vector. The vech$(\cdot)$ operator and its inverse operator math$(\cdot)$ are defined in the notation list.

There are two issues about the general model specification in (1.4):

1. There are a large amount of parameters to be estimated. The number of parameters in (1.4) is $(p+q)N^2 + N$. For example, for bivariate process $\{y_t\}$, $N = 3$ and there are 21 parameters for GARCH$(1,1)$. For trivariate $\{y_t\}$, $N = 6$ and there will be 78 parameters for GARCH$(1,1)$.

2. It is difficult to guarantee that $H_t$ is positive definite without imposing strong restrictions[1].

To overcome these issues, Engle and Kroner [1995] developed two new parameterizations for (1.4). One is called the diagonal VEC (DVEC) model. In this model, all the parameter matrices are assumed to be diagonal. Then (1.4) can be rewritten as

$$H_t = C^* + \sum_{i=1}^{q} A_i^* \odot (y_{t-i} y_{t-i}^T) + \sum_{j=1}^{p} B_j^* \odot H_{t-j}, \tag{1.5}$$

where, $C^*$, $A_i^*$'s and $B_j^*$'s are $d \times d$ symmetric matrices. It is straightforward to verify that $H_t$ is positive definite if $C^*$, $A_i^*$'s and $B_j^*$'s are positive definite. The number of parameters in (1.5) is $(p+q+1)N$. Thus the number of parameters are

[1] Francq and Zakoïan [2010] imposes some conditions under which $H_t$ in the VEC model is positive definite. In this thesis, we assume $H_t$ is positive definite without verifications.

reduced to 9 and 18 respectively for bivariate and trivariate $\{y_t\}$'s if $p = q = 1$.

The other model specification in Engle and Kroner [1995] is called the BEKK model (in the name of Baba, Engle, Kraft and Kroner). The BEKK$(p, q, k)$ model is given by

$$H_t = C + \sum_{i=1}^{q} \left( \sum_{j=1}^{k} A_{ij}^T y_{t-i} y_{t-i}^T A_{ij} \right) + \sum_{i=1}^{p} \left( \sum_{j=1}^{k} B_{ij}^T H_{t-i} B_{ij} \right), \tag{1.6}$$

where $C$, $A_{ij}$'s and $B_{ij}$'s are $d \times d$ coefficient matrices and $C$ is symmetric positive definite. In (1.6), the positivity of $H_t$ is guaranteed naturally. The number of parameters is $(p + q)kd^2 + N$. Scherrer and Ribarits [2007] defines that (1.4) is admissible if math$(c)$ is positive definite and $\forall \xi \in \mathbb{R}^d$, math$(A_i \text{vech}(\xi\xi^T))$ is positive semidefinite for $i = 1, \cdots, q$. Then they show that for the bivariate case, admissible VEC models and BEKK models are equivalent. For $d > 2$, there is a "thick" class of admissible VEC models that have no equivalent BEKK representations.

Bollerslev [1990] proposes a multivariate GARCH model in which the conditional correlation does not change over time. The constant correlation model (CCC$(p, q)$) is defined as

$$y_t = \Delta_t \xi_t,$$

where $\Delta_t$ is diagonal whose elements satisfy

$$\begin{pmatrix} \Delta^2_{t,11} \\ \vdots \\ \Delta^2_{t,dd} \end{pmatrix} = W + \sum_{i=1}^{q} A_i \begin{pmatrix} y^2_{t-i,1} \\ \vdots \\ y^2_{t-i,1} \end{pmatrix} + \sum_{j=1}^{p} B_j \begin{pmatrix} \Delta^2_{t-j,11} \\ \vdots \\ \Delta^2_{t-j,dd} \end{pmatrix},$$

W is a constant vector and $\{\xi_t\}$ is an i.i.d. sequence with mean 0 and covariance matrix

$$\begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1d} \\ \rho_{12} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{(d-1)d} \\ \rho_{1d} & \cdots & \rho_{(d-1)d} & 1 \end{pmatrix}.$$

One can easily check that the conditional covariance matrix $H_t$ is such that

$$H_{t,ij} = \rho_{ij}\Delta_{t,ii}\Delta_{t,jj},$$

and hence the conditional correlations are $\rho_{ij}$'s. The CCC model is also a subset of the VEC model.

Other special cases of the general multivariate GARCH model are summarized in Bauwens et al. [2006]. For a most recent summary on both univariate and multivariate GARCH models, see Francq and Zakoïan [2010].

Unfortunately the asymptotic theory of the multivariate GARCH model is far from coherent since many algorithms on the univariate case does not extend to multivariate models naturally. For example, Bougerol and Picard [1992]'s condition does not hold for multivariate GARCH models in general. Boussama [1998] gave a counter-example for this extension.

Jeantheau [1998] proved strong consistency for multivariate GARCH models and verify those conditions for the CCC model. Comte and Lieberman [2003] proved the asymptotic theory for the BEKK model under the assumption of finite eighth moment of $y_t$. They used the condition given by Boussama [1998] to prove stationarity and ergodicity but they did not impose any conditions to verify identifiability. Ling and McAleer [2003] shows the asymptotic theory for a class of multivariate ARMA-GARCH models with the GARCH process following the CCC specification. Hafner and Preminger [2009] proved the asymptotic theory for general multivariate GARCH$(1,1)$ under the assumption of finite sixth moment of $y_t$. However, their proof for the asymptotic normality was not actually complete. They used Markov chain technique in Meyn and Tweedie [2009] to prove stationarity and ergodicity since the GARCH$(1,1)$ model is a Markov chain. We will generalize this approach in this thesis to the GARCH$(p,q)$ case. Kristensen [2007] also gave his condition for stationarity and ergodicity using the same technique. But his condition is difficult to verify in practice.

This thesis tries to fill the gap on the asymptotic theory between univariate GARCH$(p,q)$ and general multivariate GARCH$(p,q)$ models. We study the asymptotic theory of the QMLE under mild conditions. We give some counterexamples for the parameter identifiability result in Jeantheau [1998] and provide a better necessary and sufficient condition. We prove the ergodicity of the conditional variance process on an application of theorems in Meyn and Tweedie [2009]. Under those conditions, the consistency and asymptotic normality of the QMLE can be proved by the standard compactness argument and Taylor expansion of the score function.

## 1.3 The QMLE

Parameter estimation for multivariate GARCH models is usually done by MLE, or quasi-MLE (usually Gaussian QMLE). Let $\theta$ be the parameter vector, that is,

$$\theta = (c^T, \mathrm{vec}(A_1)^T, \cdots, \mathrm{vec}(A_q)^T, \mathrm{vec}(B_1)^T, \cdots, \mathrm{vec}(B_p)^T)^T,$$

If the driving noise is i.i.d. normal, the log likelihood function is given by

$$L_n(\theta) = -\frac{1}{2n}\sum_{t=1}^{n}\left\{\log|H_t(\theta)| + y_t^T H_t^{-1}(\theta) y_t\right\} = -\frac{1}{2n}\sum_{t=1}^{n} l_t(\theta). \tag{1.7}$$

However when the i.i.d. driving noise has some other distribution distribution then (1.7) is not the log likelihood. One may still use it as an estimating method, in the sense that one may construct an estimator as $\arg\max_{\theta\in\Theta}\{L_n(\theta)\}$. In many settings this estimator is still consistent and asymptotically normal. This estimator is called the quasi-maximum likelihood estimator, or the QMLE. In the rest of this thesis we will often refer to this estimating function (1.7) as the log likelihood even though this is not technically correct.

The log likelihood (1.7) depends possibly on the infinite past. However in time series observations this is not reasonable so one really needs to condition on a finite set of initial observations. We thus define $\tilde{L}_n(\theta)$ as the log likelihood function or estimating function which is conditional on some initial values of $y_0, y_{-1}, \cdots, y_{1-q}, H_0, H_1, \cdots, H_{1-p}$. For example, these initial values can be either constants or drawn from a stationary distribution. In this thesis, we choose the initial values as

$$y_0 = y_{-1} = \cdots = y_{1-q} = y_1 \quad \text{and} \quad h_0 = h_{-1} = \cdots = h_{1-p} = c. \tag{1.8}$$

Other terms such as $\tilde{l}_t$, $\tilde{H}_t$ and $\tilde{h}_t$ can be defined analogously. We will show later in this thesis that the choice of initial values does not affect our asymptotic results. The Gaussian QMLE is defined as

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} \tilde{L}_n(\theta) = \arg\min_{\theta \in \Theta} \sum_{t=1}^{n} \tilde{l}_t(\theta), \tag{1.9}$$

where $\Theta$ is the parameter space. Note that in model (1.4), we did not assume any specific distribution on the innovation process $\{\xi_t\}$ except its mean and covariance matrix. In fact, many financial data processes heavy tails. The noise term may not actually be Gaussian so we may use the quasi-likelihood (1.9) as the *estimating* function $\tilde{L}$. It is used since the maximization problem is relatively easy to solve numerically. In this thesis we will show this estimator has the properties

- Strong consistency

$$\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0.$$

- Asymptotic normality

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathfrak{D}} N(0, \Sigma).$$

The difference between the quasi-likelihood (1.7) and the observable quasi-likelihood $\tilde{L}_n(\theta)$ is that for the former we are dealing with a sum of objects that are stationary while this is not so for the later. This is helpful in deriving some properties of the QMLE.

## 1.4 Organization of the Thesis

The remainder of this thesis is organized as follows. Chapter 2 presents the theorem to prove ergodicity and stationarity. Chapter 3 is the major chapter of this thesis. In this chapter, Section 3.2 provides the assumptions under which the GARCH model is identifiable. We also give a counter example in this chapter to show that the identifiability conditions given in Jeantheau [1998] are actually invalid. Section 3.3 is devoted to the strong consistency of the QMLE. Section 3.4 proves the asymptotic normality under the finite sixth moment of $\{y_t\}$. Chapter

4 lists various lemmas which are the intermediate results to prove the theorems in this thesis. The first section of Chapter 5 gives an example of the multivariate GARCH model which satisfies our ergodicity and identifiability assumptions. The last section of Chapter 5 addresses the scaling issue when estimating GARCH parameters in S+ FinMetrics and provides a correction in R. This is also an example of parallel computing in R using the Rmpi package. The computational codes are available from the author upon request. Some useful results in matrix algebra are collected in the appendices.

# Chapter 2

# Ergodicity and Stationarity

## 2.1 Introduction

To prove the asymptotic theory of the QMLE, we need the model to be ergodic and stationary. In this chapter, we will give conditions under which the GARCH process is ergodic and stationary. For the univariate GARCH model, Bougerol and Picard [1992] proved that the process is ergodic and strictly stationary if and only if its top Lyapunov exponent is strictly negative. The components of the matrices used to parameterize multivariate GARCH models are not necessarily positive, so this methodology cannot be extended to the multivariate case generally. Boussama [1998] gave a counter-example for this extension. Hafner and Preminger [2009] studied a GARCH(1,1) general model. We follow their methodology. However to extend this one needs a different state space and Markov representation. After finding a suitable representation, two different ones actually for different aspects, we then use the Markov chain stability theory discussed in Meyn and Tweedie [2009] to prove ergodicity and stationarity. The ergodic theorem will be given in Section 2.2 and will be prove in Section 2.3. Proposition 2.4 states some useful results on the spectral radius of the parameter matrices and this proposition is proved in Section 2.4.

The concept of ergodicity describes the way in which the chain returns to the "center" of the space, and whether it might happen in a finite mean time. Intuitively, if a Markov chain is ergodic, its $n$-step transition probability converges to some "fixed" measure. There are several forms of ergodicity in literature. In this thesis, we use the $V$-uniform ergodicity.

**Definition 2.1** ($V$-Uniform Ergodicity, Definition (16.2) in Meyn and Tweedie [2009])**.** *A Markov chain* $\mathbf{\Phi}$ *is called* $V$*-uniformly ergodic if*

$$\sup_{x\in\mathcal{X}} \frac{\sup_{v:|v|\leq V} \left| \int_{\mathcal{X}} v(w)P^n(x,dw)) - \int_{\mathcal{X}} v(w)\pi(dw) \right|}{V(x)} \to 0, \quad n\to\infty,$$

*where* $\mathcal{X}$ *is the state space,* $V:\mathcal{X}\to[1,\infty)$ *is real Borel measurable,* $P^n$ *is the* $n$*-step transition probability and* $\pi$ *is a probability measure on Borel sets of* $\mathcal{X}$*. Such* $\pi$ *is called an invariant measure.*

We choose to use $V$-uniform ergodicity because the conditions to guarantee $V$-uniform ergodicity is easier to verify than other forms of ergodicity. In particular one needs to handle an appropriate drift in the Markov representation; see (2.4) and (2.10). We also note that if $V(x) \equiv 1$ for all $x \in \mathcal{X}$, then $V$-uniform ergodicity implies the Markov chain is uniformly ergodic, that is ergodic in the usual sense and uniform for all initial conditions.

**Definition 2.2** (Strict Stationarity). *A time series $z_t$ is called strictly stationary if for any $k$, the marginal distribution of $\{z_n, z_{n+1}, \cdots, z_{n+k}\}$ does not change as $n$ varies.*

Given the existence of $\pi$, if the chain is $V$-uniformly ergodic, the transition probability will eventually converge to the invariant measure $\pi$. If the chain is initiated from the invariant measure, it is stationary. To show this, we only need to consider the first step stationarity due to the Markov property. The invariant probability measure $\pi$ is such that for any $A \in \mathcal{B}(\mathcal{X})$,

$$\pi(A) = \int_{\mathcal{X}} \pi(dw)P(w, A),$$

we can iterate to give

$$\begin{aligned}
\pi(A) &= \int_{\mathcal{X}} \left( \int_{\mathcal{X}} \pi(dx)P(x, dw) \right) P(w, A) \\
&= \int_{\mathcal{X}} \pi(dx) \int_{\mathcal{X}} P(x, dw)P(w, A) \\
&= \int_{\mathcal{X}} \pi(dw)P^2(w, A) \\
&\vdots \\
&= \int_{\mathcal{X}} \pi(dw)P^n(w, A)
\end{aligned}$$

$$= \mathrm{P}_\pi(\Phi_n \in A).$$

We can see that a Markov chain $\mathbf{\Phi}$ is strictly stationary if and only if the marginal distribution of $\Phi_n$ does not vary with time. To prove the ergodicity and strict stationarity of the GARCH process, our task is to give conditions under which the invariant measure $\pi$ exists and the chain is $V$-uniformly ergodic.

## 2.2 The Ergodicity Theorem for General Multivariate GARCH Processes

If $p = q = 1$ as in Hafner and Preminger [2009], the model (1.4) is a Markov chain. Otherwise, we need to rewrite the model into a Markov chain representation in order to make use of Markov chain technique. We define

$$Y_t = \begin{pmatrix} h_t \\ \vdots \\ h_{t-p+1} \\ \eta_t \\ \vdots \\ \eta_{t-q+1} \end{pmatrix}, \qquad w_t = \begin{pmatrix} c \\ 0 \\ \vdots \\ 0 \\ \eta_t \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

$$J = \begin{pmatrix} B_1 & \cdots & B_{p-1} & B_p & A_1 & A_2 & \cdots & A_{q-1} & A_q \\ I & & & & & & & & \\ & \ddots & & & & & 0 & & \\ & & I & & & & & & \\ & & & 0 & & & & & \\ & 0 & & & I & & & & \\ & & & & & I & & & \\ & & & & & & \ddots & & \\ & & & & & & & I & 0 \end{pmatrix}, \tag{2.1}$$

where all items in $Y_t$ and $w_t$ are $N$-dimensional vectors and all items in $J$ are $N \times N$ matrices. Thus, $Y_t$ and $w_t$ are $N(p+q)$-dimensional vectors and $J$ is a $N(p+q) \times N(p+q)$ matrix. Then (1.4) can be rewritten as

$$Y_t = w_t + JY_{t-1} = F(Y_{t-1}, \xi_t), \tag{2.2}$$

which is the Markov chain representation of (1.4). It is possible to give the transition probability explicitly for this Markov chain, but this is not needed for our purpose.

The ergodicity of $\{h_t\}$ is implied by the ergodicity of $\{Y_t\}$ since $h_t = TY_t$ is a measurable (linear) transformation, where $T$ is an $N \times N(p+q)$ matrix and $T = (I_N, 0, \cdots, 0)$. Similarly we have that $\{y_t\}$ is ergodic given that $\{Y_t\}$ is ergodic.

Consider the derivative

$$\Delta_t = \Delta(Y_{t-1}, \xi_t) = \frac{\partial Y_t}{\partial Y_{t-1}^T} = J + \frac{\partial w_t}{\partial Y_{t-1}^T} = J + \begin{pmatrix} 0 \\ \frac{\partial \eta_t}{\partial Y_{t-1}^T} \\ 0 \end{pmatrix}, \tag{2.3}$$

where the first 0 is an $N(p-1) \times N(p+q)$ null matrix and the last 0 is an $Nq \times N(p+q)$ null matrix. Applying the chain rule and based on the result in Hafner and Preminger [2009], we have that

$$\begin{aligned}
\frac{\partial \eta_t}{\partial Y_{t-1}^T} &= \frac{\partial \eta_t}{\partial h_t^T} \cdot \frac{\partial h_t}{\partial Y_{t-1}^T} \\
&= D_d^+ \frac{\partial \mathrm{vec}(H_t^{1/2} \xi_t \xi_t^T H_t^{1/2})}{\partial \mathrm{vec}^T(H_t)} D_d \cdot \frac{\partial (c + [B_1|B_2|\cdots|B_p|A_1|A_2|\cdots|A_q] Y_{t-1})}{\partial Y_{t-1}^T} \\
&= D_d^+ (\tilde{\Delta}_t \otimes I_d) D_d [B_1|B_2|\cdots|B_p|A_1|A_2|\cdots|A_q],
\end{aligned}$$

where

$$\tilde{\Delta}_t = H_t^{1/2} \xi_t \xi_t^T H_t^{-1/2}.$$

For some integer $m \geq 1$ and $t \geq m$, let

$$\gamma_m(\Delta) = \frac{1}{m} \mathbb{E} \log \left( \sup_{\bar{Y}^m} \left\| \prod_{k=1}^{m} \Delta(Y_{m-k+1}, \xi_{m-k+2}) \right\| \right),$$

where $\bar{Y}^m = \{(Y_1^T, \cdots, Y_m^T)^T \in \mathbb{R}^{(p+q)Nm}\}$.

We are now in the position to state the theorem for ergodicity and stationarity.

**Theorem 2.3** ($V$-uniform ergodicity)**.** *Consider the general multivariate GARCH model* (1.4)*. Assume that:*

**A1:** *The marginal distribution of $\{\xi_t\}$ is given by a lower semicontinuous density $f_\xi$ w.r.t. the Lebesgue measure which has support $_\xi = \{x \in \mathbb{R}^d | f_\xi(x) > 0\}$. The initial condition $Y_0$ is independent of $\{\xi_t\}$.*

**A2:** $\mathbb{E}\|\xi_t\|^{2r} < \infty$ *for some $r > 0$ ($r$ is usually small).*

**A3:** $\rho(J) < 1$.

**A4:** $\gamma_m(\Delta) < 0$ *for some integer $m \geq 1$.*

**A5:** $\Theta$ *is compact.*

*Then under Assumptions A1-A5, $\{Y_t\}$ is $V$-uniformly ergodic and the invariant measure exists. Thus, the GARCH process is asymptotically strictly stationary.*

*Proof.* See Section 2.3. □

*Remarks.* 1. Theorem 2.3 is similar in spirit to Hafner and Preminger [2009, Theorem 1]. Our proof relies on finer details and structures from Meyn and Tweedie [2009]. In particular one needs to use the matrix $J$ (2.1).

2. Since $\xi_t$ is i.i.d., a sufficient condition for Assumption A4 is $\mathbb{E}\log(\sup_{Y_1} \|\Delta(Y_1, \xi_1)\|) < 0$.

3. It is difficult to calculate $\gamma_m(\Delta)$ directly even for a small $m$. The computation of $\gamma_m(\Delta)$ usually involves Monte Carlo simulation. See Section 5.2 for an example.

4. For VEC models which have equivalent BEKK representations, a sufficient condition for Assumption A3 is $\rho(\sum_{i=1}^{q} A_i + \sum_{j=1}^{p} B_j) < 1$. More generally, we have the following Proposition. This condition, in the $p = q = 1$ is stronger than the corresponding condition in Hafner and Preminger [2009] who only requires $\rho(B_1) < 1$.

**Proposition 2.4.** *For VEC models which have equivalent BEKK representations, we have that*

1. $\rho(\sum_{j=1}^{p} B_j) < 1$ *implies* $\rho(B) < 1$, *where* $B$ *is defined in* (3.8).

2. $\rho(\sum_{i=1}^{q} A_i + \sum_{j=1}^{p} B_j) < 1$ *implies* $\rho(J) < 1$.

3. $\rho(\sum_{i=1}^{q} A_i + \sum_{j=1}^{p} B_j) < 1$ *implies* $\rho(\sum_{j=1}^{p} B_j) < 1$.

*Proof.* See Section 2.4. □

These results in Proposition 2.4 were first mentioned in Thesis of Boussama [1998], specifically in the Appendix. In Section 2.4 we write his proof but with additional details needed for our result.

## 2.3 Proof of Theorem 2.3

To prove Theorem 2.3, we introduce the following drift condition (Condition (V4) (15.28) in Meyn and Tweedie [2009]).

> There exists an extended-real-valued function $V : \mathcal{X} \to [1, \infty]$, a measurable set $C$ and constants $\beta > 0$, $b < \infty$,
>
> $$\Delta V(x) \leq -\beta V(x) + b\mathbb{I}_C(x), \quad x \in \mathcal{X}, \tag{2.4}$$
>
> where $\Delta$ is the drift operator which is defined as
>
> $$\Delta V(x) := \int P(x, dy)V(y) - V(x) = \mathbb{E}(V(\Phi_1)|\Phi_0 = x) - V(x), \quad x \in \mathcal{X}.$$

We inductively define a sequence of functions $F_t$ by

$$\begin{aligned} F_1(x, u_1) &= F(x, u_1) \\ F_{t+1}(x, u_1, \cdots, u_{t+1}) &= F(F_t(x, u_1, \cdots, u_t), u_{t+1}), \quad t \geq 1, \end{aligned} \tag{2.5}$$

where the function $F$ is defined in (2.2). This deterministic system is called the *associated control model* for (2.2).

By Theorem 16.0.1 in Meyn and Tweedie [2009], given $\{Y_t\}$ is $\psi$-irreducible and aperiodic, $\{Y_t\}$ is $V$-uniformly ergodic if and only if the drift condition (2.4) holds for some petite set[1] $C$ and some $V_0$, where $V_0$ is equivalent to $V$ in the sense that for some constant $c \geq 1$,

$$c^{-1}V \leq V_0 \leq cV. \tag{2.6}$$

By the structure of $\Delta_t = \Delta(Y_{t-1}, \xi_t)$, which is defined in (2.3), and the compactness of $\Theta$, we can choose $\xi_t = \xi^*$ sufficiently small such that

$$\rho_0 = \sup_{\bar{Y}^1} \rho(\Delta(\cdot, \xi^*)) < 1. \tag{2.7}$$

The globally attracting state of $\{Y_t\}$ exists if there exists a fixed point $Y^*$ such that $Y_t$ converges to $Y^*$ as $t \to \infty$ for the control sequence $\{\xi_t = \xi^*\}$ and any starting value $Y_0$. Here $Y^*$ depends on the choice of $\xi^*$. By Proposition 7.2.5 in Meyn and Tweedie [2009], the existence of $Y^*$ is equivalent to that the nonlinear control system (2.5) is $M$-irreducible, which is also equivalent to that $\{Y_t\}$ is $\psi$-irreducible, given that (2.5) is forward accessible (Theorem 7.2.6 in Meyn and Tweedie [2009]). Furthermore, aperiodicity follows from the fact that any cycle must contain the state $Y^*$.

Therefore, to show $\{Y_t\}$ is $V$-uniformly ergodic, it suffices to verify that

[1](Meyn and Tweedie [2009]) We call a set $C \in \mathcal{B}(\mathcal{X})$ $\nu_a$-petite if the sampled chain satisfies the bound

$$\sum_{n=0}^{\infty} P^n(x, B)a(n) \geq \nu_a(B),$$

for all $x \in C$, $B \in \mathcal{B}(\mathcal{X})$, where $\nu_a$ is a non-trivial measure on $\mathcal{B}(\mathcal{X})$ and $a = \{a(n)\}$ is a distribution or probability measure on $\mathbb{Z}_+$.

1. The globally attracting state $Y^*$ exits.
2. The associated control model (2.5) is forward accessible.
3. The drift condition (2.4) is satisfied for some function $V \geq 1$.

Furthermore, if the function $V$ we use in (2.4) is unbounded, the above three conditions make the assumptions of Theorem 8.0.2(ii) in Meyn and Tweedie [2009] satisfied and thus the chain is recurrent. By Theorem 10.4.4 in Meyn and Tweedie [2009], the chain has a unique (up to constant multiples) subinvariant measure which is invariant.

The above three topics will be discussed in the following three subsections, respectively.

### 2.3.1 The Existence of the Globally Attracting State

By the mean-value theorem, we have

$$
\begin{aligned}
\|Y_{t+1} - Y_t\| &= \|\Delta(Y_t^*, \xi^*)(Y_t - Y_{t-1})\| \\
&= \left\| \prod_{i=1}^{t} \Delta(Y_t^*, \xi^*)(Y_1 - Y_0) \right\| \\
&\leq \left\| \prod_{i=1}^{t} \Delta(Y_i^*, \xi^*) \right\| \|Y_1 - Y_0\| \\
&\leq \sup_{\bar{Y}^1} \|\Delta^t(\cdot, \xi^*)\| \, \|Y_1 - Y_0\| \\
&\leq K\rho_0^t \|Y_1 - Y_0\| \\
&\to 0, \quad \text{as } t \to \infty,
\end{aligned}
$$

where $Y_i^*$ is on the chord between $Y_{i+1}$ and $Y_i$. The last inequality holds due to (2.7) and Lemma 4.1. This proves the existence of the globally attracting state $Y^*$, i.e.,

$$Y_t \to Y^*, \text{ as } t \to \infty.$$

### 2.3.2 Forward Accessibility

Let $\{\Xi_k, \Lambda_k : k \in \mathbb{Z}_+\}$ denote the matrices

$$\begin{aligned}
\Xi_{k+1} &= \Xi_{k+1}(x_0, u_1, \cdots, u_{k+1}) := \left[\frac{\partial F}{\partial x}\right]_{(x_k, u_{k+1})}, \\
\Lambda_{k+1} &= \Lambda_{k+1}(x_0, u_1, \cdots, u_{k+1}) := \left[\frac{\partial F}{\partial u}\right]_{(x_k, u_{k+1})},
\end{aligned}$$

where $x_k = F_k(x_0, u_1, \cdots, u_k)$. Let $C_{x_0}^k = C_{x_0}^k(u_1, \cdots, u_k)$ denote the *generalized controllability matrix* (along with sequence $u_1, \cdots, u_k$)

$$C_{x_0}^k := [\Xi_k \cdots \Xi_2 \Lambda_1 | \Xi_k \cdots \Xi_3 \Lambda_2 | \cdots | \Xi_k \Lambda_{k-1} | \Lambda_k].$$

**Rank condition for multivariate control models (Condition (CM3) (7.13) in Meyn and Tweedie [2009])**

> For each initial condition $x_0 \in \mathbb{R}^N$, there exists $k \in \mathbb{Z}_+$ and a sequence $\vec{u}^0 = (u_1^0, \cdots, u_k^0) \in \mathcal{O}_\xi^k$ such that
>
> $$\text{rank} C_{x_0}^k(\vec{u}^0) = N. \qquad (2.8)$$

Proposition 7.1.4 in Meyn and Tweedie [2009] states that the control model (2.5) is forward accessible if and only if the rank condition (2.8) holds. In particular, if $\Lambda_1 = \partial F(x,e)/\partial e$ has full rank (i.e., $k=1$), condition (2.8) is satisfied.

$$\Lambda_1 = \frac{\partial F(Y_{t-1},\xi_t)}{\xi_t} = \frac{\partial w_t}{\xi_t} = \begin{pmatrix} 0 \\ \frac{\partial \eta_t}{\partial \xi_t^T} \\ 0 \end{pmatrix}.$$

It suffices to verify that $\frac{\partial \eta_t}{\partial \xi_t^T}$ has full rank for our chosen $\xi^*$. By (1), (2) and (3) in Appendix B,

$$\begin{aligned}
\left.\frac{\partial \eta_t}{\partial \xi_t^T}\right|_{\xi_t=\xi^*} &= \left.\frac{\partial}{\partial \xi_t^T}\mathrm{vech}(H_t^{1/2}\xi_t\xi_t^T H_t^{-1/2})\right|_{\xi_t=\xi^*} \\
&= D_d^+ \left.\frac{\partial}{\partial \xi_t^T}\mathrm{vec}(H_t^{1/2}\xi_t\xi_t^T H_t^{1/2})\right|_{\xi_t=\xi^*} \\
&= D_d^+(H_t^{1/2}\otimes H_t^{1/2}) \left.\frac{\partial \mathrm{vec}(\xi_t\xi_t^T)}{\partial \xi_t^T}\right|_{\xi_t=\xi^*} \\
&= D_d^+(H_t^{1/2}\otimes H_t^{1/2})(I_{d^2}+K_{dd})(\xi^*\otimes I_d) \\
&= D_d^+(H_t^{1/2}\otimes H_t^{1/2})D_d\cdot 2D_d^+(\xi^*\otimes I_d).
\end{aligned}$$

By (4) in Appendix B,

$$|D_d^+(H_t^{1/2}\otimes H_t^{1/2})D_d| = |H_t|^{(d+1)/2} \neq 0.$$

It now remains to show that $D_d^+(\xi^* \otimes I_d)$ has rank $d$. Note that $D_d^+$ only contains 1 and 0. We denote the $i^{th}$ column of $\xi^* \otimes I_d$ by $\Pi_i$. Then

$$\begin{aligned} D_d^+\Pi_i &= D_d^+\mathrm{vec}[(0,\cdots,0,\xi^*,0,\cdots,0)^T] \\ &= \mathrm{vech}[(0,\cdots,0,\xi^*,0,\cdots,0)^T] \\ &= \mathrm{vech}\begin{pmatrix} 0_{(i-1)\times d} \\ \xi^{*T} \\ 0_{(d-i)\times d} \end{pmatrix} \end{aligned}$$

We can see that the $i^{th}$ column of $D_d^+(\xi^* \otimes I_d)$ has $i$ non-zero elements, which are the first $i$ entries of the vector $\xi^*$. Furthermore, there is no more than one non-zero element on each row of $D_d^+(\xi^* \otimes I_d)$. Then we have

$$D_d^+(\xi^* \otimes I_d) = M\begin{pmatrix} \underline{\xi}_1 & 0 & \cdots & 0 \\ 0 & \underline{\xi}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \underline{\xi}_d \end{pmatrix},$$

where $M$ is an elementary matrix (see Appendix A for details) and $\underline{\xi}_i$ is an $i$-dimensional vector with all elements being the $i^{th}$ element of $\xi^*$. Since $M$ has full rank, we can have a properly chosen $\xi^*$ such that $D_d^+(\xi^* \otimes I_d)$ has rank $d$.

### 2.3.3 Drift Condition

By the mean-value theorem, there exists $\bar{Y}^*_{t-1}$ on the chord between $Y^*$ and $Y_{t-1}$ such that

$$F(Y_{t-1},\xi_t) - F(Y^*,\xi_t) = \Delta(\bar{Y}^*_{t-1},\xi_t)(Y_{t-1} - Y^*).$$

Then we have

$$\begin{aligned} Y_t &= F(Y_{t-1},\xi_t) \\ &= F(Y_{t-1},\xi_t) - F(Y^*,\xi_t) + F(Y^*,\xi_t) \\ &= F(Y^*,\xi_t) + \Delta(\bar{Y}^*_{t-1},\xi_t)(Y_{t-1} - Y^*) \\ &= \omega(\bar{Y}^*_{t-1},\xi_t) + \Delta(\bar{Y}^*_{t-1},\xi_t)Y_{t-1}, \qquad (2.9) \end{aligned}$$

where $\omega(\bar{Y}^*_{t-1},\xi_t) = F(Y^*,\xi_t) - \Delta(\bar{Y}^*_{t-1},\xi_t)Y^*$. Applying (2.9) recursively, we get

$$\begin{aligned} Y_t = {} & \omega(\bar{Y}^*_{t-1},\xi_t) + \sum_{j=1}^{m-1}\left(\prod_{k=1}^{j}\Delta(\bar{Y}^*_{t-k},\xi_{t-k+1})\right)\omega(\bar{Y}^*_{t-j-1},\xi_{t-j}) \\ & + \prod_{k=1}^{m}\Delta(\bar{Y}^*_{t-k},\xi_{t-k+1})Y_{t-m}, \end{aligned}$$

where $\bar{Y}^*_{t-k}$ on the chord between $Y^*$ and $Y_{t-k}$, $k = 1,\cdots,m$.

Define

$$\Omega = \sup_{\bar{Y}^m}\left\|\prod_{k=1}^{m}\Delta(Y_{m-k+1},\xi_{m-k+2})\right\| \quad \text{and} \quad \lambda = \mathbb{E}(\Omega^s) \text{ for some } s.$$

Consider the function $g(x) = \mathbb{E}(\Omega^x)$. We have $g(0) = 1$ and

$$\lim_{h \downarrow 0} \frac{g(h) - g(0)}{h} = \mathbb{E}(\Omega^x \log \Omega)|_{x=0} = \mathbb{E}(\log(\Omega)) < 0.$$

The last inequality results from Assumption A4. Thus we can choose $0 < s < r$ such that $\lambda < 1$, where $r$ is given is Assumption A2.

Next, consider the drift function

$$V(x) = 1 + \|x\|^s \ . \tag{2.10}$$

We observe that

$$\begin{aligned}
& \mathbb{E}(V(Y_t)|Y_{t-m} = Y_0) \\
\leq \ & 1 + \mathbb{E} \sup_{\bar{Y}^1} \|\omega(\cdot, \xi_1)\|^s \\
& + \sum_{j=1}^{m-1} \mathbb{E} \left( \sup_{\bar{Y}^j} \left\| \prod_{k=1}^{j} \Delta(\cdot, \xi_k) \right\|^s \right) \mathbb{E} \sup_{\bar{Y}^1} \|\omega(\cdot, \xi_1)\|^s + \lambda \|Y_0\|^s \\
= \ & \lambda V(Y_0) + b,
\end{aligned}$$

where

$$b = (1 - \lambda) + \mathbb{E} \sup_{\bar{Y}^1} \|\omega(\cdot, \xi_1)\|^s + \sum_{j=1}^{m-1} \mathbb{E} \left( \sup_{\bar{Y}^j} \left\| \prod_{k=1}^{j} \Delta(\cdot, \xi_k) \right\|^s \right) \mathbb{E} \sup_{\bar{Y}^1} \|\omega(\cdot, \xi_1)\|^s.$$

Thus,

$$\Delta V(Y_0) = \mathbb{E}(V(Y_t)|Y_{t-m} = Y_0) - V(Y_0) \leq (\lambda - 1)V(Y_0) + b.$$

We choose the measurable set $C$ as

$$C = \left\{ Y : V(Y) = 1 + \|Y\|^s < \frac{2}{1-\lambda} b \right\}.$$

For $Y_0 \in C$,

$$\Delta V(Y_0) \leq \frac{\lambda - 1}{2} V(Y_0) + b.$$

For $Y_0 \in C^c$,

$$\Delta V(Y_0) \leq (\lambda - 1)V(Y_0) + \frac{1-\lambda}{2} V(Y_0) = \frac{\lambda - 1}{2} V(Y_0)$$

Then (2.4) is satisfied if we assign $\beta = \frac{1-\lambda}{2}$. It remains to show that $b$ is finite, which suffices to show that both $\mathbb{E}(\sup_{\bar{Y}^1} \|\Delta(\cdot, \xi_1)\|^r)$ and $\mathbb{E}(\sup_{\bar{Y}^1} \|\omega(\cdot, \xi_1)\|^r)$ are finite. By Theorem 5.6.9 in Horn and Johnson [1985], the spectral radius is a lower bound for any matrix norm. By (5) and (6) in Appendix B, we have that

$$\|\Delta_t\| \quad \leq \quad \|J\| + \left\| \begin{pmatrix} 0 \\ \\ \frac{\partial \eta_t}{\partial Y_{t-1}^T} \\ \\ 0 \end{pmatrix} \right\|$$

$$
\begin{aligned}
&\leq \|J\| + \left\|\frac{\partial \eta_t}{\partial Y_{t-1}^T}\right\| \\
&= \|J\| + \|D_d^+(\tilde{\Delta}_t \otimes I_d)D_d[B_1|B_2|\cdots|B_p|A_1|A_2|\cdots|A_q]\| \\
&\leq C_1 + C_2\|\tilde{\Delta}_t \otimes I_d\| \\
&= C_1 + C_2\|\tilde{\Delta}_t\|\|I_d\| \\
&\leq C_1 + C_2\|\tilde{\Delta}_t\|_2 \\
&= C_1 + C_2\sqrt{\mathrm{tr}(\tilde{\Delta}_t^T\tilde{\Delta}_t)} \\
&\leq C_1 + C_2\sqrt{\frac{1}{4}[\mathrm{tr}(\tilde{\Delta}_t) + \mathrm{tr}(\tilde{\Delta}_t^T)]^2} \\
&= C_1 + C_2\mathrm{tr}(\tilde{\Delta}_t) \\
&\leq C_1 + C_2\xi_t^T\xi_t,
\end{aligned}
$$

where $C_1 = \|J\|$ and $C_2 = \|D_d^+\| \cdot \|D_d\| \cdot \|[B_1|B_2|\cdots|B_p|A_1|A_2|\cdots|A_q]\|$. We then obtain by Assumption A2 that

$$
\mathbb{E}(\sup_{\bar{Y}^1} \|\Delta(\cdot,\xi_1)\|^r) \leq C_1^r + C_2^r\mathbb{E}(\xi_t^T\xi_t)^r = C_1^r + C_3^r\mathbb{E}\|\xi_t\|^{2r} < \infty.
$$

The finiteness of $\mathbb{E}(\sup_{\bar{Y}^1}\|\omega(\cdot,\xi_1)\|^r)$ will follow.

## 2.4 Proof of Proposition 2.4

1. We apply the "vec" operator on both sides of (1.6). By Equation (1) in Appendix B, we have that

$$\mathrm{vec}(H_t) = \mathrm{vec}(C) + \sum_{i=1}^{q} \tilde{A}_i \mathrm{vec}(y_{t-i}y_{t-i}^T) + \sum_{i=1}^{p} \tilde{B}_i \mathrm{vec}(H_{t-i}), \tag{2.11}$$

where

$$\tilde{A}_i = \sum_{j=1}^{k} A_{ij} \otimes A_{ij} \quad \text{and} \quad \tilde{B}_i = \sum_{j=1}^{k} B_{ij} \otimes B_{ij}.$$

Since $H_t$ and $y_t y_t^T$ are symmetric, we left multiply the matrix $D_d^+$ on both sides of (2.11) and we can obtain that

$$\begin{aligned} & \mathrm{vech}(H_t) \\ = \;& \mathrm{vech}(C) + \sum_{i=1}^{q} D_d^+ \tilde{A}_i \mathrm{vec}(y_{t-i}y_{t-i}^T) + \sum_{i=1}^{p} D_d^+ \tilde{B}_i \mathrm{vec}(H_{t-i}) \\ = \;& \mathrm{vech}(C) + \sum_{i=1}^{q} D_d^+ \tilde{A}_i D_d \mathrm{vech}(y_{t-i}y_{t-i}^T) + \sum_{i=1}^{p} D_d^+ \tilde{B}_i D_d \mathrm{vech}(H_{t-i}), \end{aligned}$$

which is the same as (1.4) if we assign

$$A_i = D_d^+ \tilde{A}_i D_d \quad \text{and} \quad B_i = D_d^+ \tilde{B}_i D_d.$$

Suppose $\lambda$ and $u$ is one of the nonzero eigenpairs of $B$, where $u = (u_1^*, \cdots, u_p^*)^* \in \mathbb{C}^{pN}$ and $*$ denotes the conjugate transpose. We have by definition that

$$Bu = \lambda u,$$

that is,

$$\lambda u_1 = \sum_{i=1}^{p} B_i u_i \quad \text{and} \quad \lambda u_j = u_{j-1} \quad \text{for} \quad 1 < j \leq p,$$

It is therefore the case that $u_p \neq 0$ (otherwise $u = 0$) and

$$\lambda^p u_p = \left( \sum_{i=1}^{p} \lambda^{p-i} B_i \right) u_p. \tag{2.12}$$

Let $U$ be the symmetric matrix such that $\text{vech}(U) = u_p$. Thus, from (2.12), we have that

$$\begin{aligned} \text{vech}(\lambda^p U) = \lambda^p u_p &= \left( \sum_{i=1}^{p} \lambda^{p-i} B_i \right) \text{vech}(U) \\ &= \sum_{i=1}^{p} \lambda^{p-i} D_d^+ \tilde{B}_i D_d \text{vech}(U) \\ &= \sum_{i=1}^{p} \lambda^{p-i} D_d^+ \tilde{B}_i \text{vec}(U) \\ &= \sum_{i=1}^{p} \lambda^{p-i} D_d^+ \left( \sum_{j=1}^{k} B_{ij} \otimes B_{ij} \right) \text{vec}(U) \\ &= \sum_{i=1}^{p} \sum_{j=1}^{k} \lambda^{p-i} D_d^+ \text{vec}(B_{ij} U B_{ij}^T) \\ &= \sum_{i=1}^{p} \sum_{j=1}^{k} \lambda^{p-i} \text{vech}(B_{ij} U B_{ij}^T). \end{aligned}$$

Equivalently,

$$\lambda^p U = \sum_{i=1}^{p}\sum_{j=1}^{k} \lambda^{p-i} B_{ij} U B_{ij}^T. \tag{2.13}$$

We obtain (2.13) due to the fact that $\mathrm{vech}(A) = \mathrm{vech}(B)$ implies $A = B$ if both $A$ and $B$ are symmetric matrices. Note that the $\mathrm{vech}(\cdot)$ operator obeys the linear property, i.e., $\mathrm{vech}(cA) = c\mathrm{vech}(A)$ for a constant $c$.

We define a function $\varphi(\cdot)$ by

$$\varphi(X) = \sum_{i=1}^{p}\sum_{j=1}^{k} B_{ij} X B_{ij}^T,$$

whose argument is from the class from symmetric positive definite matrices. We denote the $n$-th order iterative function of $\varphi(\cdot)$ by $\varphi^n(\cdot)$, that is,

$$\varphi^n(\cdot) = \varphi(\varphi^{n-1}(\cdot)).$$

We define the matrix norm $\|\cdot\|_V$ on any arbitrary matrix $P \in \mathbb{C}^{d\times d}$ by

$$\|P\|_V = \sup\{|x^* P x| : x \in \mathbb{C}^d \quad \text{and} \quad x^* V x = 1\},$$

where $V$ is defined as

$$V = \sum_{n=0}^{\infty} \varphi^n(C),$$

where $C$ is the constant matrix in (1.6). It remains to show that $V$ is well defined. $V$ is trivially symmetric positive definite. Applying the "vec" operator on $\varphi(C)$ gives

$$\begin{aligned}
\mathrm{vec}(\varphi(C)) &= \sum_{i=1}^{p}\sum_{j=1}^{k}(B_{ij}\otimes B_{ij})\mathrm{vec}(C) \\
&= \sum_{i=1}^{p}\tilde{B}_i\mathrm{vec}(C) \\
&= D_d^{+}\left(\sum_{i=1}^{p}B_i\right)D_d\mathrm{vec}(C),
\end{aligned}$$

Suppose that

$$\mathrm{vec}(\varphi^n(C)) = D_d^{+}\left(\sum_{i=1}^{p}B_i\right)^n D_d\mathrm{vec}(C). \tag{2.14}$$

Then

$$\begin{aligned}
&\mathrm{vec}(\varphi^{n+1}(C)) \\
=\ &\mathrm{vec}(\varphi(\varphi^n(C))) \\
=\ &\mathrm{vec}\left(\varphi\left(\mathrm{mat}\left(D_d^{+}\left[\sum_{i=1}^{p}B_i\right]^n D_d\mathrm{vec}(C)\right)\right)\right) \\
=\ &\mathrm{vec}\left(\sum_{i=1}^{p}\sum_{j=1}^{k}B_{ij}\left[\mathrm{mat}\left(D_d^{+}\left[\sum_{l=1}^{p}B_l\right]^n D_d\mathrm{vec}(C)\right)\right]B_{ij}^T\right) \\
=\ &\sum_{i=1}^{p}\sum_{j=1}^{k}(B_{ij}\otimes B_{ij})\left(D_d^{+}\left[\sum_{l=1}^{p}B_l\right]^n D_d\mathrm{vec}(C)\right) \\
=\ &\sum_{i=1}^{p}\tilde{B}_i D_d^{+}\left(\sum_{l=1}^{p}B_l\right)^n D_d\mathrm{vec}(C)
\end{aligned}$$

$$
\begin{aligned}
&= D_d^+ \left( \sum_{i=1}^p D_d \tilde{B}_i D_d^+ \right) D_d D_d^+ \left( \sum_{l=l}^p B_l \right)^n D_d \mathrm{vec}(C) \\
&= D_d^+ \left( \sum_{i=1}^p B_i \right) \left( \sum_{l=l}^p B_l \right)^n D_d \mathrm{vec}(C) \\
&= D_d^+ \left( \sum_{i=1}^p B_i \right)^{n+1} D_d \mathrm{vec}(C).
\end{aligned}
$$

Therefore, (2.14) holds due to the induction. Since

$$
\begin{aligned}
\sum_{n=0}^{\infty} \|\varphi^n(C)\| &\leq \sum_{n=0}^{\infty} \|\mathrm{vec}(\varphi^n(C))\| \\
&\leq \sum_{n=0}^{\infty} \|D_d^+\| \left\| \left( \sum_{i=1}^p B_i \right)^n \right\| \|D_d \mathrm{vec}(C)\| \\
&\leq K \sum_{n=0}^{\infty} \left[ \rho \left( \sum_{i=1}^p B_i \right) \right]^n < \infty,
\end{aligned}
$$

$V$ is well defined. The last inequality holds due to Lemma 4.1. We also have that

$$
\begin{aligned}
V &= \sum_{n=0}^{\infty} \varphi^n(C) \\
&= \varphi^0(C) + \sum_{n=1}^{\infty} \varphi^n(C) \\
&= C + \sum_{n=1}^{\infty} \varphi(\varphi^{n-1}(C)) \\
&= C + \varphi \left( \sum_{n=1}^{\infty} \varphi^{n-1}(C) \right) \\
&= C + \varphi(V) \qquad (2.15)
\end{aligned}
$$

It comes that for any matrix $P \in \mathbb{C}^{d\times d}$,

$$|x^*Px| = \left|\left(\frac{x}{\sqrt{x^*Vx}}\right)^* P\frac{x}{\sqrt{x^*Vx}}\right|(x^*Vx) \leq \|P\|_V(x^*Vx). \tag{2.16}$$

This inequality holds since $\left(\frac{x}{\sqrt{x^*Vx}}\right)^* V\frac{x}{\sqrt{x^*Vx}} = 1$ (notice that $x^*Vx \neq 0$ if $x \neq 0$).

For any $x$,

$$\begin{aligned}
|\lambda|^p|x^*Ux| &\overset{(2.13)}{=} |\sum_{i=1}^{p}\sum_{j=1}^{k}\lambda^{p-i}x^*B_{ij}UB_{ij}^Tx| \\
&\leq \sum_{i=1}^{p}\sum_{j=1}^{k}|\lambda|^{p-i}|x^*B_{ij}UB_{ij}^Tx| \\
\text{(by (2.16))} \quad &\leq \sum_{i=1}^{p}\sum_{j=1}^{k}|\lambda|^{p-i}\|U\|_V(x^*B_{ij}VB_{ij}^Tx) \qquad (2.17)
\end{aligned}$$

Suppose there exists one of the eigenvalues of $B$ which is greater than or equal to 1 in modulus, denoted by $\lambda_0$. Also assume that $x_0$ is such that $|x_0^*Ux_0| = \|U\|_V$ and $x_0^*Vx_0 = 1$. Substituting $\lambda_0$ and $x_0$ into (2.17) we obtain

$$\begin{aligned}
|\lambda_0|^p &\leq \sum_{i=1}^{p}\sum_{j=1}^{k}|\lambda_0|^{p-i}(x_0^*B_{ij}VB_{ij}^Tx_0) \\
&\leq |\lambda_0|^{p-1}\sum_{i=1}^{p}\sum_{j=1}^{k}x_0^*B_{ij}VB_{ij}^Tx_0 \\
&= |\lambda_0|^{p-1}x_0^*\left(\sum_{i=1}^{p}\sum_{j=1}^{k}B_{ij}VB_{ij}^T\right)x_0
\end{aligned}$$

$$\begin{aligned} &= |\lambda_0|^{p-1} x_0^*(V - C)x_0 \\ \text{(by (2.15))} \quad &= |\lambda_0|^{p-1}(1 - x_0^* C x_0) \end{aligned}$$

$x_0^* C x_0 > 0$ since $C$ is symmetric positive definite. Therefore, $|\lambda_0| < 1$. This contradiction finalizes our proof.

2. We define a function $\tilde{\varphi}(\cdot)$ by

$$\tilde{\varphi}(X) = \sum_{i=1}^{q} \sum_{j=1}^{k} A_{ij} X A_{ij}^T + \sum_{i=1}^{p} \sum_{j=1}^{k} B_{ij} X B_{ij}^T,$$

whose argument is from the class from symmetric positive definite matrices. The matrix $\tilde{V}$ is defined as

$$\tilde{V} = \sum_{n=0}^{\infty} \varphi^n(C).$$

Similarly, we have

$$\tilde{V} = C + \tilde{\varphi}(\tilde{V}). \tag{2.18}$$

We also need to show that $\tilde{V}$ is well defined. Similar to the previous part, we have that

$$\mathrm{vec}(\tilde{\varphi}^n(C)) = D_d^+ \left( \sum_{i=1}^{q} A_i + \sum_{i=1}^{p} B_i \right)^n D_d \mathrm{vec}(C).$$

Therefore,

$$\begin{aligned}\sum_{n=0}^{\infty}\|\tilde{\varphi}^n(C)\| &\leq \sum_{n=0}^{\infty}\|\mathrm{vec}(\tilde{\varphi}^n(C))\| \\ &\leq \sum_{n=0}^{\infty}\|D_d^+\|\left\|\left(\sum_{i=1}^{q}A_i+\sum_{i=1}^{p}B_i\right)^n\right\|\|D_d\mathrm{vec}(C)\| \\ &\leq K\sum_{n=0}^{\infty}\left[\rho\left(\sum_{i=1}^{q}A_i+\sum_{i=1}^{p}B_i\right)\right]^n<\infty.\end{aligned}$$

and $\tilde{V}$ is well defined.

Suppose $\lambda$ and $u$ is one of the nonzero eigenpairs of $J$, where $u=(u_1^T,\cdots,u_p^T,u_{p+1}^T,\cdots,u_{p+q}^T)^T\in\mathbb{C}^{(p+q)N}$, then we have by definition that

$$Ju=\lambda u,$$

that is

$$\lambda u_1=\sum_{i=1}^{p}B_iu_i+\sum_{i=1}^{q}A_iu_{p+i} \tag{2.19}$$

and

$$\lambda u_j=u_{j-1}\quad\text{for } 1<j\leq p,\qquad u_j=0\quad\text{for } p+1\leq j\leq p+q.$$

Then (2.19) can be rewritten as

$$\lambda u_1=\sum_{i=1}^{p}B_iu_i.$$

It is therefore the case that $u_p \neq 0$ (otherwise $u = 0$) and

$$\lambda^p u_p = \left( \sum_{i=1}^{p} \lambda^{p-i} B_i \right) u_p. \tag{2.20}$$

Note that (2.20) and (2.12) are the same and therefore we can finish our proof by repeating the steps in the previous part.

3. By (2.18), we have

$$\tilde{V} = C + \sum_{i=1}^{q} \sum_{j=1}^{k} A_{ij} \tilde{V} A_{ij}^T + \sum_{i=1}^{p} \sum_{j=1}^{k} B_{ij} \tilde{V} B_{ij}^T = \tilde{C} + \varphi(\tilde{V}), \tag{2.21}$$

where $\tilde{C} = C + \sum_{i=1}^{q} \sum_{j=1}^{k} A_{ij} \tilde{V} A_{ij}^T$. Notice that $\tilde{C}$ is also symmetric positive definite.

Suppose $\lambda$ and $u$ is one of the nonzero eigenpairs of $\sum_{i=1}^{p} B_i$ and $U$ is such that $\mathrm{vech}(U) = u$. We have

$$\begin{aligned}
\mathrm{vech}(\lambda U) = \lambda u &= \left( \sum_{i=1}^{p} B_i \right) \mathrm{vech}(U) \\
&= \sum_{i=1}^{p} D_d^+ \tilde{B}_i D_d \mathrm{vech}(U) \\
&= \sum_{i=1}^{p} D_d^+ \tilde{B}_i \mathrm{vec}(U) \\
&= \sum_{i=1}^{p} D_d^+ \left( \sum_{j=1}^{k} B_{ij} \otimes B_{ij} \right) \mathrm{vec}(U) \\
&= \sum_{i=1}^{p} \sum_{j=1}^{k} D_d^+ \mathrm{vec}(B_{ij} U B_{ij}^T)
\end{aligned}$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{k}\mathrm{vech}(B_{ij}UB_{ij}^T).$$

Equivalently,

$$\lambda U = \sum_{i=1}^{p}\sum_{j=1}^{k} B_{ij}UB_{ij}^T. \tag{2.22}$$

We define the matrix norm $\|\cdot\|_{\tilde{V}}$ on any arbitrary matrix $P \in \mathbb{C}^{d\times d}$ by

$$\|P\|_{\tilde{V}} = \sup\{|x^*Px| : x \in \mathbb{C}^d \quad \text{and} \quad x^*\tilde{V}x = 1\},$$

It comes that

$$|x^*Px| = \left|\left(\frac{x}{\sqrt{x^*\tilde{V}x}}\right)^* P\frac{x}{\sqrt{x^*\tilde{V}x}}\right|(x^*\tilde{V}x) \le \|P\|_{\tilde{V}}(x^*\tilde{V}x). \tag{2.23}$$

For any $x$,

$$\begin{aligned}
|\lambda||x^*Ux| &\overset{(2.22)}{=} |\sum_{i=1}^{p}\sum_{j=1}^{k} x^*B_{ij}UB_{ij}^Tx| \\
&\le \sum_{i=1}^{p}\sum_{j=1}^{k} |x^*B_{ij}UB_{ij}^Tx| \\
\text{(by (2.23))} \quad &\le \sum_{i=1}^{p}\sum_{j=1}^{k} \|U\|_{\tilde{V}}(x^*B_{ij}\tilde{V}B_{ij}^Tx) \qquad (2.24)
\end{aligned}$$

Suppose there exists one of the eigenvalues of $\sum_{i=1}^{p} B_i$ which is greater than or equal to 1 in modulus, denoted by $\lambda_0$. Also assume that $x_0$ is such that $|x_0^*Ux_0| = \|U\|_{\tilde{V}}$ and $x_0^*\tilde{V}x_0 = 1$. Substituting $\lambda_0$ and $x_0$ into (2.24) we obtain

$$\begin{aligned}
|\lambda_0| &\leq \sum_{i=1}^{p}\sum_{j=1}^{k}(x_0^* B_{ij}\tilde{V}B_{ij}^T x_0) \\
&\leq \sum_{i=1}^{p}\sum_{j=1}^{k} x_0^* B_{ij}\tilde{V}B_{ij}^T x_0 \\
&= x_0^* \left(\sum_{i=1}^{p}\sum_{j=1}^{k} B_{ij}\tilde{V}B_{ij}^T\right) x_0 \\
&= |\lambda_0|^{p-1} x_0^*(\tilde{V}-\tilde{C})x_0 \\
&= |\lambda_0|^{p-1}(1 - x_0^*\tilde{C}x_0)
\end{aligned}$$

$x_0^*\tilde{C}x_0 > 0$ since $\tilde{C}$ is symmetric positive definite. Therefore, $|\lambda_0| < 1$. This contradiction finalizes our proof.

## 2.5 Conclusion and Commentary

This chapter serves as preliminary results for the next chapter. In this chapter, we give conditions under which the GARCH process is ergodic and stationary. The proof is based on the Markov chain technique in Meyn and Tweedie [2009]. This approach was first used by Hafner and Preminger [2009] on the general multivariate GARCH$(1,1)$ model. We extend it to the general multivariate GARCH$(p,q)$ case. Assumption A3 guarantees that if we iterate (2.2) to the infinite past, the infinite sum is well defined. Assumption A4 makes it possible to find the exact value of $\beta$ and $b$ for our chosen function $V$ in the drift condition (2.4). Although we assume the innovation process has unit variance, we do not need this for the ergodicity purpose. We only require the innovation process to have a finite small moment (Assumption A2). Assumption A5 is also assumed in the next chapter to prove consistency. Here, we need this assumption to obtain (2.7). Proposition (2.4) provides sufficient conditions for Assumption A3 for the VEC models with BEKK representations. This result is also useful for the next chapter when we prove the consistency.

Ergodicity is useful for the next chapter when we prove the asymptotic theory for the QMLE. For example, when we consider the limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} l_t(\theta_0),$$

the usual law of large numbers does not apply since $l_t$'s are not independent. We are able to apply the ergodic theorem instead of law of large numbers due to the ergodicity of the model. The stationarity makes the expectation not depend on time. For example, $\mathbb{E}l_t(\theta) = \mathbb{E}l_1(\theta)$ for any $t$.

# Chapter 3

# Asymptotic Theory

## 3.1 Introduction

In Section 1.3, we defined the QMLE for the model parameters. The QMLE is such that it maximizes $\tilde{L}(\theta)$, i.e., the likelihood function conditional on some initial values. This is different from the theoretical likelihood function $L(\theta)$ which depends on infinite past. Moreover, the normal density function we are using in $L(\theta)$ may or may not be consistent with the true distribution of $\xi_t$. In this situation, we still want the estimator to be consistent asymptotically normal. Consistency means that the estimator converges to the true parameter value, i.e.,

**Definition 3.1** (Consistency). *An estimator $\hat{\theta}_n$ is called consistent if*

$$\hat{\theta}_n \xrightarrow{a.s.} \theta_0.$$

Asymptotic normality means that the difference between the estimate and the true parameter converges to a normal distribution, i.e.,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N(0, \Gamma),$$

where $\Gamma$ is a positive definite matrix. We will specify $\Gamma$ later in this chapter.

Sections 3.3 and 3.4 gives conditions under which the QMLE is consistent and asymptotically normal. To achieve the asymptotic theory, another important intermediate result besides the ergodicty and stationarity is the model identifiability.

**Definition 3.2** (Identifiability)**.** *The GARCH model* (1.4) *is identifiable if* $\forall \theta, \theta_0 \in \Theta$,

$$H_t(\theta) = H_t(\theta_0) \quad P_{\theta_0} \ a.s. \Rightarrow \theta = \theta_0.$$

The rest of this chapter is organized as follows. In Section 3.2.1, we give necessary and sufficient conditions under which our GARCH model is identifiable. We also give a counter example in Section 3.2.2 to show that the sufficient conditions for identifiability in Jeantheau [1998] are invalid. Section 3.3 provides assumptions under which the QMLE is consistent. To prove this, we use a different AR(1) type representation other than the one in Chapter 2. We prove the consistency theorem using the standard compactness argument and thus the compactness assumption for the parameter space is essential. Section 3.4 proves the asymptotic normality with two additional assumptions. We only require the finite sixth moment on the observed process, which is by far the weakest assumption in literature for general multivariate GARCH models.

## 3.2 Identifiability

### 3.2.1 The Identifiability Theorem

We start this section with an important concept "matrix polynomial".

**Definition 3.3** (Matrix Polynomial)**.** *A univariate matrix polynomial $P$ of degree $p$ is defined as*

$$P(x) = \sum_{i=0}^{p} C_i x^i,$$

*where $C_i$ denotes a matrix of constant coefficients, and $C_p$ is non-zero.*

We define two matrix polynomials $\mathcal{A}(w) = \sum_{i=1}^{q} A_i w^i$ and $\mathcal{B}(w) = I_N - \sum_{j=1}^{p} B_j w^j$. Using the lag operator $L$, (1.4) can be rewritten as

$$\mathcal{B}(L)h_t = c + \mathcal{A}(L)\eta_t. \tag{3.1}$$

Hereafter, we denote the model formulation at the true parameter value by $\mathcal{A}_{\theta_0}(w)$ and $\mathcal{B}_{\theta_0}(w)$, and use $\mathcal{A}_{\theta}(w)$ and $\mathcal{B}_{\theta}(w)$ to denote the model formulation at any arbitrary parameter value.

In the univariate case, we usually assume that the two polynomials are coprime to get the identifiability. It is natural to generalize this in the multivariate case.

**Definition 3.4** (Greatest Common Left Divisor)**.** *Let $A$ and $B$ be two matrix polynomials such that their determinants are not zero. If there exist a matrix polynomial $D$ such that*

*every left divisor of $D$ is also a left divisor of $A$ and $B$, and*

*every left divisor of $A$ and $B$ is also a left divisor of $D$,*

*then $D$ is called the greatest common left divisor (g.c.l.d.) of $A$ and $B$.*

Recall that a square matrix polynomial is unimodular if its determinant is a non-zero constant. Therefore, we say that two matrix polynomials are (left) coprime if any of their greatest common left divisor is unimodular. In this sense, the greatest common left divisor is not unique since a unimodular g.c.l.d. multiplied by a unimodular matrix is still a unimodular g.c.l.d.. The condition that $\mathcal{A}$ and $\mathcal{B}$ are coprime is not sufficient for the model identifiability. We need a further condition to guarantee the identifiability.

**Theorem 3.5** (Identifiability)**.** *Assume that*

**B1:** *The model* (1.4) *has a strictly stationary and ergodic solution.*

**B2:** *The law of $\xi_t$ is such that there is no quadratic form $q$ for which $q(\xi_t) = \delta$ a.s., with some constant $\delta \in \mathbb{R}$.*

**B3:** $\forall \theta \in \Theta, \mathcal{B}_\theta$ *is invertible; $\mathcal{A}_{\theta_0}$ is also invertible.*

**B4:** *$\forall\theta \in \Theta$, $\mathcal{A}_\theta$ and $\mathcal{B}_\theta$ are (left) coprime.*

*Then under Assumptions B1-B4, the model is identifiable if and only if there exists no non-zero row vector $\alpha$ such that*

$$\alpha A_q = \alpha B_p = 0.$$

*Proof.* Section 2.2 gives conditions under which Assumption B1 holds. For detailed proof, see Section 3.2.3. □

This necessary and sufficient condition for the identifiability of multivariate time series was first introduced by Hannan [1969] to verify the identifiability of the vector ARMA model. This condition was also mentioned in Boussama [1998]. Note that identifiability is not to be confused with the model identification concept in statistics.

### 3.2.2 The Counter Example

Jeantheau [1998] provides assumptions for the CCC model (defined in (1.2)) to be identifiable. Let $P(w) = (p_{ij}(w))$ be a matrix polynomial and $d_{ij}$ be the degree of $p_{ij}(w)$. We define

$$d_j(P) = \sup_i d_{ij} \quad \text{and} \quad P_{ij}^{rc} = p_{ij,d_j},$$

which leads to the definition of column-reduced matrix.

**Definition 3.6** (Column Reduced)**.** *A polynomial matrix $P$ is column reduced if the determinant of $P^{rc}$ is not equal to zero.*

Identifiability is claimed in Jeantheau [1998] by replacing our necessary and sufficient condition with the follow additional assumption:

**B5:** Either $\mathcal{A}_{\theta_0}$ or $\mathcal{B}_{\theta_0}$ is column reduced.

Note that Jeantheau [1998] gives only sufficient conditions for identifiability.

However, we find that Jeantheau [1998]'s assumptions may not lead to the identifiability. Here is a counter-example. Let us consider a trivariate GARCH$(1, 2)$ model and let

$$A_1 = \begin{pmatrix} 0.05 & 0 & 0 \\ 0 & 0.02 & 0 \\ 0 & 0 & 0.09 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0.07 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.03 \end{pmatrix},$$

$$B = \begin{pmatrix} 0.6 & 1.2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.4 \end{pmatrix}.$$

We can easily verify that $A_1x + A_2x^2$ and $I - Bx$ are coprime.

$$\mathcal{A}_{\theta_0}^{rc} = \begin{pmatrix} 0.07 & 0 & 0 \\ 0 & 0.02 & 0 \\ 0 & 0 & 0.09 \end{pmatrix},$$

which is of full rank and thus this parameterization satisfies Jeantheau [1998]'s conditions.

However, if we redefine $A_2$ and $B$ as

$$A_2 = \begin{pmatrix} 0.07 & 0.02 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.03 \end{pmatrix}, \quad B = \begin{pmatrix} 0.6 & 0.2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.4 \end{pmatrix}$$

and let $A_1$ remain as the same, these two parameterizations produce exactly the same covariance series. And thus, Jeantheau [1998]'s conditions are invalid. The Mathematica codes for this counter example and the verification are available from the author upon request. This counter example exists because we have a non-zero row vector $\alpha = (0, 1, 0)$ such that

$$\alpha A_2 = \alpha B = 0.$$

We follow the procedure of proving the necessity part in Section 3.2.3 to construct this counter example.

### 3.2.3 Proof of Theorem 3.5

By Assumption B3, (3.1) yields

$$h_t = \mathcal{B}^{-1}(L)(c + \mathcal{A}(L)\eta_t). \tag{3.2}$$

Suppose that $h_t(\theta) = h_t(\theta_0) \quad P_{\theta_0}$ a.s. for some $t \in \mathbb{Z}$. If $\mathcal{A}_\theta(1) \neq 0$, it follows from (3.2) that

$$[\mathcal{B}_\theta^{-1}(L)\mathcal{A}_\theta(L) - \mathcal{B}_{\theta_0}^{-1}(L)\mathcal{A}_{\theta_0}(L)]\eta_t = \mathcal{B}_{\theta_0}^{-1}(1)c_0 - \mathcal{B}_\theta^{-1}(1)c.$$

If $\mathcal{B}_\theta^{-1}(L)\mathcal{A}_\theta(L) - \mathcal{B}_{\theta_0}^{-1}(L)\mathcal{A}_{\theta_0}(L) \neq 0$, there exists a set of constant matrices $D_i, i = 0, 1, \cdots, \infty$ and a constant vector $d_0$ such that $\sum_{i=0}^{\infty} D_i\eta_{t-i} = d_0$. Thus,

$$D_0\eta_t = d_0 - \sum_{i=1}^{\infty} D_i\eta_{t-i}.$$

By taking the conditional expectation given $\mathscr{F}_{t-1}$, the left hand side becomes $D_0h_t$ while the right hand side remains as the same. Hence,

$$0 = D_0(\eta_t - h_t) = D_0\text{vech}(H_t^{1/2}(\xi_t\xi_t' - I)H_t^{1/2}). \tag{3.3}$$

However, by Assumption B2, $\xi_t\xi_t' \neq I$ with a positive probability. Since $H_t^{1/2}$ is positive definite, we conclude that it is impossible that (3.3) holds. Therefore,

$$\mathcal{B}_\theta^{-1}(L)\mathcal{A}_\theta(L) = \mathcal{B}_{\theta_0}^{-1}(L)\mathcal{A}_{\theta_0}(L) \quad \text{and} \quad \mathcal{B}_{\theta_0}^{-1}(1)c_0 = \mathcal{B}_\theta^{-1}(1)c. \tag{3.4}$$

Let $M = \mathcal{B}_\theta\mathcal{B}_{\theta_0}^{-1}$, then we have

$$\mathcal{A}_\theta \quad = \quad M\mathcal{A}_{\theta_0} \tag{3.5}$$

$$\mathcal{B}_\theta \quad = \quad M\mathcal{B}_{\theta_0} \tag{3.6}$$

We want to show that $M = I$.

Please see Appendix A for some definitions and results about the decomposition of rational matrix polynomials. A rational matrix has every element as the ratio of two finite degree polynomials. Hereafter, a matrix polynomial means a matrix whose elements are all polynomials up to a finite degree. Otherwise we can rewrite it as a rational matrix polynomial. By Lemma A.5, since $M$ is a rational matrix polynomial, we can factorize $M$ as $M = SDR$, where $S$ and $R$ are unimodular whose elements are polynomials up to a certain finite order and $D$ is diagonal. $M$ is of full rank since both $\mathcal{B}_\theta$ and $\mathcal{B}_{\theta_0}$ are non-singular. Let $D = P^{-1}Q$, where

$$P = \text{diag}\{p_1, \cdots, p_N\}, \quad Q = \text{diag}\{q_1, \cdots, q_N\}.$$

and $p_i$ does not divide $q_i$ for all $i$. Then (3.5) and (3.6) yield

$$\begin{aligned} S^{-1}\mathcal{B}_\theta &= P^{-1}QR\mathcal{B}_{\theta_0} \\ S^{-1}\mathcal{A}_\theta &= P^{-1}QR\mathcal{A}_{\theta_0} \end{aligned}$$

Notice that the elements of $S^{-1}$ and $R^{-1}$ are also polynomials. Hence $P$ divides $R\mathcal{B}_{\theta_0}$, which means that for all $i$, $p_i$ divides all elements in the $i^{th}$ row of $R\mathcal{B}_{\theta_0}$

and also $P$ divides $R\mathcal{A}_{\theta_0}$. Similarly we have $Q$ divides $S^{-1}\mathcal{B}_\theta$ and $S^{-1}\mathcal{A}_\theta$. That means $P$ divides $\mathcal{A}_{\theta_0}$ and $\mathcal{A}_{\theta_0}$, and $Q$ divides $\mathcal{B}_\theta$ and $\mathcal{A}_\theta$. Thus both $P$ and $Q$ are unimodular (Assumption B4). Note that they are also diagonal, which implies that the diagonal elements of $P$ and $Q$ are all non-zero constants and therefore $M$ is unimodular. If $U$ is the coefficient matrix of $M$'s highest degree, we must have $UA_q(\theta_0) = UB_p(\theta_0) = 0$ in order to make the degree of $M\mathcal{A}_{\theta_0}$ not greater than the degree of $\mathcal{A}_\theta$. By the sufficient condition, $U$=0. Doing this procedure iteratively reduces $M$ to a constant matrix. But $\mathcal{B}_{\theta_0}(0) = \mathcal{B}_\theta(0) = I$. $M$ must be the identity matrix. Thus from (3.4) we can obtain $c = c_0$.

For the necessity part, we assume that there exists a non-zero row vector $\alpha$ such that

$$\alpha A_q = \alpha B_p = 0. \tag{3.7}$$

Let $P$ be an orthogonal matrix such that $P\alpha^T$ has zero as the first entry. Such matrix $P$ exists; for example a simple rotation matrix is one candidate. Then we can choose a vector $\beta$ such that $P\beta$ has the form $(1, 0, \cdots, 0)^T$. For any $w$, the matrix polynomial

$$P(I_N + \beta\alpha w)P^T = PP^T + (P\beta)(\alpha P^T)w = I_N,$$

which has unit determinant. Thus, the matrix polynomial $I_N + \beta\alpha w$ has determinant one since $P$ is orthogonal. We left-multiply $I_N + \beta\alpha L$ on both sides of

(3.1). Note that $(I_N + \beta\alpha L)\mathcal{A}(L)$ and $(I_N + \beta\alpha L)\mathcal{B}(L)$ still have orders $q$ and $p$ respectively due to (3.7) and they are still coprime since $I_N + \beta\alpha w$ is unimodular. This different parameterization generates the same process, which makes the identifiability invalid.

## 3.3 Consistency

Conditional on initial values, $H_t(\theta)$ can be calculated recursively, denoted by $\tilde{H}_t(\theta)$. We also define $\tilde{h}_t(\theta)$, $\tilde{l}_t(\theta)$, $\tilde{L}_t(\theta)$ analogously. It will be shown in Lemma 4.4 that the choice of initial values does not matter asymptotically.

We rewrite (1.4) in the form as

$$X_t = c_t + BX_{t-1}, \tag{3.8}$$

where

$$X_t = \begin{pmatrix} h_t \\ h_{t-1} \\ \vdots \\ h_{t-p+1} \end{pmatrix}, \qquad c_t = \begin{pmatrix} c + \sum_{i=1}^{q} A_i \eta_{t-i} \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

$$B = \begin{pmatrix} B_1 & B_2 & \cdots & B_{p-1} & B_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & I & 0 \end{pmatrix}.$$

Note that (3.8) is a different iteration than (2.2). Different iterations are used to study different aspects of the process.

**Theorem 3.7** (Consistency)**.** *Assume that*

**C1:** $\Theta$ *is compact.*

**C2:** *The model satisfies the stationarity and ergodicity assumptions given by Theorem 2.3 and the identifiability assumptions given by Theorem 3.5.*

**C3:** $\mathbb{E}(\|y_t\|^s) < \infty$ *for some* $s > 0$*.*

**C4:** $\rho(B) < 1$

*Then under Assumptions C1-C4, we have*

$$\hat{\theta}_n \xrightarrow{a.s.} \theta_0.$$

*Proof.* For any $\theta \in \Theta$ and any integer $k$, let $V_k(\theta)$ be the open ball with center $\theta$

and radius $1/k$. For any $k$, the parameter space $\Theta$ has an open cover

$$V_k(\theta_0) \bigcup \left\{ \bigcup_{\theta \in \Theta \setminus V_k(\theta_0)} V_k(\theta) \right\}.$$

By the compactness of $\Theta$, there exists $\theta_1, \cdots, \theta_j \in \Theta \setminus V_k(\theta_0)$ such that

$$\Theta \subset V_k(\theta_0) \bigcup \left\{ \bigcup_{i=1}^{j} V_k(\theta_i) \right\}.$$

Here, the choice of $j$ depends on $k$. Suppose $\hat{\theta}_n \notin V_k(\theta_0) \cap \Theta$, which implies that $\hat{\theta}_n \in \left\{ \{\cup_{i=1}^{j} V_k(\theta_i)\} \cap \Theta \right\} \setminus V_k(\theta_0)$. Without loss of generality, we assume that $\hat{\theta}_n \in \{V_k(\theta_1) \cap \Theta\} \setminus V_k(\theta_0)$. Then we have

$$\begin{aligned}
& \mathbb{E} l_1(\theta_0) \\
= \;& \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} l_t(\theta_0) && (3.9) \\
\geq \;& \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \tilde{l}_t(\theta_0) - \limsup_{n\to\infty} \sup_{\theta\in\Theta} \left| \frac{1}{n} \sum_{t=1}^{n} l_t(\theta) - \frac{1}{n} \sum_{t=1}^{n} \tilde{l}_t(\theta) \right| \\
\geq \;& \liminf_{n\to\infty} \inf_{\theta\in\Theta} \frac{1}{n} \sum_{t=1}^{n} \tilde{l}_t(\theta) && (3.10) \\
= \;& \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \tilde{l}_t(\hat{\theta}_n) && (3.11) \\
= \;& \liminf_{n\to\infty} \inf_{\theta\in\{V_k(\theta_1)\cap\Theta\}\setminus V_k(\theta_0)} \frac{1}{n} \sum_{t=1}^{n} \tilde{l}_t(\theta) && (3.12) \\
\geq \;& \liminf_{n\to\infty} \inf_{\theta\in\{V_k(\theta_1)\cap\Theta\}\setminus V_k(\theta_0)} \frac{1}{n} \sum_{t=1}^{n} l_t(\theta) - \limsup_{n\to\infty} \sup_{\theta\in\Theta} \left| \frac{1}{n} \sum_{t=1}^{n} l_t(\theta) - \frac{1}{n} \sum_{t=1}^{n} \tilde{l}_t(\theta) \right| \\
\geq \;& \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \inf_{\theta\in\{V_k(\theta_1)\cap\Theta\}\setminus V_k(\theta_0)} l_t(\theta) && (3.13)
\end{aligned}$$

$$= \mathbb{E} \inf_{\theta \in \{V_k(\theta_1) \cap \Theta\} \setminus V_k(\theta_0)} l_1(\theta) \tag{3.14}$$

$$\to \mathbb{E} l_1(\theta_1) \quad \text{as } k \to \infty \tag{3.15}$$

$$> \mathbb{E} l_1(\theta_0). \tag{3.16}$$

Equations (3.9) and (3.14) hold due to the ergodic theorem[1] (Billingsley [1995]). The ergodic theorem applies here due to Lemma 4.2 and the fact that $l_t(\theta)$ and $\tilde{l}_t(\theta)$ are measurable transformations of the stationary and ergodic process $\{y_t\}$. Inequalities (3.10) and (3.13) result from Lemma 4.4. (3.11) and (3.12) are based on the definition of the QMLE. Inequality (3.15) is true by the Beppo-Levi theorem and (3.16) results from Lemma 4.3. This contradiction indicates that $\hat{\theta}_n \in V_k(\theta_0) \cap \Theta$. The desired result follows by letting $k \to \infty$. $\square$

## 3.4 Asymptotic Normality

### 3.4.1 The Normality Thoerem

To establish the asymptotic normality of the QMLE, the following two additional assumptions are made:

**D1:** $\theta_0$ is an interior point of $\Theta$.

**D2:** $\mathbb{E}\|y_t\|^6 < \infty$.

[1] If $\{X_t\}$ is a stationary and ergodic process such that $\mathbb{E}X_t \in \mathbb{R} \cup \{+\infty\}$, then $n^{-1}\sum t = 1^n X_t$ converges almost surely to $\mathbb{E}X_1$ when $n \to \infty$.

**Theorem 3.8** (Asymptotic Normality). *Under Assumptions C1-C4 and D1-D2, we have*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N(0, J^{-1} V J^{-1}),$$

*where*

$$J = -\mathbb{E}\left(\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta^T}\right) \quad \textit{and} \quad V = \mathbb{E}\left(\frac{\partial l_t(\theta_0)}{\partial\theta}\frac{\partial l_t(\theta_0)}{\partial\theta^T}\right).$$

*Proof.* See Section 3.4.2. □

*Remarks.*

1. Comte and Lieberman [2003] studies the asymptotic normality for the BEKK model, which is a special case of the results here, with the requirement of finite eighth moment of $\{y_t\}$. This theorem reduces the moment requirement of $\{y_t\}$ from 8 in Comte and Lieberman [2003] to 6.

2. If the innovation process $\{\xi_t\}$ is indeed Gaussian, QMLE becomes regular MLE and provides the most efficiency. In this case, we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N(0, J^{-1}).$$

3. In the univariate case, the moment condition can be imposed on the innovation process $\{\xi_t\}$ (i.e., Francq and Zakoïan [2004]). However, in our multivariate case, we have to impose the moment condition on the observed process $\{y_t\}$ due to the complexity of the multivariate structure. In

the multivariate case, $\mathbb{E}\|y_t\|^k < \infty$ implies $\mathtt{E}\|\xi_t\|^k < \infty$ since

$$\mathtt{E}\|\xi_t\|^k \leq \mathtt{E}\|H_t\|^{-k/2}\|y_t\|^k \leq \gamma^{-dk/2}\mathtt{E}\|y_t\|^k < \infty,$$

where $\gamma$ is defined in Lemma 4.2. However, generally $\mathtt{E}\|\xi_t\|^k < \infty$ does not imply $\mathbb{E}\|y_t\|^k < \infty$ since $\|H_t\|$ usually has no upper bound.

4. Note that Assumption C3 is implied by D2. But we do not need Assumption D2 to prove consistency.

5. Francq and Zakoïan [2010] discussed the distribution of the QMLE when $\theta_0$ is on the boundary of $\Theta$.

### 3.4.2 Proof of Theorem 3.8

Lemma 4.6 guarantees that the matrices $V$ and $J$ are well defined. Consider the Taylor expansion on the score function around $\theta_0$.

$$0 = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial \tilde{l}_t(\hat{\theta}_n)}{\partial\theta} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta} + \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \tilde{l}_t(\theta^*)}{\partial\theta\partial\theta^T}\right)\sqrt{n}(\hat{\theta}_n - \theta_0) \quad (3.17)$$

where $\theta^*$ is between $\hat{\theta}_n$ and $\theta_0$. By (4.15),

$$\begin{aligned}\mathbb{E}\left[\frac{\partial l_t(\theta_0)}{\partial\theta_i}\middle|\mathcal{F}_{t-1}\right] &= \operatorname{tr}[(I_d - H_t^{1/2}(\theta_0)\mathbb{E}(\xi_t\xi_t^T)H_t^{-1/2}(\theta_0))\dot{H}_{t,i}(\theta_0)H_t^{-1}(\theta_0)] \\ &= \operatorname{tr}[(I_d - I_d)\dot{H}_{t,i}(\theta_0)H_t^{-1}(\theta_0)] \\ &= 0.\end{aligned}$$

We can easily extend the proof of the martingale central limit theorem in Billingsley [1961] to the multivariate case using characteristic functions. In the sense of Lemma 4.6 and the fact that $\frac{\partial l_t(\theta_0)}{\partial \theta}$ is stationary and ergodic, the conditions of the martingale central limit theorem in Billingsley [1961] are satisfied and we have that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t(\theta_0)}{\partial \theta} \xrightarrow{\mathcal{D}} N(0, V) \ .$$

By Lemma 4.7,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t(\theta_0)}{\partial \theta} + \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta} - \frac{\partial l_t(\theta_0)}{\partial \theta} \right) \xrightarrow{\mathcal{D}} N(0, V),$$

since the term in the bracket converges to zero in probability.

We now consider the Taylor expansion of $\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 l_t(\theta^*)}{\partial \theta \partial \theta^T}$ around $\theta_0$. For the $(i, j)^{th}$ element,

$$\left( \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 l_t(\theta^*)}{\partial \theta \partial \theta^T} \right)_{ij} = \left( \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta^T} \right)_{ij} + \frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta^T} \left( \frac{\partial^2 l_t(\tilde{\theta})}{\partial \theta \partial \theta^T} \right)_{ij} (\theta^* - \theta_0), \tag{3.18}$$

where $\tilde{\theta}$ is between $\theta^*$ and $\theta_0$. By the consistency, $\tilde{\theta}$ is within the neighborhood of $\theta_0$ when $n$ is sufficiently large. Then by Lemma 4.6 and the ergodic theorem,

$$\begin{aligned} \limsup_{n \to \infty} \left\| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} \left( \frac{\partial^2 l_t(\tilde{\theta})}{\partial \theta \partial \theta^T} \right)_{ij} \right\| &\leq \limsup_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in \nu(\theta_0)} \left\| \frac{\partial}{\partial \theta} \left( \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta^T} \right)_{ij} \right\| \\ &= \mathbb{E} \sup_{\theta \in \nu(\theta_0)} \left\| \frac{\partial}{\partial \theta} \left( \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta^T} \right)_{ij} \right\| < \infty. \end{aligned}$$

Thus, the second term of the right hand side of (3.18) converges to zero since $\|\theta^* - \theta_0\| \to 0$. Applying the ergodic theorem on the first term of the right hand side of (3.18) gives

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial\theta\partial\theta^T} \xrightarrow{\mathcal{P}} J.$$

By Lemma 4.7,

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \tilde{l}_t(\theta^*)}{\partial\theta\partial\theta^T} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta^*)}{\partial\theta\partial\theta^T} + \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \tilde{l}_t(\theta^*)}{\partial\theta\partial\theta^T} - \frac{\partial^2 l_t(\theta^*)}{\partial\theta\partial\theta^T}\right) \xrightarrow{\mathcal{P}} J.$$

Therefore, in view of (3.17) and the Slutsky's theorem, we finish the proof.

## 3.5 Conclusion and Commentary

In this chapter, we prove consistency and asymptotic normality of the QMLE under mild conditions. We prove the consistency using standard compactness argument (Theorem 3.7) and the asymptotic normality by the Taylor expansion of the score function (Theorem 3.8). We only assume finite sixth moment of the observed sequence $\{y_t\}$, which is by far the weakest in literature for general multivariate GARCH models. Asymptotic normality is useful for statistical inference purpose. To calculate the standard error of the estimator, one only needs to substitute the $\theta_0$ in the matrices $J$ and $V$ with the estimated value. The conditions we give for model identifiability are necessary and sufficient (Theorem 3.5). Identifiability is useful to prove Lemma 4.3, which plays an essential rule in the proof of the consistency theorem.

# Chapter 4

# Lemmas

This chapter collects the lemmas needed in Chapter 3 in order to prove the consistency and asymptotic normality of the QMLE. In particular, we prove that the difference between the theoretical likelihood function $L(\theta)$ and the observed likelihood function $\tilde{L}(\theta)$ converges to zero, and this is also true for their first and second order derivatives.

## 4.1 Lemma 4.1

**Lemma 4.1.** *For any matrix $A$, we have*

$$\|A^k\| \leq K\rho^k(A)$$

*for all $k$ and some constant $K$.*

*Proof.* $\forall \epsilon > 0$, the matrix $\dfrac{A}{\rho(A)+\epsilon}$ has spectral radius strictly less than 1, which implies that elementwise,

$$\frac{A^k}{(\rho(A)+\epsilon)^k} \to 0, \text{ as } k \to \infty.$$

Thus, applying any matrix norm on both side of the above formula gives

$$\frac{\|A^k\|}{(\rho(A)+\epsilon)^k} \to 0, \text{ as } k \to \infty.$$

Then there exists $N$ such that

$$\frac{\|A^k\|}{(\rho(A)+\epsilon)^k} < 1, \text{ for } k \geq N.$$

For $k < N$, we have

$$\|A^k\| \leq K_k(\rho(A)+\epsilon)^k,$$

for some constants $K_k,\ k = 1, \cdots, N-1$. The desired result holds by taking $K = \max\{K_1, \cdots, K_{N-1}, 1\}$ and noticing that $\epsilon$ is arbitrary. □

## 4.2 Lemma 4.2

**Lemma 4.2.** *1.* $\mathbb{E}l_t(\theta)$ *belongs to* $\mathbb{R} \cup \{+\infty\}$.

*2.* $\mathbb{E}l_t(\theta_0) < \infty$.

*Proof.* 1. Assumption C1 and the Wielandt-Hoffman theorem[1] imply that eigenvalues are continuous functions of the matrix entries and thus there exist $\gamma > 0$ such that $\lambda_{it}(\theta) \geq \gamma$ for all $i$, $t$ and $\theta$, where $\lambda_{it}(\theta), i = 1, \cdots, d$ are eigenvalues of $H_t(\theta)$. Hence,

$$\mathbb{E}l_t^-(\theta) \leq \mathbb{E}\log^- |H_t(\theta)| \leq \max\{0, -d\log\gamma\} < \infty,$$

where for a random variable $X$, $X^-$ is defined as $\max\{-X, 0\}$.

2. Note that all the eigenvalues $\lambda_{it}(\theta), i = 1, \cdots, d$ are positive. We have

$$\begin{aligned}
\mathbb{E}l_t(\theta_0) &= \mathbb{E}\log|H_t(\theta_0)| + \mathbb{E}(y_t^T H_t^{-1}(\theta_0) y_t) \\
&= \mathbb{E}\log|H_t(\theta_0)| + \mathbb{E}(\xi_t^T \xi_t) \\
&= d + \mathbb{E}\frac{2d}{s}\log|H_t(\theta_0)|^{s/2d} \\
&\leq d + \frac{2d}{s}\log\mathbb{E}|H_t(\theta_0)|^{s/2d} && (4.1) \\
&= d + \frac{2d}{s}\log\mathbb{E}\left(\prod_{i=1}^{d}\lambda_{it}(\theta_0)\right)^{s/2d} \\
&\leq d + \frac{2d}{s}\log\mathbb{E}(\max_i\{\lambda_{it}(\theta_0)\})^{s/2} \\
&= d + \frac{2d}{s}\log\mathbb{E}\|H_t(\theta_0)\|^{s/2} && (4.2) \\
&\leq d + C_1\log\mathbb{E}\|h_t(\theta_0)\|^{s/2} \\
&\leq d + C_1\log\mathbb{E}\|X_t(\theta_0)\|^{s/2}.
\end{aligned}$$

[1] For a reference, see `http://planetmath.org/encyclopedia/WielandtHoffmanTheorem.html`

Inequality (4.1) holds due to the Jensen's inequality and (4.2) is from the definition of the spectral norm. Iterating (3.8), we obtain

$$X_t = \sum_{k=0}^{\infty} B^k c_{t-k}. \tag{4.3}$$

By the compactness of the parameter space, there exist $\bar{\rho} \in (0, 1)$ such that

$$\bar{\rho} = \sup_{\theta \in \Theta} \rho(B(\theta)).$$

Hence, by Lemma 4.1 and the stationarity assumption,

$$\begin{aligned}
\mathbb{E}\|X_t(\theta_0)\|^{s/2} &= \mathbb{E}\left\|\sum_{k=0}^{\infty} B^k c_{t-k}\right\|^{s/2} \\
&\leq \sum_{k=0}^{\infty} \left\|B^k\right\|^{s/2} \mathbb{E}\|c_t\|^{s/2} \\
&\leq \sum_{k=0}^{\infty} K\bar{\rho}^{ks/2}\mathbb{E}\left\|c + \sum_{i=1}^{q} A_i \eta_{t-i}\right\|^{s/2} \\
&\leq C_2 + C_3 \mathbb{E}\|\eta_t\|^{s/2}.
\end{aligned}$$

It now only remains to show that $\mathbb{E}\|\eta_t\|^{s/2} < \infty$. By Assumption C3,

$$\begin{aligned}
\mathbb{E}\|\eta_t\|^{s/2} &\leq \mathbb{E}\|\mathrm{vec}(y_t y_t^T)\|_2^{s/2} \\
&= \mathbb{E}(y_t^T y_t)^{s/2} \\
&= \mathbb{E}\|y_t\|^s < \infty.
\end{aligned}$$

The desired result will follow.

□

## 4.3 Lemma 4.3

**Lemma 4.3.** $\mathbb{E}(l_t(\theta_0)) < \mathbb{E}(l_t(\theta))$ *for all* $\theta \neq \theta_0$.

*Proof.*

$$\begin{aligned}
& \mathbb{E}(l_t(\theta)) - \mathbb{E}(l_t(\theta_0)) \\
= \; & \mathbb{E}\log\frac{|H_t(\theta)|}{|H_t(\theta_0)|} + \mathbb{E}(y_t^T H_t^{-1}(\theta) y_t) - \mathbb{E}(y_t^T H_t^{-1}(\theta_0) y_t) \\
= \; & \mathbb{E}\log\frac{|H_t(\theta)|}{|H_t(\theta_0)|} + \mathbb{E}[\mathrm{tr}(y_t^T H_t^{-1}(\theta) y_t)] - \mathbb{E}(\xi_t^T \xi_t) \\
= \; & \mathbb{E}\log\frac{|H_t(\theta)|}{|H_t(\theta_0)|} + \mathbb{E}[\mathrm{tr}(\xi_t^T H_t^{1/2}(\theta_0) H_t^{-1}(\theta) H_t^{1/2}(\theta_0)\xi_t)] - d \\
= \; & \mathbb{E}\log\frac{|H_t(\theta)|}{|H_t(\theta_0)|} + \mathrm{tr}[\mathbb{E}(\xi_t\xi_t^T H_t^{1/2}(\theta_0) H_t^{-1}(\theta) H_t^{1/2}(\theta_0))] - d \\
= \; & \mathbb{E}\log\frac{|H_t(\theta)|}{|H_t(\theta_0)|} + \mathrm{tr}[\mathbb{E}(\xi_t\xi_t^T)\mathbb{E}(H_t^{1/2}(\theta_0) H_t^{-1}(\theta) H_t^{1/2}(\theta_0))] - d \\
= \; & \mathbb{E}\log\frac{|H_t(\theta)|}{|H_t(\theta_0)|} + \mathbb{E}[\mathrm{tr}(H_t^{1/2}(\theta_0) H_t^{-1}(\theta) H_t^{1/2}(\theta_0))] - d \\
= \; & \mathbb{E}\log\frac{|H_t(\theta)|}{|H_t(\theta_0)|} + \mathbb{E}[\mathrm{tr}(H_t(\theta_0) H_t^{-1}(\theta))] - d \\
> \; & \mathbb{E}\log\frac{|H_t(\theta)|}{|H_t(\theta_0)|} + \mathbb{E}[\log|H_t(\theta_0) H_t^{-1}(\theta))| + d] - d = 0. \qquad (4.4)
\end{aligned}$$

Inequality (4.4) holds due to Inequality (7) in Appendix B. □

## 4.4 Lemma 4.4

**Lemma 4.4.** $\lim_{n\to\infty} \sup_{\theta\in\Theta} |\frac{1}{n}\sum_{t=1}^{n} l_t(\theta) - \frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\theta)| = 0,\ a.s..$

*Proof.* Iterating (3.8), we obtain

$$X_t = B^t X_0 + \sum_{i=1}^{q} B^{t-i} c_i + \sum_{i=q+1}^{t} B^{t-i} c_i. \tag{4.5}$$

Analogously,

$$\tilde{X}_t = B^t \tilde{X}_0 + \sum_{i=1}^{q} B^{t-i} \tilde{c}_i + \sum_{i=q+1}^{t} B^{t-i} c_i. \tag{4.6}$$

Hence for $t \geq 1$, we have almost surely that,

$$\begin{aligned}
\sup_{\theta\in\Theta} \|h_t - \tilde{h}_t\| &\leq \sup_{\theta\in\Theta} \|X_t - \tilde{X}_t\| \\
&= \sup_{\theta\in\Theta} \left\| B^t(X_0 - \tilde{X}_0) + \sum_{k=1}^{q} B^{t-k}(c_k - \tilde{c}_k) \right\| \\
&\leq K\bar{\rho}^t \sup_{\theta\in\Theta} \left( \left\| \sum_{k=0}^{\infty} B^k c_{-k} - \tilde{X}_0 \right\| + K \sum_{k=1}^{q} \bar{\rho}^{-k} \|c_k - \tilde{c}_k\| \right) \\
&\leq O(\bar{\rho}^t). \qquad (4.7)
\end{aligned}$$

Inequality (4.7) holds since each norm inside of the supremum has finite expectation. Consider the function $l_t(h_t) = \log \|H_t\| + y_t^T H_t^{-1} y_t$, by (1), (9), (10) and (11) in Appendix B, we have

$$\frac{\partial l_t(h_t)}{\partial h_t^T} = \frac{\partial \log |H_t|}{\partial h_t^T} + \frac{\partial}{\partial h_t^T} \mathrm{vec}(y_t^T H_t^{-1} y_t)$$

$$
\begin{aligned}
&= \mathrm{vech}^T\left(\frac{\partial \log|H_t|}{\partial H_t}\right) + (y_t^T \otimes y_t^T)\frac{\partial}{\partial h_t^T}\mathrm{vec}(H_t^{-1}) \\
&= \mathrm{vech}^T(H_t^{-1}) + (y_t^T \otimes y_t^T) D_d \frac{\partial \mathrm{vech}(H_t^{-1})}{\partial \mathrm{vech}^T(H_t)} \\
&= \mathrm{vech}^T(H_t^{-1}) - (y_t^T \otimes y_t^T)(H_t^{-1} \otimes H_t^{-1}) D_d \\
&= \mathrm{vech}^T(H_t^{-1}) - (H_t^{-1} y_t \otimes H_t^{-1} y_t)^T D_d.
\end{aligned}
$$

By the mean value theorem, for some positive number $s$,

$$
\begin{aligned}
\mathbb{E}\sup_{\theta\in\Theta}|l_t - \tilde{l}_t|^{s/2} &= \mathbb{E}\sup_{\theta\in\Theta}|l_t(h_t) - l_t(\tilde{h}_t)|^{s/2} = \mathbb{E}\sup_{\theta\in\Theta}\left|\frac{\partial l_t(\bar{h}_t)}{\partial h_t^T}(h_t - \tilde{h}_t)\right|^{s/2} \\
&\leq \mathbb{E}\sup_{\theta\in\Theta}\left\|\frac{\partial l_t(\bar{h}_t)}{\partial h_t^T}\right\|^{s/2} \|h_t - \tilde{h}_t\|^{s/2} \\
&\leq (C_1 + C_2\mathbb{E}\|y_t\|^s)\bar{\rho}^{st/2} = O(\bar{\rho}^{st/2}),
\end{aligned}
$$

where $\bar{h}_t$ is between the chord of $h_t$ and $\tilde{h}_t$. By the Markov inequality, for any $\epsilon > 0$,

$$
\begin{aligned}
\sum_{t=1}^{\infty}\mathbb{P}\left(\sup_{\theta\in\Theta}|l_t - \tilde{l}_t| > \epsilon\right) &= \sum_{t=1}^{\infty}\mathbb{P}\left(\sup_{\theta\in\Theta}|l_t - \tilde{l}_t|^{s/2} > \epsilon^{s/2}\right) \\
&\leq \sum_{t=1}^{\infty}\frac{\mathbb{E}\sup_{\theta\in\Theta}|l_t - \tilde{l}_t|^{s/2}}{\epsilon^{s/2}} \\
&\leq \sum_{t=1}^{\infty}\frac{O(\bar{\rho}^{st/2})}{\epsilon^{s/2}} < \infty.
\end{aligned}
$$

By the Borel-Cantelli lemma, we have $\sup_{\theta\in\Theta}|l_t - \tilde{l}_t| \to 0$, a.s.. And the desired result follows by the Césaro's mean theorem since

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n} l_t(\theta)-\frac{1}{n}\sum_{t=1}^{n}\tilde{l}_t(\theta)\right|\le\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|l_t(\theta)-\tilde{l}_t(\theta)|.$$

□

## 4.5 Lemma 4.5

**Lemma 4.5.** *1.* $\mathbb{E}\left\|\dot{H}_{t,i}(\theta)\right\|^3<\infty$, *where* $\dot{H}_{t,i}(\theta)=\dfrac{\partial H_t(\theta)}{\partial\theta_i}$.

*2.* $\mathbb{E}\left\|\ddot{H}_{t,ij}(\theta)\right\|^2<\infty$, *where* $\ddot{H}_{t,ij}(\theta)=\dfrac{\partial^2 H_t(\theta)}{\partial\theta_i\partial\theta_j}$.

*3.* $\mathbb{E}\left\|\dddot{H}_{t,ijk}(\theta)\right\|<\infty$, *where* $\dddot{H}_{t,ijk}(\theta)=\dfrac{\partial^3 H_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}$.

*Proof.* It suffices to show that $\mathbb{E}\left\|\dfrac{\partial X_t}{\partial\theta^T}\right\|^3<\infty$, $\mathbb{E}\left\|\dfrac{\partial}{\partial\theta^T}\mathrm{vec}\left(\dfrac{\partial X_t}{\partial\theta^T}\right)\right\|^2<\infty$ and $\mathbb{E}\left\|\dfrac{\partial}{\partial\theta_i}\left[\dfrac{\partial}{\partial\theta^T}\mathrm{vec}\left(\dfrac{\partial X_t}{\partial\theta^T}\right)\right]\right\|<\infty$. We consult the formulas in Appendix B various times when calculating the derivatives.

1. By (4.3), we have that

$$\frac{\partial X_t}{\partial c^T}=\sum_{k=0}^{\infty}B^k\underline{1},\tag{4.8}$$

where $\underline{1}=(I_N,0_{N\times N},\cdots,0_{N\times N})^T$,

$$\frac{\partial X_t}{\partial\mathrm{vec}^T(A_i)}=\sum_{k=0}^{\infty}B^k\underline{\epsilon}_{t-k-i},\tag{4.9}$$

where

$$\underline{\epsilon}_{t-k-i} = (\epsilon_{t-k-i}^T, 0_{N^2\times N}, \cdots, 0_{N^2\times N})^T$$

and

$$\begin{aligned}
\epsilon_{t-k-i} &= \frac{\partial A_i \eta_{t-k-i}}{\partial \mathrm{vec}^T(A_i)} = \frac{\partial \mathrm{vec}(A_i \eta_{t-k-i})}{\partial \mathrm{vec}^T(A_i)} = \frac{\partial \mathrm{vec}(I_N A_i \eta_{t-k-i})}{\partial \mathrm{vec}^T(A_i)} \\
&= (\eta_{t-k-i}^T \otimes I_N)\frac{\partial \mathrm{vec}(A_i)}{\partial \mathrm{vec}^T(A_i)} = \eta_{t-k-i}^T \otimes I_N,
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial X_t}{\partial \mathrm{vec}^T(B_i)} \\
=& \sum_{k=1}^{\infty} \frac{\partial \mathrm{vec}(I_{Np} B^k c_{t-k})}{\partial \mathrm{vec}^T(B_i)} \\
=& \sum_{k=1}^{\infty} (c_{t-k}^T \otimes I_{Np}) \frac{\partial \mathrm{vec}(B^k)}{\partial \mathrm{vec}^T(B_i)} \\
=& \sum_{k=1}^{\infty} (c_{t-k}^T \otimes I_{Np}) \frac{\partial \mathrm{vec}(B^k)}{\partial \mathrm{vec}^T(B)} \cdot \frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_i)} \\
=& \sum_{k=1}^{\infty} (c_{t-k}^T \otimes I_{Np}) \left( \sum_{l=0}^{k-1} (B^T)^{k-1-l} \otimes B^l \right) \frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_i)}. \qquad (4.10)
\end{aligned}$$

It is worth pointing out that $\partial \mathrm{vec}(B)/\partial \mathrm{vec}^T(B_i)$ is a matrix with elements 1's and 0's and does not depend on any model parameters. This is useful when we calculate the higher order derivatives. By Lemma 4.1 and the stationarity conditions,

$$\mathbb{E} \left\| \frac{\partial X_t}{\partial c^T} \right\|^3 \leq \mathbb{E} \left( \sum_{k=0}^{\infty} \|B^k\| \cdot \|\underline{1}\| \right)^3 \leq \left( \sum_{k=0}^{\infty} K \bar{\rho}^k \right)^3 < \infty,$$

$$
\begin{aligned}
\mathbb{E}\left\|\frac{\partial X_t}{\partial \mathrm{vec}^T(A_i)}\right\|^3 &\leq \mathbb{E}\left(\sum_{k=0}^{\infty}\|B^k\|\cdot\|\underline{\epsilon}_{t-k-i}\|\right)^3 \\
&\leq \mathbb{E}\left(\sum_{k=0}^{\infty}\|B^k\|\cdot\|y_{t-k-i}\|^2\right)^3 \\
&\leq C_1\left(\sum_{k=0}^{\infty}K\bar{\rho}^k\right) < \infty,
\end{aligned}
$$

where $C_1 = \max\{\mathbb{E}\|y_t\|^6, \mathbb{E}\|y_{t_1}\|^4\|y_{t_2}\|^2, \mathbb{E}\|y_{t_1}\|^2\|y_{t_2}\|^2\|y_{t_3}\|^2\}$. All the expectations are finite by Assumption D2 and the Hölder's inequality since

$$
\mathbb{E}\|y_{t_1}\|^4\|y_{t_2}\|^2 \leq (\mathbb{E}\|y_{t_1}\|^6)^{2/3}(\mathbb{E}\|y_{t_2}\|^6)^{1/3} < \infty,
$$

and

$$
\mathbb{E}\|y_{t_1}\|^2\|y_{t_2}\|^2\|y_{t_3}\|^2 \leq (\mathbb{E}(\|y_{t_1}\|^6))^{1/3}(\mathbb{E}(\|y_{t_2}\|^6))^{1/3}(\mathbb{E}(\|y_{t_3}\|^6))^{1/3} < \infty.
$$

Furthermore,

$$
\begin{aligned}
&\mathbb{E}\left\|\frac{\partial X_t}{\partial \mathrm{vec}^T(B_i)}\right\|^3 \\
\leq\ & \mathbb{E}\left(\sum_{k=1}^{\infty}\|c_{t-k}\|\left\|\sum_{l=0}^{k-1}(B^T)^{k-1-l}\otimes B^l\right\|\left\|\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_i)}\right\|\right)^3 \\
\leq\ & \mathbb{E}\left(\sum_{k=1}^{\infty}\|c_{t-k}\|\left(\sum_{l=0}^{k-1}\|B^{k-1-l}\|\cdot\|B^l\|\right)\left\|\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_i)}\right\|\right)^3 \\
\leq\ & C_1\mathbb{E}\left(\sum_{k=1}^{\infty}\|c_{t-k}\|\left(\sum_{l=0}^{k-1}K^2\bar{\rho}^{k-1}\right)\right)^3
\end{aligned}
$$

$$
\begin{aligned}
&= C_2 \mathbb{E}\left(\sum_{k=1}^{\infty} k\bar{\rho}^{k-1}\|c_{t-k}\|\right)^3 \\
&= C_3 \left(\sum_{k=1}^{\infty} k\bar{\rho}^{k-1}\right)^3 < \infty,
\end{aligned}
$$

where $C_3/C_2 = \max\{\mathbb{E}\|c_t\|^3, \mathbb{E}\|c_{t_1}\|^2\|c_{t_2}\|, \mathbb{E}(\|c_{t_1}\| \cdot \|c_{t_2}\| \cdot \|c_{t_3}\|)\}$. It remains to show that all these expectations are finite.

$$
\begin{aligned}
\mathbb{E}\|c_t\|^3 &\leq \mathbb{E}\left(\|c\| + \sum_{i=1}^{q} \|A_i\| \cdot \|y_{t-i}\|^2\right)^3 \\
&= C_1 + C_2\mathbb{E}\|y_t\|^2 + C_3\mathbb{E}\|y_{t_1}\|^2\|y_{t_2}\|^2 + C_4\mathbb{E}\|y_t\|^4 \\
&\qquad + C_5\mathbb{E}\|y_{t_1}\|^2\|y_{t_2}\|^2\|y_{t_3}\|^2 + C_6\mathbb{E}\|y_{t_1}\|^4\|y_{t_2}\|^2 + C_7\mathbb{E}\|y_t\|^6 \\
&< \infty.
\end{aligned}
$$

By the Hölder's inequality,

$$
\mathbb{E}\|c_{t_1}\|^2\|c_{t_2}\| \leq (\mathbb{E}\|c_{t_1}\|^3)^{2/3}(\mathbb{E}\|c_{t_2}\|^3)^{1/3} < \infty,
$$

and

$$
\mathbb{E}\|c_{t_1}\|\|c_{t_2}\|\|c_{t_3}\| \leq (\mathbb{E}(\|c_{t_1}\|^3))^{1/3}(\mathbb{E}(\|c_{t_2}\|^3))^{1/3}(\mathbb{E}(\|c_{t_3}\|^3))^{1/3} < \infty.
$$

2. It follows from the previous part that

$$\frac{\partial}{\partial c^T}\text{vec}\left(\frac{\partial X_t}{\partial c^T}\right) = 0,$$

$$\frac{\partial}{\partial \text{vec}^T(A_i)}\text{vec}\left(\frac{\partial X_t}{\partial c^T}\right) = 0,$$

$$\frac{\partial}{\partial \text{vec}^T(A_j)}\text{vec}\left(\frac{\partial X_t}{\partial \text{vec}^T(A_i)}\right) = 0,$$

$$\begin{aligned}
&\frac{\partial}{\partial \text{vec}^T(B_j)}\text{vec}\left(\frac{\partial X_t}{\partial c^T}\right) \\
= &\sum_{k=0}^{\infty}\frac{\partial \text{vec}(B^k \underline{1})}{\partial \text{vec}^T(B_j)} \\
= &\sum_{k=1}^{\infty}(\underline{1}^T \otimes I_{Np})\frac{\partial \text{vec}(B^k)}{\partial \text{vec}^T(B_j)} \\
= &\sum_{k=1}^{\infty}(\underline{1}^T \otimes I_{Np})\frac{\partial \text{vec}(B^k)}{\partial \text{vec}^T(B)}\frac{\partial \text{vec}(B)}{\partial \text{vec}^T(B_j)} \\
= &\sum_{k=1}^{\infty}(\underline{1}^T \otimes I_{Np})\left(\sum_{l=0}^{k-1}(B^T)^{k-1-l}\otimes B^l\right)\frac{\partial \text{vec}(B)}{\partial \text{vec}^T(B_j)},
\end{aligned} \tag{4.11}$$

$$\begin{aligned}
&\frac{\partial}{\partial \text{vec}^T(B_j)}\text{vec}\left(\frac{\partial X_t}{\partial \text{vec}^T(A_i)}\right) \\
= &\sum_{k=0}^{\infty}\frac{\partial \text{vec}(B^k \underline{\epsilon}_{t-k-i})}{\partial \text{vec}^T(B_j)} \\
= &\sum_{k=1}^{\infty}(\underline{\epsilon}_{t-k-i}^T \otimes I_{Np})\frac{\partial \text{vec}(B^k)}{\partial \text{vec}^T(B_j)} \\
= &\sum_{k=1}^{\infty}(\underline{\epsilon}_{t-k-i}^T \otimes I_{Np})\frac{\partial \text{vec}(B^k)}{\partial \text{vec}^T(B)}\frac{\partial \text{vec}(B)}{\partial \text{vec}^T(B_j)}
\end{aligned}$$

$$= \sum_{k=1}^{\infty}(\underline{\epsilon}_{t-k-i}^{T} \otimes I_{Np}) \left(\sum_{l=0}^{k-1}(B^T)^{k-1-l} \otimes B^l\right) \frac{\partial \text{vec}(B)}{\partial \text{vec}^T(B_j)}, \quad (4.12)$$

$$\begin{aligned}
& \frac{\partial}{\partial \text{vec}^T(B_j)} \text{vec}\left(\frac{\partial X_t}{\partial \text{vec}^T(B_i)}\right) \\
= & \sum_{k=2}^{\infty}(c_{t-k}^T \otimes I_{Np}) \frac{\partial}{\partial \text{vec}^T(B_j)} \text{vec}\left[\left(\sum_{l=0}^{k-1}(B^T)^{k-1-l} \otimes B^l\right) \frac{\partial \text{vec}(B)}{\partial \text{vec}^T(B_i)}\right] \\
= & \sum_{k=2}^{\infty}(c_{t-k}^T \otimes I_{Np}) \left(\frac{\partial \text{vec}^T(B)}{\partial \text{vec}(B_i)} \otimes I_{Np}\right) \sum_{l=0}^{k-1} \frac{\partial \text{vec}((B^T)^{k-1-l} \otimes B^l)}{\partial \text{vec}^T(B_j)} \\
= & \sum_{k=2}^{\infty}(c_{t-k}^T \otimes I_{Np}) \left(\frac{\partial \text{vec}^T(B)}{\partial \text{vec}(B_i)} \otimes I_{Np}\right) \sum_{l=0}^{k-1}(I_{Np} \otimes K_{Np,Np} \otimes I_{Np}) \times \\
& \left(\frac{\partial \text{vec}((B^T)^{k-1-l})}{\partial \text{vec}^T(B_j)} \otimes \text{vec}(B^l) + \text{vec}((B^T)^{k-1-l}) \otimes \frac{\partial \text{vec}(B^l)}{\partial \text{vec}^T(B_j)}\right) \\
= & \sum_{k=2}^{\infty}(c_{t-k}^T \otimes I_{Np}) \left(\frac{\partial \text{vec}^T(B)}{\partial \text{vec}(B_i)} \otimes I_{Np}\right) \sum_{l=0}^{k-1}(I_{Np} \otimes K_{Np,Np} \otimes I_{Np}) K_{Np,Np} \\
& \times \left(\frac{\partial \text{vec}(B^{k-1-l})}{\partial \text{vec}^T(B_j)} \otimes \text{vec}(B^l) + \text{vec}(B^{k-1-l}) \otimes \frac{\partial \text{vec}(B^l)}{\partial \text{vec}^T(B_j)}\right) \\
= & \sum_{k=2}^{\infty}(c_{t-k}^T \otimes I_{Np}) \left(\frac{\partial \text{vec}^T(B)}{\partial \text{vec}(B_i)} \otimes I_{Np}\right) \sum_{l=0}^{k-1}(I_{Np} \otimes K_{Np,Np} \otimes I_{Np}) K_{Np,Np} \\
& \times \left[\left(\frac{\partial \text{vec}(B^{k-1-l})}{\partial \text{vec}^T(B)} \frac{\partial \text{vec}(B)}{\partial \text{vec}^T(B_j)}\right) \otimes \text{vec}(B^l)\right. \\
& \left. + \text{vec}(B^{k-1-l}) \otimes \left(\frac{\partial \text{vec}(B^l)}{\partial \text{vec}^T(B)} \frac{\partial \text{vec}(B)}{\partial \text{vec}^T(B_j)}\right)\right] \\
= & \sum_{k=2}^{\infty}(c_{t-k}^T \otimes I_{Np}) \left(\frac{\partial \text{vec}^T(B)}{\partial \text{vec}(B_i)} \otimes I_{Np}\right) \sum_{l=0}^{k-1}(I_{Np} \otimes K_{Np,Np} \otimes I_{Np}) K_{Np,Np} \\
& \times \left\{\left[\left(\sum_{m=0}^{k-2-l}(B^T)^{k-2-l-m} \otimes B^m\right) \frac{\partial \text{vec}(B)}{\partial \text{vec}^T(B_j)}\right] \otimes \text{vec}(B^l)\right. \\
& \left. + \text{vec}(B^{k-1-l}) \otimes \left[\left(\sum_{m=0}^{l-1}(B^T)^{l-1-m} \otimes B^m\right) \frac{\partial \text{vec}(B)}{\partial \text{vec}^T(B_j)}\right]\right\} \\
= & \sum_{k=2}^{\infty}(c_{t-k}^T \otimes I_{Np}) \left(\frac{\partial \text{vec}^T(B)}{\partial \text{vec}(B_i)} \otimes I_{Np}\right) \Gamma(B, k, j). \quad (4.13)
\end{aligned}$$

Therefore,

$$
\begin{aligned}
\mathbb{E}\left\|\frac{\partial}{\partial \mathrm{vec}^T(B_j)}\mathrm{vec}\left(\frac{\partial X_t}{\partial c^T}\right)\right\|^2 &\leq \left(C_1\sum_{k=1}^{\infty}\sum_{l=0}^{k-1}K\bar{\rho}^{k-1-l}K\bar{\rho}^{l}\right)^2 \\
&\leq \left(C_2\sum_{k=1}^{\infty}k\bar{\rho}^{k-1}\right)^2 < \infty,
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\left\|\frac{\partial}{\partial \mathrm{vec}^T(B_j)}\mathrm{vec}\left(\frac{\partial X_t}{\partial \mathrm{vec}^T(A_i)}\right)\right\|^2 \\
\leq\ & C_1\mathbb{E}\left(\sum_{k=1}^{\infty}\|\epsilon_{t-k-i}\|\sum_{l=0}^{k-1}\|B^{k-1-l}\|\cdot\|B^l\|\right)^2 \\
\leq\ & C_1\mathbb{E}\left(\sum_{k=1}^{\infty}\|y_{t-k-i}\|^2\sum_{l=0}^{k-1}K\bar{\rho}^{k-1-l}K\bar{\rho}^{l}\right)^2 \\
\leq\ & C_2\mathbb{E}\left(\sum_{k=1}^{\infty}\|y_{t-k-i}\|^2 k\bar{\rho}^{k-1}\right)^2 \\
=\ & C_3\left(\sum_{k=1}^{\infty}k\bar{\rho}^{k-1}\right)^2 < \infty,
\end{aligned}
$$

where $C_3/C_2 = \max\{\mathbb{E}\|y_t\|^4, \mathbb{E}\|y_{t_1}\|^2\|y_{t_2}\|^2\}$.

$$
\begin{aligned}
&\|\Gamma(B,k,j)\| \\
\leq\ & C_1\sum_{l=0}^{k-1}\left(\sum_{m=0}^{k-2-l}\|B^{k-2-l-m}\|\cdot\|B^m\|\cdot\|B^l\|\right. \\
&\qquad \left. +\|B^{k-1-l}\|\cdot\sum_{m=0}^{l-1}\|B^{l-1-m}\|\cdot\|B^m\|\right) \\
\leq\ & C_1\left(\sum_{l=0}^{k-1}\left(\sum_{m=0}^{k-2-l}K\bar{\rho}^{k-2-l-m}K\bar{\rho}^{m}K\bar{\rho}^{l}\right.\right.
\end{aligned}
$$

$$
\begin{aligned}
& +K\bar{\rho}^{k-1-l}\sum_{m=0}^{l-1} K\bar{\rho}^{l-1-m}K\bar{\rho}^{m}\Big)\Big) \\
\leq\ & C_2\left(\sum_{l=0}^{k-1}(k-1-l)\bar{\rho}^{k-2} + l\bar{\rho}^{k-2}\right) \\
=\ & O(k(k-1)\bar{\rho}^{k-2}).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\mathbb{E}\left\|\frac{\partial}{\partial \mathrm{vec}^T(B_j)}\mathrm{vec}\left(\frac{\partial X_t}{\partial \mathrm{vec}^T(B_i)}\right)\right\|^2 &\leq C_3\mathbb{E}\left(\sum_{k=2}^{\infty}\|c_{t-k}\|\cdot\|\Gamma(B,k,j)\|\right)^2 \\
&= C_4\left(\sum_{k=2}^{\infty}\|\Gamma(B,k,j)\|\right)^2 \\
&= C_4\left(\sum_{k=2}^{\infty}O(k(k-1)\bar{\rho}^{k-2})\right)^2 < \infty,
\end{aligned}
$$

where $C_4/C_3 = \max\{\mathbb{E}\|c_t\|^2, \mathbb{E}\|c_{t_1}\|\|c_{t_2}\|\}$.

3. Only the following third order derivatives are non-zero.

$$
\begin{aligned}
& \frac{\partial}{\partial \mathrm{vec}^T(B_j)}\mathrm{vec}\left[\frac{\partial}{\partial \mathrm{vec}^T(B_i)}\mathrm{vec}\left(\frac{\partial X_t}{\partial c^T}\right)\right] \\
=\ & \sum_{k=2}^{\infty}\left(\frac{\partial \mathrm{vec}^T(B)}{\partial \mathrm{vec}(B_i)}\otimes \underline{1}^T\otimes I_{Np}\right)\sum_{l=0}^{k-1}\frac{\partial}{\partial \mathrm{vec}^T(B_j)}\mathrm{vec}((B^T)^{k-1-l}\otimes B^l) \\
=\ & \sum_{k=2}^{\infty}\left(\frac{\partial \mathrm{vec}^T(B)}{\partial \mathrm{vec}(B_i)}\otimes \underline{1}^T\otimes I_{Np}\right)\sum_{l=0}^{k-1}(I_{Np}\otimes K_{Np,Np}\otimes I_{Np})K_{Np,Np}\times \\
& \left\{\left[\left(\sum_{m=0}^{k-2-l}(B^T)^{k-2-l-m}\otimes B^m\right)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_j)}\right]\otimes \mathrm{vec}(B^l)\right)
\end{aligned}
$$

$$+ \operatorname{vec}(B^{k-1-l}) \otimes \left[ \left( \sum_{m=0}^{l-1} (B^T)^{l-1-m} \otimes B^m \right) \frac{\partial \operatorname{vec}(B)}{\partial \operatorname{vec}^T(B_j)} \right] \Bigg\}$$

$$= \sum_{k=2}^{\infty} \left( \frac{\partial \operatorname{vec}^T(B)}{\partial \operatorname{vec}(B_i)} \otimes \underline{1}^T \otimes I_{Np} \right) \Gamma(B, k, j),$$

$$\frac{\partial}{\partial \operatorname{vec}^T(B_r)} \operatorname{vec} \left[ \frac{\partial}{\partial \operatorname{vec}^T(B_j)} \operatorname{vec} \left( \frac{\partial X_t}{\partial \operatorname{vec}^T(A_i)} \right) \right]$$

$$= \sum_{k=2}^{\infty} \left( \frac{\partial \operatorname{vec}^T(B)}{\partial \operatorname{vec}(B_j)} \otimes \underline{\epsilon}_{t-k-i}^T \otimes I_{Np} \right) \sum_{l=0}^{k-1} \frac{\partial}{\partial \operatorname{vec}^T(B_r)} \operatorname{vec}((B^T)^{k-1-l} \otimes B^l)$$

$$= \sum_{k=2}^{\infty} \left( \frac{\partial \operatorname{vec}^T(B)}{\partial \operatorname{vec}(B_j)} \otimes \underline{\epsilon}_{t-k-i}^T \otimes I_{Np} \right) \sum_{l=0}^{k-1} (I_{Np} \otimes K_{Np,Np} \otimes I_{Np}) K_{Np,Np} \times$$

$$\Bigg\{ \left[ \left( \sum_{m=0}^{k-2-l} (B^T)^{k-2-l-m} \otimes B^m \right) \frac{\partial \operatorname{vec}(B)}{\partial \operatorname{vec}^T(B_r)} \right] \otimes \operatorname{vec}(B^l)$$

$$+ \operatorname{vec}(B^{k-1-l}) \otimes \left[ \left( \sum_{m=0}^{l-1} (B^T)^{l-1-m} \otimes B^m \right) \frac{\partial \operatorname{vec}(B)}{\partial \operatorname{vec}^T(B_r)} \right] \Bigg\}$$

$$= \sum_{k=2}^{\infty} \left( \frac{\partial \operatorname{vec}^T(B)}{\partial \operatorname{vec}(B_j)} \otimes \underline{\epsilon}_{t-k-i}^T \otimes I_{Np} \right) \Gamma(B, k, r),$$

$$\frac{\partial}{\partial c^T} \operatorname{vec} \left[ \frac{\partial}{\partial \operatorname{vec}^T(B_j)} \operatorname{vec} \left( \frac{\partial X_t}{\partial \operatorname{vec}^T(B_i)} \right) \right]$$

$$= \sum_{k=2}^{\infty} \Gamma^T(B, k, j) \left( \frac{\partial \operatorname{vec}(B)}{\partial \operatorname{vec}^T(B_i)} \otimes I_{Np} \right) \frac{\partial}{\partial c^T} \operatorname{vec}(c_{t-k}^T \otimes I_{Np})$$

$$= \sum_{k=2}^{\infty} \Gamma^T(B, k, j) \left( \frac{\partial \operatorname{vec}(B)}{\partial \operatorname{vec}^T(B_i)} \otimes I_{Np} \right)$$

$$\times (I_{Np} \otimes K_{1,Np} \otimes I_{Np}) \left[ \frac{\partial c_{t-k}}{\partial c^T} \otimes \operatorname{vec}(I_{Np}) \right]$$

$$= \sum_{k=2}^{\infty} \Gamma^T(B, k, j) \left( \frac{\partial \operatorname{vec}(B)}{\partial \operatorname{vec}^T(B_i)} \otimes I_{Np} \right) (I_{Np} \otimes K_{1,Np} \otimes I_{Np}) \left[ \underline{1} \otimes \operatorname{vec}(I_{Np}) \right],$$

$$
\begin{aligned}
&\frac{\partial}{\partial \mathrm{vec}^T(A_r)}\mathrm{vec}\left[\frac{\partial}{\partial \mathrm{vec}^T(B_j)}\mathrm{vec}\left(\frac{\partial X_t}{\partial \mathrm{vec}^T(B_i)}\right)\right] \\
=\;& \sum_{k=2}^{\infty}\Gamma^T(B,k,j)\left(\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_i)}\otimes I_{Np}\right)\frac{\partial}{\partial \mathrm{vec}^T(A_r)}\mathrm{vec}(c_{t-k}^T\otimes I_{Np}) \\
=\;& \sum_{k=2}^{\infty}\Gamma^T(B,k,j)\left(\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_i)}\otimes I_{Np}\right) \\
&\times(I_{Np}\otimes K_{1,Np}\otimes I_{Np})\left[\frac{\partial c_{t-k}}{\partial \mathrm{vec}^T(A_r)}\otimes \mathrm{vec}(I_{Np})\right] \\
=\;& \sum_{k=2}^{\infty}\Gamma^T(B,k,j)\left(\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_i)}\otimes I_{Np}\right) \\
&\times(I_{Np}\otimes K_{1,Np}\otimes I_{Np})\left[\underline{\epsilon}_{t-k}\otimes \mathrm{vec}(I_{Np})\right].
\end{aligned}
$$

Let $B_{uv,r}$ denote the $(u,v)^{th}$ element of $B_r$ and $C_{Np} = (I_{Np}\otimes K_{Np,Np}\otimes I_{Np})K_{Np,Np}$.

$$
\begin{aligned}
&\frac{\partial}{\partial B_{uv,r}}\left[\frac{\partial}{\partial \mathrm{vec}^T(B_j)}\mathrm{vec}\left(\frac{\partial X_t}{\partial \mathrm{vec}^T(B_i)}\right)\right] \\
=\;& \sum_{k=2}^{\infty}(c_{t-k}^T\otimes I_{Np})\left(\frac{\partial \mathrm{vec}^T(B)}{\partial \mathrm{vec}(B_i)}\otimes I_{Np}\right)\frac{\partial \Gamma(B,k,j)}{\partial B_{uv,r}} \\
=\;& \sum_{k=3}^{\infty}\left(\frac{\partial \mathrm{vec}^T(B)}{\partial \mathrm{vec}(B_i)}\otimes \underline{1}^T\otimes I_{Np}\right)\sum_{l=0}^{k-1}C_{Np}\times \\
&\left\{\left[\frac{\partial}{\partial B_{uv,r}}\left(\sum_{m=0}^{k-2-l}(B^T)^{k-2-l-m}\otimes B^m\right)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_j)}\right]\otimes \mathrm{vec}(B^l)\right. \\
&+\left[\left(\sum_{m=0}^{k-2-l}(B^T)^{k-2-l-m}\otimes B^m\right)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_j)}\right]\otimes\frac{\partial \mathrm{vec}(B^l)}{\partial B_{uv,r}} \\
&+\frac{\partial \mathrm{vec}(B^{k-1-l})}{\partial B_{uv,r}}\otimes\left[\left(\sum_{m=0}^{l-1}(B^T)^{l-1-m}\otimes B^m\right)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_j)}\right] \\
&\left.+\,\mathrm{vec}(B^{k-1-l})\otimes\left[\frac{\partial}{\partial B_{uv,r}}\left(\sum_{m=0}^{l-1}(B^T)^{l-1-m}\otimes B^m\right)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_j)}\right]\right\} \\
=\;& \sum_{k=3}^{\infty}\left(\frac{\partial \mathrm{vec}^T(B)}{\partial \mathrm{vec}(B_i)}\otimes \underline{1}^T\otimes I_{Np}\right)\sum_{l=0}^{k-1}C_{Np}\times
\end{aligned}
$$

$$\left\{\left[\left(\sum_{m=0}^{k-2-l}\frac{\partial(B^T)^{k-2-l-m}}{\partial B_{uv,r}}\otimes B^m\right)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_j)}\right]\otimes \mathrm{vec}(B^l)\right.$$

$$+\left[\left(\sum_{m=0}^{k-2-l}(B^T)^{k-2-l-m}\otimes\frac{\partial B^m}{\partial B_{uv,r}}\right)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_j)}\right]\otimes \mathrm{vec}(B^l)$$

$$+\left[\left(\sum_{m=0}^{k-2-l}(B^T)^{k-2-l-m}\otimes B^m\right)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_j)}\right]$$

$$\otimes\left(\sum_{m=0}^{l-1}(B^T)^{l-1-m}\otimes B^m\right)\frac{\partial \mathrm{vec}(B)}{\partial B_{uv,r}}$$

$$+\left(\sum_{m=0}^{k-2-l}(B^T)^{k-2-l-m}\otimes B^m\right)\frac{\partial \mathrm{vec}(B)}{\partial B_{uv,r}}$$

$$\otimes\left[\left(\sum_{m=0}^{l-1}(B^T)^{l-1-m}\otimes B^m\right)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_j)}\right]$$

$$+\mathrm{vec}(B^{k-1-l})\otimes\left[\left(\sum_{m=0}^{l-1}\frac{\partial(B^T)^{l-1-m}}{\partial B_{uv,r}}\otimes B^m\right)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_j)}\right]$$

$$\left.+\ \mathrm{vec}(B^{k-1-l})\otimes\left[\left(\sum_{m=0}^{l-1}(B^T)^{l-1-m}\otimes\frac{\partial B^m}{\partial B_{uv,r}}\right)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_j)}\right]\right\}$$

$$=\sum_{k=3}^{\infty}\left(\frac{\partial \mathrm{vec}^T(B)}{\partial \mathrm{vec}(B_i)}\otimes\underline{1}^T\otimes I_{Np}\right)\sum_{l=0}^{k-1}C_{Np}\times$$

$$\left\{\left[\left(\sum_{m=0}^{k-2-l}\left(\left(\sum_{s=0}^{k-3-l-m}(B^T)^{k-3-l-m-s}\otimes B^s\right)\frac{\partial B}{\partial B_{uv,r}}\right)^T\otimes B^m\right)\right.\right.$$

$$\left.\times\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_j)}\right]\otimes \mathrm{vec}(B^l)$$

$$+\left[\left(\sum_{m=0}^{k-2-l}(B^T)^{k-2-l-m}\right.\right.$$

$$\left.\left.\otimes\left[\left(\sum_{s=0}^{m-1}(B^T)^{m-1-s}\otimes B^s\right)\frac{\partial B}{\partial B_{uv,r}}\right]\right)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_j)}\right]\otimes \mathrm{vec}(B^l)$$

$$+\left[\left(\sum_{m=0}^{k-2-l}(B^T)^{k-2-l-m}\otimes B^m\right)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_j)}\right]$$

$$\otimes\left(\sum_{m=0}^{l-1}(B^T)^{l-1-m}\otimes B^m\right)\frac{\partial \mathrm{vec}(B)}{\partial B_{uv,r}}$$

$$+\left(\sum_{m=0}^{k-2-l}(B^T)^{k-2-l-m}\otimes B^m\right)\frac{\partial \text{vec}(B)}{\partial B_{uv,r}}$$
$$\otimes\left[\left(\sum_{m=0}^{l-1}(B^T)^{l-1-m}\otimes B^m\right)\frac{\partial \text{vec}(B)}{\partial \text{vec}^T(B_j)}\right]$$
$$+\text{vec}(B^{k-1-l})\otimes\left[\left(\sum_{m=0}^{l-1}\left(\left(\sum_{s=0}^{l-2-m}(B^T)^{l-2-m-s}\otimes B^s\right)\frac{\partial B}{\partial B_{uv,r}}\right)^T\right.\right.$$
$$\left.\left.\otimes B^m\right)\frac{\partial \text{vec}(B)}{\partial \text{vec}^T(B_j)}\right]$$
$$+\text{vec}(B^{k-1-l})\otimes\left[\left(\sum_{m=0}^{l-1}(B^T)^{l-1-m}\otimes\left[\left(\sum_{s=0}^{m-1}(B^T)^{m-1-s}\right.\right.\right.\right.$$
$$\left.\left.\left.\left.\otimes B^s\right)\frac{\partial B}{\partial B_{uv,r}}\right]\right)\frac{\partial \text{vec}(B)}{\partial \text{vec}^T(B_j)}\right]\Bigg\}.$$

Therefore,

$$\begin{aligned}&\mathbb{E}\left\|\frac{\partial}{\partial \text{vec}^T(B_j)}\text{vec}\left[\frac{\partial}{\partial \text{vec}^T(B_i)}\text{vec}\left(\frac{\partial X_t}{\partial c^T}\right)\right]\right\|\\ \leq\;& C_1\sum_{k=2}^{\infty}\|\Gamma(B,k,j)\|\leq C_2\sum_{k=2}^{\infty}k(k-1)\bar{\rho}^{k-2}<\infty,\end{aligned}$$

$$\begin{aligned}&\mathbb{E}\left\|\frac{\partial}{\partial \text{vec}^T(B_r)}\text{vec}\left[\frac{\partial}{\partial \text{vec}^T(B_j)}\text{vec}\left(\frac{\partial X_t}{\partial \text{vec}^T(A_i)}\right)\right]\right\|\\ \leq\;& C_1\sum_{k=2}^{\infty}\mathbb{E}\|\underline{\epsilon}_{t-k-i}\|\cdot\|\Gamma(B,k,j)\|\leq C_2\mathbb{E}\|y_t\|^2\sum_{k=2}^{\infty}k(k-1)\bar{\rho}^{k-2}<\infty,\end{aligned}$$

$$\begin{aligned}&\mathbb{E}\left\|\frac{\partial}{\partial c^T}\text{vec}\left[\frac{\partial}{\partial \text{vec}^T(B_j)}\text{vec}\left(\frac{\partial X_t}{\partial \text{vec}^T(B_i)}\right)\right]\right\|\\ \leq\;& C_1\sum_{k=2}^{\infty}\|\Gamma(B,k,j)\|\leq C_2\sum_{k=2}^{\infty}k(k-1)\bar{\rho}^{k-2}<\infty,\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}\left\|\frac{\partial}{\partial \mathrm{vec}^T(A_r)}\mathrm{vec}\left[\frac{\partial}{\partial \mathrm{vec}^T(B_j)}\mathrm{vec}\left(\frac{\partial X_t}{\partial \mathrm{vec}^T(B_i)}\right)\right]\right\| \\
\leq\ & C_1\sum_{k=2}^{\infty}\mathbb{E}\|\underline{\epsilon}_{t-k}\|\cdot\|\Gamma(B,k,j)\| \leq C_2\mathbb{E}\|y_t\|^2\sum_{k=2}^{\infty}k(k-1)\bar{\rho}^{k-2}<\infty,
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}\left\|\frac{\partial}{\partial B_{uv,r}}\left[\frac{\partial}{\partial \mathrm{vec}^T(B_j)}\mathrm{vec}\left(\frac{\partial X_t}{\partial \mathrm{vec}^T(B_i)}\right)\right]\right\| \\
\leq\ & C_1\sum_{k=3}^{\infty}\sum_{l=0}^{k-1}\left\{\sum_{m=0}^{k-2-l}\left[\left(\sum_{s=0}^{k-3-l-m}\bar{\rho}^{k-3-l-m-s}\bar{\rho}^{s}\right)\bar{\rho}^{m}\right]\bar{\rho}^{l}\right. \\
&+\sum_{m=0}^{k-2-l}\left[\bar{\rho}^{k-2-l-m}\left(\sum_{s=0}^{m-1}\bar{\rho}^{m-1-s}\bar{\rho}^{s}\right)\right]\bar{\rho}^{l} \\
&+\left(\sum_{m=0}^{k-2-l}\bar{\rho}^{k-2-l-m}\bar{\rho}^{m}\right)\left(\sum_{m=0}^{l-1}\bar{\rho}^{l-1-m}\bar{\rho}^{m}\right) \\
&+\left(\sum_{m=0}^{k-2-l}\bar{\rho}^{k-2-l-m}\bar{\rho}^{m}\right)\left(\sum_{m=0}^{l-1}\bar{\rho}^{l-1-m}\bar{\rho}^{m}\right) \\
&+\bar{\rho}^{k-1-l}\left[\sum_{m=0}^{l-1}\left(\sum_{s=0}^{l-2-m}\bar{\rho}^{l-2-m-s}\bar{\rho}^{s}\right)\bar{\rho}^{m}\right] \\
&\left.+\bar{\rho}^{k-1-l}\left[\sum_{m=0}^{l-1}\bar{\rho}^{l-1-m}\left(\sum_{s=0}^{m-1}\bar{\rho}^{m-1-s}\bar{\rho}^{s}\right)\right]\right\} \\
=\ & C_2\sum_{k=3}^{\infty}k(k-1)(k-2)\bar{\rho}^{k-3}<\infty.
\end{aligned}$$

□

## 4.6 Lemma 4.6

**Lemma 4.6.** *1.* $\mathbb{E}\left\|\frac{\partial l_t(\theta_0)}{\partial\theta}\frac{\partial l_t(\theta_0)}{\partial\theta^T}\right\|<\infty.$

*2.* $\mathbb{E}\left\|\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta^T}\right\|<\infty.$

*3. There exists a neighborhood $\nu(\theta_0)$ such that for all $i$, $j$ and $k$,*

$$\mathbb{E} \sup_{\theta\in\nu(\theta_0)} \left| \frac{\partial^3 l_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k} \right| < \infty.$$

*Proof.* 1. By (8) and (14) in Appendix B,

$$\begin{aligned}
& \frac{\partial l_t(\theta)}{\partial\theta_i} \\
= \;& \frac{\partial}{\partial\theta_i}\log|H_t(\theta)| + \frac{\partial}{\partial\theta_i}\mathrm{tr}(y_t y_t^T H_t^{-1}(\theta)) \\
= \;& |H_t^{-1}(\theta)|\frac{\partial}{\partial\theta_i}|H_t(\theta)| + \mathrm{tr}\left(y_t y_t^T \frac{\partial}{\partial\theta_i} H_t^{-1}(\theta)\right) \\
= \;& |H_t^{-1}(\theta)|\frac{\partial|H_t(\theta)|}{\partial\mathrm{vec}^T(H_t(\theta))}\frac{\partial\mathrm{vec}(H_t(\theta))}{\partial\theta_i} - \mathrm{tr}(y_t y_t^T H_t^{-1}(\theta)\dot{H}_{t,i}(\theta)H_t^{-1}(\theta)) \\
= \;& |H_t^{-1}(\theta)|\mathrm{vec}^T\left(\frac{\partial|H_t(\theta)|}{\partial H_t(\theta)}\right)\mathrm{vec}\left(\frac{\partial H_t(\theta)}{\partial\theta_i}\right) \\
& -\mathrm{tr}(y_t y_t^T H_t^{-1}(\theta)\dot{H}_{t,i}(\theta)H_t^{-1}(\theta)) \\
= \;& |H_t^{-1}(\theta)|\mathrm{vec}^T(|H_t(\theta)|H_t^{-1}(\theta))\mathrm{vec}(\dot{H}_{t,i}(\theta)) \\
& -\mathrm{tr}(y_t y_t^T H_t^{-1}(\theta)\dot{H}_{t,i}(\theta)H_t^{-1}(\theta)) \\
= \;& \mathrm{tr}(H_t^{-1}(\theta)\dot{H}_{t,i}(\theta)) - \mathrm{tr}(y_t y_t^T H_t^{-1}(\theta)\dot{H}_{t,i}(\theta)H_t^{-1}(\theta)) \\
= \;& \mathrm{tr}[(I_d - y_t y_t^T H_t^{-1}(\theta))\dot{H}_{t,i}(\theta)H_t^{-1}(\theta)]. \qquad (4.14)
\end{aligned}$$

When $\theta = \theta_0$, we have

$$\frac{\partial l_t(\theta_0)}{\partial\theta_i} = \mathrm{tr}[(I_d - H_t^{1/2}(\theta_0)\xi_t\xi_t^T H_t^{-1/2}(\theta_0))\dot{H}_{t,i}(\theta_0)H_t^{-1}(\theta_0)] = \mathrm{tr}(\Upsilon_{t,i}). \qquad (4.15)$$

The following three results are useful in our proof:

(a)

$$
\begin{aligned}
& \|H_t^{1/2}(\theta_0)\xi_t\xi_t^T H_t^{-1/2}(\theta_0)\| \\
\leq\ & \|H_t^{1/2}(\theta_0)\xi_t\xi_t^T H_t^{-1/2}(\theta_0)\|_2 \\
=\ & \Big\{\mathrm{tr}[(H_t^{1/2}(\theta_0)\xi_t\xi_t^T H_t^{-1/2}(\theta_0))^T H_t^{1/2}(\theta_0)\xi_t\xi_t^T H_t^{-1/2}(\theta_0)]\Big\}^{1/2} \\
\leq\ & \frac{1}{2}\Big\{\mathrm{tr}[(H_t^{1/2}(\theta_0)\xi_t\xi_t^T H_t^{-1/2}(\theta_0))^T] + \mathrm{tr}(H_t^{1/2}(\theta_0)\xi_t\xi_t^T H_t^{-1/2}(\theta_0))\Big\} \\
=\ & \mathrm{tr}(H_t^{1/2}(\theta_0)\xi_t\xi_t^T H_t^{-1/2}(\theta_0)) \\
=\ & \mathrm{tr}(\xi_t^T\xi_t) \\
=\ & \|\xi_t\|^2
\end{aligned}
$$

(b) $\mathbb{E}\|\xi_t\|^6 \leq \mathbb{E}\|H^{-1/2}\|^6\|y_t\|^6 \leq \dfrac{1}{\gamma^3}\mathbb{E}\|y_t\|^6 < \infty$, where $\gamma$ is defined in Lemma 4.2

By Lemma 4.5, Formulas (5), (15) and (16) in Appendix B, the independence between $\xi_t$ and $H_t$ and the Cauchy-Schwarz inequality, we can obtain that

$$
\begin{aligned}
& \mathbb{E}\left|\frac{\partial l_t(\theta_0)}{\partial\theta_i}\frac{\partial l_t(\theta_0)}{\partial\theta_j}\right| \\
=\ & \mathbb{E}|\mathrm{tr}(\Upsilon_{t,i})\mathrm{tr}(\Upsilon_{t,j})| \\
=\ & \mathbb{E}|\mathrm{tr}(\Upsilon_{t,i}\otimes\Upsilon_{t,j})| \\
\leq\ & C_1\mathbb{E}\|\Upsilon_{t,i}\otimes\Upsilon_{t,j}\|
\end{aligned}
$$

$$
\begin{aligned}
&= C_1\mathbb{E}(\|\Upsilon_{t,i}\|\cdot\|\Upsilon_{t,j}\|)\\
&\leq C_2\mathbb{E}\left[(1+\|\xi_t\|)^2\|\dot{H}_{t,i}(\theta_0)\|\cdot\|\dot{H}_{t,j}(\theta_0)\|\right]\\
&= C_2\mathbb{E}(1+\|\xi_t\|)^2\mathbb{E}\left(\|\dot{H}_{t,i}(\theta_0)\|\cdot\|\dot{H}_{t,j}(\theta_0)\|\right)\\
&\leq C_2(1+2\mathbb{E}\|\xi_t\|^2+\mathbb{E}\|\xi_t\|^4)\left[\mathbb{E}\|\dot{H}_{t,i}(\theta_0)\|^2\right]^{1/2}\left[\mathbb{E}\|\dot{H}_{t,j}(\theta_0)\|^2\right]^{1/2}<\infty.
\end{aligned}
$$

And the desired result follows.

2. By Formula (14) in Appendix B and the product rule,

$$
\begin{aligned}
&\frac{\partial l_t^2(\theta)}{\partial\theta_i\partial\theta_j}\\
&= \operatorname{tr}\left[\frac{\partial}{\partial\theta_j}\dot{H}_{t,i}(\theta)H_t^{-1}(\theta)-y_ty_t^T\frac{\partial}{\partial\theta_j}H_t^{-1}(\theta)\dot{H}_{t,i}(\theta)H_t^{-1}(\theta)\right]\\
&= \operatorname{tr}\left[\ddot{H}_{t,ij}(\theta)H_t^{-1}(\theta)+\dot{H}_{t,i}(\theta)\frac{\partial H_t^{-1}(\theta)}{\partial\theta_j}-y_ty_t^T\left(\frac{\partial H_t^{-1}(\theta)}{\partial\theta_j}\dot{H}_{t,i}(\theta)H_t^{-1}(\theta)\right.\right.\\
&\qquad\left.\left.+H_t^{-1}(\theta)\ddot{H}_{t,ij}(\theta)H_t^{-1}(\theta)+H_t^{-1}(\theta)\dot{H}_{t,i}(\theta)\frac{\partial H_t^{-1}(\theta)}{\partial\theta_j}\right)\right]\\
&= \operatorname{tr}\left[\ddot{H}_{t,ij}(\theta)H_t^{-1}(\theta)+\dot{H}_{t,i}(\theta)H_t^{-1}(\theta)\dot{H}_{t,j}(\theta)H_t^{-1}(\theta)+y_ty_t^TH_t^{-1}(\theta)\right.\\
&\qquad\left.\left(\dot{H}_{t,j}(\theta)\dot{H}_{t,i}(\theta)-\ddot{H}_{t,ij}(\theta)+\dot{H}_{t,i}(\theta)H_t^{-1}(\theta)\dot{H}_{t,j}(\theta)\right)H_t^{-1}(\theta)\right]. \qquad (4.16)
\end{aligned}
$$

When $\theta=\theta_0$, we have

$$
\begin{aligned}
\frac{\partial l_t^2(\theta_0)}{\partial\theta_i\partial\theta_j} &= \operatorname{tr}\left[\ddot{H}_{t,ij}(\theta_0)H_t^{-1}(\theta_0)+\dot{H}_{t,i}(\theta_0)H_t^{-1}(\theta_0)\dot{H}_{t,j}(\theta_0)H_t^{-1}(\theta_0)\right.\\
&\qquad+H_t^{1/2}(\theta_0)\xi_t\xi_t^TH_t^{-1/2}(\theta_0)\left(\dot{H}_{t,j}(\theta_0)\dot{H}_{t,i}(\theta_0)-\ddot{H}_{t,ij}(\theta_0)\right.\\
&\qquad\left.\left.+\dot{H}_{t,i}(\theta_0)H_t^{-1}(\theta_0)\dot{H}_{t,j}(\theta_0)\right)H_t^{-1}(\theta_0)\right].
\end{aligned}
$$

By Lemma 4.5, the independence between $\xi_t$ and $H_t$ and the Cauchy-Schwarz inequality, we can obtain that

$$
\begin{aligned}
& \mathbb{E}\left|\frac{\partial l_t^2(\theta_0)}{\partial\theta_i\partial\theta_j}\right| \\
= \; & \mathbb{E}\Big|\mathrm{tr}\Big[\ddot{H}_{t,ij}(\theta_0)H_t^{-1}(\theta_0) + \dot{H}_{t,i}(\theta_0)H_t^{-1}(\theta_0)\dot{H}_{t,j}(\theta_0)H_t^{-1}(\theta_0) \\
& \quad +H_t^{1/2}(\theta_0)\xi_t\xi_t^T H_t^{-1/2}(\theta_0)\Big(\dot{H}_{t,j}(\theta_0)\dot{H}_{t,i}(\theta_0) - \ddot{H}_{t,ij}(\theta_0) \\
& \qquad +\dot{H}_{t,i}(\theta_0)H_t^{-1}(\theta_0)\dot{H}_{t,j}(\theta_0)\Big) H_t^{-1}(\theta_0)\Big]\Big|. \\
\le \; & C_1\mathbb{E}\Big\|\Big[\ddot{H}_{t,ij}(\theta_0)H_t^{-1}(\theta_0) + \dot{H}_{t,i}(\theta_0)H_t^{-1}(\theta_0)\dot{H}_{t,j}(\theta_0)H_t^{-1}(\theta_0) \\
& \quad +H_t^{1/2}(\theta_0)\xi_t\xi_t^T H_t^{-1/2}(\theta_0)\Big(\dot{H}_{t,j}(\theta_0)\dot{H}_{t,i}(\theta_0) - \ddot{H}_{t,ij}(\theta_0) \\
& \qquad +\dot{H}_{t,i}(\theta_0)H_t^{-1}(\theta_0)\dot{H}_{t,j}(\theta_0)\Big) H_t^{-1}(\theta_0)\Big]\Big\|. \\
\le \; & C_2\mathbb{E}\|\ddot{H}_{t,ij}(\theta_0)\| + C_3\mathbb{E}\|\dot{H}_{t,i}(\theta_0)\|\cdot\|\dot{H}_{t,j}(\theta_0)\| \\
& \quad +C_4\mathbb{E}\|\xi_t\|^2\Big(\mathbb{E}\|\dot{H}_{t,j}(\theta_0)\|\cdot\|\dot{H}_{t,i}(\theta_0)\| + \mathbb{E}\|\ddot{H}_{t,ij}(\theta_0)\| \\
& \qquad \Big(+C_5\mathbb{E}\|\dot{H}_{t,j}(\theta_0)\|\cdot\|\dot{H}_{t,i}(\theta_0)\|\Big) \\
\le \; & C_6 + C_7\Big(\mathbb{E}\|\dot{H}_{t,i}(\theta_0)\|^2\Big)^{1/2}\Big(\mathbb{E}\|\dot{H}_{t,j}(\theta_0)\|^2\Big)^{1/2} < \infty.
\end{aligned}
$$

And the desired result follows.

3. By Formula (14) in Appendix B and the product rule,

$$
\begin{aligned}
& \frac{\partial^3 l_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k} \\
= \; & \mathrm{tr}\left\{\frac{\partial}{\partial\theta_k}\Big[\ddot{H}_{t,ij}(\theta)H_t^{-1}(\theta) + \dot{H}_{t,i}(\theta)H_t^{-1}(\theta)\dot{H}_{t,j}(\theta)H_t^{-1}(\theta) + y_t y_t^T H_t^{-1}(\theta)\right.
\end{aligned}
$$

$$
\begin{aligned}
& \left(\dot{H}_{t,j}(\theta)\dot{H}_{t,i}(\theta) - \ddot{H}_{t,ij}(\theta) + \dot{H}_{t,i}(\theta)H_t^{-1}(\theta)\dot{H}_{t,j}(\theta)\right) H_t^{-1}(\theta)\Big]\Big\} \\
= \; & \operatorname{tr}\Big\{\Big(\dddot{H}_{t,ijk}(\theta) - \ddot{H}_{t,ij}(\theta)H_t^{-1}(\theta)\dot{H}_{t,k}(\theta) + \ddot{H}_{t,ik}(\theta)H_t^{-1}(\theta)\dot{H}_{t,j}(\theta) \\
& -\dot{H}_{t,i}(\theta)H_t^{-1}(\theta)\dot{H}_{t,k}(\theta)H_t^{-1}(\theta)\dot{H}_{t,j}(\theta) + \dot{H}_{t,i}H_t^{-1}(\theta)\ddot{H}_{t,jk} \\
& -\dot{H}_{t,i}(\theta)H_t^{-1}(\theta)\dot{H}_{t,j}(\theta)H_t^{-1}(\theta)\dot{H}_{t,k}(\theta)\Big) H_t^{-1}(\theta) - y_t y_t^T H_t^{-1}(\theta) \\
& \Big[\dot{H}_{t,k}(\theta)H_t^{-1}(\theta)\left(\dot{H}_{t,j}(\theta)\dot{H}_{t,i}(\theta) - \ddot{H}_{t,ij}(\theta) + \dot{H}_{t,i}(\theta)H_t^{-1}(\theta)\dot{H}_{t,j}(\theta)\right) \\
& \quad -\ddot{H}_{t,jk}(\theta)\dot{H}_{t,i}(\theta) - \dot{H}_{t,j}(\theta)\ddot{H}_{t,ik}(\theta) + \dddot{H}_{t,ijk}(\theta) \\
& \quad -\ddot{H}_{t,ik}(\theta)H_t^{-1}(\theta)\dot{H}_{t,j}(\theta) + \dot{H}_{t,i}(\theta)H_t^{-1}(\theta)\dot{H}_{t,k}(\theta)H_t^{-1}(\theta)\dot{H}_{t,j}(\theta) \\
& \quad -\dot{H}_{t,i}(\theta)H_t^{-1}(\theta)\ddot{H}_{t,jk}(\theta) \\
& \quad +\left(\dot{H}_{t,j}(\theta)\dot{H}_{t,i}(\theta) - \ddot{H}_{t,ij}(\theta) + \dot{H}_{t,i}(\theta)H_t^{-1}(\theta)\dot{H}_{t,j}(\theta)\right) \\
& \quad H_t^{-1}(\theta)\dot{H}_{t,k}(\theta)\Big] H_t^{-1}(\theta)\Big\} \\
= \; & \operatorname{tr}[\Psi_{1,t} - y_t y_t^T H_t^{-1}(\theta)\Psi_{2,t}]. \qquad (4.17)
\end{aligned}
$$

We wish to use the same technique as in the previous parts in order to reduce the moment requirement on $\|y_t\|$. We need to show that the difference between $\sup_{\theta\in\nu(\theta_0)} \|y_t y_t^T H_t^{-1}(\theta)\|$ and $\|y_t y_t^T H_t^{-1}(\theta_0)\|$ is arbitrarily small. Suppose $\nu(\theta_0) = \|\hat{\theta}_n - \theta_0\| < \epsilon$.

$$
\begin{aligned}
& \sup_{\theta\in\nu(\theta_0)} \|y_t y_t^T H_t^{-1}(\theta)\| \\
\leq \; & \|y_t y_t^T H_t^{-1}(\theta_0)\| + \sup_{\theta\in\nu(\theta_0)} \|y_t y_t^T [H_t^{-1}(\theta) - H_t^{-1}(\theta_0)]\| \\
\leq \; & \|\xi_t\|^2 + \sup_{\theta\in\nu(\theta_0)} \|y_t y_t^T H_t^{-1}(\theta)[H_t(\theta_0) - H_t(\theta)]H_t^{-1}(\theta_0)\|
\end{aligned}
$$

$$
\begin{aligned}
&\le \|\xi_t\|^2 + \frac{1}{\gamma^2}\sup_{\theta\in\nu(\theta_0)} \|y_t y_t^T[H_t(\theta)-H_t(\theta_0)]\| \\
&\le \|\xi_t\|^2 + C_1 \sup_{\theta\in\nu(\theta_0)} \|y_t\|^2\|X_t(\theta)-X_t(\theta_0)\| \\
&\le \|\xi_t\|^2 + C_1 \sup_{\theta\in\nu(\theta_0)} \sum_{k=0}^{\infty} \|y_t\|^2\|B^k(\theta)c_{t-k}(\theta)-B^k(\theta_0)c_{t-k}(\theta_0)\| \\
&= \|\xi_t\|^2 + C_1 \sup_{\theta\in\nu(\theta_0)} \sum_{k=0}^{\infty} \|y_t\|^2\|(B^k(\theta)-B^k(\theta_0))c_{t-k}(\theta) \\
&\quad +B^k(\theta_0)(c_{t-k}(\theta)-c_{t-k}(\theta_0))\| \\
&= \|\xi_t\|^2 + C_1 \sup_{\theta\in\nu(\theta_0)} \sum_{k=0}^{\infty} \|y_t\|^2 \left\|\sum_{l=0}^{k-1}[B^{k-1-l}(\theta)(B(\theta)-B(\theta_0))B^l(\theta_0)]c_{t-k}(\theta)\right. \\
&\quad \left. +B^k(\theta_0)(c_{t-k}(\theta)-c_{t-k}(\theta_0))\right\| \\
&\le \|\xi_t\|^2 + C_2\epsilon \sum_{k=0}^{\infty} \|y_t\|^2 \left[k\bar{\rho}^{k-1}\sup_{\theta\in\nu(\theta_0)}\|c_{t-k}(\theta)\| + \bar{\rho}^k\left(1+\sum_{i=1}^{q}\|y_{t-k-i}\|^2\right)\right].
\end{aligned}
$$

The summation converges almost surely since it has finite expectation. For example,

$$
\mathbb{E}\|y_t\|^2\|y_{t-k-i}\|^2 \le \left(\mathbb{E}\|y_t\|^4\right)^{1/2}\left(\mathbb{E}\|y_{t-k-i}\|^4\right)^{1/2}.
$$

By noticing $\epsilon$ is arbitrarily small, we have almost surely that

$$
\sup_{\theta\in\nu(\theta_0)} \|y_t y_t^T H_t^{-1}(\theta)\| \le \|\xi_t\|^2 + o(1),
$$

which is independent of $\Psi_{2,t}$ in (4.17). Both $\|\Psi_{1,t}\|$ and $\|\Psi_{2,t}\|$ have finite expectations due to the Hölder's inequality, Lemma 4.5 and the fact that

$\|H_t^{-1}\| \leq 1/\gamma$. For instance,

$$\mathbb{E}\|\dot{H}_{t,i}\dot{H}_{t,j}\dot{H}_{t,k}\| \leq \left(\mathbb{E}\|\dot{H}_{t,i}\|^3\right)^{1/3}\left(\mathbb{E}\|\dot{H}_{t,j}\|^3\right)^{1/3}\left(\mathbb{E}\|\dot{H}_{t,k}\|^3\right)^{1/3} < \infty,$$

$$\mathbb{E}\|\ddot{H}_{t,ij}\dot{H}_{t,k}\| \leq \left(\mathbb{E}\|\ddot{H}_{t,ij}\|^2\right)^{1/2}\left(\mathbb{E}\|\dot{H}_{t,k}\|^2\right)^{1/2} < \infty.$$

Thus,

$$\begin{aligned}
& \mathbb{E}\sup_{\theta\in\nu(\theta_0)}\left|\frac{\partial^3 l_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| \\
\leq\ & C_3\left[\mathbb{E}\sup_{\theta\in\nu(\theta_0)}\|\Psi_{1,t}\| + \mathbb{E}\sup_{\theta\in\nu(\theta_0)}\|y_t y_t^T H_t^{-1}(\theta)\|\|\Psi_{2,t}\|\right] \\
\leq\ & C_3\left[\mathbb{E}\sup_{\theta\in\nu(\theta_0)}\|\Psi_{1,t}\| + \mathbb{E}\sup_{\theta\in\nu(\theta_0)}(\|\xi_t\|^2 + o(1))\cdot\mathbb{E}\sup_{\theta\in\nu(\theta_0)}\|\Psi_{2,t}\|\right] \\
<\ & \infty.
\end{aligned}$$

□

## 4.7 Lemma 4.7

**Lemma 4.7.** *1.* $\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\theta} - \frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}\right\| \xrightarrow{\mathcal{P}} 0$ *as* $n\to\infty$.

*2.* $\sup_{\theta\in\nu(\theta_0)}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta^T} - \frac{\partial^2\tilde{l}_t(\theta)}{\partial\theta\partial\theta^T}\right\| \xrightarrow{\mathcal{P}} 0$ *as* $n\to\infty$.

*Proof.* 1. Given the initial values we chose in (1.8),

$$\tilde{X}_0 = (c^T, \cdots, c^T)^T \quad \text{and} \quad \tilde{\eta}_0 = \tilde{\eta}_1 = \cdots = \tilde{\eta}_{1-q} = \text{vech}(y_1 y_1^T).$$

In view of (4.5), (4.6) (4.8) (4.9) and (4.10), we have almost surely that, for $t \geq M$, where $M$ is a sufficiently large integer,

$$\begin{aligned}
\left\| \frac{\partial X_t}{\partial c^T} - \frac{\partial \tilde{X}_t}{\partial c^T} \right\| &= \left\| \sum_{k=1}^{q} B^{t-k} \left( \frac{\partial c_k}{\partial c^T} - \frac{\partial \tilde{c}_k}{\partial c^T} \right) + B^t \left( \frac{\partial X_0}{\partial c^T} - \frac{\partial \tilde{X}_0}{\partial c^T} \right) \right\| \\
&= \left\| B^t \left( \sum_{k=0}^{\infty} B^k \underline{1} - \underline{I} \right) \right\| \leq O(\bar{\rho}^t),
\end{aligned}$$

where $\underline{I} = (I^T, \cdots, I^T)^T$

$$\begin{aligned}
&\left\| \frac{\partial X_t}{\partial \text{vec}^T(A_i)} - \frac{\partial \tilde{X}_t}{\partial \text{vec}^T(A_i)} \right\| \\
= &\left\| \sum_{k=1}^{q} B^{t-k} \left( \frac{\partial c_k}{\partial \text{vec}^T(A_i)} - \frac{\partial \tilde{c}_k}{\partial \text{vec}^T(A_i)} \right) \right. \\
&\left. + B^t \left( \frac{\partial X_0}{\partial \text{vec}^T(A_i)} - \frac{\partial \tilde{X}_0}{\partial \text{vec}^T(A_i)} \right) \right\| \\
= &\left\| B^t \left( \sum_{k=0}^{\infty} B^k \underline{\epsilon}_{-k-i} \right) \right\| \leq O(\bar{\rho}^t),
\end{aligned}$$

$$\begin{aligned}
&\left\| \frac{\partial X_t}{\partial \text{vec}^T(B_i)} - \frac{\partial \tilde{X}_t}{\partial \text{vec}^T(B_i)} \right\| \\
= &\left\| \left[ \sum_{k=1}^{q} ((c_k - \tilde{c}_k)^T \otimes I_{Np}) \left( \sum_{l=0}^{t-k-1} (B^T)^{t-k-1-l} \otimes B^l \right) \frac{\partial \text{vec}(B)}{\partial \text{vec}^T(B_i)} \right] \right. \\
&\left. + \frac{\partial B^t X_0}{\partial \text{vec}^T(B_i)} \right\| \\
= &\left\| \sum_{k=1}^{q} ((c_k - \tilde{c}_k)^T \otimes I_{Np}) \left( \sum_{l=0}^{t-k-1} (B^T)^{t-k-1-l} \otimes B^l \right) \frac{\partial \text{vec}(B)}{\partial \text{vec}^T(B_i)} \right. \\
&\left. + \sum_{k=1}^{\infty} c_{-k}^T \otimes I_{Np} \left( \sum_{l=0}^{t+k-1} (B^T)^{t+k-1-l} \otimes B^l \right) \frac{\partial \text{vec}(B)}{\partial \text{vec}^T(B_i)} \right\|
\end{aligned}$$

$$\leq \quad O(t\bar{\rho}^t).$$

Thus,

$$\|\dot{H}_{t,i} - \dot{\tilde{H}}_{t,i}\| \leq O(t\bar{\rho}^t).$$

By (4.7), almost surely,

$$\|H_t^{-1} - \tilde{H}_t^{-1}\| \leq \|H_t^{-1}\|\|H_t - \tilde{H}_t\|\|\tilde{H}_t^{-1}\| \leq \frac{1}{\gamma}O(\bar{\rho}^t)\frac{1}{\gamma} = O(\bar{\rho}^t).$$

In view of (4.14), almost surely, for $t \geq M$,

$$\begin{aligned}
\|\dot{H}_{t,i}H_t^{-1} - \dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}\| &\leq \|\dot{H}_{t,i}\|\|H_t^{-1} - \tilde{H}_t^{-1}\| + \|\dot{H}_{t,i} - \dot{\tilde{H}}_{t,i}\|\|\tilde{H}_t^{-1}\| \\
&\leq \|\dot{H}_{t,i}\|O(\bar{\rho}^t) + \frac{1}{\gamma}O(t\bar{\rho}^t) \\
&= \|\dot{H}_{t,i}\|O(\bar{\rho}^t) + O(t\bar{\rho}^t),
\end{aligned}$$

and

$$\begin{aligned}
&\|H_t^{-1}\dot{H}_{t,i}H_t^{-1} - \tilde{H}_t^{-1}\dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}\| \\
\leq\ &\|H_t^{-1} - \tilde{H}_t^{-1}\|\|\dot{H}_{t,i}\|\|H_t^{-1}\| + \|\tilde{H}_t^{-1}\|\|\dot{H}_{t,i}H_t^{-1} - \dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}\| \\
\leq\ &O(\bar{\rho}^t)\|\dot{H}_{t,i}\|\frac{1}{\gamma} + \frac{1}{\gamma}[\|\dot{H}_{t,i}\|O(\bar{\rho}^t) + O(t\bar{\rho}^t)] \\
=\ &\|\dot{H}_{t,i}\|O(\bar{\rho}^t) + O(t\bar{\rho}^t).
\end{aligned}$$

Therefore, by (4.14),

$$
\begin{aligned}
& \left\| \frac{\partial l_t(\theta_0)}{\partial \theta} - \frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta} \right\| \\
\leq\ & C_1(\|\dot{H}_{t,i}(\theta_0)H_t^{-1}(\theta_0) - \dot{\tilde{H}}_{t,i}(\theta_0)\tilde{H}_t^{-1}(\theta_0)\| \\
& \quad +\|y_t\|^2\|H_t^{-1}(\theta_0)\dot{H}_{t,i}(\theta_0)H_t^{-1}(\theta_0) - \tilde{H}_t^{-1}(\theta_0)\dot{\tilde{H}}_{t,i}(\theta_0)\tilde{H}_t^{-1}(\theta_0)\|) \\
=\ & [\|\dot{H}_{t,i}\|O(\bar{\rho}^t) + O(t\bar{\rho}^t)] + \|y_t\|^2[\|\dot{H}_{t,i}\|O(\bar{\rho}^t) + O(t\bar{\rho}^t)].
\end{aligned}
$$

For any $\epsilon > 0$, by the Markov inequality,

$$
\begin{aligned}
& \mathbb{P}\left( \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t(\theta_0)}{\partial \theta} - \frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta} \right\| > \epsilon \right) \\
\leq\ & \mathbb{P}\left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\| \frac{\partial l_t(\theta_0)}{\partial \theta} - \frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta} \right\| > \epsilon \right) \\
\leq\ & \frac{\frac{1}{\sqrt{n}} \sum_{t=M}^{n} \mathbb{E} \left\| \frac{\partial l_t(\theta_0)}{\partial \theta} - \frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta} \right\|}{\epsilon} + o(1) \\
\leq\ & \frac{1}{\epsilon\sqrt{n}} \sum_{t=M}^{n} \mathbb{E}[\|\dot{H}_{t,i}\|O(\bar{\rho}^t) + O(t\bar{\rho}^t)] + \mathbb{E}\|y_t\|^2[\|\dot{H}_{t,i}\|O(\bar{\rho}^t) + O(t\bar{\rho}^t)] \\
\leq\ & \frac{1}{\epsilon\sqrt{n}} \sum_{t=M}^{n} O(t\bar{\rho}^t) + O(\bar{\rho}^t)[\mathbb{E}\|y_t\|^4]^{1/2}[\mathbb{E}\|\dot{H}_{t,i}\|^2]^{1/2} \\
\leq\ & \frac{1}{\epsilon\sqrt{n}} \sum_{t=M}^{n} O(t\bar{\rho}^t) \to 0.
\end{aligned}
$$

This finalizes our proof of the first part.

2. In view of (4.5), (4.6) (4.11) (4.12), (4.13) and the results from the previous part of this lemma, we have almost surely that, for $t \geq M$, where $M$ is a

sufficiently large integer,

$$\begin{aligned}
&\left\|\frac{\partial}{\partial \mathrm{vec}^T(B_i)}\mathrm{vec}\left(\frac{\partial X_t}{\partial c^T}\right)-\frac{\partial}{\partial \mathrm{vec}^T(B_i)}\mathrm{vec}\left(\frac{\partial \tilde{X}_t}{\partial c^T}\right)\right\| \\
= \quad & \left\|\sum_{k=0}^{\infty}\frac{\partial \mathrm{vec}(B^{t+k}\underline{1})}{\partial \mathrm{vec}^T(B_i)}-\frac{\partial \mathrm{vec}(B^t\underline{I})}{\partial \mathrm{vec}^T(B_i)}\right\| \\
= \quad & \left\|\left[\sum_{k=0}^{\infty}(\underline{1}^T\otimes I_{Np})\left(\sum_{l=0}^{t+k-1}(B^T)^{t+k-1-l}\otimes B^l\right)\right.\right. \\
& \quad \left.\left. -(\underline{1}^T\otimes I_{Np})\left(\sum_{l=0}^{t-1}(B^T)^{t-1-l}\otimes B^l\right)\right]\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_i)}\right\| \\
\leq \quad & O(t\bar{\rho}^t),
\end{aligned}$$

$$\begin{aligned}
&\left\|\frac{\partial}{\partial \mathrm{vec}^T(B_j)}\mathrm{vec}\left(\frac{\partial X_t}{\partial \mathrm{vec}^T(A_i)}\right)-\frac{\partial}{\partial \mathrm{vec}^T(B_j)}\mathrm{vec}\left(\frac{\partial \tilde{X}_t}{\partial \mathrm{vec}^T(A_i)}\right)\right\| \\
= \quad & \left\|\frac{\partial \mathrm{vec}\left(B^t\left(\sum_{k=0}^{\infty}B^k\underline{\epsilon}_{-k-i}\right)\right)}{\partial \mathrm{vec}^T(B_j)}\right\| \\
= \quad & \left\|\sum_{k=0}^{\infty}(\underline{\epsilon}_{-k-i}^T\otimes I_{Np})\left(\sum_{l=0}^{t+k-1}(B^T)^{t+k-1-l}\otimes B^l\right)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_i)}\right\| \\
\leq \quad & O(t\bar{\rho}^t),
\end{aligned}$$

$$\begin{aligned}
&\left\|\frac{\partial}{\partial \mathrm{vec}^T(B_j)}\mathrm{vec}\left(\frac{\partial X_t}{\partial \mathrm{vec}^T(B_i)}\right)-\frac{\partial}{\partial \mathrm{vec}^T(B_j)}\mathrm{vec}\left(\frac{\partial \tilde{X}_t}{\partial \mathrm{vec}^T(B_i)}\right)\right\| \\
= \quad & \left\|\frac{\partial}{\partial \mathrm{vec}^T(B_j)}\times\right. \\
& \quad \mathrm{vec}\left[\sum_{k=1}^{q}(c_k-\tilde{c}_k)^T\otimes I_{Np})\left(\sum_{l=0}^{t-k-1}(B^T)^{t-k-1-l}\otimes B^l\right)\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_i)}\right] \\
& \quad +\frac{\partial}{\partial \mathrm{vec}^T(B_j)}\times
\end{aligned}$$

$$
\begin{aligned}
& \operatorname{vec}\left[\sum_{k=1}^{\infty}(c_{-k}^{T}) \otimes I_{Np}\left(\sum_{l=0}^{t+k-1}(B^{T})^{t+k-1-l} \otimes B^{l}\right) \frac{\partial \operatorname{vec}(B)}{\partial \operatorname{vec}^{T}(B_{i})}\right]\Bigg\| \\
= & \Bigg\| \sum_{k=1}^{q}\left(\frac{\partial \operatorname{vec}^{T}(B)}{\partial \operatorname{vec}(B_{i})} \otimes (c_{k}-\tilde{c}_{k})^{T} \otimes I_{Np}\right) \\
& \times\left[\sum_{l=0}^{t-k-1} \frac{\partial}{\partial \operatorname{vec}^{T}(B_{j})} \operatorname{vec}((B^{T})^{t-k-1-l} \otimes B^{l})\right] \\
& +\sum_{k=1}^{\infty}\left(\frac{\partial \operatorname{vec}^{T}(B)}{\partial \operatorname{vec}(B_{i})} \otimes c_{-k}^{T} \otimes I_{Np}\right) \\
& \times\left[\sum_{l=0}^{t+k-1} \frac{\partial}{\partial \operatorname{vec}^{T}(B_{j})} \operatorname{vec}((B^{T})^{t+k-1-l} \otimes B^{l})\right]\Bigg\| \\
= & \Bigg\| \sum_{k=1}^{q}\left(\frac{\partial \operatorname{vec}^{T}(B)}{\partial \operatorname{vec}(B_{i})} \otimes (c_{k}-\tilde{c}_{k})^{T} \otimes I_{Np}\right)(I_{Np} \otimes K_{Np,Np} \otimes I_{Np}) \times \\
& \sum_{l=0}^{t-k-1}\left[\frac{\partial \operatorname{vec}((B^{T})^{t-k-1-l})}{\partial \operatorname{vec}^{T}(B_{j})} \otimes \operatorname{vec}(B^{l})\right. \\
& \left. +\operatorname{vec}((B^{T})^{t-k-1-l}) \otimes \frac{\partial \operatorname{vec}(B^{l})}{\partial \operatorname{vec}^{T}(B_{j})}\right] \\
& +\sum_{k=1}^{\infty}\left(\frac{\partial \operatorname{vec}^{T}(B)}{\partial \operatorname{vec}(B_{i})} \otimes c_{-k}^{T} \otimes I_{Np}\right)(I_{Np} \otimes K_{Np,Np} \otimes I_{Np}) \times \\
& \sum_{l=0}^{t+k-1}\left[\frac{\partial \operatorname{vec}((B^{T})^{t+k-1-l})}{\partial \operatorname{vec}^{T}(B_{j})} \otimes \operatorname{vec}(B^{l})\right. \\
& \left. +\operatorname{vec}((B^{T})^{t+k-1-l}) \otimes \frac{\partial \operatorname{vec}(B^{l})}{\partial \operatorname{vec}^{T}(B_{j})}\right]\Bigg\| \\
= & \Bigg\| \sum_{k=1}^{q}\left(\frac{\partial \operatorname{vec}^{T}(B)}{\partial \operatorname{vec}(B_{i})} \otimes (c_{k}-\tilde{c}_{k})^{T} \otimes I_{Np}\right)(I_{Np} \otimes K_{Np,Np} \otimes I_{Np}) \times \\
& \sum_{l=0}^{t-k-1}\left[\left(\sum_{m=0}^{t-k-2-l}[(B^{T})^{t-k-2-l-m} \otimes B^{m}] \frac{\partial \operatorname{vec}(B)}{\partial \operatorname{vec}^{T}(B_{j})}\right)^{T} \otimes \operatorname{vec}(B^{l})\right. \\
& \left. +\operatorname{vec}((B^{T})^{t-k-1-l}) \otimes \frac{\partial \operatorname{vec}(B^{l})}{\partial \operatorname{vec}^{T}(B_{j})}\right] \\
& +\sum_{k=1}^{\infty}\left(\frac{\partial \operatorname{vec}^{T}(B)}{\partial \operatorname{vec}(B_{i})} \otimes c_{-k}^{T} \otimes I_{Np}\right)(I_{Np} \otimes K_{Np,Np} \otimes I_{Np}) \times
\end{aligned}
$$

$$\sum_{l=0}^{t+k-1}\left[\left(\sum_{m=0}^{t+k-2-l}[(B^T)^{t+k-2-l-m}\otimes B^m]\frac{\partial \mathrm{vec}(B)}{\partial \mathrm{vec}^T(B_j)}\right)^T \otimes \mathrm{vec}(B^l)\right.$$
$$\left.+\mathrm{vec}((B^T)^{t+k-1-l})\otimes\frac{\partial \mathrm{vec}(B^l)}{\partial \mathrm{vec}^T(B_j)}\right]\Bigg\|$$
$$\leq \; O(t^2\bar{\rho}^t).$$

Thus,

$$\|\ddot{H}_{t,ij}-\ddot{\tilde{H}}_{t,ij}\|\leq O(t^2\bar{\rho}^t).$$

In view of (4.14) and the results from the previous part of this lemma, we have almost surely that, for $t\geq M$,

$$\begin{aligned}
\|\ddot{H}_{t,ij}H_t^{-1}-\ddot{\tilde{H}}_{t,ji}\tilde{H}_t^{-1}\| &\leq \|\ddot{H}_{t,ij}\|\|H_t^{-1}-\tilde{H}_t^{-1}\|+\|\ddot{H}_{t,ij}-\ddot{\tilde{H}}_{t,ij}\|\|\tilde{H}_t^{-1}\| \\
&\leq \|\ddot{H}_{t,ij}\|O(\bar{\rho}^t)+\frac{1}{\gamma}O(t^2\bar{\rho}^t) \\
&= \|\ddot{H}_{t,ij}\|O(\bar{\rho}^t)+O(t^2\bar{\rho}^t),
\end{aligned}$$

$$\begin{aligned}
&\|\dot{H}_{t,i}H_t^{-1}\dot{H}_{t,j}H_t^{-1}-\dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1}\| \\
\leq\;& \|\dot{H}_{t,i}-\dot{\tilde{H}}_{t,i}\|\|H_t^{-1}\|\|\dot{H}_{t,j}\|\|H_t^{-1}\| \\
&+\|\dot{\tilde{H}}_{t,i}\|\|H_t^{-1}\dot{H}_{t,j}H_t^{-1}-\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1}\| \\
\leq\;& O(\bar{\rho}^t)\frac{1}{\gamma^2}\|\dot{H}_{t,j}\|+\|\dot{\tilde{H}}_{t,i}\|[\|\dot{H}_{t,j}\|O(\bar{\rho}^t)+O(t\bar{\rho}^t)] \\
=\;& [\|\dot{H}_{t,j}\|+\|\dot{\tilde{H}}_{t,i}\|\|\dot{H}_{t,j}\|]O(\bar{\rho}^t)+\|\dot{\tilde{H}}_{t,i}\|O(t\bar{\rho}^t),
\end{aligned}$$

$$
\begin{aligned}
& \|H_t^{-1}\dot{H}_{t,i}\dot{H}_{t,j}H_t^{-1} - \tilde{H}_t^{-1}\dot{\tilde{H}}_{t,i}\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1}\| \\
\leq\ & \|H_t^{-1} - \tilde{H}_t^{-1}\|\|\dot{H}_{t,i}\|\|\dot{H}_{t,j}\|\|H_t^{-1}\| \\
& \quad +\|\tilde{H}_t^{-1}\|(\|\dot{H}_{t,i} - \dot{\tilde{H}}_{t,i}\|\|\dot{H}_{t,j}\|\|H_t^{-1}\| + \|\dot{\tilde{H}}_{t,i}\|\|\dot{H}_{t,j}H_t^{-1} - \dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1}\|) \\
\leq\ & O(\bar{\rho}^t)\|\dot{H}_{t,i}\|\|\dot{H}_{t,j}\|\frac{1}{\gamma} \\
& \quad +\frac{1}{\gamma}\left(O(t\bar{\rho}^t)\|\dot{H}_{t,j}\|\frac{1}{\gamma} + \|\dot{\tilde{H}}_{t,i}\|[\|\dot{H}_{t,j}\|O(\bar{\rho}^t) + O(t\bar{\rho}^t)]\right) \\
=\ & [\|\dot{H}_{t,i}\|\|\dot{H}_{t,j}\| + \|\dot{\tilde{H}}_{t,i}\|\|\dot{H}_{t,j}\|]O(\bar{\rho}^t) + [\|\dot{\tilde{H}}_{t,i}\| + \|\dot{H}_{t,j}\|]O(t\bar{\rho}^t),
\end{aligned}
$$

$$
\begin{aligned}
& \|H_t^{-1}\ddot{H}_{t,ij}H_t^{-1} - \tilde{H}_t^{-1}\ddot{\tilde{H}}_{t,ij}\tilde{H}_t^{-1}\| \\
\leq\ & \|H_t^{-1} - \tilde{H}_t^{-1}\|\|\ddot{H}_{t,ij}\|\|H_t^{-1}\| + \|\tilde{H}_t^{-1}\|\|\ddot{H}_{t,ij}H_t^{-1} - \ddot{\tilde{H}}_{t,ij}\tilde{H}_t^{-1}\| \\
\leq\ & O(\bar{\rho}^t)\|\ddot{H}_{t,ij}\|\frac{1}{\gamma} + \frac{1}{\gamma}[\|\ddot{H}_{t,ij}\|O(\bar{\rho}^t) + O(t^2\bar{\rho}^t)] \\
=\ & \|\ddot{H}_{t,ij}\|O(\bar{\rho}^t) + O(t^2\bar{\rho}^t),
\end{aligned}
$$

$$
\begin{aligned}
& \|H_t^{-1}\dot{H}_{t,i}H_t^{-1}\dot{H}_{t,j}H_t^{-1} - \tilde{H}_t^{-1}\dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1}\| \\
\leq\ & \|H_t^{-1} - \tilde{H}_t^{-1}\|\|\dot{H}_{t,i}\|\|H_t^{-1}\|\|\dot{H}_{t,j}\|\|H_t^{-1}\| \\
& \quad +\|\tilde{H}_t^{-1}\|\|\dot{H}_{t,i}H_t^{-1}\dot{H}_{t,j}H_t^{-1} - \dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1}\| \\
\leq\ & O(\bar{\rho}^t)\|\dot{H}_{t,i}\|\|\dot{H}_{t,j}\|\frac{1}{\gamma^2} + [\|\dot{H}_{t,j}\| + \|\dot{\tilde{H}}_{t,i}\|\|\dot{H}_{t,j}\|]O(\bar{\rho}^t) + \|\dot{\tilde{H}}_{t,i}\|O(t\bar{\rho}^t) \\
=\ & [\|\dot{H}_{t,j}\| + \|\dot{\tilde{H}}_{t,i}\|\|\dot{H}_{t,j}\|]O(\bar{\rho}^t) + \|\dot{\tilde{H}}_{t,i}\|O(t\bar{\rho}^t).
\end{aligned}
$$

Similar as Lemma 4.5, we can show that

$$\mathbb{E}\left\|\dot{\tilde{H}}_{t,i}(\theta)\right\|^3 < \infty \quad \text{and} \quad \mathbb{E}\left\|\ddot{\tilde{H}}_{t,ij}(\theta)\right\|^2 < \infty.$$

Therefore, by (4.16),

$$\begin{aligned}
&\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta^T} - \frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta\partial\theta^T}\right\| \\
\leq\ & [\|\ddot{H}_{t,ij}\| + \|\dot{H}_{t,j}\| + \|\dot{\tilde{H}}_{t,i}\|\|\dot{H}_{t,j}\|]O(\bar{\rho}^t) + \|\dot{\tilde{H}}_{t,i}\|O(t\bar{\rho}^t) + O(t^2\bar{\rho}^t) \\
& +\|y_t\|^2\left\{[\|\dot{H}_{t,i}\|\|\dot{H}_{t,j}\| + \|\dot{\tilde{H}}_{t,i}\|\|\dot{H}_{t,j}\| + \|\ddot{H}_{t,ij}\| + \|\dot{H}_{t,j}\|]O(\bar{\rho}^t)\right. \\
& \left. +[\|\dot{\tilde{H}}_{t,i}\| + \|\dot{H}_{t,j}\|]O(t\bar{\rho}^t) + O(t^2\bar{\rho}^t)\right\}.
\end{aligned}$$

Applying the Hölder's inequality yields

$$\mathbb{E}\|\dot{H}_{t,i}\|\|\dot{H}_{t,j}\| \leq [\mathbb{E}\|\dot{H}_{t,i}\|^2]^{1/2}[\mathbb{E}\|\dot{H}_{t,j}\|^2]^{1/2} < \infty,$$

$$\mathbb{E}\|y_t\|^2\|\dot{H}_{t,i}\|\|\dot{H}_{t,j}\| \leq [\mathbb{E}\|y_t\|^6]^{1/3}[\mathbb{E}\|\dot{H}_{t,i}\|^3]^{1/3}[\mathbb{E}\|\dot{H}_{t,j}\|^3]^{1/3} < \infty,$$

$$\mathbb{E}\|y_t\|^2\|\dot{H}_{t,i}\| \leq [\mathbb{E}\|\|y_t\|^4]^{1/2}[\mathbb{E}\|\dot{H}_{t,i}\|^2]^{1/2} < \infty,$$

$$\mathbb{E}\|y_t\|^2\|\ddot{H}_{t,ij}\| \leq [\mathbb{E}\|\|y_t\|^4]^{1/2}[\mathbb{E}\|\ddot{H}_{t,ij}\|^2]^{1/2} < \infty.$$

The terms with tilde have the similar results. For any $\epsilon > 0$, by the Markov inequality,

$$
\begin{aligned}
& \mathbb{P}\left(\sup_{\theta\in\nu(\theta_0)}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta^T}-\frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta\partial\theta^T}\right\|>\epsilon\right)\\
\leq\ & \mathbb{P}\left(\sup_{\theta\in\nu(\theta_0)}\frac{1}{n}\sum_{t=1}^{n}\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta^T}-\frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta\partial\theta^T}\right\|>\epsilon\right)\\
\leq\ & \sup_{\theta\in\nu(\theta_0)}\frac{\frac{1}{n}\sum_{t=M}^{n}\mathbb{E}\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta^T}-\frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta\partial\theta^T}\right\|}{\epsilon}+o(1)\\
\leq\ & \frac{1}{\epsilon n}\sum_{t=M}^{n}\mathbb{E}\left\{\left[\|\ddot{H}_{t,ij}\|+\|\dot{H}_{t,j}\|+\|\dot{\tilde{H}}_{t,i}\|\|\dot{H}_{t,j}\|\right]O(\bar{\rho}^t)\right.\\
& \quad \left.+\|\dot{\tilde{H}}_{t,i}\|O(t\bar{\rho}^t)+O(t^2\bar{\rho}^t)\right\}\\
& +\mathbb{E}\left(\|y_t\|^2\left\{\left[\|\dot{H}_{t,i}\|\|\dot{H}_{t,j}\|+\|\dot{\tilde{H}}_{t,i}\|\|\dot{H}_{t,j}\|+\|\ddot{H}_{t,ij}\|+\|\dot{H}_{t,j}\|\right]O(\bar{\rho}^t)\right.\right.\\
& \quad \left.\left.+[\|\dot{\tilde{H}}_{t,i}\|+\|\dot{H}_{t,j}\|]O(t\bar{\rho}^t)+O(t^2\bar{\rho}^t)\right\}\right)\\
\leq\ & \frac{1}{\epsilon n}\sum_{t=M}^{n}O(t^2\bar{\rho}^t)\to 0.
\end{aligned}
$$

This finalizes our proof of this lemma.

$\square$

# Chapter 5

# Numeric Examples

## 5.1 Introduction

This chapter consists of two sections besides the introduction. Section 5.2 gives a set of model parameters and verify that they satisfy the ergodicity and identifiability assumptions in Chapters 2 and 3. In particular, we show how to calculate $\gamma_m(\Delta)$ using Monte Carlo simulation. It is difficult to verify whether all $\theta \in \Theta$ satisfy our assumptions. Instead we only verify that the true parameter $\theta_0$ satisfies our assumption and thus we do not verify the compactness assumption. When estimating GARCH parameters in S+ FinMetrics using the normal estimating function, the estimates are consistent. But the standard errors are not calculated properly. If we use non-Gaussian estimating functions, we have to scale the estimates to make them consistent. Section 5.3 addresses this scaling issue and provides corrections in R. Details about S+ FinMetrics can be found in Zivot and Wang [2006].

## 5.2 A Multivariate GARCH (1,1) Model Which Satisfies the Ergodicity and Identifiability Assumptions

Consider a bivariate GARCH(1, 1) model. Here, $d = 2$, $N = 3$, $p = q = 1$. We assume the innovations are Gaussian. Let the true parameters be

$$c = \begin{pmatrix} 0.03 \\ 0.01 \\ 0.04 \end{pmatrix} \quad A = \begin{pmatrix} 0.06 & 0 & 0 \\ 0 & 0.02 & 0 \\ 0 & 0 & 0.07 \end{pmatrix} \quad B = \begin{pmatrix} 0.009 & 0 & 0 \\ 0 & 0.005 & 0 \\ 0 & 0 & 0.01 \end{pmatrix}.$$

We will verify that this model satisfies our ergodicity and identifiability assumptions, i.e., Theorem 2.3 and Theorem 3.5. We will also show that the estimator is consistent and asymptotically normal by simulation. All computations are done in S+ FinMetrics and the codes are available from the author upon request.

### 5.2.1 Ergodicity

The eigenvalues of the matrix $J$ (defined in (2.1)) are

```
> eigen(J)$values
```

```
[1]  0.2696224 -0.2596224  0.2494903 -0.2404903  0.1439435 -0.1389435
```

Thus, $\rho(J) = 0.2696 < 1$ and Assumption A3 is fulfilled. Next, in order to verify Assumption A4, we will show that $\gamma_2(\Delta) < 0$ by Monte Carlo simulation. Since $p = q = 1$, we have $Y_1 = h_1 = (h_{1,1}, h_{1,2}, h_{1,3})$ and $Y_2 = h_2 = (h_{2,1}, h_{2,2}, h_{2,3})$.

*Remarks.* Our intuition says $\gamma_m(\Delta)$ decreases as $m$ increases. While $\gamma_1$ may be negative, we have decided to calculate $\gamma_2$. Since below we see it is negative, it is sufficient for our purposes. One may also have used for example $\gamma_4$, but the supremum in the integrand will be more complicated to approximate, hence we have decided to calculate $\gamma_2(\Delta)$.

The approximation of $\gamma_2(\Delta)$ involves the following two major steps.

1. We use the sample mean to approximate the expectation. Particularly, during each replication, we simulate a normal random vector $\xi^T = (\xi_1^T, \xi_2^T)^T$, i.e., four independent standard normal random numbers. Then for each simulated $\xi$, we compute the supremum using the procedures in the next step. We replicate this for $M = 500$ times and use the average to approximate the expectation.

2. For each simulated $\xi$, we discretize the domain of $h$ and consider all the possible values of $h$ to obtain the supremum. By the definition of $\Delta$ in Section 2.2, $\Delta$ is invariant on the scale of $h$. That is, if we change $H_t$ to $C \odot H_t$, where $C$ is a $d \times d$ constant matrix, the value of $\Delta$ remains the same.

Let $h = (h_1, h_2, h_3)^T$. Without loss of generality, we can only consider the values such that $h$ is on the unit ball.

(a) We put 100 equally spaced points on the interval $[0, 1]$. Hence, there are $10,000$ possible combinations for $(h_1, h_3)$.

(b) We need to eleminate those possibilities where $h_1^2 + h_3^2 > 1$. These points are beyond the unit ball no matter what value $h_2$ takes.

(c) $h_2$ can be calculated by $h_2 = \sqrt{1 - h_1^2 - h_3^2}$.

(d) We need to eleminate those possibilities where $h_1 h_3 \leq h_2^2$. These points invalidate the positivity of $H_t$.

(e) The supremum can be approximated by inserting all the valid combinations of $(h_1, h_2, h_3)$ into the equation and compare the values of the norm.

After (d), there are only 1761 possible combinations of $(h_1, h_2, h_3)$ remaining for consideration. After trying all the possible combinations of $h_1 = (h_{1,1}, h_{1,2}, h_{1,3})$ and $h_2 = (h_{2,1}, h_{2,2}, h_{2,3})$, we can compute that $\gamma_2(\Delta) = -0.148 < 0$. Assumption A4 is satisfied.

### 5.2.2 Identifiability

It is difficult to verify Assumptions B3 and B4 for any arbitrary $\theta$ within the parameter space. But one can easily verify that

$$\mathcal{A}_{\theta_0} = A = \begin{pmatrix} 0.06 & 0 & 0 \\ 0 & 0.02 & 0 \\ 0 & 0 & 0.07 \end{pmatrix}, \quad \mathcal{B}_{\theta_0} = I_3 - B = \begin{pmatrix} 0.991 & 0 & 0 \\ 0 & 0.995 & 0 \\ 0 & 0 & 0.99 \end{pmatrix}$$

are invertible and $\mathcal{A}_{\theta_0}$ and $\mathcal{B}_{\theta_0}$ are coprime. The matrix $[A_q(\theta_0)|B_p(\theta_0)] = [A|B]$ has rank 3. The identifiability assumptions are satisfied.

## 5.3 Scaling Problems When Fitting GARCH Models in S+ FinMetrics

In this section, we focus on the univariate GARCH model (1.3). In the model definition, we assume that the innovations have unit variance. However, in practice, in order to improve the goodness-of-fit, we may wish to use heavy-tailed innovations, which may invalidate the unit variance assumption. For example, a $t(\nu)$ distribution has variance $\nu/(\nu-2)$ for $\nu > 2$, where $\nu$ is the degree of freedom. Hence, we need to scale the innovations in order to fulfill the model assumption, which will lead to the scaling of model parameters.

Suppose that $\tilde{\xi}_t = \xi_t/d$ have unit variance, where $d$ is the scaling parameter.

The new conditional variance is $\tilde{\sigma}_t^2 = d^2\sigma_t^2$ since $y_t = \sigma_t\xi_t = \tilde{\sigma}_t\tilde{\xi}_t$. We multiply $d^2$ on both sides of (1.3),

$$\tilde{\sigma}_t^2 = d^2\sigma_t^2 = d^2c + \sum_{i=1}^{p}(d^2\alpha_i)y_{t-i}^2 + \sum_{j=1}^{q}\beta_j(d^2\sigma_{t-j}^2).$$

The new parameter vector

$$\begin{aligned}\tilde{\theta} &= (\tilde{c}, \tilde{\alpha}_1, \cdots, \tilde{\alpha}_p, \tilde{\beta}_1, \cdots, \tilde{\beta}_q)^T \\ &= (d^2c, d^2\alpha_1, \cdots, d^2\alpha_p, \beta_1, \cdots, \beta_q)^T,\end{aligned}$$

and the GARCH parameter $\beta_j$'s do not need to be scaled. In model fitting, $d$ can be estimated by the standard deviation of the residuals, i.e.,

$$\hat{d}_n = \left(\frac{1}{n-1}\sum_{t=1}^{n}\hat{\xi}_t^2\right)^{1/2}.$$

To demonstrate the scaling issue and provide an algorithm to modify the results given by S+FinMetrics, we simulate GARCH series with different innovations and fit GARCH models using various kernels. We will discuss four cases:

- normal innovations, normal kernel;
- $t$ innovations, normal kernel;
- normal innovations, $t(5)$ kernel;

- $t(6)$ innovations, $t(5)$ kernel.

### 5.3.1 GARCH Series Simulation

The FinMetric function `simulate` can be used to simulate GARCH series. This function can only be used on "garch" or "mgarch" object. We can simulate GARCH series in general using the following algorithm:

1. Choose parameter values $c = c^{(0)}$, $\alpha_i = \alpha_i^{(0)}$ and $\beta_j = \beta_j^{(0)}$.

2. Choose initial values. For example, the initial values can be chosen as (1.8).

3. For $t = 1, \cdots, n$, compute $\sigma_t^2$ using (1.3).

4. Compute $y_t$ by $y_t = \sigma_t \xi_t$, where $\xi_t$'s are i.i.d. standard normal or $t$ random numbers.

Multivariate GARCH models can be simulated analogously. We may wish to remove the first few entries to allow the series to "warm up". In this section, we simulate GARCH$(1, 1)$ with parameters $c = 0$, $\alpha = 0.3$ and $\beta = 0.6$.

### 5.3.2 normal innovations, normal kernel

Table 5.1 shows the result for normal innovations and normal kernel based on 500 replications, where

**mean of $\hat{\alpha}$ or $\hat{\beta}$:** the average of the 500 parameter estimations.

**mean of $se_{\hat{\alpha}}$ or $se_{\hat{\beta}}$:** the average of the 500 standard errors given by S+FinMetrics.

**sd of $\hat{\alpha}$ or $\hat{\beta}$:** the standard deviation of the 500 parameter estimations, which can be treated as the true standard errors of the estimators.

| | | | |
|---|---|---|---|
| mean of $\hat{\alpha}$ | 0.29796 | mean of $\hat{\beta}$ | 0.59841 |
| mean of $se_{\hat{\alpha}}$ | 0.020239 | mean of $se_{\hat{\beta}}$ | 0.023249 |
| sd of $\hat{\alpha}$ | 0.020419 | sd of $\hat{\beta}$ | 0.023385 |

**Table 5.1:** Normal Innovation, Normal Kernel

We can see that if we use the normal kernel to estimate the parameters of GARCH models whose innovation come from the normal distribution, both coefficient estimations and standard errors match the true values. No scaling is needed. Figures 5.1-5.4 are the density plots and the normal QQ-plot of $\hat{\alpha}$ and $\hat{\beta}$, which indicate that they are consistent and asymptotically normal.

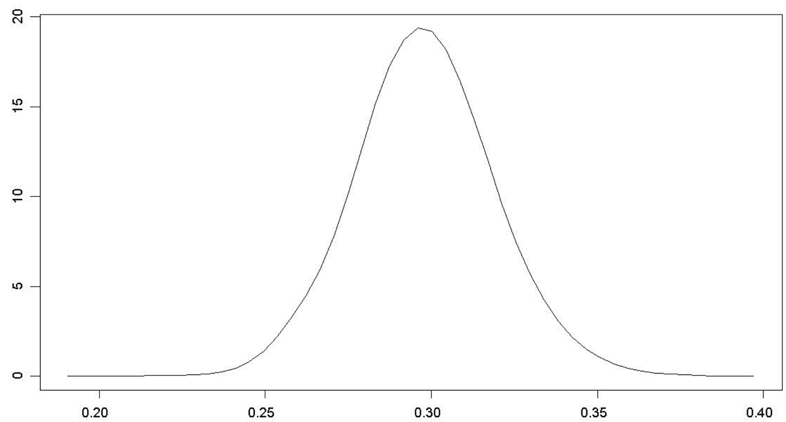


**Figure 5.1:** Density Plot of $\hat{\alpha}$

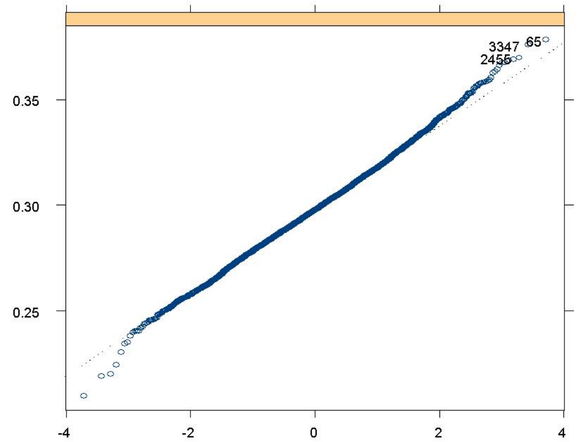


**Figure 5.2:** QQ Plot of $\hat{\alpha}$

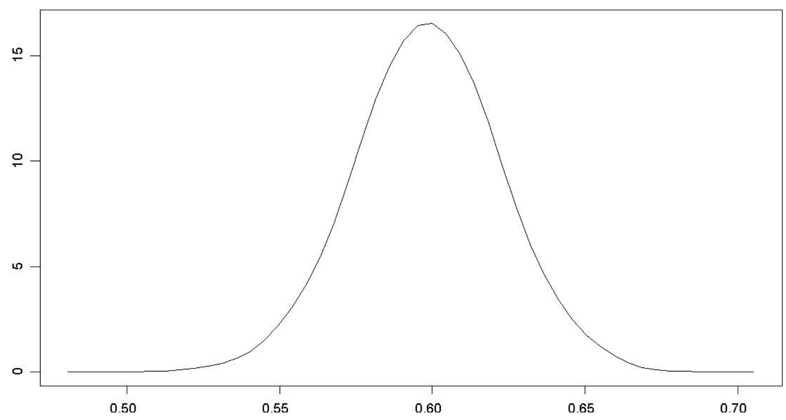


**Figure 5.3:** Density Plot of $\hat{\beta}$

### 5.3.3 $t$ innovations, normal kernel

We generate GARCH series using three different innovations: $t(6)$, $t(12)$, $t(25)$ and estimate the parameters using the normal kernel. Results are collected in Table 5.2.

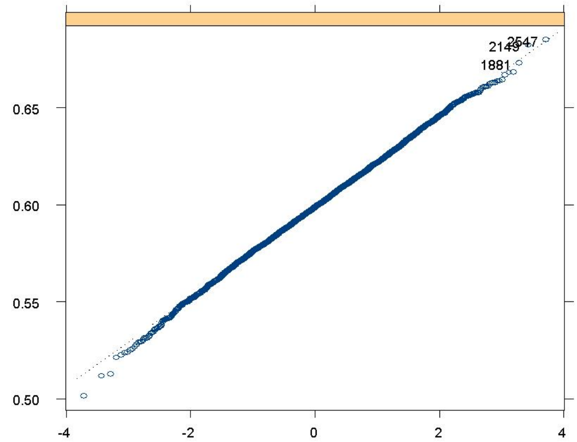


**Figure 5.4:** QQ Plot of $\hat{\beta}$

| | $t(6)$ | $t(12)$ | $t(25)$ |
|---|---|---|---|
| mean of $\hat{\alpha}$ | 0.29821 | 0.30016 | 0.29720 |
| mean of $\text{se}_{\hat{\alpha}}$ | 0.014181 | 0.017637 | 0.018940 |
| sd of $\hat{\alpha}$ | 0.031806 | 0.024784 | 0.022450 |
| mean of $\hat{\beta}$ | 0.59701 | 0.59756 | 0.59786 |
| mean of $\text{se}_{\hat{\beta}}$ | 0.015584 | 0.019632 | 0.021633 |
| sd of $\hat{\beta}$ | 0.033746 | 0.026456 | 0.024833 |

**Table 5.2:** Results of $t$ Innovation, Normal Kernel by S+ FinMetrics

The parameter estimates are still consistent since we are using the normal kernel. However, the standard errors given by FinMetrics are different from the true ones, especially when the kernel is more distinct from normal (e.g., $t(6)$). In order to verify this, we implement the fitting procedures in R using the algorithm in Francq and Zakoïan [2004] to calculate the Hessian matrix. The results from the R program (Table 5.3) are close to the true ones, which means the standard errors given in S+ FinMetrics for this case are inaccurate. The R codes are available from the author upon request. For parameter estimation in R, we use the R function *nlm* to maximize the likelihood function. Replicating this function in R is very computationally intensive. The Rmpi package, developed by Dr H. Yu, allows one to create R programs which run cooperatively in parallel across multiple machines, or multiple CPUs on one machine, to accomplish a goal more quickly than running a single program on one machine.

| | $t(6)$ | $t(12)$ | $t(25)$ |
|---|---|---|---|
| mean of $\hat{\alpha}$ | 0.28972 | 0.29288 | 0.29186 |
| mean of $se_{\hat{\alpha}}$ | 0.028847 | 0.023231 | 0.021311 |
| sd of $\hat{\alpha}$ | 0.035462 | 0.023107 | 0.018762 |
| mean of $\hat{\beta}$ | 0.61057 | 0.60463 | 0.60267 |
| mean of $se_{\hat{\beta}}$ | 0.031317 | 0.026191 | 0.024642 |
| sd of $\hat{\beta}$ | 0.034141 | 0.027624 | 0.026601 |

**Table 5.3:** Results of $t$ Innovation, Normal Kernel by R

Figure 5.5 is the density plot of $\hat{\alpha}$ from different innovations, where the blue, red and yellow lines denote the density of $\hat{\alpha}_1$ from $t(6)$, $t(12)$ and $t(25)$ innovations, respectively.

### 5.3.4 normal innovations, $t(5)$ kernel

From Table 5.4, the estimation of $\alpha$ is no longer consistent. We have to scale the ARCH parameter since we are using a heavy tailed kernel. After each fitting, we multiply the estimation of $\alpha_1$ by the inverse of the variance of $\hat{\eta}_t$. We can see that the estimate is close to the true one after scaling.

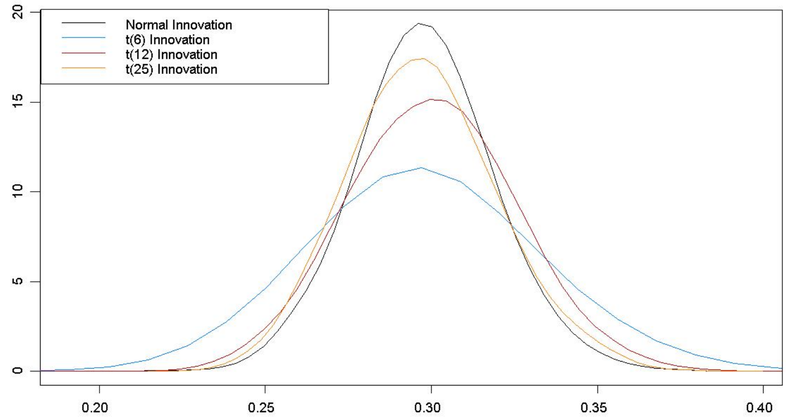


**Figure 5.5:** Density Plot of $\hat{\alpha}$ from Different Innovations

| | | | |
|---|---|---|---|
| mean of $\hat{\alpha}$ | 0.36652 | mean of $\hat{\beta}$ | 0.59727 |
| mean of $\mathrm{se}_{\hat{\alpha}}$ | 0.035078 | mean of $\mathrm{se}_{\hat{\beta}}$ | 0.032773 |
| sd of $\hat{\alpha}$ | 0.027357 | sd of $\hat{\beta}$ | 0.025587 |
| mean of scaled $\hat{\alpha}_1$ | 0.29964 | | |

**Table 5.4:** Normal Innovations, $t(5)$ Kernel

### 5.3.5 $t(6)$ innovations, $t(5)$ kernel

The results are shown in Table 5.5. The scaling parameter is close to one compared with the normal-$t(5)$ case since the two $t$ distributions are close to each other.

| | | | |
|---|---|---|---|
| mean of $\hat{\alpha}$ | 0.31429 | mean of $\hat{\beta}$ | 0.59881 |
| mean of $\mathrm{se}_{\hat{\alpha}}$ | 0.027837 | mean of $\mathrm{se}_{\hat{\beta}}$ | 0.028328 |
| sd of $\hat{\alpha}$ | 0.028835 | sd of $\hat{\beta}$ | 0.027452 |
| mean of scaled $\hat{\alpha}_1$ | 0.29973 | | |

**Table 5.5:** $t(6)$ Innovations, $t(5)$ Kernel

## 5.4 Conclusion and Commentary

In this chapter, we first give an example of a multivariate GARCH parameterization such that it satisfies the ergodicity and identifiability assumptions we gave in previous chapters. In particular, we showed that $\gamma_2(\Delta) < 0$ by Monte Carlo simulation. We then addressed the scaling issue in S+ FinMetrics when estimating GARCH parameters. S+ FinMetrics provides reasonable results when we use normal innovation and normal kernel. For heavy tail innovation and normal kernel, parameter estimation in S+ FinMetrics are acceptable but the algorithm of calculating the standard error is wrong. We have to scale both the parameter estimation and the standard deviation if we use a heavy tail kernel. The variance of the standardized residuals can be used as the scaling parameter.

# Chapter 6

# Concluding Remarks

In this thesis, we prove the asymptotic theory of the QMLE for general multivariate GARCH models under mild conditions. We give some counterexamples for the parameter identifiability result in Jeantheau [1998] and provide a better necessary and sufficient condition. We prove the ergodicity of the conditional variance process on an application of theorems by Meyn and Tweedie [2009]. Under those conditions, the consistency and asymptotic normality of the QMLE are proved by the standard compactness argument and Taylor expansion of the score function. We only require finite sixth moment on the observed sequence. We extend Francq and Zakoïan [2004]'s results and technique from univariate GARCH models to the multivariate case. We generalize the multivariate GARCH$(1,1)$ results in Hafner and Preminger [2009] to multivariate GARCH$(p,q)$. The results in this thesis for the general case covers Comte and Lieberman [2003]'s results for BEKK, and we reduce their moment requirement from eight to six. We also give numeric examples on verifying the assumptions and the scaling issue when estimating GARCH parameters in S+ FinMetrics.

My future work on the multivariate GARCH models includes

1. fitting multivariate GARCH models using real data and studying the efficiency of the estimator;

2. examining and comparing the performance of different types of multivariate GARCH models;

3. developing a better parameter estimation algorithm and an R package; and

4. using multivariate GARCH models on risky assets and derivative pricing.

# Bibliography


L. Bauwens, S. Laurent, and J. Rombouts. Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109, Feb 2006. 9

I. Berkes and L. Horvàth. The rate of consistency of the quasi-maximum likelihood estimator. *Statistics & Probability Letters*, 61(2):133–143, 2003. 5

I. Berkes and L. Horvàth. The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics*, 32(2):633–655, 2004. 5

I. Berkes, L. Horvàth, and P. Kokoszka. GARCH processes: structure and estimation. *Bernoulli*, 9(2):201–227, 2003. 5

P. Billingsley. The Lindeberg-Lévy theorem for martingales. *Proceedings of the American Mathematical Society*, 12(5):788–792, 1961. 61

P. Billingsley. *Probability and Measure.* New York: J. Wiley & Sons,, 3rd edition, 1995. ISBN 0471007102. 58

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986. 3

T. Bollerslev. Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Review of Economics and Statistics*, 72 (3):498–505, 1990. 8

T. Bollerslev, R. Engle, and J. Wooldrige. A capital asset pricing model with time varying covariances. *Journal of Political Economy*, 96(1):116–131, 1998. 6

P. Bougerol and N. Picard. Stationarity of GARCH processes and of some non-negative time series. *Journal of Econometrics*, 52(1-2):115–127, 1992. 5, 10, 15

F. Boussama. *Ergodicité, mélange et estimation sans les modèles GARCH*. PhD thesis, Université Paris 7, 1998. 10, 15, 22, 49

F. Comte and O. Lieberman. Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis*, 84(1):61–84, 2003. 10, 59, 113

R. Engle and F. Kroner. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150, 1995. 7, 8

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, 1982. 3

C. Francq and J. Zakoïan. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4):605–637, 2004. 5, 59, 109, 113

C. Francq and J. Zakoïan. *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons Ltd, 2010. ISBN 978-0-470-68391-0. 7, 9, 60

G. C. Goodwin, S. F. Graebe, and M. E. Salgado. *Control System Design*. Prentice Hall, 2001. ISBN 0139586539, 9780139586538. 119, 121

C. M. Hafner and A. Preminger. On asymptotic theory for multivariate GARCH models. *Journal of Multivariate Analysis*, 100(9):2044–2054, 10 2009. 10, 15, 18, 20, 21, 22, 43, 113

E. J. Hannan. The identification of vector mixed atoregressive-moving average systems. *Biometrika*, 56(1):223–225, Mar 1969. 49

R. A. Horn and C. A. Johnson. *Matrix Analysis*. Press Syndicate of the University of Cambridge, 1985. ISBN 0521305861. 31

T. Jeantheau. Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, 14:70–86, 1998. iii, 10, 13, 46, 49, 50, 51, 113

D. Kristensen. *Uniform Ergodicity of a Class of Markov Chains with Applications to Time Series Models*. Columbia University, 2007. 10

S. Lee and B. E. Hansen. Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory*, 10(1):29–52, 1994. 5

S. Ling and M. McAleer. Asymptotic theory for a new vector ARMA-GARCH model. *Econometric Theory*, 19(2):280–310, 2003. 10

R. L. Lumsdaine. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica*, 64(3):575–596, 1996. 5

H. Lütkepohl. *Handbook of Matrices.* John Wiley & Sons, 1996. ISBN 9780471970156. 122

S. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* Cambridge University Press, $2^{nd}$ edition, 2009. ISBN 9780521731829. iii, 10, 11, 16, 21, 23, 24, 25, 26, 27, 43, 113

D. B. Nelson. Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6(3):318–334, 1990. 5

W. Scherrer and E. Ribarits. On the parameterization of multivariate GARCH models. *Econometric Theory*, 23(3):464–484, 2007. 8

A. A. Weiss. Asymptotic theory for ARCH models: estimation and testing. *Econometric Theory*, 2(1):107–131, 1986. 5

E. Zivot and J. Wang. *Modeling Financial Times Series with S-PLUS*, volume 13. Springer Science+Business Media, Inc., 2nd edition, 2006. ISBN 9780387279657. 101

# Appendix A

# Decomposition of Rational Matrix Polynomials

The following materials in this appendix are from Goodwin et al. [2001].

Let us introduce the a set of definitions related to the factorization of matrix polynomials and then the important Smith-McMillan Lemma:

**Definition A.1** (Rank). *The rank of a polynomial matrix is the rank of the matrix almost everywhere in its argument.*

**Definition A.2** (Elementary Operation). *An elementary operation on a polynomial matrix is one of the following three operations:*

1. *interchange of two rows or two columns;*

2. *multiplication of one row or one column by a constant;*

3. *addition of one row (column) to another row (column) times a polynomial.*

**Definition A.3** (Elementary Matrix)**.** *A left (right) elementary matrix is a matrix such that, when it multiplies from the left (right) a polynomial matrix, then it performs a row (column) elementary operation on the polynomial matrix. All elementary matrices are unimodular.*

**Definition A.4** (Equivalent Matrices)**.** *Two polynomial matrices $U(w)$ and $V(w)$ are equivalent matrices, if there exist sets of left and right elementary matrices, $\{L_1(w), L_2(w), \cdots, L_s(w)\}$ and $\{R_1(w), R_2(w), \cdots, R_t(w)\}$, respectively, such that*

$$U(w) = L_s(w)\cdots L_1(w)V(w)R_1(w)\cdots R_t(w).$$

**Lemma A.5** (Smith-McMillan Lemma)**.** *Let $X(w) = (X_{ij}(w))$ be a $d \times d$ matrix polynomial, where $X_{ij}(w)$'s are rational polynomials:*

$$X(w) = \frac{Y(w)}{K(w)},$$

*where $Y(w)$ is a $d \times d$ matrix polynomial of rank $r$ and $K(w)$ is the least common multiple of the denominators of all elements $X_{ij}(w)$. Then $X(w)$ is equivalent to a matrix $\Pi(w)$, with*

$$\Pi(w) = diag\left\{\frac{\epsilon_1(w)}{\delta_1(w)}, \frac{\epsilon_2(w)}{\delta_2(w)}, \cdots, \frac{\epsilon_r(w)}{\delta_r(w)}, 0, \cdots, 0\right\},$$

*where* $\{\epsilon_i(w), \delta_i(w)\}$ *is a pair of monic and coprime polynomials for* $i = 1, \cdots, r$. *Furthermore,* $\epsilon_i(w)$ *is a factor of* $\epsilon_{i+1}(w)$ *and* $\delta_i(w)$ *is a factor of* $\delta_{i-1}(w)$.

*Proof.* See Goodwin et al. [2001]. □

# Appendix B

# Some Useful Results in Matrix Algebra

The following results are from Lütkepohl [1996].

1. $\mathrm{vec}(ABC) = (C^T \otimes A)\mathrm{vec}(B)$.
2. $X(m \times n)$: $\dfrac{\partial \mathrm{vec}(XX^T)}{\partial \mathrm{vec}(X)^T} = (I_{m^2} + K_{mm})(X \otimes I_m)$.
3. $D_m^+ K_{mm} = D_m^+$.
4. $A(m \times m)$: $|D_m^+(A \otimes A)D_m| = |A|^{m+1}$.
5. $\|A \otimes B\| = \|A\|\|B\|$.
6. $A, B(m \times m)$ positive semidefinite: $\mathrm{tr}(AB) \le \frac{1}{4}(\mathrm{tr}(A) + \mathrm{tr}(B))^2$.
7. $A(m \times m)$ positive definite: $\log|A| \le \mathrm{tr}(A) - m$. The equality holds if and only if $A = I_m$.

8. $\dfrac{\partial |X|}{\partial X} = |X| X^{-1}$.

9. $|X| > 0$: $\dfrac{\partial \log |X|}{\partial X} = (X^T)^{-1}$.

10. $X(m \times m)$ nonsingular: $\dfrac{\partial \mathrm{vech}(X^{-1})}{\partial \mathrm{vech}^T(X)} = -D_m^+(X^{-1} \otimes X^{-1}) D_m$.

11. $(A \otimes B)(C \otimes D) = AC \otimes BD$.

12. $\dfrac{\partial \mathrm{vec}(X^i)}{\partial \mathrm{vec}^T(X)} = \sum_{j=0}^{i-1} (X^T)^{i-1-j} \otimes X^j$, $i = 1, 2, \cdots$.

13. $x(m \times 1)$, $Y(x)(n \times p)$, $Z(x)(q \times r)$:

$$\begin{aligned} & \frac{\partial [\mathrm{vec}(Y) \otimes \mathrm{vec}(Z)]}{\partial x^T} \\ = \quad & (I_p \otimes K_{rn} \otimes I_q) \left[ \frac{\partial \mathrm{vec}(Y)}{\partial x^T} \otimes \mathrm{vec}(Z) + \mathrm{vec}(Y) \otimes \frac{\partial \mathrm{vec}(Z)}{\partial x^T} \right]. \end{aligned}$$

14. $x \in \mathbb{R}$, $A(x)$ nonsingular: $\dfrac{dA(x)^{-1}}{dx} = -A(x)^{-1} \dfrac{dA(x)}{dx} A(x)^{-1}$.

15. $A, B(m \times n)$: $|\mathrm{tr}(AB)| \leq \|A\|_2 \|B\|_2 \leq \min(m,n) \|A\| \|B\|$.

16. $\mathrm{tr}(AB) = \mathrm{tr}(A \otimes B)$.

# Curriculum Vitae

## Name

Weibin Jiang

## Post-secondary Education and Degrees

**September 2007 - Present** Ph.D. candidate, Department of Statistical & Actuarial Sciences, the University of Western Ontario, London, ON, Canada.

**September 2006 - August 2007** Master of Science, Department of Statistical & Actuarial Sciences, the University of Western Ontario, London, ON, Canada.

**September 2001 - June 2005** Honors Bachelor of Science, Department of Financial Mathematics, School of Mathematical Sciences, Peking University, Beijing, China.

## Honors and Awards

**March 2011** Robert and Ruth Lumsden Graduate Awards in Science.

**May 2010** The 2010 Student Research Presentation Award, awarded by Statistical Society of Canada at the 2010 annual meeting in Quebec City, QC.

**September 2006 - Present** Western Graduate Research Scholarship at the University of Western Ontario.

## Related Work Experience

**September 2006 - August 2011** Teaching assistant and research assistant at the University of Western Ontario.

**July 2005 - August 2006** China Continent Property & Casualty Insurance Co. Ltd.