2-13-2018 2:30 PM

Modelling the Common Risk among Equities Using a New Time Series Model

Jingjia Chu, *The University of Western Ontario*

Supervisor: Kulperger, Reg, *The University of Western Ontario*
Joint Supervisor: Yu, Hao, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Statistics and Actuarial Sciences

Abstract

A new additive structure of multivariate GARCH model is proposed where the dynamic changes of the conditional correlation between the stocks are aggregated by the common risk term. The observable sequence is divided into two parts, a common risk term and an individual risk term, both following a GARCH type structure. The conditional volatility of each stock will be the sum of these two conditional variance terms. All the conditional volatility of the stock can shoot up together because a sudden peak of the common volatility is a sign of the system shock.

We provide sufficient conditions for strict stationarity and ergodicity of the model. The ergodicity of the model cannot be studied in the standard way because of the nonlinearity. After reforming the original mathematical representation of the model into a complicated Markovian structure, the systematic theory for Markov chain from Meyn and Tweedie (2009) is applied.

All the parameters in the model are identifiable in terms of the second conditional moments under mild assumptions. Then there exists a unique solution of parameters in the domain which maximizes the likelihood function for a sufficiently large sample size. The choice of starting values is unimportant within the parameter space defined by the ergodicity theorem. Under some general assumptions we proposed, without specifying the distribution of the innovation, different initial values will lead to the same estimates asymptotically. Once both assumptions for ergodicity and identifiability are satisfied, the quasi maximum likelihood (QML) has become a reasonable method to estimate parameters in practice. The sufficient conditions for the strong consistency and asymptotic normality of the QML estimator are proposed.

The Monte Carlo simulation example is included in this thesis to demonstrate how to verify the assumptions in the strict stationarity and asymptotic normality theorems. The numeric issues for the estimating process in practice are addressed with possible solutions.

Keywords: Common risk, Conditional Volatility, Conditional Correlation, Ergodicity, GARCH, Multivariate Time Series, Underlying Driven Process, Asymptotic Normality, Consistency.

Dedicated to my family for their love, support and encouragement.

Acknowlegements

The research included in this dissertation could not have been performed without the assistance and support of many individuals. I would like to extend my gratitude first and foremost to my thesis supervisors, Dr. Reg Kulperger and Dr. Hao Yu, for mentoring me over my Ph.D. research. Their everlasting patience, academic enthusiasm, and immense knowledge sustain the cumulative progress of my research. Their guidance helped me in all the time of my study and work.

I would like to take this opportunity to thank my examiners: Dr. Adam Kolkiewicz, Dr. Stephen Sapp, Dr. Marcos Escobar-Anel and Dr. Rogemar Mamon for their willingness and effort to my defense.

I am grateful for the advice I received from Dr. Duncan Murdoch during my Master's program. I am also indebted to my instructors at the University of Western Ontario, including but not limited to Dr. John Braun, Dr. Ian McLeod, Dr. Wenqing He, Dr. David Stanford and Dr. Mark Reesor. It was them who showed me the path into the kingdom of statistics when I started my graduate study. I also want to convey my appreciation to my mentors of statistical consultation, Dr. David Bellhouse and Dr. Bethany White. Their encouragement and assistance are highly appreciated and will always be remembered.

Besides that, I would also like to extend my deepest gratitude to my parents, grandparents and Tianwei without whose love, support, and understanding I could never have completed this doctoral degree.

Last but not least, I would like to thank all my friends and colleagues, Na Li, Bin Luo, Xin Liu, Xin Wang, Jiang Wu, Yaofei Xiong, Yuzhou Zhang, Heng Xiong, Shen Shan, Chen Yang and others.

Special thanks to two of my non-human friends, Bilbo and Rosie. They have been playing a very important role during my journey at Western.

Contents

Abstract **i**

Dedication **iii**

Acknowlegements **iv**

List of Figures **vii**

List of Tables **ix**

1 Introduction **1**

1.1 Heteroskedasticity Models 3

1.2 Factor models 8

1.3 Model Specification 11

1.4 Statistical Theory of multivariate models 16

1.5 Organization of the Thesis 18

2 Ergodicity and Stationarity **19**

2.1 Introduction 19

2.2 Markovian Process and Nonlinear State Space Model 22

2.3 Ergodicity and Stationarity Theorem 26

2.4 Proof of the Theorem 28

3 Gaussian QMLE and its Asymptotic Theory **38**

3.1 Gaussian Quasi-Maximum Likelihood Estimator 38

3.2 Strong Consistency 41

3.2.1 Proof of Theorem 3.2.2 44

3.2.2 Lemmas 46

3.3 Asymptotic Normality 61

3.3.1 Proof of Theorem 3.3.1 62

3.3.2 Lemmas 72

4 Simulation Study **114**

4.1 Introduction 114

4.2 Monte Carlo Study Preparation 115

4.3 Simulated Results 119

4.4 Numeric Issues with Solutions 129

4.4.1 The Scale Difference 129

4.4.2 Computational Speed 134

4.4.3 Initial values and Starting Point 135

5 Concluding Remarks **140**

Bibliography **142**

A Useful algebra results **149**

B Some Definitions in Markov Chain **152**

C Other Mathematical definitions **154**

Curriculum Vitae **155**

List of Figures

1.1 The original daily closing prices and the log returns 2

1.2 The plot on the left: the solid black line shows the sample density of WFC log return, the blue dashed line shows the corresponding normal density. The plot on the right: the sample autocorrelation of WFC log return. . . 3

1.3 The sample autocorrelation plots of transformed WFC log returns (absolute values on the left and squared values on the right). 4

4.1 Log return of IBM and CSCO from 1995-01-01 to 2007-12-31 116

4.2 The simulated paths: the top three are the simulated $\boldsymbol{x}_t$ (the black solid line represents $x_{1,t}$ and the red dashed line represents $x_{2,t}$) and the three below are the corresponding $\boldsymbol{\sigma}_t$ (the black line represents $\sigma^2_{1,t}$, the red line represents $\sigma^2_{2,t}$ and the blue line represents $\sigma^2_{0,t}$) 121

4.3 The histogram of $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2$ when $K_2 = 1000$. The blue lines represent the true values. 121

4.4 The histogram of $\hat{\rho}, \hat{\omega}_1, \hat{\omega}_2, \hat{\beta}_0$ when $K_2 = 1000$. The blue lines represent the true values. 122

4.5 The histogram of $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2$ when $K_2 = 20000$. The blue lines represent the true values. 122

4.6 The histogram of $\hat{\rho}, \hat{\omega}_1, \hat{\omega}_2, \hat{\beta}_0$ when $K_2 = 20000$. The blue lines represent the true values. 123

4.7 qqnorm of rescaled $\hat{\rho_{1,2}}$ with different sample sizes K_2 129

4.8 qqnorm of rescaled $\hat{\omega}_1$ with different sample sizes K_2 130

4.9 qqnorm of rescaled $\hat{\beta}_1$ with different sample sizes K_2 130

4.10 qqnorm of rescaled $\hat{\beta}_0$ with different sample sizes K_2 131

4.11 The black solid line: kernel density of rescaled $\hat{\rho}_{1,2}$ with different sample sizes K_2. The blue dashed line: standard normal using the same mean and sd from the estimates 131

4.12 The black solid line: kernel density of rescaled $\hat{\omega}_1$ with different sample sizes K_2. The blue dashed line: standard normal using the same mean and sd from the estimates 132

4.13 The black solid line: kernel density of rescaled $\hat{\beta}_1$ with different sample sizes K_2. The blue dashed line: standard normal using the same mean and sd from the estimates 132

4.14 The black solid line: kernel density of rescaled $\hat{\beta}_0$ with different sample sizes K_2. The blue dashed line: standard normal using the same mean and sd from the estimates 133

4.15 Violin plot for 1000 computation time using the same target function written in R and Rcpp 136

List of Tables

4.1 The numeric estimate from IBM and CSCO centered log return and the 'True' value used for the bootstrap simulations (rounded to two decimal digits) 119

4.2 The absolute biases with different sample sizes (rounded to two decimal digits) 125

4.3 The relative absolute biases with different sample size (%) 125

4.4 RMSE of the estimates (rounded to two decimal digits) 126

4.5 SD of $\sqrt{K_2}(\hat{\boldsymbol{\theta}}_{K_2}-\boldsymbol{\theta}_0)$ and estimated asymptotic SD (rounded to two decimal digits) 128

4.6 RASY with different sample sizes (%) 128

4.7 Kurtosis (skewness) of $\sqrt{K_2}(\hat{\boldsymbol{\theta}}_{K_2}-\boldsymbol{\theta}_0)$ (rounded to one decimal digits) . . 133

4.8 Estimates and corresponding values of the negative likelihood from Windows and Linux system: $\hat{\boldsymbol{\theta}}_1$ is the estimate from Windows system and $\hat{\boldsymbol{\theta}}_2$ is the estimate from Windows system. 137

4.9 Estimates and corresponding values of the negative likelihood from Windows and Linux system using $\boldsymbol{\theta}_{start,a}$ and $\boldsymbol{\theta}_{start,b}$. 138

Chapter 1

Introduction

The log returns are commonly used in econometrics for some reasons. The raw prices are restricted to be positive whereas the log returns can be any real numbers. Let $S_0, S_1, \ldots$ be a sequence of daily stock closing prices. Then the log return x_t (or return in the following sections) is defined as

$$x_t = \log(S_t/S_{t-1}) \approx \frac{S_t - S_{t-1}}{S_{t-1}}.$$

The right hand side of the approximation sign is obtained by Taylor expansion. The log returns can be interpreted as continuously compounded returns and the log return values do not depend on monetary units of the original asset prices (see Figure 1.1 and 4.1). Moreover, the weekly or monthly log returns can be easily computed by summing up the daily returns. Most of the observations plotted in Figure 4.1 fall into a relatively narrow range with only few above 5% or below -5%.

As a measure of riskiness in financial securities, it is necessary to estimate the volatility of the log returns instead of the raw prices for the financial modeling. Though the volatility of the log returns does not tell which direction the log return goes, it can tell us how far on average the returns move. It can be used in derivative pricing and risk control.

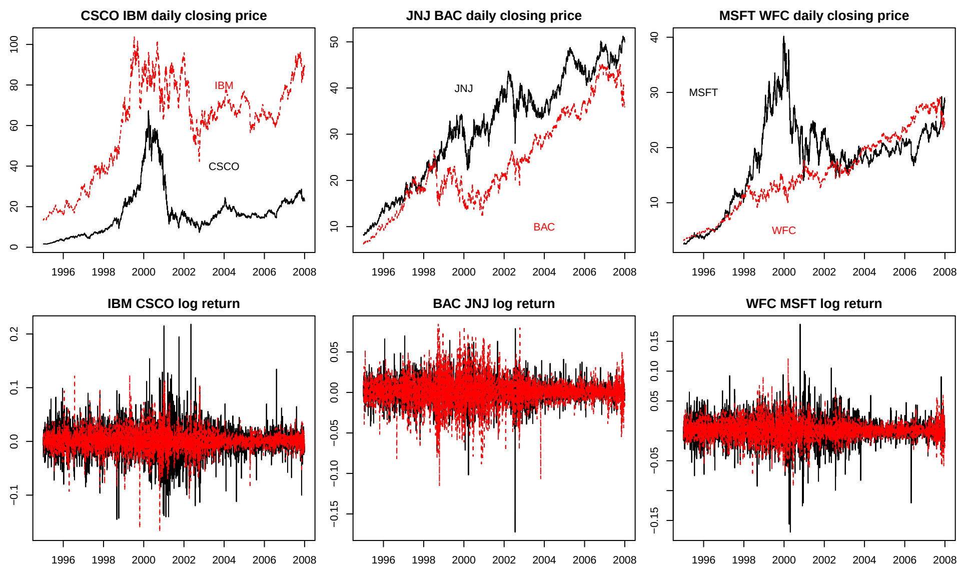

Figure 1.1: The original daily closing prices and the log returns

In early studies of financial models, the log returns are assumed to be independent and identically distributed with a mean and a variance which remain the same over time. This type of structure is motivated by the Black-Scholes-Merton (BSM) model (Black and Scholes, 1973; Merton, 1973) which is an important framework to derive the option prices. In the BSM model,

$$dS_t/S_t = \mu dt + \sigma dW_t.$$

After discretizing the time interval in this formula, this formula leads to the conclusion that the daily log returns follow an independent and identically normal distribution with mean $(\mu - \frac{1}{2}\sigma^2)$ and variance σ^2. Therefore, the volatility of the log returns can be estimated by the sample standard deviation. If $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$, then

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Such an assumption does not hold in practice for all kinds of different reasons. If the log returns are normally distributed, the sample density will be close to the normal density. However, the sample density in Figure 1.2 has a much higher peak and fatter tails

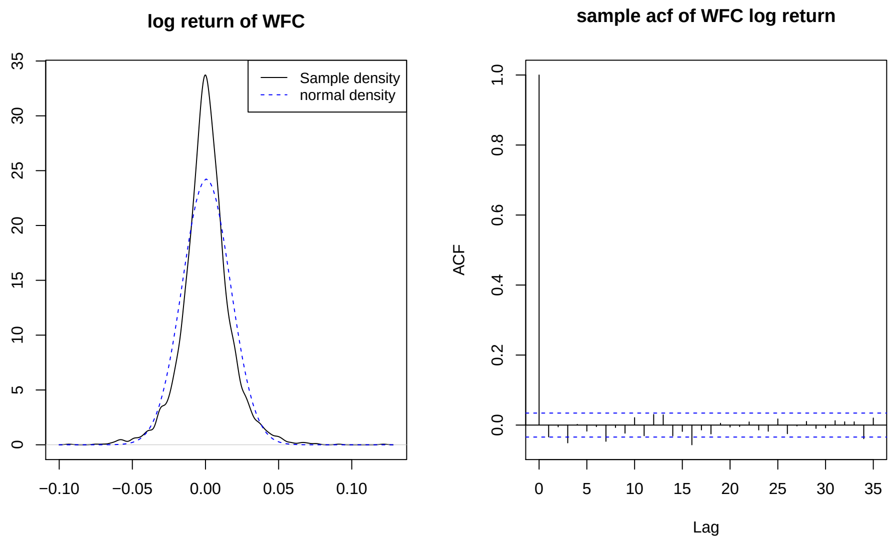

Figure 1.2: The plot on the left: the solid black line shows the sample density of WFC log return, the blue dashed line shows the corresponding normal density. The plot on the right: the sample autocorrelation of WFC log return.

compared to the corresponding theoretical normal density. This phenomenon is known as leptokurtosis. Another noticeable violation is that a large value tends to be followed by another large value, and a small value tends to be followed by another small value, regardless of signs. This characteristic of financial time series is called volatility clustering. One more evidence of such a feature is based on the sample autocorrelation functions. Though the sample autocorrelations of the log return sequence are mostly within the confidence bands around 0, the sample autocorrelations of the transformed sequences, both the absolute values and the squared values, decay to 0 slowly in Figures 1.2 and 1.3. All of these suggest that the second order of the log returns or volatilities is changing dynamically depending on the previous values.

1.1 Heteroskedasticity Models

The conditional heteroskedasticity models have played an important role in financial world today by taking the nature of the financial log return series into consideration.

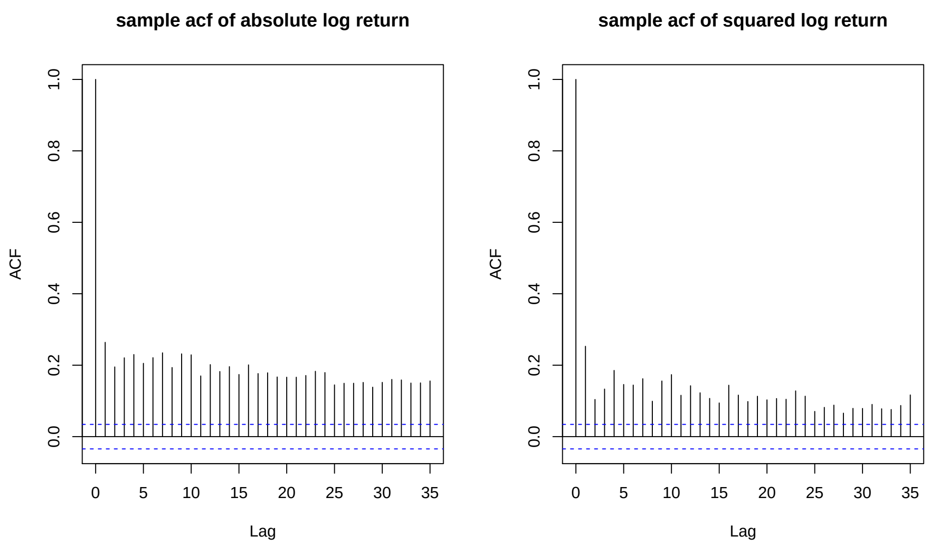

Figure 1.3: The sample autocorrelation plots of transformed WFC log returns (absolute values on the left and squared values on the right).

Assume $\{x_t : t > 0\}$ is the observed process and let $\mathcal{F}_t$ be a set (σ-field) generated by $\{x_t, x_{t-1}, \ldots\}$, then the general form of the conditional heteroskedasticity model is written in a multiplicative structure. The variance of the log return depends on the observations up to one-step before the current time. In mathematical equations,

$$\begin{cases} x_t = \epsilon_t \sigma_t \\ E(x_t | \mathcal{F}_{t-1}) = 0 \\ E(x_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2 \end{cases} \quad . \tag{1.1}$$

The innovations $\{\epsilon_t : t \in T\}$ are i.i.d. random noise with mean 0 and variance 1. Moreover, the innovations are independent of $\mathcal{F}_{t-1}$, and σ_t's are $\mathcal{F}_{t-1}$ adapted.

Engle (1982) introduced the autoregressive conditional heteroskedasticity (ARCH) model with the unique ability of capturing volatility clustering in financial time series at the time. The ARCH(q) model defines the conditional variance of x_t to be

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i x_{t-i}^2 .$$

However, the selected lag q tends to be large when the model is applied to real market data. Subsequently, Bollerslev (1986) extended the formula of σ_t^2 by adding its autoregressive terms, then the number of terms on the right hand side can be notably reduced. The conditional variance of the univariate GARCH(p,q) model is defined as

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i x_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2.$$

If the backshift operator is used in the representation, the univariate GARCH(p,q) model can be written as an ARCH(∞) model. The ARCH(∞) model is

$$\sigma_t^2 = \phi_0 + \sum_{i=1}^{\infty} \phi_i \epsilon_{t-i}^2.$$

The detailed equalities of the coefficients can be found in Francq and Zakoian (2010).

Other extensions of the univariate GARCH model try to characterize the asymmetry effect, which include exponential GARCH model (Nelson, 1991), threshold GARCH model (Zakoian, 1994), double threshold (G)ARCH model (Li and Li, 1996), dynamic asymmetric GARCH model (Caporin and McAleer, 2006). The theories and applications of univariate (G)ARCH type models are well developed, while the multivariate cases are much harder in general.

When there is more than one time series, it becomes necessary to understand the co-movements of the returns. It is well known that the volatilities of stock returns are correlated with each other. In contrast to the univariate cases, the multivariate volatility estimations based on a GARCH dependence are much more flexible. There are two possible ways to build a parametric model in the multivariate GARCH models. One is to model the conditional second moment directly and the other one is to model the conditional correlation along with the marginal conditional variance for each sequence together.

The multivariate GARCH models are specified based on the first two conditional moments as well as the univariate cases. A multivariate volatility model, called half-Vec (*vech*) GARCH model (Bollerslev et al., 1988), is also one of the most general forms of multivariate GARCH models. Let *vech* denote the vector-half operator, which stacks the lower triangular elements of an $m \times m$ matrix as a vector with length $m \times (m+1)/2$. Then

$$\begin{cases} \boldsymbol{x_t} = H_t^{1/2} \boldsymbol{\epsilon_t}, \\ \boldsymbol{h_t} = \boldsymbol{c} + \sum_{i=1}^{q} A_i \boldsymbol{\eta_{t-i}} + \sum_{j=1}^{p} B_j \boldsymbol{h_{t-j}}, \end{cases} \tag{1.2}$$

where

$$\boldsymbol{h_t} = \mathrm{vech}(H_t),$$

$$\boldsymbol{\eta_{t-i}} = \mathrm{vech}(\boldsymbol{\epsilon_t}\boldsymbol{\epsilon_t}^\intercal),$$

$$\boldsymbol{\epsilon_t} \overset{\text{i.i.d.}}{\sim} (0, I_m),$$

and A_i, B_j are $m \times m$ coefficient matrices.

In this class of models, the conditional covariance matrix is modeled directly. The number of parameters in the general m-dimensional case is $\left\{(p+q)\left[\frac{m(m+1)}{2}\right]^2 + \frac{m(m+1)}{2}\right\}$. It increases at a rate proportional to m^4, which makes it difficult to get the estimations.

Another famous class of the multivariate GARCH models built on H_t is the BEKK model (Bollerslev et al., 1988; Engle and Kroner, 1995). The conditional covariance matrix is considered as

$$H_t = CC^\intercal + \sum_{k=1}^{K}\sum_{i=1}^{q} A_{ik}\boldsymbol{\epsilon_t}\boldsymbol{\epsilon_t}^\intercal A_{ik}^\intercal + \sum_{k=1}^{K}\sum_{i=1}^{p} B_{ik} H_{t-i} B_{ik}^\intercal$$

where C, A_{ik} and B_{ik} are m by m matrices. C is a triangular matrix, A_{ik} and B_{ik} are not necessarily symmetric. The number of parameters is $(p+q)Km^2 + \frac{m(m+1)}{2}$, which is much smaller than the *Vech* version.

There are simpler ways of specifying H_t by using the method in the second category mentioned above. The constant correlation coefficient (CCC) GARCH model is presented by Bollerslev (1990), who assumes that the conditional correlation matrix $\boldsymbol{R}$ is time-invariant, where

$$R = \begin{pmatrix} 1 & \rho_{1,2} & \dots & \rho_{1,m} \\ \rho_{1,2} & 1 & \dots & \rho_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,m} & \rho_{2,m} & \dots & 1 \end{pmatrix}_{m \times m}$$

The number of parameters is reduced to $O(m^2)$ from $O(m^4)$ in the *Vech* GARCH model. The model is defined as

$$\begin{cases} \boldsymbol{x_t} = H_t^{1/2} \boldsymbol{\epsilon_t}, \\ H_t = S_t R S_t, \\ \boldsymbol{\Delta_t} = c + \sum_{i=1}^{q} A_i \boldsymbol{x_{t-i}}^2 + \sum_{j=1}^{p} B_j \boldsymbol{\Delta_{t-j}}, \end{cases} \tag{1.3}$$

where $\boldsymbol{\Delta_t}$ is a $\boldsymbol{m}$ dimensional vector of diagonal elements of the conditional covariance matrix $\boldsymbol{H_t}$, $\boldsymbol{S_t}$ is the diagonal matrix of the elements in $\sqrt{\Delta_t}$, and the square vector $\boldsymbol{x_{t-i}}^2$ is $(x_{1,t-i}^2, \cdots, x_{m,t-i}^2)^\intercal$.

A less restrictive time-variant conditional correlation version, called the dynamic correlation coefficient (DCC) GARCH, is studied by Engle (2002), Tse and Tsui (2002). The conditional correlation is changed to be dynamic in the structure of $\boldsymbol{H_t}$.

$$H_t = S_t R_t S_t$$

where the elements in $\boldsymbol{R_t}$, $\rho_{ij,t} = \dfrac{q_{ij,t}}{\sqrt{q_{ii,t} q_{jj,t}}}$. The terms in both the denominator and the numerator can be written as a weighted average of their past values and the element in matrix $\boldsymbol{\epsilon_t \epsilon_t}^\intercal$, may or may not with a constant. In matrix form,

$$Q_t = (1 - \lambda)(\boldsymbol{\epsilon_t \epsilon_t}^\intercal) + \lambda Q_{t-1}$$

or

$$Q_t = S(1-\alpha-\beta) + \alpha(\boldsymbol{\epsilon_t}\boldsymbol{\epsilon_t}^\intercal) + \beta Q_{t-1}.$$

Both CCC-GARCH and DCC-GARCH models are built by modelling the conditional variance of each series and the conditional correlation between series.

There are additional extensions to the multivariate GARCH models as described above, e.g. the generalized orthogonal GARCH (Van der Weide, 2002) and the vector ARMA-GARCH model (Ling and McAleer, 2003).

1.2 Factor models

The strong positive association between the equity variance and several explanatory variables is confirmed by Christie (1982). The volatilities of equities are driven by the same underlying process which is related to some variables besides the returns. A successful class of the multivariate models is the capital asset model and its extension, factor models. The asset pricing model (Treynor, 1962, 1961; Sharpe, 1964) has been introduced by economists by comparing the sensitivity, β's, of the series with the overall market risk. Later, Fama and French expanded the variables in the asset returns model to a three factor model (Fama and French, 1993) and a five factor model (Fama and French, 2015). In the earliest setup, there is only one factor which is the market return. The model is

$$\mathbb{E}x_i = x_f + \beta_i(\mathbb{E}x_m - x_f)$$

where x_i is the return of asset i, x_f is the risk-free rate of interest and β_i is the sensitivity of the expected excess asset returns to the expected excess market returns. In such a setup, the correlation between two expected returns, $\rho_{i,j} = \beta_i\beta_j$, is a constant over time. These models only consider the relative risk between the individual series and the general market performance. They treat the market index or some overall market variables as the common risk factors which do not take the dynamic change along with time into consideration.

One way to improve the model is to change the static factors into hidden dynamic factors. The generalized factor model (Forni et al., 2000) assumes the individual log return is a linear combination of K factors with an idiosyncratic risk,

$$x_{i,t} = \sum_{j=1}^{K} \beta_j f_{t,j} + \eta_{i,t}, \quad i = 1, 2, \ldots, m.$$

or in a matrix form

$$\boldsymbol{x_t} = B\boldsymbol{f_t} + \boldsymbol{\eta_t}$$

where B is a loading matrix with m rows and K columns, the idiosyncratic risk are correlated with a covariance matrix $\boldsymbol{\Omega}$. The factors have the following conditional specification

$$\mathbb{E}_{t-1}(\boldsymbol{f_t}) = \boldsymbol{0},$$

$$\mathbb{E}_{t-1}(\boldsymbol{f_t}\boldsymbol{f_t}^\intercal) = \Lambda_t$$

where Λ_t is a positive definite matrix.

The conditional covariance matrix of $\boldsymbol{x_t}$ is

$$H_t = B\Lambda_t B^\intercal + \Omega.$$

The identifiability in the dynamic factor models is a problem since any full rank square matrix $\boldsymbol{T}$ can be used to premultiply the factor $\boldsymbol{f_t}$, then the conditional second moment remains the same which means

$$H_t = (BT)(T^{-1}\Lambda_t(T^\intercal)^{-1})(BT)^\intercal.$$

Therefore, the solution of the parameters is not unique based on the conditional second moment. The number of parameters increases at a rate proportional to m^2 because the number of elements in B is $m(m+1)/2$.

The factor GARCH model (Vrontos et al., 2003) is a special case of the factor model with $\Omega = 0$ and the number of factors equals to the dimensions of the observed process, $K = m$. Moreover, Λ_t is assumed to be a diagonal matrix with GARCH specified diagonal elements $\sigma^2_{i,t}$ where

$$\sigma^2_{i,t} = \omega_i + \alpha_i f^2_{i,t-1} + \beta_i \sigma^2_{i,t-1}, \quad i = 1, \ldots, m.$$

Hafner and Preminger (2009a) solve the identification problem by putting constraints on the constant terms ω_i's in the GARCH structure such that all of them are set to be one in the equation above.

Though the dynamic factor model has many good features, it has some problems in practice. In the general dynamic factor model settings, the number of factors K need to be specified at the beginning (see Hallin and Liška, 2007). As in the principal component analysis, the factors may lack practical interpretations.

Another way to improve the static factor model is to build the model with economic reasoning. The risk can be divided in to systematic and idiosyncratic risk factors with an additive structure. Using the idea from Vasicek (1987),

$$x_{i,t} = \rho_i \epsilon_{0,t} + \sqrt{1 - \rho_i^2}\epsilon_{i,t}$$

where $\epsilon_{0,t}$ is the systematic risk and $\epsilon_{i,t}$'s are the idiosyncratic risks. The systematic risk and each of the idiosyncratic risks are independent of each other. In this setting, the correlation between assets i and j remains constant over time, which is $\rho_i \rho_j$. Berd et al. (2007) modified the model into a one-factor (G)ARCH model that

$$x_{i,t} = br_{0,t} + \sigma\epsilon_{i,t}$$

where $b \geq 0$ and $r_{0,t}$ is the market risk factor (systematic risk). The market risk factor $r_{0,t}$ has a conditional distribution with mean 0 and variance $\sigma^2_{0,t}$, where $\sigma^2_{0,t}$ has the structure of a univariate GARCH conditional variance. The model has been introduced without any statistical property.

Generally speaking, the factor models with loading matrices have identifiability problem. If the conditional distribution of the factors remains the same over time (e.g. the constant conditional mean and variance), then the loadings can be simply determined by the unique solution of a linear regression. However, if the conditional distribution of the factors is changing dynamically, the number of solutions of the loading matrix can be infinity without additional constraints.

1.3 Model Specification

With information flowing around the world instantaneously, most markets (Asian, European, and American) will react to the same events (good news or bad news). Currently, most stock prices will go up or down together following big events (random shocks). Carr and Wu (2009) found that a common stochastic variance risk factor exists among the stocks by using the market option premiums. We want to introduce a simple common risk model which keeps the GARCH structure and involves the stock returns only. We propose a new additive GARCH type model by using a common risk term to characterize the internal relationship among series explicitly. The common risk term could be used as an indicator of the shock among series. The conditional correlations aggregated by this common risk term are changing dynamically.

The univariate GARCH model has a huge success in financial practice, while most of the multivariate GARCH models extensions do not have a simple way to capture the common risk among different stocks. The goal of this section is to propose a model which defines the common conditional variance term directly. The model will have a structure similar to the univariate GARCH model, and it will also have characteristics similar to the univariate GARCH model. The idea is borrowed from the factor models to use an additive structure. Not like the class of DCC GARCH models, asymptotic theorems can be provided in this new setting.

Some empirical studies show that GARCH model with $p = q = 1$ is the most commonly used one in applied econometrics, so we only define the model as an extension of GARCH(1,1) process here. The model could be easily generalized based on GARCH(p,q) case, although the statistical study of the generalized model will be much harder.

Consider an $\mathbb{R}^m$-valued stochastic process $\{\boldsymbol{x_t}, t \in \mathbb{Z}\}$ on a probability space $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{P}})$ and a multidimensional parameter $\boldsymbol{\theta}$ in the parameter space $\boldsymbol{\Theta}$ where

$$\boldsymbol{\theta} = (\rho_{1,2}, \cdots, \rho_{m-1,m}, \omega_1, \cdots, \omega_m, \alpha_1, \cdots, \alpha_m, \beta_1, \cdots, \beta_m, \beta_{01}, \cdots, \beta_{0m})^\intercal,$$

which belongs to a parameter space of the form

$$\boldsymbol{\Theta} \subset [-1, 1]^{\frac{m(m-1)}{2}} \times [0, \infty)^{4m}.$$

Denote a $m + 1$ dimensional real valued innovation process by $\{\boldsymbol{\epsilon_t}, t \in \mathbb{Z}\}$, and the information set (σ-field) denote the information set available at time t by $\mathcal{F}_t = \sigma(\boldsymbol{\epsilon_t}, \boldsymbol{\epsilon_{t-1}}, \ldots)$. Assume that the innovations are independent and identically distributed with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$ which is also a correlation matrix,

$$\Sigma = \begin{pmatrix} R_{m\times m} & \mathbf{0}_{m\times 1} \\ \mathbf{0}_{1\times m} & 1 \end{pmatrix}$$

where $\boldsymbol{R}$ is the same matrix defined above. The elements in the $\boldsymbol{R}$ matrix represent the internal connections between shocks. When $\boldsymbol{R}$ is an identity matrix, the model becomes a two factor model with the loading coefficients set to be 1.

The innovation at each time t can be divided into two parts as $\boldsymbol{\epsilon_t}^\intercal = (\boldsymbol{\epsilon}^\intercal_{\boldsymbol{t,ind}}, \epsilon_{0,t})$ where

$\boldsymbol{\epsilon_{t,ind}} = (\epsilon_{1,t}, \epsilon_{2,t}, \ldots, \epsilon_{m,t})^\intercal$. The first part is an m-dimensional vector of correlated individual shocks $\boldsymbol{\epsilon_{t,ind}}$ and the second part is a univariate independent common shock term $\epsilon_{0,t}$.

We say that $\boldsymbol{x_t}$ is a common risk model with an additive GARCH structure if, for all $t \in \mathbb{Z}$, we have

$$\begin{cases} x_{1,t} = \epsilon_{1,t}\sigma_{1,t} + \epsilon_{0,t}\sigma_{0,t} \\ x_{2,t} = \epsilon_{2,t}\sigma_{2,t} + \epsilon_{0,t}\sigma_{0,t} \\ \qquad \cdots\cdots \\ x_{m,t} = \epsilon_{m,t}\sigma_{m,t} + \epsilon_{0,t}\sigma_{0,t} \end{cases} \tag{1.4}$$

where $\sigma_{1,t}, \ldots, \sigma_{m,t}$ are following a GARCH structure and $\sigma_{0,t}$ is related to all of them,

$$\begin{cases} \sigma^2_{1,t} = \omega_1 + \alpha_1 x^2_{1,t-1} + \beta_1 \sigma^2_{1,t-1} \\ \sigma^2_{2,t} = \omega_2 + \alpha_2 x^2_{2,t-1} + \beta_2 \sigma^2_{2,t-1} \\ \qquad \cdots\cdots \\ \sigma^2_{m,t} = \omega_m + \alpha_m x^2_{m,t-1} + \beta_m \sigma^2_{m,t-1} \\ \sigma^2_{0,t} = \beta_{01}\sigma^2_{1,t} + \cdots + \beta_{0m}\sigma^2_{m,t}. \end{cases} \tag{1.5}$$

Introduce the following notations,

$$D_t = diag\{\sigma_{1,t}, \sigma_{2,t} \cdots, \sigma_{m,t}\}, \mathbf{1} = (1, 1, \cdots, 1)^\intercal.$$

Then (1.4) could be written in a matrix form:

$$\boldsymbol{x_t} = D_t \boldsymbol{\epsilon_{t,ind}} + \sigma_{0,t}\epsilon_{0,t}\mathbf{1}. \tag{1.6}$$

So the model could either be specified by (1.5) and (1.6) together or (1.4) and (1.5) together.

The main idea of this setup has been published in Chu et al. (2016) but in a slightly different mathematical form. The identifiability theorem in Chapter 3 of this thesis was also included in that proceeding manuscript without proof, whereas this thesis gives additional details. Aside from the reason to obtain the parameter identifiability, the structure of the covariance matrix Σ has its own interpretation. For a bivariate case, this model allows two stock returns to be correlated through the individual innovations even in the absence of the common shock. An independent external shock is applied to the two stocks in the same way which drives the returns simultaneously.

The number of parameters is increasing at the rate $O(m^2)$ which is similar to the CCC-GARCH model. We could partition the vector of unknown parameters into two parts: the parameters in the innovations correlation matrix Σ and the coefficients in (1.5). The number of total parameters is $s = s_1 + 3m + 1$, where $s_1 = \dfrac{m(m-1)}{2}$ is the number of parameters in $\boldsymbol{R}$. There is no redundancy in defining the dependency among the returns through the matrix Σ and the common factor $\sigma^2_{0,t}$ in the model. The process is the same as a CCC-GARCH(1,1) process when $\beta_{01} = \beta_{02} = \cdots = \beta_{0m} = 0$. Therefore, CCC-GARCH(1,1) is a special case of this process. Note that for $\rho_{1,2} = \cdots = \rho_{m-1,m} = 0$, the process further reduces to m independent univariate GARCH(1,1) series.

The conditional covariance matrix of $\boldsymbol{x}_t$, $H_t = \mathrm{cov}(\boldsymbol{x}_t|\mathcal{F}_{t-1})$, can be computed from the definition,

$$H_t = \begin{pmatrix} \sigma^2_{0,t} + \sigma^2_{1,t} & \sigma^2_{0,t} + \rho_{1,2}\sigma_{1,t}\sigma_{2,t} & \dots & \sigma^2_{0,t} + \rho_{1,m}\sigma_{1,t}\sigma_{m,t} \\ \sigma^2_{0,t} + \rho_{1,2}\sigma_{1,t}\sigma_{2,t} & \sigma^2_{0,t} + \sigma^2_{2,t} & \dots & \sigma^2_{0,t} + \rho_{2,m}\sigma_{2,t}\sigma_{m,t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2_{0,t} + \rho_{1,m}\sigma_{1,t}\sigma_{m,t} & \sigma^2_{0,t} + \rho_{2,m}\sigma_{2,t}\sigma_{m,t} & \dots & \sigma^2_{0,t} + \sigma^2_{m,t} \end{pmatrix}.$$

H_t can be written as the sum of two parts: $H_t = \sigma^2_{0,t} J + D_t R D_t$ where

$$J = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}_{m\times m} = \mathbf{1}\mathbf{1}^\top.$$

Each term on the left hand side (1.5) has its own lower bound,

$$\sigma_{1,t}^2 \geq \omega_1, \ldots, \sigma_{m,t}^2 \geq \omega_m \text{ and } \sigma_{0,t}^2 \geq \beta_{01}\omega_1 + \cdots + \beta_{0m}\omega_m.$$

Hence, the conditional variance of each component in $\boldsymbol{x_t}$ at time t has different lower bound as well.

The conditional correlation between series i and series j can be represented by the elements in $\boldsymbol{H}_t$ matrix

$$\begin{aligned}
\rho_{ij,t} &= \frac{\operatorname{cov}(x_{i,t}, x_{j,t})}{\sqrt{\operatorname{var}(x_{i,t}) \operatorname{var}(x_{j,t})}} \\
&= \frac{\sigma_{0,t}^2 + \rho_{i,j}\sigma_{i,t}\sigma_{j,t}}{\sqrt{(\sigma_{0,t}^2 + \sigma_{i,t}^2)(\sigma_{0,t}^2 + \sigma_{j,t}^2)}} \\
&= \frac{1 + \rho_{i,j}\left(\dfrac{\sigma_{i,t}}{\sigma_{0,t}}\right)\left(\dfrac{\sigma_{j,t}}{\sigma_{0,t}}\right)}{\sqrt{1 + \left(\dfrac{\sigma_{i,t}}{\sigma_{0,t}}\right)^2}\sqrt{1 + \left(\dfrac{\sigma_{j,t}}{\sigma_{0,t}}\right)^2}}.
\end{aligned}$$

From the equations above, the conditional correlation matrix tends to be $\boldsymbol{J}$ when the common term $\sigma_{0,t}$ is much larger than both $\sigma_{i,t}$ and $\sigma_{j,t}$. In this case, the common risk term is dominant, and all the log return series are nearly perfectly correlated. On the contrary, the conditional correlation matrix will be approaching the constant correlation matrix $\boldsymbol{R}$ when the common risk term is much smaller than $\sigma_{i,t}$ and $\sigma_{j,t}$ or is really close to 0. Then, the conditional correlation will become time invariant which is the same as a CCC-GARCH model. Mathematically,

$$\boldsymbol{R}_t \to \boldsymbol{J} \text{ when } \sigma_{0,t} \to \infty,$$

$$\boldsymbol{R}_t \to \boldsymbol{R} \text{ when } \sigma_{0,t} \to 0.$$

Based on the specification in (1.5), R_t cannot be neither J or R as $\sigma^2_{0,t}$ is a linear combination of $\sigma^2_{i,t}$ for $i = 1, 2, \cdots, m$. Nonetheless, the expression for $\sigma^2_{0,t}$ can have other possibilities which might involve terms like $\sigma^2_{0,t-1}$, $x^2_{i,t-1}$ or $\sigma^2_{i,t-1}$. This is only a demonstration of the potential of the model with the additive structure in (1.4).

1.4 Statistical Theory of multivariate models

Since the model proposed in the previous section follows a GARCH type structure, the main tool used in this thesis can be borrowed from the GARCH models. The theories in this thesis include the stationarity and ergodicity theorem, the consistency theorem and the asymptotic normality theorem.

The strict stationarity of the univariate GARCH(1,1) model is proved in Klüppelberg et al. (2004) by rewriting the conditional variance $\sigma^2_{i,t}$ as an infinite sum of the past squared innovations $\{\epsilon^2_t\}$. Whether the process is strictly stationary depends on γ where $\gamma = \mathbb{E} \log \alpha\eta^2_t + \beta$. The necessary and sufficient conditions for a univariate GARCH(p, q) model is provided in Nelson and Cao (1992) and Tsai and Chan (2008). In the general univariate GARCH (p,q), the latest p conditional variance can be combined with the last q observed points to construct a linear representation in vector form. The top Lyapunov exponent of the spectral radius of the coefficient matrix in the new form is used to control the strong consistency.

The local consistency and asymptotic normality theorem of the QMLE in the univariate GARCH(1,1) model is provided in Lumsdaine (1996). A local neighbourhood of the true parameter is defined, then the consistency theorem is built inside this neighbourhood under very restrictive assumptions. Lee and Hansen (1994) provide the strong consistency among the whole parameter space in the univariate GARCH(1,1) setting under the strictly stationary and ergodic assumption. The finite moment requirement is only needed with the 4th moment to obtain the asymptotic normality. The consistency and asymptotic normality theorems in a univariate GARCH(p,q) model is very similar to the GARCH(1,1) case which uses the top Lyapunov exponent and the finite 4th moment plus some other conditions (see Berkes et al., 2003; Francq and Zakoïan, 2004).

The theories of the multivariate GARCH models are very complicated. The methodology in univariate cases cannot be extended to multivariate GARCH models. The results of the multivariate models are studied case by case including both the stationary theory and the asymptotic normality theory. Jeantheau (1998) proves the strong consistency theory of the estimator under the strict stationary and the identifiable assumptions. As an example, the detailed conditions to check the strict stationarity and the identifiability for the CCC-GARCH model are included in the paper. Comte and Lieberman (2003) continues the next stage of the asymptotic theory which is the normality of the QMLE with the finite 8th moment. The Markov chain theory in Meyn and Tweedie (2009) is used to prove the stationarity and ergodicity of the BEKK GARCH model (Boussama et al., 2011), the *Vech* GARCH model (see Hafner and Preminger, 2009b for the GARCH(1,1) case, Jiang (2011) for the general GARCH (p,q) case) and the factor GARCH model (Hafner and Preminger, 2009a). For the asymptotic normality, both Hafner and Preminger (2009b) and Jiang (2011) prove it under the finite 6th moment while Hafner and Preminger (2009a) requires only the finite 4th moment with the factor GARCH setting. All of the proof of the asymptotic normality mentioned above follow the framework in Chapter 4 of Amemiya (1985).

Francq and Zakoian (2010) provide a comprehensive summary of some multivariate GARCH models in Chapter 11, including the BEKK, *Vech* and CCC GARCH models. The stationarity theory is proved for *Vech* and CCC GARCH model while the estimation theory with asymptotic results is only provided for CCC GARCH. For the DCC-GARCH model, the theoretically sound statistical inference procedures do not yet exist, as noted in Caporin and McAleer (2006).

1.5 Organization of the Thesis

The notations and the new common underlying risk model have been introduced in Section 1.3 above as well as some properties. The terms on the left hand side of (1.5) can have other specifications. The main idea in this class of models is that both of these parts are determined by the previous values iteratively. The similar theoretical derivation in this thesis can be followed under moderately varying assumptions.

Chapter 2 states the stationarity and ergodicity theorem of our model, which is crucial in time series. The sufficient but not necessary conditions are provided in the theorem. In Chapter 3, the parameter estimation method using Gaussian quasi-maximum likelihood is discussed. The strong consistency and the asymptotic normality of the estimator, two important large sample results, are provided in this chapter. One of the assumptions in this chapter can be substituted by the conditions in the previous chapter. Chapter 4 addresses the results of the Monte Carlo simulation study based on a bivariate example. The example chosen in this chapter satisfies all the assumptions in Chapter 2 and Chapter 3. Furthermore, some numeric issues and possible solutions are discussed in this chapter. The last chapter, Chapter 5, concludes all the results and presents future work.

Chapter 2

Ergodicity and Stationarity

2.1 Introduction

The asymptotic theory of the quasi maximum likelihood estimator needs to be established if the observed time series is stable in some kind of form. Moreover, the reliability of the forecasting highly depends on this kind of stability. If a process is stable, the statistics obtained from the sample could be used to describe the future behavior of the process. In addition, some of the statistical properties of the process will remain the same in the future as the observed past. The type we would like to choose to describe the stability in this multivariate time series model is the stationarity and geometric ergodicity.

The stationarity of a stochastic process has two different forms, the strict (or strong) stationarity and the weak (or second order) stationarity. The strong stationarity is defined in terms of the joint distribution while the weak one is based on the first and second moments. Their definitions can be found in almost any textbook on time series analysis, see Tsay (2010) and Francq and Zakoian (2010) as examples.

Definition 2.1 (Weak Stationarity)

The process $\{Y_t\}$ is weakly stationary if,

1. *the mean of the process does not change over time which means*

$$\mathbb{E}Y_t = \mu \text{ for all } t \in \mathbb{Z},$$

2. *the autocovariance of the process is finite and it does not change when time is shifted*

$$cov(Y_t, Y_{t+l}) = \gamma(l) \text{ for all } l, t \in \mathbb{Z} \text{ and } \gamma(0) < \infty.$$

Definition 2.2 (Strict Stationarity)

The process $\{Y_t\}$ *is said to be strictly stationary if the joint distribution of* $(Y_{t_1}, \cdots, Y_{t_k})$ *is identical to that of* $(Y_{t_1+l}, \cdots, Y_{t_k+l})$ *for all* $l \in \mathbb{Z}$ *, where* k *is an arbitrary positive integer and* $(t_1, ..., t_k)$ *is any collection of* k *integers. It is written as*

$$(Y_{t_1}, \cdots, Y_{t_k}) \stackrel{d}{=} (Y_{t_1+l}, \cdots, Y_{t_k+l}),$$

where $\stackrel{d}{=}$ *means equal in distribution.*

On the one hand, the weak stationarity does not imply strict stationarity since the higher moments of the process may depend on time t, but on the other hand, a strictly stationary process with a finite second moment is weakly stationary. A measurable function of a strictly stationary variable is still strictly stationary, but this is not true for the weakly stationary variables. In this section, we will prove the strict stationarity of the process in Section 1.3.

The definition of an ergodic process is very technical. Intuitively, if a process is ergodic, the initial values will be irrelevant in the long run. The ergodicity is the requisite to apply the ergodic theorem, which is a law of large numbers of the stochastic process. By the ergodic theorem, the average of time series converges to the same limit as the ensemble average when the sample size gets large, and this limit could be considered as the center of the process, then the process will return to the center in a finite time on average or we can say that the expected return time to the center is finite.

The form of ergodicity used in Hafner and Preminger (2009b) is the V-geometric ergodicity which means that the difference between the t-step transition probability measure $P^t(y,\cdot)$ and the stationary probability measure converges at a geometric rate under the V-norm distance measure. The form we choose here is the V-uniform ergodicity which is a special case of the V-geometric ergodicity.

Definition 2.3 (V-Uniform Ergodicity, Ch16 in Meyn and Tweedie (2009)) *Consider an ergodic Chain Y on the state space S, and let $P(y,\cdot)$ be the transition probability, $\pi(\cdot)$ be its stationary distribution. V is a positive function such that $V : S \to [1,\infty)$. The chain is said to be V-uniform ergodic if*

$$\left|\left|\left|P^t(y,\cdot) - \pi(\cdot)\right|\right|\right|_V \to 0, \quad t \to \infty$$

where $P^t(y,\cdot)$ is the t-step transition probability and the V-norm distance between P_1 and P_2 is defined as

$$|||P_1 - P_2|||_V := \sup_{y\in S} \frac{||P_1(y,\cdot) - P_2(y,\cdot)||_V}{V(y)} = \sup_{y\in S} \frac{\sup_{|g|\leq V} |P_1(y,g) - P_2(y,g)|}{V(y)}.$$

From the definitions listed above, we could see that both the stationarity and ergodicity require the process to remain unchanged but in different ways. The stationarity requires that some properties of the process do not change over time, while the ergodicity demands the behavior of the sequence stays the same not only over time but also over the defined state space. Hence the stationarity is a necessary but not a sufficient condition for ergodicity (Bendat and Piersol, 2010).

When the parameters in our defined model satisfy certain conditions, they will generate the process which has the desired statistical properties. We provide the sufficient conditions for a common risk process to be stationary and ergodic in this chapter.

In the appendix of Hafner and Preminger (2009b), a technique is provided to show us how to rewrite a multivariate GARCH(1,1) model as a suitable state space expression when they prove Theorem 1, and Jiang (2011) extends this theorem into a general multivariate GARCH(p,q) model. The same methodology in their paper is used to proof the stationarity and ergodic theorem in this thesis. The observed process is combined with the unobservable conditional volatility terms to form a higher dimensional recursive formula. Then, this specification can be reduced to a lower dimensional representation which depends on the volatilities and innovations only. Therefore, the new specification can be treated as a chain to use the theorems in Meyn and Tweedie (2009).

2.2 Markovian Process and Nonlinear State Space Model

The first step is to change the model in Section 1.3 into a special form such that the classical well-developed theory of Markov Chains can be applied.

The general model specified by (1.5) and (1.4) can be rewritten in a different form in order to apply the standard theory of Markovian structures. The last term in (1.5), $\sigma^2_{0,t-1}$, is replaced by other terms. Hence, the stochastic process at time t, $\boldsymbol{x_t}$, can be written as a function of previous conditional variances $\sigma^2_{1,t-1}, \sigma^2_{2,t-1}, \cdots, \sigma^2_{m,t-1}$ and the current innovation variables $\boldsymbol{\epsilon_t}$.

$$\begin{cases} x_{1,t} = \epsilon_{0,t}\sqrt{\beta_{01}[\omega_1 + \alpha_1 x^2_{1,t-1} + \beta_1\sigma^2_{1,t-1}] + \cdots + \beta_{0m}[\omega_m + \alpha_m x^2_{m,t-1} + \beta_m\sigma^2_{m,t-1}]} \\ \qquad + \epsilon_{1,t}\sqrt{\omega_1 + \alpha_1 x^2_{1,t-1} + \beta_1\sigma^2_{1,t-1}} \\ \qquad \cdots\cdots \\ x_{m,t} = \epsilon_{0,t}\sqrt{\beta_{01}[\omega_1 + \alpha_1 x^2_{1,t-1} + \beta_1\sigma^2_{1,t-1}] + \cdots + \beta_{0m}[\omega_m + \alpha_m x^2_{m,t-1} + \beta_m\sigma^2_{m,t-1}]} \\ \qquad + \epsilon_{m,t}\sqrt{\omega_m + \alpha_m x^2_{m,t-1} + \beta_m\sigma^2_{m,t-1}} \\ \sigma^2_{1,t} = \omega_1 + \alpha_1 x^2_{1,t-1} + \beta_1\sigma^2_{1,t-1} \\ \qquad \cdots\cdots \\ \sigma^2_{m,t} = \omega_m + \alpha_m x^2_{m,t-1} + \beta_m\sigma^2_{m,t-1}, \end{cases} \tag{2.1}$$

Then, a $2m$-dimensional Markovian process $\{\boldsymbol{Z_t}\}$ is formed where $\boldsymbol{Z_t} = (Z_{1,t}, \cdots, Z_{2m,t})^\intercal = (x_{1,t}, \cdots, x_{m,t}, \sigma^2_{1,t}, \cdots, \sigma^2_{m,t})^\intercal$ at each time point t. Then

$$\boldsymbol{Z_t} = G(\boldsymbol{Z_{t-1}}, \boldsymbol{\epsilon_t}) = \begin{pmatrix} g_1(\boldsymbol{Z_{t-1}}, \boldsymbol{\epsilon_t}) \\ \vdots \\ g_m(\boldsymbol{Z_{t-1}}, \boldsymbol{\epsilon_t}) \\ g_{m+1}(\boldsymbol{Z_{t-1}}, \boldsymbol{\epsilon_t}) \\ \vdots \\ g_{2m}(\boldsymbol{Z_{t-1}}, \boldsymbol{\epsilon_t}) \end{pmatrix}. \tag{2.2}$$

where $g_1, \cdots, g_{2m}$ are some deterministic non-linear functions.

Meyn and Tweedie (2009) summarized the tools to study stochastic processes following different kinds of chain structures. Our model written in (2.1) fits into the framework of a multidimensional nonlinear state space model defined in Chapter 2.2.2 of Meyn and Tweedie (2009).

Definition 2.4 (Nonlinear State Space Model or NSS(F))

A stochastic process $Y = \{Y_t\}$ is called a nonlinear state space model driven by F with control set O_w or NSS(F) if

NSS1 for each $t \geq 0$, Y_t and W_t are random variables on $\mathbb{R}^n$ and $\mathbb{R}^p$ respectively, satisfying inductively for $t \geq 1$,

$$Y_t = F(Y_{t-1}, W_t)$$

for some smooth (C^∞) function $F : S \times O_w \to S$, where S is an open subset of $\mathbb{R}^n$ and O_w is an open subset of $\mathbb{R}^p$;

NSS2 the random variables $\{W_t\}$ form an i.i.d. disturbance sequence on $\mathbb{R}^p$, whose marginal distribution Γ possesses a density γ_w which is supported on an open set O_w.

Define the sequence of inductive mappings: $\{F_t : S \times O_w \to S : t \geq 0\}$:

$$F_0(y_0) = y_0, \quad F_1(y_0, u_1) = F(y_0, u_1),$$

$$F_t(y_0, u_1, u_2, \cdots, u_t) = F(F_{t-1}(y_0, u_1, u_2, \cdots, u_{t-1}), u_t) \text{ for } t > 1$$

At each time t, the latter half of the multivariate function $\boldsymbol{G}$ in (2.2) from g_{m+1} to g_{2m} is completely determined by the past $\boldsymbol{Z_{t-1}}$ which does not involve any randomness from the innovation in any sense. We could use the relationship between $\boldsymbol{Z_t}$ and $(\boldsymbol{Z_{t-1}}, \boldsymbol{\epsilon_t})$ in the first half of this $\boldsymbol{G}$ function and substitute the corresponding terms in the second half. Therefore,

$$\begin{cases} \sigma_{1,t}^2 = \omega_1 + \alpha_1 \left(\epsilon_{1,t-1}\sigma_{1,t-1} + \epsilon_{0,t-1}\sqrt{\beta_{01}\sigma_{1,t-1}^2 + \cdots + \beta_{0m}\sigma_{m,t-1}^2}\right)^2 + \beta_1\sigma_{1,t-1}^2 \\ \sigma_{2,t}^2 = \omega_2 + \alpha_2 \left(\epsilon_{2,t-1}\sigma_{2,t-1} + \epsilon_{0,t-1}\sqrt{\beta_{01}\sigma_{1,t-1}^2 + \cdots + \beta_{0m}\sigma_{m,t-1}^2}\right)^2 + \beta_2\sigma_{2,t-1}^2 \\ \quad \cdots\cdots \\ \sigma_{m,t}^2 = \omega_m + \alpha_m \left(\epsilon_{m,t-1}\sigma_{m,t-1} + \epsilon_{0,t-1}\sqrt{\beta_{01}\sigma_{1,t-1}^2 + \cdots + \beta_{0m}\sigma_{m,t-1}^2}\right)^2 + \beta_m\sigma_{m,t-1}^2. \end{cases} \tag{2.3}$$

After all these operations, the original $2m$ dimensional Markov model above is reduced into an m dimensional formulation. This new m dimensional representation still follows a Markovian structure.

Define

$$\boldsymbol{Y_t} = F(\boldsymbol{Y_{t-1}}, \boldsymbol{W_t}), \tag{2.4}$$

where $\boldsymbol{Y_t} = (\sigma_{1,t}, \sigma_{2,t}, \cdots, \sigma_{m,t})^\intercal$ and $\boldsymbol{W_t} = \boldsymbol{\epsilon_{t-1}}$. This process can be treated as a homogeneous Markov chain. The process $\{\boldsymbol{Y_t}\}$ written in (2.4) is an NSS(F) where F is a smooth function with $S \in (0, \infty)^m$ and an $m+1$ dimensional innovation process $\boldsymbol{W_t}$ satisfying *NSS2* on the control set.

Definition 2.5 (The Associated Control Model CM(F))

CM1 The deterministic system

$$y_t = F_t(y_0, u_1, u_2, \cdots, u_t), \qquad t \in \mathbb{N}^+,$$

where the sequence of maps $\{F_t : S \times O_w^t \to S, t \in \mathbb{N}\}$ *has been defined above, is called the associated control system for the NSS(F) model, denoted by CM(F), given the deterministic control sequence* $\{u_1, u_2, \cdots, u_{t-1}, u_t, t \in \mathbb{N}^+\}$ *lies in the control set* $O_w \subseteq \mathbb{R}^p$.

For an m dimensional vector $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_m)^\intercal$ and another $m + 1$ dimensional vector $\boldsymbol{U} = (U_1, \cdots, U_m, U_{m+1})$, we define functions $f_1, \ldots, f_m$ as

$$f_1(\boldsymbol{Y}, \boldsymbol{U}) = \sqrt{\omega_1 + \alpha_1 \left(U_1 Y_1 + U_{m+1} \sqrt{\beta_{01} Y_1^2 + \cdots + \beta_{0m} Y_m^2}\right)^2 + \beta_1 Y_1^2}$$
$$f_2(\boldsymbol{Y}, \boldsymbol{U}) = \sqrt{\omega_2 + \alpha_2 \left(U_2 Y_2 + U_{m+1} \sqrt{\beta_{01} Y_1^2 + \cdots + \beta_{0m} Y_m^2}\right)^2 + \beta_2 Y_2^2}$$
$$\cdots\cdots$$
$$f_m(\boldsymbol{Y}, \boldsymbol{U}) = \sqrt{\omega_m + \alpha_m \left(U_m Y_m + U_{m+1} \sqrt{\beta_{01} Y_1^2 + \cdots + \beta_{0m} Y_m^2}\right)^2 + \beta_m Y_m^2}.$$

It is easy to see that these functions are the components of the function $\boldsymbol{F}$ in (2.4), which means

$$\boldsymbol{Y_t} = F(\boldsymbol{Y_{t-1}}, \boldsymbol{W_t}) = \begin{pmatrix} f_1(\boldsymbol{Y_{t-1}}, \boldsymbol{W_t}) \\ \vdots \\ f_m(\boldsymbol{Y_{t-1}}, \boldsymbol{W_t}) \end{pmatrix}. \tag{2.5}$$

The irreducibility of the original stochastic process could be studied based on the associated control model driven by $\boldsymbol{F}$. The control model associated with F defined in (2.4) is denoted by $\boldsymbol{y_t}$ given the control sequence $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_{t-1}}, \boldsymbol{u_t}\}$ where $\boldsymbol{y_t} = (y_{1,t}, y_{2,t}, \cdots, y_{m,t})$. Hence the sequence of function $\{\boldsymbol{F_t}\}$ is specified by

$$F_1(\boldsymbol{y}_0, \boldsymbol{u_1}) = F(\boldsymbol{y}_0, \boldsymbol{u_1}),$$

$$\boldsymbol{y_t} = F_t(\boldsymbol{y_0}, \boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_{t-1}}, \boldsymbol{u_t}) \quad t > 1.$$

If we could prove the geometric ergodicity of the process $\{\boldsymbol{Y_t}\}$, then it is natural that $\{\boldsymbol{Z}_{t-1}\}$ and the original process $\{\boldsymbol{x_t}\}$ are ergodic as well (Proposition 4 of Carrasco and Chen 2002).

2.3 Ergodicity and Stationarity Theorem

The main result of this chapter is included in this section. The observable process $\boldsymbol{x}_t$ will be stationary and ergodic under the four assumptions. These four assumptions are sufficient but not necessary for the conclusion in this theory.

Theorem 2.3.1 (Geometric Ergodicity)

Consider the stochastic process $\boldsymbol{X}_t$ defined by (1.4) and (1.5) and its matching dimension reduced form in (2.5). Assume that:

A1 *The marginal distribution of $\{\boldsymbol{\epsilon_t}\}$ is given by a lower semi continuous density $f_{\boldsymbol{\epsilon_t}}$ w.r.t. the Lebesgue measure which has support of an open set on $\mathbb{R}^{m+1}$. The initial value $\boldsymbol{Y_0}$ in (2.5) is independent of $\{\boldsymbol{\epsilon_t}\}$;*

A2 *$\alpha_i > 0$ and $\omega_i > 0$ for $i = 1, 2, \ldots, m$;*

A3 *$\alpha_i + \beta_i < 1$ for $i = 1, 2, \ldots, m$;*

A4 *There exist a positive integer p_1 and a positive number $s_3 \leq 2$ such that $\sup_{\bar{y}} \mathbb{E}\left[\|B(\bar{y}, \boldsymbol{\epsilon_t})\|_{p_1}^{s_3}\right] < 1$, where $\boldsymbol{B}$ is the partial derivative of function $\boldsymbol{F(Y, W)}$ with respect to the first variable $\boldsymbol{Y}$, and the matrix norm $\|\cdot\|$ is the operator norm corresponding to a given vector p-norm.*

Under Assumptions $A1-A4$, the process $\{\boldsymbol{Y_t}\}$ is geometrically ergodic and a time invariant measure $\boldsymbol{\pi}$ exists. Then the original process $\{\boldsymbol{X_t}\}$ is also geometrically ergodic. If the process $\{\boldsymbol{X_t}\}$ starts from the stationary distribution, it becomes a strictly stationary process.

The first assumption, $A1$, is a regular constraint with respect to the distribution of the innovations. It is easily satisfied by all well defined continuous densities such as the normal density and the student-t density. Assumptions $A1$ and $A2$ are needed for forward accessibility which means the chain can be controlled using a certain control sequence for all possible states in the space defined. In addition, Assumptions $A2$ and $A3$ lead to the conclusion that a universal attracting point exists for any starting state.

The last assumption is associated with the drift of the chain. When this assumption is satisfied, the mean of the one-step drift is controlled by the current state. If the current state is far from the attracting state, the chain has a tendency to pull the process back near the attracting point in the next step. The last assumption seems really abstract and the verification of such an assumption is not easy. The detailed steps of verification are included in Chapter 4, which includes performing a Monte Carlo simulation.

There are two existence conditions included in the last assumption, s_3 and p_1. This assumption could be modified. The existence of p_1 might be weakened by using the spectral radius of the matrix since it is the lower bound of the matrix induced norm. Both Hafner and Preminger (2009b) and Jiang (2011) proposed the ergodic theory for multivariate *Vech* GARCH model based on similar assumptions. Assumption 2.3 in Hafner and Preminger (2009b) and Assumption $A4$ in Jiang (2011) are special cases of Assumption $A4$ here. In their setting, the integer p_1 has been set to be 2 and the existence of s_3 was proved by the continuity of an exponential function.

The moment conditions, Assumption 2.2 in Hafner and Preminger (2009b) and Assumption $A2$ in Jiang (2011), are needed here as well. These conditions are embedded in the model setup since the second moment of $\boldsymbol{\epsilon_t}$ is Σ which is finite.

2.4 Proof of the Theorem

Some of the definitions from Meyn and Tweedie (2009), which are used in this chapter, are provided in Appendix B.

The drift condition V4 as (15.28) in Meyn and Tweedie (2009) is essential when we want to prove the theorem in this chapter. We start this section by introducing the concept of geometric drift towards some set C which is the drift condition V4 in Meyn and Tweedie (2009).

Geometric Drift Towards C

There exist an extended-real values function $V : \mathcal{Z} \to [1, \infty]$, a measurable set C, and constants $\delta > 0$, $\nu < \infty$, such that

$$\Delta V(z) \leq -\delta V(z) + \nu \mathbb{1}_C(z), \quad z \in \mathcal{Z}, \tag{2.6}$$

where $\Delta V(z)$ is the one step 'mean-drift' on a chain which is defined as

$$\Delta V(z) := \int P(z, dy)V(y) - V(z) = \mathbb{E}[V(X_{t+1})|X_t = z] - V(z), \quad z \in Z. \tag{2.7}$$

The proof of Theorem 2.3.1 will be divided into two main parts,

- the NSS(F) model is a ψ-irreducible aperiodic T-chain in Lemma 2.4.1 below,
- there is a V function which satisfies the drift condition above on a petite set C in Lemma 2.4.4 below.

The existence of such a V function will lead to the geometric ergodicity of the process $\{\boldsymbol{Y}_t\}$, given that $\{\boldsymbol{Y}_t\}$ is a ψ-irreducible aperiodic chain (Theorem 15.0.1 and Theorem 16.0.1 of Meyn and Tweedie, 2009).

If the function V used in the geometric drift condition is unbounded on the whole state space but bounded on C, then the chain is positive recurrent with invariant probability measure π (Theorem 15.0.10 of Meyn and Tweedie, 2009). As in (i) of Theorem 15.0.10 of Meyn and Tweedie (2009), for all $x \in C$, $\left|P^t(y,C) - P^\infty(C)\right| \to 0$ when $t \to \infty$ such that $P^\infty = \pi$.

The same assumptions lead to the V-uniform ergodicity of the process $\{\boldsymbol{Y_t}\}$, which is a more general form of ergodicity than the geometric ergodicity (see Chapter 16 of Meyn and Tweedie, 2009 and Jiang (2011)).

Lemma 2.4.1 *Under Assumptions $A1-A3$, the CM(F) associated with NSS(F) in (2.4) is a ψ-irreducible aperiodic T-chain.*

Proof Under Assumptions $A1$ and $A2$, the NSS(F) is a T-chain (Definition B.1 in Appendix B) since the CM(F) is forward accessible by Proposition 7.1.4 of Meyn and Tweedie (2009) (the proof of forward accessibility refers to Lemma 2.4.2 below). An equivalent form of the $\boldsymbol{M}$-irreducibility of the CM(F) (Definition B.4 in AppendixB), which is the existence of a fixed globally attracting point $\boldsymbol{y}^*$ (Theorem 7.2.5 of Meyn and Tweedie, 2009), has been proved in Lemma 2.4.3 under Assumption $A3$. The control sequence $\{\boldsymbol{y}_t\}$ will converge to this $\boldsymbol{y}^*$ as $t \to \infty$ under a control sequence $\{\boldsymbol{u}_t = \boldsymbol{u}^*\}$ from any possible initial state $\boldsymbol{y}_0$. So the $\boldsymbol{M}$-irreducible CM(F) leads to the conclusion that the NSS(F) is ψ-irreducible (Theorem 7.2.5 and Theorem 7.2.6 of Meyn and Tweedie, 2009).

Such a control model has a minimal set $\boldsymbol{M}$ (the same set in $\boldsymbol{M}$-irreducible) which can be uniquely (in some sense) partitioned into finite disjoint closed sets $\boldsymbol{Q} = \{Q_i : 1 \leq i \leq l\}$ for an integer $l \geq 1$ (i.e. $\boldsymbol{M} = \bigcup_{i=1}^{l} Q_i$) and $\boldsymbol{Q}$ is a periodic orbit (Theorem 7.3.3 of Meyn and Tweedie, 2009). Since there is a globally attracting point $\boldsymbol{y}^*$ contained in the minimal set $\boldsymbol{M}$, $\boldsymbol{y}^*$ is reachable at almost any time. In other words, $\boldsymbol{y}^*$ belongs to each Q_i, so we conclude that l is 1 and the minimal set is aperiodic.

By Theorem 7.3.5 of Meyn and Tweedie (2009), the NSS(F) model is a ψ-irreducible aperiodic T-chain if the CM(F) is an M-irreducible chain and its unique minimal set $\boldsymbol{M}$ is aperiodic. □

Therefore, the proof of Theorem 2.3.1 is equivalent to verifying the following conditions:

1. The associated control model driven by F, $\boldsymbol{y_t}$, is forward accessible under Assumptions $A1$ and $A2$.

2. The globally attracting point of the CM(F) $\boldsymbol{y}^*$ exists if Assumption $A3$ is satisfied.

3. The drift condition (2.6) is satisfied under Assumption $A4$ with an unbounded drift function V with a petite set $\boldsymbol{C}$.

These three conditions are verified in different lemmas below.

Lemma 2.4.2 *If Assumptions* $A1$ *and* $A2$ *are satisfied, then the associated control model driven by F defined in (2.5) is forward accessible.*

Proof For a given initial value in the support, $\boldsymbol{y_0} \in S$, and a control sequence $\{\boldsymbol{u_t} : \boldsymbol{u_t} \in O_w, t \in \mathbb{N}^+\}$, let $\{B_{t+1} : t \in \mathbb{N}\}$ denote the partial derivative matrix of function $\boldsymbol{F}$ with respect to the first variable and let $\{A_{t+1} : t \in \mathbb{N}\}$ denote the partial derivative matrix of function $\boldsymbol{F}$ with respect to the second variable, both evaluated at $(\boldsymbol{y_t}, \boldsymbol{u_{t+1}})$. Mathematically,

$$B_{t+1} = B_{t+1}(\boldsymbol{y_0}, \boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_{t+1}}) = B(\boldsymbol{y_t}, \boldsymbol{u_{t+1}}) := \left[\frac{\partial F}{\partial \boldsymbol{y}}\right]_{(\boldsymbol{y_t}, \boldsymbol{u_{t+1}})},$$

$$A_{t+1} = A_{t+1}(\boldsymbol{y_0}, \boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_{t+1}}) = A(\boldsymbol{y_t}, \boldsymbol{u_{t+1}}) := \left[\frac{\partial F}{\partial \boldsymbol{u}}\right]_{(\boldsymbol{y_t}, \boldsymbol{u_{t+1}})},$$

where $\boldsymbol{y_t} = F_t(\boldsymbol{y_0}, \boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_t})$.

A short notation $\gamma(\boldsymbol{y_t})$ is used to replace the expression $\sqrt{\beta_{01} y_{1,t}^2 + \cdots + \beta_{0m} y_{m,t}^2}$ in all the equations below. Hence, the complicated expressions contained in the elements of the matrices A_{t+1} and B_{t+1} can be simplified in writing.

As we can see, the elements of A_{t+1} and B_{t+1} are functions of $\boldsymbol{y_t}$ and $\boldsymbol{u_{t+1}}$. B_{t+1} is an $m \times m$ square matrix while A_{t+1} is an m by $m + 1$ matrix. Moreover, matrix A_{t+1}

can be divided into an m dimensional diagonal matrix $AA(\boldsymbol{y_t}, \boldsymbol{u_{t+1}})$ and a column vector $AB(\boldsymbol{y_t}, \boldsymbol{u_{t+1}})$.

Specifically,

$$B(\boldsymbol{y_t}, \boldsymbol{u_{t+1}}) = \begin{pmatrix} bb_{1,1} & bb_{1,2} & \dots & bb_{1,m-1} & bb_{1,m} \\ bb_{2,1} & bb_{2,2} & \dots & bb_{2,m-1} & bb_{2,m} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ bb_{m-1,1} & bb_{m-1,2} & \dots & bb_{m-1,m-1} & bb_{m-1,m} \\ bb_{m,1} & bb_{m,2} & \dots & bb_{m,m-1} & bb_{m,m} \end{pmatrix}$$

and

$$A(\boldsymbol{y_t}, \boldsymbol{u_{t+1}}) = \left(AA(\boldsymbol{y_t}, \boldsymbol{u_{t+1}})_{m\times m} AB(\boldsymbol{y_t}, \boldsymbol{u_{t+1}})_{m\times 1}\right) = \begin{pmatrix} aa_{1,1} & 0 & \dots & 0 & ab_{1,m+1} \\ 0 & aa_{2,2} & \dots & 0 & ab_{2,m+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & aa_{m,m} & ab_{m,m+1} \end{pmatrix},$$

where

$$bb_{i,i} = \frac{\alpha_i(u_{i,t+1}y_{i,t} + u_{m+1,t+1}\gamma(\boldsymbol{y_t}))(u_{i,t+1} + \frac{u_{m+1,t+1}\beta_{0i}y_{i,t}}{\gamma(\boldsymbol{y_t})}) + \beta_i y_{i,t}}{f_i(\boldsymbol{y_t}, \boldsymbol{u_{t+1}})}$$

for $i = 1, \cdots, m$,

$$bb_{i,j} = \frac{\alpha_i(u_{i,t+1}y_{i,t} + u_{m+1,t+1}\gamma(\boldsymbol{y_t}))u_{m+1,t+1}\beta_{0j}y_{j,t}}{f_i(\boldsymbol{y_t}, \boldsymbol{u_{t+1}})\gamma(\boldsymbol{y_t})}$$

for $i \neq j$ and $i, j = 1, \cdots, m$ and

$$aa_{i,i} = \frac{\alpha_i(u_{i,t+1}y_{i,t} + u_{m+1,t+1}\gamma(\boldsymbol{y_t}))y_{i,t}}{f_i(\boldsymbol{y_t}, \boldsymbol{u_{t+1}})}. \tag{2.8}$$

$$ab_{i,m+1} = \frac{\alpha_i(u_{i,t+1}y_{i,t} + u_{m+1,t+1}\gamma(\boldsymbol{y_t}))\gamma(\boldsymbol{y_t})}{f_i(\boldsymbol{y_t}, \boldsymbol{u_{t+1}})}$$

for $i = 1, \cdots, m$.

Denote the generalized controllability matrix along with the control sequence $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_t}$ by $C^t_{\boldsymbol{y_0}}, t = 1, 2, 3, \ldots,$

$$C^t_{\boldsymbol{y_0}} := [B_t \cdots B_2 A_1 | B_t \cdots B_3 A_2 | \cdots | B_t A_{t-1} | A_t].$$

The non-linear control model driven by F is forward accessible if and only if for each initial value $\boldsymbol{y_0} \in S$, there exist $t \in \mathbb{N}^+$ and a sequence of control variables $\overrightarrow{\boldsymbol{u}}^0 = (\boldsymbol{u_1}^0, \cdots, \boldsymbol{u_t}^0) \in O^t_w$ such that the rank of $C^t_{\boldsymbol{y_0}}$ is full (Proposition 7.1.4 of Meyn and Tweedie, 2009).

In this transformed setup, we need to find a t and a control sequence which satisfies

$$\text{rank} C^t_{\boldsymbol{y_0}}(\overrightarrow{\boldsymbol{u}}^0) = m. \tag{2.9}$$

Starting from $t = 1$, we will move on to $t = 2$ if we cannot find a $\overrightarrow{\boldsymbol{u}}^0 = (\boldsymbol{u_1})$ to satisfy the rank condition in (2.9) under Assumption $A2$.

When $t = 1$,

$$C^t_{\boldsymbol{y_0}} = (A_1)_{m\times(m+1)} = (AA(\boldsymbol{y_0}, \boldsymbol{u_1})_{m\times m} AB(\boldsymbol{y_0}, \boldsymbol{u_1})_{m\times 1}),$$

where AA is the diagonal matrix. If AA is a full rank matrix with rank m, we could conclude that the CM(F) is forward accessible because the condition in (2.9) is met.

Then, the sufficient condition for the forward accessibility is changed to find a suitable $\boldsymbol{u_1}$ such that the diagonal elements of matrix AA defined above in (2.8) are non-zero. From the model setup and Assumption $A2$, it is easy to see $\alpha_i > 0$, $y_{i,0} > 0$ and $f_i(\boldsymbol{y_0}, \boldsymbol{u_1}) > 0$. Hence, any $m + 1$ dimensional $\boldsymbol{u_1}$ satisfies

$$(u_{i,1} y_{i,0} + u_{m+1,1}\gamma(\boldsymbol{y_0})) \neq 0 \text{ for } i = 1, 2, \ldots, m \tag{2.10}$$

would work here. □

We will determine the value of this control variable $\boldsymbol{u_1}$ in the next part of this section so that it can serve the purpose of both the 1st and the 2nd conditions in the proof of this theorem.

In the rest of this chapter, we are going to measure the distances between vectors and work with the mean value theorem in multivariate cases. The partial derivative of the multidimensional function F with respect to a vector is a matrix. Thus, the first question we need to answer is how to define an appropriate vector norm as well as a matrix norm.

The vector norm used in this thesis is the L_p norm and the matrix norm is chosen to be the corresponding induced operator norm. For $p \geq 1$, the L_p norm of a vector $\boldsymbol{y} = (y_1, \ldots, y_n)$ is

$$||\boldsymbol{y}||_p := \left(\sum_{i=1}^{n} |y_i|^p \right)^{1/p}.$$

The operator norm of a $m \times n$ matrix A corresponding to a given vector p-norm is defined as

$$\begin{aligned} ||A||_p : &= \sup_{\boldsymbol{y}} \left\{ \frac{||A\boldsymbol{y}||_p}{||\boldsymbol{y}||_p} : \boldsymbol{y} \in \mathbb{R}^n, ||\boldsymbol{y}||_p \neq 0 \right\} \\ &= \sup_{\boldsymbol{y}} \left\{ ||A\boldsymbol{y}||_p : \boldsymbol{y} \in \mathbb{R}^n, ||\boldsymbol{y}||_p = 1 \right\}. \end{aligned}$$

There is no easy way to calculate the induced matrix norm for a general p except for these special values, $1, 2$ and ∞. However, if A is a diagonal matrix, the p-norm of A is always the absolute value of the largest diagonal element by the definition of the operator norm. The searching process in this section takes advantage of this property, which helps reduce the complication of the induced matrix norms.

Lemma 2.4.3 *There exists a globally attracting point in the associated control model driven by F under Assumption A3.*

Proof Given an induced matrix norm $||\cdot||_p$, the mean value theorem of multidimensional variables has been applied on the difference between the control points $\boldsymbol{y_{t+1}}$ and $\boldsymbol{y_t}$ under the control sequence $\{\boldsymbol{u}_t = \boldsymbol{u}^*\}$. There exists a m dimensional vector $\boldsymbol{y_t}^*$ between these two points, such that

$$
\begin{aligned}
\|\boldsymbol{y}_{t+1} - \boldsymbol{y}_t\|_p &= \|F(\boldsymbol{y}_t, \boldsymbol{u}^*) - F(\boldsymbol{y}_{t-1}, \boldsymbol{u}^*)\|_p \\
&= \left\| (\boldsymbol{y}_t - \boldsymbol{y}_{t-1}) \cdot \left[\frac{\partial F}{\partial \boldsymbol{y}}\right]_{(\boldsymbol{y}_t{}^*, \boldsymbol{u}^*)} \right\|_p \\
&= \|(\boldsymbol{y}_t - \boldsymbol{y}_{t-1}) \cdot B(\boldsymbol{y}_t{}^*, \boldsymbol{u}^*)\|_p \\
&\leq \|\boldsymbol{y}_t - \boldsymbol{y}_{t-1}\|_p \, \|B(\boldsymbol{y}_t{}^*, \boldsymbol{u}^*)\|_p \, .
\end{aligned}
\tag{2.11}
$$

where p is a fixed positive integer.

If

$$\|B(\boldsymbol{y}_t, \boldsymbol{u}^*)\|_p < 1 \tag{2.12}$$

is true for any $\boldsymbol{y}_t$ in the space with a fixed $\boldsymbol{u}^*$, then we can find a constant ρ_0 on the cord of $\sup_{\boldsymbol{y}} \|B(\boldsymbol{y}_t, \boldsymbol{u}^*)\|_p$ and 1, and apply the inequality in (2.11) iteratively. So,

$$\|\boldsymbol{y}_{t+1} - \boldsymbol{y}_t\|_p \leq \|\boldsymbol{y}_t - \boldsymbol{y}_{t-1}\|_p \, \rho_0 \leq \ldots \leq \rho_0^t \, \|\boldsymbol{y}_1 - \boldsymbol{y}_0\|_p \, .$$

The change between time steps $\|\boldsymbol{y}_{t+1} - \boldsymbol{y}_t\|_p$, given the control variable $\boldsymbol{u}^*$, is approaching 0 when t goes to infinity, i.e.

$$\boldsymbol{y}_t \rightarrow \boldsymbol{y}^* \;\; as \; t \rightarrow \infty.$$

This proves the existence of the globally attracting point $\boldsymbol{y}^*$ in the control model.

In order to prove the existence of the globally attracting point, we only need to find a control variable $\boldsymbol{u}^*$ which satisfies the condition in (2.12). Since both p and $\boldsymbol{u}^*$ are unknown, a straightforward way to simplify the condition is to set the last element of $\boldsymbol{u}^*$, u^*_{m+1}, to be zero. Then the matrix $\boldsymbol{B}$ is changed to be a diagonal matrix, the value of p will not have any effect on the result of $\|\boldsymbol{B}(\boldsymbol{y}_t, \boldsymbol{u}^*)\|_p$. Therefore, the condition in (2.12) becomes

$$\|B(\boldsymbol{y}_t, \boldsymbol{u}^*)\|_p = \max \left\{ \frac{\alpha_1 u_1^{*2} y_1 + \beta_1 y_1}{\sqrt{\omega_1 + \alpha_1 u_1^{*2} y_1^2 + \beta_1 y_1^2}}, \ldots, \frac{\alpha_m u_m^{*2} y_m + \beta_m y_m}{\sqrt{\omega_m + \alpha_m u_m^{*2} y_m^2 + \beta_1 y_m^2}} \right\} < 1. \tag{2.13}$$

Then the inequality above could be further simplified as

$$(\alpha_i u_i^{*2} + \beta_i)(\alpha_i u_i^{*2} + \beta_i - 1)y_i^2 - \omega_i < 0$$

for $i = 1, \dots, m$. The first m elements in $\boldsymbol{u}^*$ could be chosen as any combination of

$$-\sqrt{\frac{1-\beta_1}{\alpha_1}} \leq u_1^* \leq \sqrt{\frac{1-\beta_1}{\alpha_1}}, \dots, -\sqrt{\frac{1-\beta_m}{\alpha_m}} \leq u_m^* \leq \sqrt{\frac{1-\beta_m}{\alpha_m}},$$

so that $\alpha_i u_i^{*2} + \beta_i - 1 \leq 0$, which makes the inequality in (2.13) true for any $\boldsymbol{y}$ in the state space.

By Assumption $A3$, for any $\boldsymbol{\theta}$ in the parameter space, we could choose the control variable to be $\boldsymbol{u}^* = (1, \dots, 1, 0)$, which will satisfy the condition in (2.12) and lead to the existence of a globally attracting point.

This control variable does not only satisfy (2.13) but also fulfill the full rank condition (2.10) in Lemma 2.4.2. □

The last piece of the puzzle is to find the fixed constant p so that both the vector norm and the induced matrix norm can be defined.

The drift condition $V4$ in Meyn and Tweedie (2009) needs to be verified under the original stochastic model $\boldsymbol{Y_t}$ instead of the associated control model $\boldsymbol{y_t}$ in the previous subsections.

Lemma 2.4.4 *The drift condition (2.6) is satisfied under Assumption $A4$ with an unbounded drift function V with a petite set $\boldsymbol{C}$.*

Define a function V as

$$V(\boldsymbol{Y}) = 1 + \|\boldsymbol{Y}\|_{p_1}^{s_3}, \tag{2.14}$$

where s_3 and p_1 are the same numbers as the ones stated in Assumption $A4$.

Proof Suppose the globally attracting point in Lemma 2.4.3 is denoted by $\boldsymbol{y}^*$. By the mean value theorem, we can get

$$\begin{aligned}
V(\boldsymbol{Y_t}) &= 1 + \|\boldsymbol{Y_t}\|_{p_1}^{s_3} \\
&= 1 + \|F(\boldsymbol{Y_{t-1}}, \boldsymbol{W_t}) - F(\boldsymbol{y}^*, \boldsymbol{W_t}) + F(\boldsymbol{y}^*, \boldsymbol{W_t})\|_{p_1}^{s_3} \\
&= 1 + \|B(\boldsymbol{Y_{t-1}}', \boldsymbol{W_t}) \cdot (\boldsymbol{Y_{t-1}} - \boldsymbol{y}^*) + F(\boldsymbol{y}^*, \boldsymbol{W_t})\|_{p_1}^{s_3} \qquad \boldsymbol{Y_{t-1}}' \text{ is between } \boldsymbol{Y_{t-1}} \text{ and } \boldsymbol{y}^* \\
&= 1 + \|F(\boldsymbol{y}^*, \boldsymbol{W_t}) - B(\boldsymbol{Y_{t-1}}', \boldsymbol{W_t})\boldsymbol{y}^* + B(\boldsymbol{Y_{t-1}}', \boldsymbol{W_t})\boldsymbol{Y_{t-1}}\|_{p_1}^{s_3}.
\end{aligned} \tag{2.15}$$

By Minkowski's inequality,

$$\begin{aligned}
&E[V(\boldsymbol{Y_{t+1}})|\boldsymbol{Y_t} = \boldsymbol{y}] \\
=&1 + \mathbb{E}\,\|F(\boldsymbol{y}^*, \boldsymbol{W_{t+1}}) - B(\boldsymbol{y_t}', \boldsymbol{W_{t+1}})\boldsymbol{y}^* + B(\boldsymbol{y_t}', \boldsymbol{W_{t+1}})\boldsymbol{y}\|_{p_1}^{s_3} \\
\le&1 + \mathbb{E}\,\|F(\boldsymbol{y}^*, \boldsymbol{W_{t+1}}) - B(\boldsymbol{y_t}', \boldsymbol{W_{t+1}})\boldsymbol{y}^*\|_{p}^{s_3} + \|\boldsymbol{y}\|_p^s\,\mathbb{E}\,\|B(\boldsymbol{y_t}', \boldsymbol{W_{t+1}})\|_{p_1}^{s_3} \\
\le&1 + \mathbb{E}\,\|F(\boldsymbol{y}^*, \boldsymbol{W_{t+1}})\|_{p_1}^{s_3} + \|\boldsymbol{y}^*\|_p^{s_3}\,\mathbb{E}\,\|B(\boldsymbol{y_t}', \boldsymbol{W_{t+1}})\|_{p_1}^{s_3} + \|\boldsymbol{y}\|_{p_1}^{s_3}\,\mathbb{E}\,\|B(\boldsymbol{y_t}', \boldsymbol{W_{t+1}})\|_{p_1}^{s_3}.
\end{aligned}$$

For any given $\boldsymbol{y}$, $\boldsymbol{y_t}'$ is a fixed point between $\boldsymbol{y}$ and the globally attracting point $\boldsymbol{y}^*$ without any randomness. Assumption $A4$ is equivalent to that $\mathbb{E}\,\|B(\boldsymbol{y_t}', \boldsymbol{W_{t+1}})\|_{p_1}^{s_3} < 1$ is true for any points $\boldsymbol{y_t}'$ within that interval and for any possible $\boldsymbol{y}$. The value in Assumption $A4$, $\sup_{\bar{y}}\mathbb{E}\left[\|B(\bar{y}, \boldsymbol{W_{t+1}})\|_{p_1}^{s_3}\right]$, is denoted by λ in the following statements.

So,

$$E[V(\boldsymbol{Y_{t+1}})|\boldsymbol{Y_t} = \boldsymbol{y}] \le \lambda V(\boldsymbol{y}) + \nu \tag{2.16}$$

where

$$\nu = 1 - \lambda + \mathbb{E}\,\|F(\boldsymbol{y}^*, \boldsymbol{W_{t+1}})\|_{p_1}^{s_3} + \|\boldsymbol{y}^*\|_{p_1}^{s_3}\,\mathbb{E}\,\|B(\boldsymbol{y_t}', \boldsymbol{W_{t+1}})\|_{p_1}^{s_3}.$$

Then drift function defined in (2.7) becomes

$$\Delta V(\boldsymbol{y}) = E[V(\boldsymbol{Y_{t+1}})|\boldsymbol{Y_t} = \boldsymbol{y}] - V(\boldsymbol{y}) \le (\lambda - 1)V(\boldsymbol{y}) + \nu.$$

The measurable set $\boldsymbol{C}$ is chosen as

$$C = \left\{ \boldsymbol{Y} : V(\boldsymbol{Y}) = 1 + \|\boldsymbol{Y}\|_{p_1}^{s_3} \le \frac{2}{1-\lambda}\nu \right\}. \tag{2.17}$$

We can see that $\dfrac{2}{1-\lambda} \ge 2$, it is easy to tell that $\boldsymbol{C}$ is not an empty set.

Since $\boldsymbol{C}$ is a union of closed intervals on real numbers, $\boldsymbol{C}$ is a compact set. By Theorem 6.2.5 in Meyn and Tweedie (2009), $\boldsymbol{C}$ has to be a petite set since the chain $\boldsymbol{Y}_t$ is a ψ-irreducible T-chain.

For $\boldsymbol{Y} \in \boldsymbol{C}$, since $\lambda < 1$ and $V(\boldsymbol{Y}) > 0$,

$$\Delta V(\boldsymbol{Y}) \le \frac{\lambda - 1}{2} V(\boldsymbol{Y}) + \nu.$$

For $\boldsymbol{Y} \notin \boldsymbol{C}$, so $V(\boldsymbol{Y})\dfrac{1-\lambda}{2} > \nu$,

$$\Delta V(\boldsymbol{Y}) \le \frac{\lambda - 1}{2} V(\boldsymbol{Y}).$$

Let $\delta = \dfrac{1-\lambda}{2}$ and the measurable set $\boldsymbol{C}$ to be the one in (2.17). The task left in the drift condition (2.6) is to verify that ν is finite. The finiteness of ν is equivalent to the finiteness of $\mathbb{E}\,\|F(\boldsymbol{y}^*, \boldsymbol{W}_{t+1})\|_{p_1}^{s_3}$ and $\mathbb{E}\,\|B(\boldsymbol{Y}_t{}', \boldsymbol{W}_{t+1})\|_{p_1}^{s_3}$. By the model setup, $\mathbb{E}\,\|\boldsymbol{W}_{t+1}\|^2$ is $\boldsymbol{\Sigma}$ which is finite, so $\mathbb{E}\,\|\boldsymbol{W}_{t+1}\|_{p_1}^{s_3}$ is finite when $s_3 \le 2$. This model setup would lead to the finiteness of $\mathbb{E}\,\|F(\boldsymbol{y}^*, \boldsymbol{W}_{t+1})\|_{p_1}^{s_3}$.

From the previous proof, we can get

$$\mathbb{E}\,\|B(\boldsymbol{y}_t{}', \boldsymbol{W}_{t+1})\|_{p_1}^{s_3} \le \lambda < 1.$$

The finiteness of ν follows. □

Chapter 3

Gaussian QMLE and its Asymptotic Theory

In this chapter, a parameter estimation method called Gaussian quasi-maximum likelihood (QML) will be used to find the parameter values. The estimator (QMLE) will converge as the sample size increases, which leads to two important asymptotic results. The first one is the estimator converges to the true value almost surely, and the other one is, after choosing a scale related to the sample size, the difference between the estimator and the true parameter converges to a normal distribution. In brief, we will establish the strong consistency and the asymptotic normality theorem under certain conditions in this chapter.

3.1 Gaussian Quasi-Maximum Likelihood Estimator

A distribution must be specified for the innovation $\boldsymbol{\epsilon}_t$ process in order to form the likelihood function. The maximum likelihood (ML) method is particularly useful in statistical inferences because it usually provides an estimator which is both consistent and asymptotically normal. The quasi-maximum likelihood (QML) method could draw statistical inferences based on an even misspecified distribution of the innovations while the ML method assumes that the true distribution of the innovations is the specified distribution. ML method essentially is a special case of the QML method with no specification error.

The observations $\boldsymbol{x_t}$'s are assumed to follow a realization of an m-dimensional common risk process with an unknown true parameter

$$\boldsymbol{\theta_0} = (\rho_{1,2}^{(0)}, \cdots, \rho_{m-1,m}^{(0)}, \omega_1^{(0)}, \cdots, \omega_m^{(0)}, \alpha_1^{(0)}, \cdots, \alpha_m^{(0)}, \beta_1^{(0)}, \cdots, \beta_m^{(0)}, \beta_{01}^{(0)}, \cdots, \beta_{0m}^{(0)})^\intercal,$$

which belongs to a parameter space of the form

$$\boldsymbol{\Theta} \subset [-1, 1]^{\frac{m(m-1)}{2}} \times [0, \infty)^{4m}. \tag{3.1}$$

Under the assumption of normally distributed driving innovations , $\boldsymbol{\epsilon_t}$'s, we could estimate $\boldsymbol{\theta_0}$ by constructing the Gaussian quasi likelihood function based on the one-step ahead density of conditional distribution $\boldsymbol{x_t}|\mathcal{F}_{t-1}$.

The observations in (1.4) can be written as linear combinations of normally distributed variables given the past. Therefore, the conditional distribution of the observations $\boldsymbol{x_t}$'s are multivariate normal too, e.g. $\boldsymbol{x_t}|\mathcal{F}_{t-1} \sim N(\boldsymbol{0}, H_t)$. The model in (1.4) and (1.5) can be revised to a different form as

$$\begin{cases} \boldsymbol{x}_t|\mathcal{F}_{t-1} = H_t^{1/2}\boldsymbol{\xi_t} \\ H_t = \begin{pmatrix} \sigma_{0,t}^2 + \sigma_{1,t}^2 & \sigma_{0,t}^2 + \rho_{1,2}\sigma_{1,t}\sigma_{2,t} & \dots & \sigma_{0,t}^2 + \rho_{1,m}\sigma_{1,t}\sigma_{m,t} \\ \sigma_{0,t}^2 + \rho_{1,2}\sigma_{1,t}\sigma_{2,t} & \sigma_{0,t}^2 + \sigma_{2,t}^2 & \dots & \sigma_{0,t}^2 + \rho_{2,m}\sigma_{2,t}\sigma_{m,t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{0,t}^2 + \rho_{1,m}\sigma_{1,t}\sigma_{m,t} & \sigma_{0,t}^2 + \rho_{2,m}\sigma_{2,t}\sigma_{m,t} & \dots & \sigma_{0,t}^2 + \sigma_{m,t}^2 \end{pmatrix}, \end{cases} \tag{3.2}$$

where the innovations $\boldsymbol{\xi_t}$ are a sequence of i.i.d m-dimensional standard normal variables. Then the quasi log likelihood function for n observations is conditional on an initial information set $\mathcal{F}_0$, up to an additive constant and a constant scale, given by

$$L_n(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{t=1}^{n}\{\log|H_t(\boldsymbol{\theta})| + \boldsymbol{x_t}^\intercal H_t(\boldsymbol{\theta})^{-1}\boldsymbol{x_t}\} = -\frac{1}{n}\sum_{t=1}^{n} l_t(\boldsymbol{\theta}). \tag{3.3}$$

where $\mathcal{F}_t = \sigma\{\boldsymbol{\epsilon_t}, \boldsymbol{\epsilon_{t-1}}, \ldots\}$.

Theoretically, $\{x_{1,0}, \cdots, x_{m,0}, \sigma_{1,0}, \cdots, \sigma_{m,0}\}$, which are drawn from their stationary distribution or depends on the infinite past. However, we do not possibly know their stationary distribution in practice, which makes this likelihood impossible to work with. What we can do is to work with the likelihood function conditional on some finite given initial values or a finite past. Define $\tilde{L}_n(\boldsymbol{\theta})$ as the target function or the quasi log likelihood which is conditional on a set of initial values $\{\tilde{x}_{1,0}, \cdots, \tilde{x}_{m,0}, \tilde{\sigma}_{1,0}, \cdots, \tilde{\sigma}_{m,0}\}$. The choice of the the initial value is almost arbitrary, the only constraint is that $\tilde{\sigma}_{1,0}, \cdots, \tilde{\sigma}_{m,0}$ need to take some non negative values.

Based on this initial value, we define $\tilde{\sigma}_{1,t}, \cdots, \tilde{\sigma}_{m,t}$ for $t \geq 2$ iteratively as

$$\begin{cases} \tilde{\sigma}^2_{1,t} = \omega_1 + \alpha_1 x^2_{1,t-1} + \beta_1\tilde{\sigma}^2_{1,t-1} \\ \tilde{\sigma}^2_{2,t} = \omega_2 + \alpha_2 x^2_{2,t-1} + \beta_2\tilde{\sigma}^2_{2,t-1} \\ \qquad \cdots\cdots \\ \tilde{\sigma}^2_{m,t} = \omega_m + \alpha_m x^2_{m,t-1} + \beta_m\tilde{\sigma}^2_{m,t-1}. \end{cases}$$

For $t = 1$, we just change the values $x_{1,0}, \cdots, x_{m,0}$ to $\tilde{x}_{1,0}, \cdots, \tilde{x}_{m,0}$ in the iteration above. Then, other terms $\tilde{H}_t(\boldsymbol{\theta}), \tilde{l}_t(\boldsymbol{\theta})$ and $\tilde{L}_n(\boldsymbol{\theta})$ can be defined analogously,

$$\tilde{L}_n(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{t=1}^{n}\{\log|\tilde{H}_t(\boldsymbol{\theta})| + \boldsymbol{x_t}^\intercal \tilde{H}_t(\boldsymbol{\theta})^{-1}\boldsymbol{x_t}\} = -\frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\boldsymbol{\theta}). \tag{3.4}$$

This $\tilde{L}_n$ is called the observed likelihood which is also a conditional quasi likelihood, and the corresponding filtration $\tilde{\mathcal{F}}_t = \sigma\{\boldsymbol{\epsilon_t}, \boldsymbol{\epsilon_{t-1}}, \ldots, \boldsymbol{\epsilon_0}\}$ for $t \geq 0$. It is found that $\tilde{H}_1$ is a fixed matrix since all the elements are determined by $\tilde{x}_{1,0}, \cdots, \tilde{x}_{m,0}$. The similarity between these two likelihood functions is visible, we will explain more on their difference. (3.4) is conditional on any possible initial values whereas (3.3) is conditional on the stationary distribution (or a random variable based on infinite past). $\tilde{L}_n(\boldsymbol{\theta})$ is a statistic which can be calculated from the observable data, whereas it is not possible to get $L_n(\boldsymbol{\theta})$ based on the observations. The impact of any two different initial sets will vanish when the sample size approaches to infinity. Despite that, the choice of the initial value does have its practical effect on other aspects when we solve the optimization problem numerically, such as computational cost, efficiency, etc. $\tilde{L}_n$ in (3.4) is the quasi likelihood function in this chapter while it may have different meanings in another context.

The QML estimator is defined on the workable function in (3.4) as

$$\begin{aligned} \hat{\boldsymbol{\theta}}_n &= \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \tilde{L}_n(\boldsymbol{\theta}) \\ &= \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{n}\sum_{t=1}^{n}\{\log|\tilde{H}_t(\boldsymbol{\theta})| + \boldsymbol{x}_t{}^{\top}\tilde{H}_t(\boldsymbol{\theta})^{-1}\boldsymbol{x}_t\} \\ &= \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{n}\sum_{t=1}^{n}\tilde{l}_t(\boldsymbol{\theta}). \end{aligned} \tag{3.5}$$

We are investigating the statistical property of this estimator in the rest of this chapter.

3.2 Strong Consistency

The first asymptotic result we are interested in is the consistency since the consistency describes the relationship between the sample size and how far the estimate is from the true value. As the sample size increase indefinitely, the estimate can be arbitrarily close to the true value in some sense. In this section, we show that the estimate converges to the true value almost surely as the sample size approaches infinity.

We start this section with the concept of parameter identifiability, which is crucial for the strong consistency.

Definition 3.1 (Parameter Identifiability)

Suppose that $\boldsymbol{H}_t(\boldsymbol{\theta})$ *be the conditional second moment of* $\boldsymbol{x_t}$*,* $\boldsymbol{\Theta}$ *be the parameter space. Then* $\boldsymbol{H}_t(\boldsymbol{\theta})$ *is identifiable if* $\forall \boldsymbol{\theta_1}, \boldsymbol{\theta_2} \in \boldsymbol{\Theta}$ *and all* $t \in \mathbb{Z}$

$$\boldsymbol{H}_t(\boldsymbol{\theta_1}) = \boldsymbol{H}_t(\boldsymbol{\theta_2}) \quad a.s. \Rightarrow \quad \boldsymbol{\theta_1} = \boldsymbol{\theta_2}.$$

Since the model is a parametric model, the parameter identifiability and the model identifiability are the same in this setting. This definition of identifiability is consistent with Jeantheau (1998). We note that we are dealing with a stochastic process whose conditional distribution belongs to a location-scale family. The location has been set to be 0, and the conditional scale is $\boldsymbol{H}_t(\boldsymbol{\theta})$.

The principal term $\sigma_{0,t}$ contributes to all the conditional volatilities at the same time. It is necessary to study the condition of parameter identification since the parameter estimates are based on maximizing the likelihood function. If the parameters are not identifiable, we could end up with different estimates when we choose different initial searching points.

Theorem 3.2.1 (Model Identifiability Theorem)

Assume that:

A5 *The law of* $\boldsymbol{\epsilon_t}$ *is such that there is no quadratic form* $\boldsymbol{q}$ *for which* $q(\boldsymbol{\epsilon_t}) = c$ *a.s. with some* $c \in \mathbb{R}$.

Under Assumptions $A1-A5$*, if* $m \geq 2$*, then the conditional second moment matrix* $\boldsymbol{H}_t(\boldsymbol{\theta})$ *is identifiable. There exists a unique solution of* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *which maximizes the quasi likelihood function if* $\boldsymbol{n}$ *is sufficiently large. In other words, the model is identifiable.*

Remark In Lemma 3.2.4, we require $\boldsymbol{m}$ to be equal or greater than 2.

Proof The fundamental step to prove this theorem is Lemma 3.2.3 below. If Assumptions $A5$ and $A2$ are satisfied, Lemma 3.2.4 tells us that the parameters are identifiable from $\boldsymbol{H}_t$, the conditional second moment of $\boldsymbol{x_t}$. Suppose that $\boldsymbol{\theta_0}$ is the true value of the parameters, then $\mathbb{E}(L_n(\boldsymbol{\theta_0})) > \mathbb{E}(L_n(\boldsymbol{\theta}))$ for all $\boldsymbol{\theta} \neq \boldsymbol{\theta_0}$ and all n by Lemma 3.2.5. Under Assumptions $A1 - A4$, the process $\boldsymbol{x_t}$ is ergodic and stationary, therefore there will be a unique solution of $\boldsymbol{\theta_0}$ in the parameter space $\boldsymbol{\Theta}$ which maximizes the likelihood function when the sample size n is sufficiently large. □

The first assumption is necessary for the invertibility so that the current σ's in (3.3) can be written as a function of the infinite past. Assumption $A5$ is a mild constraint on the innovation distribution as Assumption $A1$, which can be easily satisfied by a wide range of well-defined distributions. As explained in Chapter 2, Assumptions $A1 - A4$ are not equivalent to a stationary and ergodic process since equivalence means necessary and sufficient conditions. Assumptions $A1$, $A3$ and $A4$ can be substituted by the conclusion of Theorem 2.3.1, the observable process $\boldsymbol{x_t}$ is stationary and ergodic while Assumption $A2$ is always needed.

Theorem 3.2.2 (Consistency Theorem)
Consider the stochastic process $\boldsymbol{x_t}$ defined by (1.4) and (1.5) with true parameter $\boldsymbol{\theta_0}$ satisfying the following assumptions:

$B1$ *The parameter space $\boldsymbol{\Theta}$ is compact;*

$B2$ *The observed sequence $\{\boldsymbol{x_t}\}$ is strictly stationary and ergodic;*

$B3$ *For the observed sequence $\{\boldsymbol{x_t}\}$, there exists a positive constant ν_1 such that $\mathbb{E}\,\|\boldsymbol{x_t}\|^{\nu_1}$ is finite;*

$B4$ *The model is identifiable.*

Under Assumptions $B1 - B4$ the quasi maximum likelihood estimator in (3.5) is strongly consistent, i.e.

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0 \quad \text{as } n \to \infty.$$

Remark

1. $||\cdot||$ in Assumption $B3$ is the Euclidean norm of a vector, e.g. $p = 2$ as a L_2 norm. Since all the p-norms are equivalent, $\mathbb{E}\,||\boldsymbol{x_t}||_p^{\nu}$ is also finite for any integer p.

2. Assumption $B3$ is related to the stochastic equicontinuity of $L_n(\boldsymbol{\theta_0})$ when $n \to \infty$.

3. Though the uniqueness of $\hat{\boldsymbol{\theta}}_n$ is not guaranteed when n is finite, Lemma 3.2.8 tells us that every sequence of $\hat{\boldsymbol{\theta}}_n$ converges to $\boldsymbol{\theta}_0$ almost surely.

The assumptions we proposed are similar to the ones in Hafner and Preminger (2009b), Hafner and Preminger (2009a) and Liu (2011). The norms in this theorem and all context below are the L_2 vector norm and the corresponding induced matrix norm. The first assumption, $B1$, is needed for the compactness argument so that we can use the finite subcover to finish the proof. We assume that the stationary solution with a finite moment is observed in Assumptions $B2$ and $B3$. The stationarity assumption is crucial to apply the ergodic theorem. The existence of ν_1 is the key while the value of ν_1 is not that important. The difference between (3.3) and (3.4) will approach zero with the increased sample sizes if such a ν_1 exists . As mentioned in the identifiability theorem, we can use Assumptions $A2$ and $A5$ to replace the last Assumption $B4$.

3.2.1 Proof of Theorem 3.2.2

Hafner and Preminger (2009a) state their consistency theorem as Theorem 2 and provide the proof in their Appendix A. The proof of the consistency theorem in this thesis uses a similar compactness argument.

For any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and any positive number c, let $V_c(\boldsymbol{\theta})$ be the open ball with center $\boldsymbol{\theta}$ and radius c and $\bar{V}_c(\boldsymbol{\theta})$ be the closed ball with the same center and radius.

We will prove the theory using contradiction method. Suppose that $\hat{\boldsymbol{\theta}}_n \nrightarrow \boldsymbol{\theta}_0$ a.s., then there exists at least one ω within the sample space such that $\lim_{n\to\infty} \hat{\boldsymbol{\theta}}_n(\omega) \neq \boldsymbol{\theta}_0$. If we use Λ_c to denote the space $\mathbf{\Theta}\backslash V_c(\boldsymbol{\theta_0})$, then by Heine-Borel theorem, Λ is also a compact space. We can see that $\left\|\hat{\boldsymbol{\theta}}_n(\omega) - \boldsymbol{\theta}_0\right\| \geq c$ infinite often, or $\hat{\boldsymbol{\theta}}_n(\omega) \in \Lambda$ infinite often, for any arbitrarily small c. For each of these ω's stated above, we can find a subsequence $\hat{\boldsymbol{\theta}}_{n_l} \to \boldsymbol{\theta}_1$ where $\boldsymbol{\theta}_1 \in \Lambda$ (see Chapter 7 in Kolmogorov and Fomin, 1970). Since $\hat{\boldsymbol{\theta}}_{n_l}(\omega)$ is the MLE with sample size n_l, we will focus on the sequence of increased sample sizes $n_1, n_2, \ldots$. For a positive integer k, there exists k' such that $\left\|\hat{\boldsymbol{\theta}}_{n_l}(\omega) - \boldsymbol{\theta}_1\right\| \leq 1/k$ for any $l \geq k'$. In other words, $\hat{\boldsymbol{\theta}}_{n_l} \in \bar{V}_{1/k}(\boldsymbol{\theta}_1) \cap \Lambda$ for any $l \geq k'$.

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta_0}} l_t(\boldsymbol{\theta_0}) \\
&= \lim_{\substack{l \geq k' \\ l \to \infty}} \frac{1}{n_l} \sum_{t=1}^{n_l} l_t(\boldsymbol{\theta_0}) \\
&= \lim_{\substack{l \geq k' \\ l \to \infty}} \frac{1}{n_l} \sum_{t=1}^{n_l} \tilde{l}_t(\boldsymbol{\theta_0}) \\
&\geq \liminf_{l \to \infty} \inf_{\boldsymbol{\theta} \in \mathbf{\Theta}} \frac{1}{n_l} \sum_{t=1}^{n_l} \tilde{l}_t(\boldsymbol{\theta}) \\
&= \liminf_{\substack{l \geq k' \\ l \to \infty}} \frac{1}{n_l} \sum_{t=1}^{n_l} \tilde{l}_t(\hat{\boldsymbol{\theta}}_{n_l}) && (3.6) \\
&\geq \liminf_{\substack{l \geq k' \\ l \to \infty}} \frac{1}{n_l} \sum_{t=1}^{n_l} l_t(\hat{\boldsymbol{\theta}}_{n_l}) - \limsup_{\substack{l \geq k' \\ l \to \infty}} \sup_{\boldsymbol{\theta} \in \mathbf{\Theta}} \left| \frac{1}{n_l} \sum_{t=1}^{n_l} l_t(\boldsymbol{\theta}) - \frac{1}{n_l} \sum_{t=1}^{n_l} \tilde{l}_t(\boldsymbol{\theta}) \right| \\
&\geq \liminf_{\substack{l \geq k' \\ l \to \infty}} \inf_{\boldsymbol{\theta} \in \bar{V}_{1/k}(\boldsymbol{\theta}_1) \cap \Lambda} \frac{1}{n_l} \sum_{t=1}^{n_l} l_t(\boldsymbol{\theta}) \\
&\geq \liminf_{\substack{l \geq k' \\ l \to \infty}} \frac{1}{n_l} \sum_{t=1}^{n_l} \inf_{\boldsymbol{\theta} \in \bar{V}_{1/k}(\boldsymbol{\theta}_1) \cap \Lambda} l_t(\boldsymbol{\theta}) && (3.7)
\end{aligned}
$$

The first two equations hold because of Lemma 3.2.7 and the ergodic theorem. (3.6) is obtained by the definition of the quasi MLE. The second inequality results from Lemma 3.2.8 and the last line is true because of the assumption we started with. Now we will focus on the last term in the inequality above.

We can see that, for any $\boldsymbol{\theta}$ within the compact set $\bar{V}_{1/k}(\boldsymbol{\theta}_1) \cap \Lambda$, the sequence $\{l_t(\boldsymbol{\theta})\}_t$ is stationary and ergodic since every $l_t(\boldsymbol{\theta})$ is a measurable transformation of the observable sequence, i.e. $l_t(\boldsymbol{\theta}) = f(\boldsymbol{\theta}, \boldsymbol{y_t}, \boldsymbol{y_{t-1}}, \ldots)$ and f is a measurable function. By the uniform ergodic theorem (Theorem A.1 and Exercise 7.3 in Francq and Zakoian, 2010), we can conclude that $\{\inf_{\boldsymbol{\theta} \in \bar{V}_{1/k}(\boldsymbol{\theta}_1) \cap \Lambda} l_t(\boldsymbol{\theta})\}_t$ is also a stationary and ergodic sequence. The modified ergodic theorem can be applied, so

$$\liminf_{\substack{l \geq k' \\ l \to \infty}} \frac{1}{n_l} \sum_{t=1}^{n_l} \inf_{\boldsymbol{\theta} \in \bar{V}_{1/k}(\boldsymbol{\theta}_1) \cap \Lambda} l_t(\boldsymbol{\theta}) = \mathbb{E} \inf_{\boldsymbol{\theta} \in \bar{V}_{1/k}(\boldsymbol{\theta}_1) \cap \Lambda} l_t(\boldsymbol{\theta}).$$

Thus, (3.7) above becomes

$$\mathbb{E}_{\boldsymbol{\theta_0}} l_t(\boldsymbol{\theta_0}) \geq \mathbb{E} \inf_{\boldsymbol{\theta} \in \bar{V}_{1/k}(\boldsymbol{\theta}_1) \cap \Lambda} l_t(\boldsymbol{\theta}) \to \mathbb{E} l_t(\boldsymbol{\theta_1})$$

as $k \to \infty$. The convergence is obtained by Beppo Levi's theorem since $\mathbb{E} l_t(\boldsymbol{\theta_1})$ is well defined (Lemma 3.2.6) and $\mathbb{E} \inf_{\boldsymbol{\theta} \in \bar{V}_{1/k}(\boldsymbol{\theta}_1) \cap \Lambda} l_t(\boldsymbol{\theta})$ is monotone increasing to $\mathbb{E} l_t(\boldsymbol{\theta_1})$ as k goes to infinity.

To conclude, if $\hat{\boldsymbol{\theta}}_n \notin V_c(\boldsymbol{\theta_0}) \cap \boldsymbol{\Theta}$ for arbitrarily small c, it will lead to $\mathbb{E} l_t(\boldsymbol{\theta}_0) \geq \mathbb{E} l_t(\boldsymbol{\theta}_1)$, which contradicts that $\boldsymbol{\theta_0}$ is the unique minimum of $\mathbb{E} l_t$ in Lemma 3.2.5. Therefore, $\hat{\boldsymbol{\theta}}_n$ must be within $V_c(\boldsymbol{\theta_0}) \cap \boldsymbol{\Theta}$. The strong consistency $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0$ follows as $c \to 0$.

3.2.2 Lemmas

Lemma 3.2.3 *If U is a $r \times d$ matrix and $\boldsymbol{V}_{r \times 1}$ is an $\mathcal{F}_{t-1}$-measurable vector,*

$$\text{if } U \begin{pmatrix} x_{1,t}^2 \\ x_{2,t}^2 \\ \vdots \\ x_{d,t}^2 \end{pmatrix} = \boldsymbol{V} \Rightarrow U = 0 \text{ and } \boldsymbol{V} = 0$$

Proof Similar to the proof for Lemma 3.1 in Jeantheau (1998).

The information contained in the $\sigma(\mathcal{F}_{t-1})$ includes $\boldsymbol{x_{t-1}}, \sigma_{1,t-1}, \cdots, \sigma_{m,t-1}, \sigma_{0,t-1}$ and

$\sigma_{1,t}, \cdots, \sigma_{m,t}, \sigma_{0,t}$. The first line of the equation could be written as

$$\sum_{i=1}^{d} U_{1i} h_i^2(\sigma_{1,t}, \cdots, \sigma_{m,t}, \sigma_{0,t}, \boldsymbol{\epsilon_t}) = \sum_{i=1}^{d} U_{1i} k_i(\sigma_{1,t}, \cdots, \sigma_{m,t}, \sigma_{0,t}, \boldsymbol{\epsilon_t}) = V_1$$

where h_i functions are corresponding to the rules specified in (1.4) (non degenerate function).

Let function $k_i = h_i^2$, $i = 1, 2, \cdots, d$, then k_i's are the quadratic functions of $\boldsymbol{\epsilon_t}$ for given $\sigma_{1,t}, \cdots, \sigma_{m,t}, \sigma_{0,t}$.

$U_{1,i}$ and V_1 are constants with respect to $\sigma(\mathcal{F}_{t-1})$ while the innovation term at time t, $\boldsymbol{\epsilon_t}$, is independent of $\sigma(\mathcal{F}_{t-1})$. Let μ be the measure of $(\sigma_{1,t}, \cdots, \sigma_{m,t}, \sigma_{0,t}, U_{11}, \cdots, U_{1d}, V_1)$. We get,

$$\begin{aligned} 1 =& P(\sum_{i=1}^{d} U_{1i} k_i(\sigma_{1,t}, \cdots, \sigma_{m,t}, \sigma_{0,t}, \boldsymbol{\epsilon_t}) = V_1) \\ =& \int P(\sum_{i=1}^{d} u_{1i} k_i(\sigma_{1,t}, \cdots, \sigma_{m,t}, \sigma_{0,t}, \boldsymbol{\epsilon_t}) = v_1) d\mu(\sigma_{1,t}, \cdots, \sigma_{m,t}, \sigma_{0,t}, u_{11}, \cdots, u_{1d}, v_1). \end{aligned}$$

Then,

$$P(\sum_{i=1}^{d} u_{1i} k_i(\sigma_{1,t}, \cdots, \sigma_{m,t}, \sigma_{0,t}, \boldsymbol{\epsilon_t}) = v_1) = 1 \quad \mu \text{ a.s.}$$

$u_{11} = \cdots = u_{1d} = v_1$ μ *a.s.* because of the Assumption $A6$ above. Hence, $U_{11} = \cdots = U_{1d} = V_1$, P *a.s.* It means that all the coefficients U_{1i} equal to 0 almost surely, so does V_1. For other elements of U and V, it is also true. □

Lemma 3.2.4 *Under Assumptions A2, A5 and A6, the conditional second moment matrix in (3.2) is identifiable when $m \geq 2$. If $\boldsymbol{\theta_1}, \boldsymbol{\theta_2} \in \boldsymbol{\Theta}$,*

$$H_t(\boldsymbol{\theta_1}) = H_t(\boldsymbol{\theta_2}) \quad a.s. \Rightarrow \quad \boldsymbol{\theta_1} = \boldsymbol{\theta_2}.$$

Remark The proof below, specifically (3.12) and (3.13), requires U_1 and U_2 having at least two distinguish rows. This only happens when m is equal or greater than 2.

Proof Assume $H_t(\boldsymbol{\theta_1}) = H_t(\boldsymbol{\theta_2})$ where

$$\boldsymbol{\theta_1} := (\rho_{1,2}, \cdots, \rho_{m-1,m}, \omega_1, \cdots, \omega_m, \alpha_1, \cdots, \alpha_m, \beta_1, \cdots, \beta_m, \alpha_{01}, \cdots, \alpha_{0m}, \beta_{01}, \cdots, \beta_{0m})^T$$

$$\boldsymbol{\theta_2} := (\rho'_{1,2}, \cdots, \rho'_{m-1,m}, \omega'_1, \cdots, \omega'_m, \alpha'_1, \cdots, \alpha'_m, \beta'_1, \cdots, \beta'_m, \alpha'_{01}, \cdots, \alpha'_{0m}, \beta'_{01}, \cdots, \beta'_{0m})^T.$$

Use the standard backshift operator B ($B^i x_t^2 = x_{t-i}^2$ for any integer i), (1.5) can be written in a more compact way:

$$\begin{cases} (1-\beta_1 B)\sigma_{1,t}^2 = \omega_1 + \alpha_1 B x_{1,t}^2 \\ \qquad\qquad \cdots\cdots \\ (1-\beta_m B)\sigma_{m,t}^2 = \omega_m + \alpha_m B x_{m,t}^2 \end{cases} \tag{3.8}$$

By Assumption $A6$, the above equations are invertible.

$$\begin{cases} \sigma_{1,t}^2 = \dfrac{\omega_1}{1-\beta_1} + \dfrac{\alpha_1 B x_{1,t}^2}{(1-\beta_1 B)} \\ \qquad \cdots\cdots \\ \sigma_{m,t}^2 = \dfrac{\omega_m}{1-\beta_m} + \dfrac{\alpha_m B x_{m,t}^2}{(1-\beta_m B)} \\ \sigma_{0,t}^2 = \beta_{01}\sigma_{1,t}^2 + \cdots + \beta_{0m}\sigma_{m,t}^2 \\ \qquad = \beta_{01}\left[\dfrac{\omega_1}{1-\beta_1} + \dfrac{\alpha_1 B x_{1,t}^2}{(1-\beta_1 B)}\right] + \cdots + \beta_{0m}\left[\dfrac{\omega_m}{1-\beta_m} + \dfrac{\alpha_m B x_{m,t}^2}{(1-\beta_m B)}\right] \end{cases} \tag{3.9}$$

The diagonal elements in the conditional covariance matrix H_t can be expressed as following:

$$\begin{aligned} H_{ii,t} &= \sigma_{i,t}^2 + \sigma_{0,t}^2 \\ &= \beta_{01}\sigma_{1,t}^2 + \cdots + (\beta_{0i}+1)\sigma_{i,t}^2 \cdots + \beta_{0m}\sigma_{m,t}^2 \\ &= \beta_{01}\left[\frac{\omega_1}{1-\beta_1} + \frac{\alpha_1 B x_{1,t}^2}{(1-\beta_1 B)}\right] + \cdots + (\beta_{0i}+1)\left[\frac{\omega_i}{1-\beta_i} + \frac{\alpha_i B x_{i,t}^2}{(1-\beta_i B)}\right] + \\ &\quad \cdots + \beta_{0m}\left[\frac{\omega_m}{1-\beta_m} + \frac{\alpha_m B x_{m,t}^2}{(1-\beta_m B)}\right]. \end{aligned} \tag{3.10}$$

$H_{ii,t}(\boldsymbol{\theta_1}) = H_{ii,t}(\boldsymbol{\theta_2})$ holds for all possible past values of $\boldsymbol{x_t}, \boldsymbol{x_{t-1}}, \boldsymbol{x_{t-2}}, \cdots$. It is easy to see the constant terms on both sides are equal,

$$\begin{aligned}
&\beta_{01}\frac{\omega_1}{1-\beta_1} + \cdots + (1+\beta_{0i})\frac{\omega_i}{1-\beta_i} + \cdots + \beta_{0m}\frac{\omega_m}{1-\beta_m} \\
=&\beta'_{01}\frac{\omega'_1}{1-\beta'_1} + \cdots + (1+\beta'_{0i})\frac{\omega'_i}{1-\beta'_i} + \cdots + \beta'_{0m}\frac{\omega'_m}{1-\beta'_m}.
\end{aligned} \tag{3.11}$$

From (3.10), we can extract the diagonal equations $H_{ii,t}(\boldsymbol{\theta_1}) - H_{ii,t}(\boldsymbol{\theta_2}) = 0$ for $i = 1, \ldots, m$. Let

$$U_1 = \begin{pmatrix} (\beta_{01}+1)\alpha_1 - (\beta'_{01}+1)\alpha'_1 & \beta_{02}\alpha_2 - \beta'_{02}\alpha'_2 & \cdots & \beta_{0m}\alpha_m - \beta'_{0m}\alpha'_m \\ \beta_{01}\alpha_1 - \beta'_{01}\alpha'_1 & (\beta_{02}+1)\alpha_2 - (\beta'_{02}+1)\alpha'_2 & \cdots & \beta_{0m}\alpha_m - \beta'_{0m}\alpha'_m \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{01}\alpha_1 - \beta'_{01}\alpha'_1 & \beta_{02}\alpha_2 - \beta'_{02}\alpha'_2 & \cdots & (\beta_{0m}+1)\alpha_m - (\beta_{0m}+1)'\alpha'_m \end{pmatrix}$$

Then,

$$U_1 \begin{pmatrix} x^2_{1,t-1} \\ x^2_{2,t-1} \\ \vdots \\ x^2_{m,t-1} \end{pmatrix} = \boldsymbol{V}_1. \tag{3.12}$$

The vector on the right hand side of the above equation $\boldsymbol{V}_1$ is a function of $\boldsymbol{x^2_{t-2}}, \boldsymbol{x^2_{t-3}}, \cdots$, so it is an $\mathcal{F}_{t-2}$-measurable vector. According to Lemma 3.2.3, this equation leads to the conclusion $\boldsymbol{U}_1 = 0$ and $\boldsymbol{V}_1 = \boldsymbol{0}$.(i.e. $U_1\boldsymbol{x^2_{t-1}} = \boldsymbol{V}_1 \Rightarrow U_1 = 0$). Therefore, we can have $\alpha_i = \alpha'_i$ and $\beta_{0i} = \beta'_{0i}$ based on Assumption $A2$.

Similarly, let U_2 denote the following matrix

$$\begin{pmatrix} (\beta_{01}+1)\alpha_1\beta_1 - (\beta'_{01}+1)\alpha'_1\beta'_1 & \beta_{02}\alpha_2\beta_2 - \beta'_{02}\alpha'_2\beta'_2 & \cdots & \beta_{0m}\alpha_m\beta_m - \beta'_{0m}\alpha'_m\beta'_m \\ \beta_{01}\alpha_1\beta_1 - \beta'_{01}\alpha'_1\beta'_1 & (\beta_{02}+1)\alpha_2\beta_2 - (\beta'_{02}+1)\alpha'_2\beta'_2 & \cdots & \beta_{0m}\alpha_m\beta_m - \beta'_{0m}\alpha'_m\beta'_m \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{01}\alpha_1\beta_1 - \beta'_{01}\alpha'_1\beta'_1 & \beta_{02}\alpha_2\beta_2 - \beta'_{02}\alpha'_2\beta'_2 & \cdots & (\beta_{0m}+1)\alpha_m\beta'_m - (\beta_{0m}+1)'\alpha'_m\beta'_m \end{pmatrix}$$

then

$$U_2 \begin{pmatrix} x^2_{1,t-1} \\ x^2_{2,t-1} \\ \vdots \\ x^2_{m,t-1} \end{pmatrix} = \boldsymbol{V}_2. \tag{3.13}$$

and $\boldsymbol{V}_2$ is an $\mathcal{F}_{t-3}$-measurable vector $\Rightarrow U_2 = 0 \Rightarrow \beta_i = \beta_i'$.

The next step to go back to the constant term in (3.11) based on the identities we got above.

$$\begin{cases} (\beta_{01}+1)\dfrac{\omega_1}{1-\beta_1} + \beta_{02}\dfrac{\omega_2}{1-\beta_2} + \cdots + \beta_{0m}\dfrac{\omega_m}{1-\beta_m} = (\beta_{01}+1)\dfrac{\omega_1'}{1-\beta_1} + \beta_{02}\dfrac{\omega_2'}{1-\beta_2} + \cdots + \beta_{0m}\dfrac{\omega_m'}{1-\beta_m} \\ \beta_{01}\dfrac{\omega_1}{1-\beta_1} + (1+\beta_{02})\dfrac{\omega_2}{1-\beta_2} + \cdots + \beta_{0m}\dfrac{\omega_m}{1-\beta_m} = \beta_{01}\dfrac{\omega_1'}{1-\beta_1} + (1+\beta_{02})\dfrac{\omega_2'}{1-\beta_2} + \cdots + \beta_{0m}\dfrac{\omega_m'}{1-\beta_m} \\ \cdots \\ \beta_{01}\dfrac{\omega_1}{1-\beta_1} + \beta_{02}\dfrac{\omega_2}{1-\beta_2} + \cdots + (1+\beta_{0m})\dfrac{\omega_m}{1-\beta_m} = \beta_{01}\dfrac{\omega_1'}{1-\beta_1} + (1+\beta_{02})\dfrac{\omega_2'}{1-\beta_2} + \cdots + (1+\beta_{0m})\dfrac{\omega_m'}{1-\beta_m} \end{cases} \tag{3.14}$$

We could subtract the second line from the first line,

$$\frac{\omega_1}{1-\beta_1} - \frac{\omega_2}{1-\beta_2} = \frac{\omega_1'}{1-\beta_1} - \frac{\omega_2'}{1-\beta_2}$$

Since $1-\beta_1 > 0$,

$$\omega_2 - \omega_2' = \frac{1-\beta_2}{1-\beta_1}(\omega_1 - \omega_1').$$

The difference between $\omega_i - \omega_i'$ could be expressed in terms of $\omega_1 - \omega_1'$,

$$\begin{cases} \omega_2 - \omega_2' = \dfrac{1-\beta_2}{1-\beta_1}(\omega_1 - \omega_1') \\ \omega_3 - \omega_3' = \dfrac{1-\beta_3}{1-\beta_1}(\omega_1 - \omega_1') \\ \cdots \\ \omega_m - \omega_m' = \dfrac{1-\beta_m}{1-\beta_1}(\omega_1 - \omega_1'). \end{cases}$$

Plug all these equalities back into the first line of (3.14),

$$\left(\frac{1+\beta_{01}}{1-\beta_1} + \frac{\beta_{02}}{1-\beta_2}\cdot\frac{1-\beta_2}{1-\beta_1} + \cdots + \frac{\beta_{0m}}{1-\beta_m}\cdot\frac{1-\beta_m}{1-\beta_1}\right)(\omega_1 - \omega_1') = 0.$$

Then,

$$\frac{1+\beta_{01}+\beta_{02}\cdots+\beta_{0m}}{1-\beta_1}(\omega_1 - \omega_1') = 0.$$

Since $1+\beta_{01}+\beta_{02}\cdots+\beta_{0m} > 0$ and $1-\beta_1 > 0$, we can get that $\omega_1 = \omega_1'$. Like the manner, each time we can convert the first line in (3.14) into different positive numbers multiply $\omega_i - \omega_i'$. Then, we could end up with the conclusion that all the constants are identifiable $\omega_i = \omega_i'$ for $i = 1, 2, \cdots, m$.

Up to this point, all the parameters in $\sigma_{1,t}, \cdots, \sigma_{m,t}, \sigma_{0,t}$ have been identifiable. The identifiability of the parameters in the correlation matrix $(\rho_{1,2}, \cdots, \rho_{m-1,m}) = (\rho_{1,2}', \cdots, \rho_{m-1,m}')$ will follow from the equality of the non-diagonal terms in $H_t(\boldsymbol{\theta_1}) = H_t(\boldsymbol{\theta_2})$. □

Lemma 3.2.5 *At each time index t, $\mathbb{E}(l_t(\boldsymbol{\theta_0})) < \mathbb{E}(l_t(\boldsymbol{\theta}))$ for all $\boldsymbol{\theta} \neq \boldsymbol{\theta_0}$*

Proof

$$\begin{aligned}
& \mathbb{E}_{\boldsymbol{\theta_0}}(l_t(\boldsymbol{\theta})) - \mathbb{E}_{\boldsymbol{\theta_0}}(l_t(\boldsymbol{\theta_0})) \\
= \; & \mathbb{E}_{\boldsymbol{\theta_0}} \log\frac{|H_t(\boldsymbol{\theta})|}{|H_t(\boldsymbol{\theta_0})|} + \mathbb{E}_{\boldsymbol{\theta_0}}(\boldsymbol{x_t}'H_t^{-1}(\boldsymbol{\theta})\boldsymbol{x_t}) - \mathbb{E}_{\boldsymbol{\theta_0}}(\boldsymbol{x_t}'H_t(\boldsymbol{\theta_0})^{-1}\boldsymbol{x_t}) \\
= \; & \mathbb{E}_{\boldsymbol{\theta_0}} \log\frac{|H_t(\boldsymbol{\theta})|}{|H_t(\boldsymbol{\theta_0})|} + \mathbb{E}_{\boldsymbol{\theta_0}}(\boldsymbol{\xi_t}'H_t^{1/2}(\boldsymbol{\theta_0})H_t^{-1}(\boldsymbol{\theta})H_t^{1/2}(\boldsymbol{\theta_0})\boldsymbol{\xi_t}) - \mathbb{E}_{\boldsymbol{\theta_0}}(\boldsymbol{\xi_t}'\boldsymbol{\xi_t}) \\
= \; & \mathbb{E}_{\boldsymbol{\theta_0}} \log\frac{|H_t(\boldsymbol{\theta})|}{|H_t(\boldsymbol{\theta_0})|} + \mathbb{E}_{\boldsymbol{\theta_0}}[tr(\boldsymbol{\xi_t}'H_t^{1/2}(\boldsymbol{\theta_0})H_t^{-1}(\boldsymbol{\theta})H_t^{1/2}(\boldsymbol{\theta_0})\boldsymbol{\xi_t})] - m \\
= \; & \mathbb{E}_{\boldsymbol{\theta_0}} \log\frac{|H_t(\boldsymbol{\theta})|}{|H_t(\boldsymbol{\theta_0})|} + \mathbb{E}_{\boldsymbol{\theta_0}}[tr(\boldsymbol{\xi_t}\boldsymbol{\xi_t}'H_t^{1/2}(\boldsymbol{\theta_0})H_t^{-1}(\boldsymbol{\theta})H_t^{1/2}(\boldsymbol{\theta_0}))] - m \\
= \; & \mathbb{E}_{\boldsymbol{\theta_0}} \log\frac{|H_t(\boldsymbol{\theta})|}{|H_t(\boldsymbol{\theta_0})|} + \mathbb{E}_{\boldsymbol{\theta_0}}[tr(H_t^{1/2}(\boldsymbol{\theta_0})H_t^{-1}(\boldsymbol{\theta})H_t^{1/2}(\boldsymbol{\theta_0}))] - m \\
= \; & \mathbb{E}_{\boldsymbol{\theta_0}} \log\frac{|H_t(\boldsymbol{\theta})|}{|H_t(\boldsymbol{\theta_0})|} + \mathbb{E}_{\boldsymbol{\theta_0}}[tr(H_t^{1/2}(\boldsymbol{\theta_0})H_t^{1/2}(\boldsymbol{\theta_0})H_t^{-1}(\boldsymbol{\theta}))] - m \\
= \; & \mathbb{E}_{\boldsymbol{\theta_0}} \log\frac{|H_t(\boldsymbol{\theta})|}{|H_t(\boldsymbol{\theta_0})|} + \mathbb{E}_{\boldsymbol{\theta_0}}[tr(H_t(\boldsymbol{\theta_0})H_t^{-1}(\boldsymbol{\theta}))] - m \\
> \; & \mathbb{E}_{\boldsymbol{\theta_0}} \log\frac{|H_t(\boldsymbol{\theta})|}{|H_t(\boldsymbol{\theta_0})|} + \mathbb{E}_{\boldsymbol{\theta_0}}[\log|H_t(\boldsymbol{\theta_0})H_t^{-1}(\boldsymbol{\theta})| + m] - m = 0
\end{aligned}$$

If A_{m*m} is a positive definite matrix, then $\log|A| \leq tr(A) - m$ where the equal sign holds if and only if $A = I_m$. The inequality in last line holds due to this statement from Lemma A.1 of Bollerslev and Wooldridge (1992). Then by Lemma 3.2.4, $H_t(\boldsymbol{\theta_0})H_t^{-1}(\boldsymbol{\theta}) \neq I$ for all t if and only if $\boldsymbol{\theta} \neq \boldsymbol{\theta_0}$. □

Lemma 3.2.6 *At each time index t, $\mathbb{E}(l_t(\boldsymbol{\theta}))$ is well defined in $\mathbb{R} \cup \{+\infty\}$ for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The theoretical average of the time series converges, which means*

$$\frac{1}{n}\sum_{t=1}^{n} l_t(\boldsymbol{\theta}) \to \mathbb{E}_{\boldsymbol{\theta_0}} l_t(\boldsymbol{\theta}) \quad a.s..$$

Proof Since $H_t(\boldsymbol{\theta})$ is a conditional covariance matrix for each time index t, it is positive definite by the definition. Let us denote the eigenvalues of the m by m square matrix $H_t(\boldsymbol{\theta})$ by $\{\lambda_{it}(\boldsymbol{\theta})\}_{i=1}^{m}$, then all the eigenvalues $\lambda_{it}(\boldsymbol{\theta})$ are positive for $i = 1, \ldots, m$.

The eigenvalues are continuous functions of the matrix elements based on the compact parameter space assumption and the Wielandt-Haffman theorem. Therefore, we could find a positive real number γ such that $\lambda_{it}(\boldsymbol{\theta}) \geq \gamma$ for $\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all i, t. The constant γ needs to satisfy that $0 < \gamma \leq \inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \lambda_{it}(\boldsymbol{\theta})$. Hence,

$$|H_t(\boldsymbol{\theta})| = \prod_{i=1}^{m} \lambda_{it}(\boldsymbol{\theta}) \geq \gamma^m > 0$$

$$\mathbb{E}_{\boldsymbol{\theta_0}} l_t^-(\boldsymbol{\theta}) \leq \mathbb{E}_{\boldsymbol{\theta_0}} \log^- |H_t(\boldsymbol{\theta})| \leq \max\{0, -m\log\gamma\} < +\infty$$

$\mathbb{E}(l_t(\boldsymbol{\theta}))$ is well defined.

Next, the sequence $\{l_t(\boldsymbol{\theta})\}_t$ is stationary and ergodic since it is noted as a measurable transformation of the strictly stationary sequence $\{\boldsymbol{x_t}, \boldsymbol{x_{t-1}}, \ldots\}$. We could apply the standard ergodic theorem for stationary series to $l_t(\boldsymbol{\theta})$ from Doob (1990). Therefore, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\frac{1}{n}\sum_{t=1}^{n} l_t(\boldsymbol{\theta}) \to \mathbb{E}_{\boldsymbol{\theta_0}} l_t(\boldsymbol{\theta}) \quad a.s.,$$

where $l_t(\boldsymbol{\theta})$ is the theoretical function depending on the infinite past. □

Lemma 3.2.7 *At each time index t, $\mathbb{E}_{\boldsymbol{\theta_0}} l_t(\boldsymbol{\theta_0}) < +\infty$.*

Proof The eigenvalues of $\boldsymbol{H}_t(\boldsymbol{\theta})$ are denoted by $\{\lambda_{it}(\boldsymbol{\theta})\}_{i=1}^m$, the same notation used in the proof of previous lemma.

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta_0}} l_t(\boldsymbol{\theta_0}) &= \mathbb{E}_{\boldsymbol{\theta_0}} \log |H_t(\boldsymbol{\theta_0})| + \mathbb{E}_{\boldsymbol{\theta_0}}(\boldsymbol{x_t}' H_t(\boldsymbol{\theta_0})^{-1} \boldsymbol{x_t}) \\
&= \mathbb{E}_{\boldsymbol{\theta_0}} \log |H_t(\boldsymbol{\theta_0})| + \mathbb{E}_{\boldsymbol{\theta_0}}(\boldsymbol{\xi_t}'\boldsymbol{\xi_t}) \\
&= \frac{2m}{v_1} \mathbb{E}_{\boldsymbol{\theta_0}} \log |H_t(\boldsymbol{\theta_0})|^{v_1/2m} + m \\
&\leq \frac{2m}{v_1} \log \mathbb{E}_{\boldsymbol{\theta_0}} |H_t(\boldsymbol{\theta_0})|^{v_1/2m} + m \\
&= \frac{2m}{v_1} \log \mathbb{E}_{\boldsymbol{\theta_0}} \left(\prod_{i=1}^m \lambda_{it}(\boldsymbol{\theta_0}) \right)^{v_1/2m} + m \\
&\leq \frac{2m}{v_1} \log \mathbb{E}_{\boldsymbol{\theta_0}} \left(\max_{i=1,2,\cdots,m} \lambda_{it}(\boldsymbol{\theta_0}) \right)^{v_1/2} + m \\
&\leq C_1 \log \mathbb{E}_{\boldsymbol{\theta_0}} \|H_t(\boldsymbol{\theta_0})\|_p^{v_1/2} + m \\
&\leq C_2 \log \mathbb{E}_{\boldsymbol{\theta_0}} \|h_t(\boldsymbol{\theta_0})\|_p^{v_1/2} + m
\end{aligned} \tag{3.15}
$$

The process is strictly stationary under the true value of parameter $\boldsymbol{\theta_0}$. The existence of a finite v_1th moment of the observed sequence in Assumption $B3$ leads to the finite $v_1/2$th moment of $h_t(\boldsymbol{\theta_0})$ regardless what induced matrix norm we choose. Hence, the last line in the above inequality is finite. □

Lemma 3.2.8 $\limsup_{n\to\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{n} \sum_{t=1}^n l_t(\boldsymbol{\theta}) - \frac{1}{n} \sum_{t=1}^n \tilde{l}_t(\boldsymbol{\theta}) \right| = 0$ *a.s..*

Proof In the following proof, $C_1, C_2, \ldots$ and $a_1, b_1, c_1, a_2, b_2, c_2 \ldots$ will represent some finite constants and they may have different values in different inequalities below. The difference between $l_t(\boldsymbol{\theta})$ and $\tilde{l}_t(\boldsymbol{\theta})$ will be measured at the exponent $v_1/8$ where v_1 is defined in Assumption $B3$.

$$
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |l_t(\boldsymbol{\theta}) - \tilde{l}_t(\boldsymbol{\theta})|^{v_1/8} \tag{3.16}
$$

$$
\begin{aligned}
&= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\log|H_t(\boldsymbol{\theta})| - \log|\tilde{H}_t(\boldsymbol{\theta})| + \boldsymbol{x_t}' H_t(\boldsymbol{\theta})^{-1}\boldsymbol{x_t} - \boldsymbol{x_t}'\tilde{H}_t(\boldsymbol{\theta})^{-1}\boldsymbol{x_t}\right|^{\nu_1/8} \\
&\le C_1 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\log|H_t(\boldsymbol{\theta})| - \log|\tilde{H}_t(\boldsymbol{\theta})|\right|^{\nu_1/8} + C_1 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\boldsymbol{x_t}'\left[H_t(\boldsymbol{\theta})^{-1} - \tilde{H}_t(\boldsymbol{\theta})^{-1}\right]\boldsymbol{x_t}\right|^{\nu_1/8}
\end{aligned}
$$

The first job is to show that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|H_t^{-1}(\boldsymbol{\theta})\right\|_p$ and $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\tilde{H}_t^{-1}(\boldsymbol{\theta})\right\|_p$ are finite.

The constant γ in Lemma 3.2.6 needs to satisfy that $0 < \gamma \le \inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \lambda_{it}(\boldsymbol{\theta})$. Therefore, the eigenvalues of $H_t(\boldsymbol{\theta})^{-1}$ are $\{\lambda_{it}(\boldsymbol{\theta})^{-1}\}_{i=1}^m$ and the positive real number γ satisfies

$$
\frac{1}{\gamma} \ge \frac{1}{\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \lambda_{it}(\boldsymbol{\theta})} \ge \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|H_t^{-1}(\boldsymbol{\theta})\right\|_2 \ge \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|H_t^{-1}(\boldsymbol{\theta})\right\|_p
$$

for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all i, t and $p \ge 2$.

Similarly, we could find a finite absolute boundary for $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\tilde{H}_t^{-1}(\boldsymbol{\theta})\right\|_p$ given the same norm p. Hence, there exists a positive number a_1 such that both $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|\tilde{H}_t^{-1}(\boldsymbol{\theta})\right\|_p$ and $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|H_t^{-1}(\boldsymbol{\theta})\right\|_p$ are smaller than a_1.

Next, it is time to study the two terms in the last line of (3.16). The first term, which is the absolute difference between $\log|H_t(\boldsymbol{\theta})|$ and $\log|\tilde{H}_t(\boldsymbol{\theta})|$, could be bounded above.

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\log|H_t(\boldsymbol{\theta})| - \log|\tilde{H}_t(\boldsymbol{\theta})|\right|^{\nu_1/8} \\
&= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left(\log\left|I_m + \left[H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta})\right]\tilde{H}_t^{-1}(\boldsymbol{\theta})\right|\right)^{\nu_1/8} \\
&\le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left(m\log\left\|I_m + \left[H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta})\right]\tilde{H}_t^{-1}(\boldsymbol{\theta})\right\|_p\right)^{\nu_1/8} \\
&\le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left[m\log\left(\|I_m\|_p + \left\|\left[H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta})\right]\tilde{H}_t^{-1}(\boldsymbol{\theta})\right\|_p\right)\right]^{\nu_1/8} \qquad (3.17)\\
&\le \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left[m\log\left(1 + \left\|\left[H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta})\right]\tilde{H}_t^{-1}(\boldsymbol{\theta})\right\|_p\right)\right]^{\nu_1/8} \\
&\le m^{\nu_1/8} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta})\right\|_p^{\nu_1/8} \left\|\tilde{H}_t^{-1}(\boldsymbol{\theta})\right\|_p^{\nu_1/8} \\
&\le m^{\nu_1/8} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta})\right\|_p^{\nu_1/8} a_1^{\nu_1/8}
\end{aligned}
$$

Using (8) in Appendix A, we could get the first inequality. The triangle inequality leads to the second inequality, while (1) in Appendix A points to the fourth inequality.

The second term in the last line of (3.16) can be bounded above as well. Applying (7), (8) and (9) in Appendix A, we have

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \boldsymbol{x_t}' \left[H_t(\boldsymbol{\theta})^{-1} - \tilde{H}_t(\boldsymbol{\theta})^{-1} \right] \boldsymbol{x_t} \right|^{\nu_1/8} \\
=& \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| tr(\boldsymbol{x_t}' \tilde{H}_t(\boldsymbol{\theta})^{-1} \left[H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta}) \right] H_t(\boldsymbol{\theta})^{-1} \boldsymbol{x_t}) \right|^{\nu_1/8} \\
=& \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| tr(\tilde{H}_t(\boldsymbol{\theta})^{-1} \left[H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta}) \right] H_t(\boldsymbol{\theta})^{-1} \boldsymbol{x_t} \boldsymbol{x_t}') \right|^{\nu_1/8} \\
=& C_2 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \tilde{H}_t(\boldsymbol{\theta})^{-1} \right\|_p^{\nu_1/8} \left\| H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta}) \right\|_p^{\nu_1/8} \left\| H_t(\boldsymbol{\theta})^{-1} \right\|_p^{\nu_1/8} \|\boldsymbol{x_t}\boldsymbol{x_t}'\|_p^{\nu_1/8} \\
\le& C_3 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \tilde{H}_t(\boldsymbol{\theta})^{-1} \right\|_p^{\nu_1/8} \left\| H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta}) \right\|_p^{\nu_1/8} \left\| H_t(\boldsymbol{\theta})^{-1} \right\|_p^{\nu_1/8} \|\boldsymbol{x_t}\|_p^{\nu_1/4} \\
\le& C_3 a_1^{\nu_1/8} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta}) \right\|_p^{\nu_1/8} a_1^{\nu_1/8} \|\boldsymbol{x_t}\|_p^{\nu_1/4} \\
\le& C_4 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta}) \right\|_p^{\nu_1/8} \|\boldsymbol{x_t}\|_p^{\nu_1/4} .
\end{aligned}
\tag{3.18}
$$

Now, based on the results above, (3.16) can be further simplified as

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |l_t(\boldsymbol{\theta}) - \tilde{l}_t(\boldsymbol{\theta})|^{\nu_1/8} \\
\le& C_5 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta}) \right\|_p^{\nu_1/8} + C_4 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta}) \right\|_p^{\nu_1/8} \|\boldsymbol{x_t}\|_p^{\nu_1/4}
\end{aligned}
\tag{3.19}
$$

Our interest lies in the difference between $\tilde{H}_t(\boldsymbol{\theta})$ and $H_t(\boldsymbol{\theta})$. The diagonal terms and non-diagonal terms of H_t need to be considered separately. Let *vech* be the Half-vectorization of the symmetric conditional covariance matrix H_t. By (5) in Appendix A,

$$
\|vech(H_t)\|_p^{\nu_1/8} = \left(\sum_{i=1}^{m} H_{ii,t}^p + \sum_{\substack{i,j=1,\cdots,m \\ i<j}} H_{ij,t}^p \right)^{\nu_1/8p} \le \left(\sum_{i=1}^{m} |H_{ii,t}| + \sum_{\substack{i,j=1,\cdots,m \\ i<j}} |H_{ij,t}| \right)^{\nu_1/8}
$$

Then for any $t > 1$, we have

$$
\begin{aligned}
&\sup_{\theta\in\Theta}\left\|H_t-\tilde{H}_t\right\|_p^{v_1/8}\\
\le& C_6\sup_{\theta\in\Theta}\left\|vech(H_t)-vech(\tilde{H}_t)\right\|_p^{v_1/8}\\
\le& C_6\sup_{\theta\in\Theta}\left(\sum_{i=1}^m|H_{ii,t}-\tilde{H}_{ii,t}|+\sum_{\substack{i,j=1,\cdots,m\\ i<j}}|H_{ij,t}-\tilde{H}_{ij,t}|\right)^{v_1/8}\\
\le& C_6\sup_{\theta\in\Theta}\left(\sum_{i=1}^m|\sigma_{i,t}^2+\sigma_{0,t}^2-\tilde{\sigma}_{i,t}^2-\tilde{\sigma}_{0,t}^2|+\sum_{\substack{i,j=1,\cdots,m\\ i<j}}|\rho_{i,j}\sigma_{i,t}\sigma_{j,t}-\rho_{i,j}\tilde{\sigma}_{i,t}\tilde{\sigma}_{j,t}|\right)^{v_1/8}\\
\le& C_6\sup_{\theta\in\Theta}\left(\sum_{i=1}^m|\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2|+m|\sigma_{0,t}^2-\tilde{\sigma}_{0,t}^2|+\sum_{\substack{i,j=1,\cdots,m\\ i<j}}|\rho_{i,j}|\cdot|\sigma_{i,t}\sigma_{j,t}-\tilde{\sigma}_{i,t}\tilde{\sigma}_{j,t}|\right)^{v_1/8}\\
\le& C_6\sup_{\theta\in\Theta}\left(\sum_{i=1}^m(m\beta_{0i}+1)|\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2|+\sum_{\substack{i,j=1,\cdots,m\\ i<j}}|\rho_{i,j}|\cdot|\sigma_{i,t}\sigma_{j,t}-\tilde{\sigma}_{i,t}\tilde{\sigma}_{j,t}|\right)^{v_1/8}\\
\le& C_6\sup_{\theta\in\Theta}\left(\sum_{i=1}^m(m\beta_{0i}+1)^{v_1/8}|\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2|^{v_1/8}+\sum_{\substack{i,j=1,\cdots,m\\ i<j}}|\rho_{i,j}|^{v_1/8}\cdot|\sigma_{i,t}\sigma_{j,t}-\tilde{\sigma}_{i,t}\tilde{\sigma}_{j,t}|^{v_1/8}\right)
\end{aligned}
\tag{3.20}
$$

We could get the difference between $\sigma_{i,t}^2$ and $\tilde{\sigma}_{i,t}^2$ by iterating (1.5).

$$
\begin{aligned}
\sigma_{i,t}^2 =&\sigma_{i,0}^2\beta_i^t+\sum_{j=0}^{t-2}\beta_i^j(\omega_i+\alpha_i x_{i,t-1-j}^2)+\beta_i^{t-1}\omega_i+\alpha_i\beta_i^{t-1}x_{i,0}^2\\
=&\frac{\omega_i}{1-\beta_i}+\sum_{j=0}^{\infty}\alpha_i\beta_i^j x_{i,t-1-j}^2
\end{aligned}
\tag{3.21}
$$

$$
\tilde{\sigma}_{i,t}^2 =\tilde{\sigma}_{i,0}^2\beta_i^t+\sum_{j=0}^{t-2}\beta_i^j(\omega_i+\alpha_i x_{i,t-1-j}^2)+\beta_i^{t-1}\omega_i+\alpha_i\beta_i^{t-1}\tilde{x}_{i,0}^2 \tag{3.22}
$$

The $v_1/8$th moment of $\sigma_{i,t}^2$ based on the infinite past is finite because of the finiteness of $\mathbb{E}\,\|\boldsymbol{x_t}\|^{v_1}$ in Assumption $B3$.

$$
\begin{aligned}
\mathbb{E}\sup_{\theta\in\Theta}(\sigma_{i,t}^2)^{v_1/2} =&\mathbb{E}\sup_{\theta\in\Theta}(\frac{\omega_i}{1-\beta_i}+\sum_{j=0}^{\infty}\alpha_i\beta_i^j x_{i,t-1-j}^2)^{v_1/2} \\
\le&\sup_{\theta\in\Theta}\left[\frac{\omega_i^{v_1/2}}{(1-\beta_i)^{v_1/2}}+\sum_{j=0}^{\infty}\alpha_i^{v_1/2}\beta_i^{jv_1/2}\mathbb{E}(x_{i,t-1-j}^2)^{v_1/2}\right] \\
\le&\sup_{\theta\in\Theta}\left[\frac{\omega_i^{v_1/2}}{(1-\beta_i)^{v_1/2}}+\frac{\alpha_i^{v_1/2}}{1-\beta_i^{v_1/2}}C_7\right] \\
\le& a_2<\infty
\end{aligned}
\tag{3.23}
$$

We could get similar finiteness result for some lower moments of $\sigma_{i,t}^2$,

$$
\begin{aligned}
&\mathbb{E}\sup_{\theta\in\Theta}(\sigma_{i,t}^2)^{v_1/4}\le a_3 \\
&\mathbb{E}\sup_{\theta\in\Theta}(\sigma_{i,t}^2)^{v_1/8}\le a_4.
\end{aligned}
\tag{3.24}
$$

All a_2, a_3 and a_4 are finite constants, which satisfy the above inequalities for any i from 1 to m.

From Assumption $B1$ and the model setup, there exist b_1 and ρ_0 such that $0\le\beta_i\le b_1<1$ for $i=1,2,\cdots,m$ and $|\rho_{i,j}|\le\rho_0<1$.

Then,

$$
\begin{aligned}
|\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2|^{v_1/2}\le&(|\sigma_{i,0}^2-\tilde{\sigma}_{i,0}^2|\beta_i^t)^{v_1/2}+(\alpha_i\beta_i^{t-1}|x_{i,0}^2-\tilde{x}_{i,0}^2|)^{v_1/2} \\
\le&\beta_i^{tv_1/2}[(\sigma_{i,0}^2)^{v_1/2}+(\tilde{\sigma}_{i,0}^2)^{v_1/2}+\alpha_i\beta_i^{-v_1/2}(x_{i,0}^2)^{v_1/2}+\alpha_i\beta_i^{-v_1/2}(\tilde{x}_{i,0}^2)^{v_1/2}] \\
\le& M_i\beta_i^{tv_1/2}\le M_i b_1^{tv_1/2}
\end{aligned}
\tag{3.25}
$$

M_i is a random variable that depends on the infinite past values $\{x_{i,t},t\le 0\}$. The expectation of M_i is finite which could be obtained from (3.23).

$$
\begin{aligned}
\mathbb{E}\sup_{\theta\in\Theta}|\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2|^{v_1/2}\le&\mathbb{E}\sup_{\theta\in\Theta}(|\sigma_{i,0}^2-\tilde{\sigma}_{i,0}^2|\beta_i^t)^{v_1/2}+\mathbb{E}\sup_{\theta\in\Theta}(\alpha_i\beta_i^{t-1}|x_{i,0}^2-\tilde{x}_{i,0}^2|)^{v_1/2} \\
\le&\mathbb{E}\sup_{\theta\in\Theta}M_i\beta_i^{tv_1/2}\le a_5 b_1^{tv_1/2}
\end{aligned}
\tag{3.26}
$$

Related results can be easily obtained following the steps above because of the existence of the finite numbers a_2, a_3 and a_4 in (3.23) and (3.24). So we have

$$\begin{aligned}
&\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2|^{\nu_1/4}\leq a_6 b_1^{t\nu_1/4}\\
&\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2|^{\nu_1/8}\leq a_7 b_1^{t\nu_1/8}.
\end{aligned}\tag{3.27}$$

The finite constants a_5, a_6 and a_7 are assumed to be universally applied to $i = 1, 2, \cdots, m$.

The differences between the cross terms are much more complicated than the diagonal terms. Then,

$$\begin{aligned}
|\sigma_{i,t}\sigma_{j,t}-\tilde{\sigma}_{i,t}\tilde{\sigma}_{j,t}|^{\nu_1/4}=&|\sigma_{i,t}\sigma_{j,t}-\tilde{\sigma}_{i,t}\sigma_{j,t}+\tilde{\sigma}_{i,t}\sigma_{j,t}-\tilde{\sigma}_{i,t}\tilde{\sigma}_{j,t}|^{\nu_1/4}\\
\leq&\left(\sigma_{j,t}|\sigma_{i,t}-\tilde{\sigma}_{i,t}|+\tilde{\sigma}_{i,t}|\sigma_{j,t}-\tilde{\sigma}_{j,t}|\right)^{\nu_1/4}\\
\leq&\left(\frac{\sigma_{j,t}}{\sigma_{i,t}+\tilde{\sigma}_{i,t}}\cdot|\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2|+\frac{\tilde{\sigma}_{i,t}}{\sigma_{j,t}+\tilde{\sigma}_{j,t}}\cdot|\sigma_{j,t}^2-\tilde{\sigma}_{j,t}^2|\right)^{\nu_1/4}\\
\leq&C_8(\sigma_{j,t}^2)^{\nu_1/4}|\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2|^{\nu_1/4}+C_9(\tilde{\sigma}_{i,t}^2)^{\nu_1/4}\left|\sigma_{j,t}^2-\tilde{\sigma}_{j,t}^2\right|^{\nu_1/4}.
\end{aligned}\tag{3.28}$$

From the representations of $\tilde{\sigma}_{i,t}^2$ in (3.22) above, we could bound this last line by adding up the expectation of ${\boldsymbol{x_t}}^2$,

$$\begin{aligned}
&(\tilde{\sigma}_{i,t}^2)^{\nu_1/2}\\
\leq&[(\tilde{\sigma}_{i,0}^2\beta_i^t)^{\nu_1/2}+\sum_{k=0}^{t-2}\beta_i^{k\nu_1/4}(\omega_i^{\nu_1/2}+\alpha_i^{\nu_1/2}x_{i,t-1-k}^{\nu_1})+\beta_i^{(t-1)\nu_1/2}\omega_i^{\nu_1/2}+\alpha_i^{\nu_1/2}\beta_i^{(t-1)\nu_1/2}\tilde{x}_{i,0}^{\nu_1}]\\
\leq&(\tilde{\sigma}_{i,0}^2\beta_i^t)^{\nu_1/2}+\sum_{k=0}^{t-2}\beta_i^{k\nu_1/4}(\omega_i^{\nu_1/2}+\alpha_i^{\nu_1/2}\,\|\boldsymbol{x_{t-1-k}}\|^{\nu_1})+\beta_i^{(t-1)\nu_1/2}\omega_i^{\nu_1/2}+\alpha_i^{\nu_1/2}\beta_i^{(t-1)\nu_1/2}\tilde{x}_{i,0}^{\nu_1}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}(\tilde{\sigma}_{i,t}^2)^{\nu_1/2}\\
\leq&\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}[(\tilde{\sigma}_{i,0}^2\beta_i^t)^{\nu_1/2}+\sum_{k=0}^{t-2}\beta_i^{k\nu_1/4}(\omega_i^{\nu_1/2}+\alpha_i^{\nu_1/2}\,\|\boldsymbol{x_{t-1-k}}\|^{\nu_1})+\beta_i^{(t-1)\nu_1/2}\omega_i^{\nu_1/2}+\alpha_i^{\nu_1/2}\beta_i^{(t-1)\nu_1/2}\tilde{x}_{i,0}^{\nu_1}]\\
\leq&(\tilde{\sigma}_{i,0}^2)^{\nu_1/2}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\beta_i^{t\nu_1/2}+\sum_{k=0}^{t-2}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\beta_i^{k\nu_1/2}\alpha_i^{\nu_1/2}\mathbb{E}\,\|\boldsymbol{x_{t-1-k}}\|^{\nu_1}+a_8\leq a_9<\infty
\end{aligned}$$

The upper bound of $\mathbb{E}\sup_{\theta\in\Theta}(\tilde{\sigma}_{i,t}^2)^{\nu_1/4}$ would be obtained likewise,

$$\mathbb{E}\sup_{\theta\in\Theta}(\tilde{\sigma}_{i,t}^2)^{\nu_1/4} \leq a_{10}. \tag{3.29}$$

Now we proof that the expectation of the cross term decays exponentially. Using (3.23), (3.26) and (3.29), we have

$$\begin{aligned}
&\mathbb{E}\sup_{\theta\in\Theta}|\rho_{i,j}|^{\nu_1/4}|\sigma_{i,t}\sigma_{j,t}-\tilde{\sigma}_{i,t}\tilde{\sigma}_{j,t}|^{\nu_1/4}\\
\leq& C_8\sup_{\theta\in\Theta}|\rho_{i,j}|^{\nu_1/4}\mathbb{E}\sup_{\theta\in\Theta}(\sigma_{j,t}^2)^{\nu_1/4}|\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2|^{\nu_1/4}+C_9\sup_{\theta\in\Theta}|\rho_{i,j}|^{\nu_1/4}\mathbb{E}\sup_{\theta\in\Theta}(\tilde{\sigma}_{i,t}^2)^{\nu_1/4}|\sigma_{j,t}^2-\tilde{\sigma}_{j,t}^2|^{\nu_1/4}\\
\leq& C_{10}[\mathbb{E}\sup_{\theta\in\Theta}(\sigma_{j,t}^2)^{\nu_1/2}\mathbb{E}\sup_{\theta\in\Theta}|\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2|^{\nu_1/2}]^{1/2}+C_{11}[\mathbb{E}\sup_{\theta\in\Theta}(\tilde{\sigma}_{i,t}^2)^{\nu_1/2}\mathbb{E}\sup_{\theta\in\Theta}|\sigma_{j,t}^2-\tilde{\sigma}_{j,t}^2|^{\nu_1/2}]^{1/2}\\
\leq& C_{10}[a_2\cdot a_5 b_1^{t\nu_1/2}]^{1/2}+C_{11}[a_9 a_5 b_1^{t\nu_1/2}]^{1/2}\\
=& O(b_1^{t\nu_1/4}).
\end{aligned}$$

Similarly, by (3.24), (3.27) and (3.29), we can get

$$\mathbb{E}\sup_{\theta\in\Theta}|\rho_{i,j}|^{\nu_1/8}|\sigma_{i,t}\sigma_{j,t}-\tilde{\sigma}_{i,t}\tilde{\sigma}_{j,t}|^{\nu_1/8} \leq C_{14}b_1^{t\nu_1/8} = O(b_1^{t\nu_1/8}). \tag{3.30}$$

Thus,

$$\begin{aligned}
&\sum_{t=1}^{\infty}\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|H_t-\tilde{H}_t\right\|_p^{\nu_1/8}\right]\\
\leq& C_6\sum_{t=1}^{\infty}\mathbb{E}\sup_{\theta\in\Theta}\left(\sum_{i=1}^{m}(m\beta_{0i}+1)^{\nu_1/8}|\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2|^{\nu_1/8}+\sum_{\substack{i,j=1,\cdots,m\\ i<j}}|\rho_{i,j}|^{\nu_1/8}\cdot|\sigma_{i,t}\sigma_{j,t}-\tilde{\sigma}_{i,t}\tilde{\sigma}_{j,t}|^{\nu_1/8}\right)\\
\leq& C_6\sum_{t=1}^{\infty}\left(\sum_{i=1}^{m}(m+1)^{\nu_1/8}\mathbb{E}\sup_{\theta\in\Theta}|\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2|^{\nu_1/8}+\sum_{\substack{i,j=1,\cdots,m\\ i<j}}\mathbb{E}\sup_{\theta\in\Theta}|\rho_{i,j}|^{\nu_1/8}|\sigma_{i,t}\sigma_{j,t}-\tilde{\sigma}_{i,t}\tilde{\sigma}_{j,t}|^{\nu_1/8}\right)\\
\leq& C_6\sum_{t=1}^{\infty}\left(m(m+1)^{\nu_1/8}a_7 b_1^{t\nu_1/8}+\frac{m(m-1)}{2}C_{14}b_1^{t\nu_1/8}\right)
\end{aligned}$$

$$\leq \sum_{t=1}^{\infty} O(b_1^{t\nu_1/8}) \leq d_1 < \infty \tag{3.31}$$

by using the fact that b_1 is smaller than 1.

Cauchy-Schwarz inequality and (3) in Appendix A are used one more time on the second term in (3.19) to study the summation of this term from $t = 1$ to infinity. Thus,

$$\begin{aligned}
&\sum_{t=1}^{\infty} \mathbb{E}\left[\sup_{\theta\in\Theta} \left\|H_t - \tilde{H}_t\right\|_p^{\nu_1/8} \|\boldsymbol{x_t}\|_p^{\nu_1/4}\right] \\
\leq& \sum_{t=1}^{\infty}(\mathbb{E}\sup_{\theta\in\Theta} \left\|H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta})\right\|_p^{\nu_1/4} \mathbb{E}\,\|\boldsymbol{x_t}\|_p^{\nu_1/2})^{1/2} \\
\leq& C_{16}\sum_{t=1}^{\infty}(\mathbb{E}\sup_{\theta\in\Theta} \left\|H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta})\right\|_p^{\nu_1/4})^{1/2} \\
\leq& C_{17}\sum_{t=1}^{\infty}\left(\sum_{i=1}^{m}(m+1)^{\nu_1/4}\mathbb{E}\sup_{\theta\in\Theta}|\sigma_{i,t}^2 - \tilde{\sigma}_{i,t}^2|^{\nu_1/4} + \sum_{\substack{i,j=1,\cdots,m\\ i<j}} \mathbb{E}\sup_{\theta\in\Theta}|\rho_{i,j}|^{\nu_1/4}|\sigma_{i,t}\sigma_{j,t} - \tilde{\sigma}_{i,t}\tilde{\sigma}_{j,t}|^{\nu_1/4}\right)^{1/2} \\
\leq& C_{17}\sum_{t=1}^{\infty}\left[m(m+1)^{\nu_1/4}a_6 b_1^{t\nu_1/4} + \frac{m(m-1)}{2}C_{12}b_1^{t\nu_1/4}\right]^{1/2} \\
\leq& \sum_{t=1}^{\infty} O(b_1^{t\nu_1/8}) \leq d_2 < \infty
\end{aligned} \tag{3.32}$$

To complete the proof of the strong consistency, for any $c > 0$, by Markov inequality, (3.31) and (3.32) a few lines above,

$$\begin{aligned}
&\sum_{t=1}^{\infty} P(\sup_{\theta\in\Theta}|l_t(\boldsymbol{\theta}) - \tilde{l}_t(\boldsymbol{\theta})| > c) \\
=& \sum_{t=1}^{\infty} P(\sup_{\theta\in\Theta}|l_t(\boldsymbol{\theta}) - \tilde{l}_t(\boldsymbol{\theta})|^{\nu_1/8} > c^{\nu_1/8}) \\
\leq& \sum_{t=1}^{\infty} \frac{\mathbb{E}\sup_{\theta\in\Theta}|l_t(\boldsymbol{\theta}) - \tilde{l}_t(\boldsymbol{\theta})|^{\nu_1/8}}{c^{\nu_1/8}} \\
\leq& \sum_{t=1}^{\infty} \frac{\mathbb{E}\sup_{\theta\in\Theta}|l_t(\boldsymbol{\theta}) - \tilde{l}_t(\boldsymbol{\theta})|^{\nu_1/8}}{c^{\nu_1/8}} \\
\leq& \sum_{t=1}^{\infty} \frac{\mathbb{E}[C_5\sup_{\theta\in\Theta}\left\|H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta})\right\|_p^{\nu_1/8} + C_4\sup_{\theta\in\Theta}\left\|H_t(\boldsymbol{\theta}) - \tilde{H}_t(\boldsymbol{\theta})\right\|_p^{\nu_1/8}\|\boldsymbol{x_t}\|_p^{\nu_1/4}]}{c^{\nu_1/8}}
\end{aligned}$$

$$\leq \frac{C_5 d_1}{c^{\nu_1/4}} + \frac{C_4 d_2}{c^{\nu_1/4}}$$

$$< \infty.$$

By the 1st Borel-Cantelli lemma, we could conclude that $\sup_{\boldsymbol{\theta}} |l_t(\boldsymbol{\theta}) - \tilde{l}_t(\boldsymbol{\theta})| \to 0$ a.s..

We have thus shown the desired result by the Cesàro mean theorem since

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

□

3.3 Asymptotic Normality

After establishing the strong consistency in the last section, it is time to move on to study the convergence speed of the quasi maximum likelihood estimator. More specifically, the distribution of the difference between the estimator $\hat{\boldsymbol{\theta}}_n$ and the true parameter $\boldsymbol{\theta}_0$ would approach a normal distribution with a certain speed under some conditions.

The convergence speed generally is a monotone increasing function of the sample size n. For instance, if $\{X_1, X_2, \ldots\}$ is a sequence of independently and identically distributed random varaibles with mean μ and finite variance σ^2. Then, by the classical central limit thoery, an estimator of the mean denoted by $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$, converges to a normal distrbution, i.e. $\sqrt{n}\,(\hat{\mu}_n - \mu) \xrightarrow{D} N(0, \sigma^2)$.

We can state another asymptotic result if the following assumptions are satisfied.

$C1$ The observed sequence $\boldsymbol{x}_t$ has a finite 8th moment.

$C2$ The parameter $\boldsymbol{\theta_0}$ is an interior point of $\boldsymbol{\Theta}$.

Theorem 3.3.1 (Asymptotic Normality)

Under Assumptions $B1 - B2$, $B4$ and $C1 - C2$, the QMLE $\hat{\boldsymbol{\theta}}_n$ defined by (3.5) has an asymptotic normal distribution around the true value $\boldsymbol{\theta}_0$, which means

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, J^{-1}VJ^{-1})$$

where

$$J = -\mathbb{E}\left[\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\intercal}\right] \text{ and } V = \mathbb{E}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^\intercal}\right].$$

Assumption $\boldsymbol{C}2$ is very common in proving the asymptotic distribution under some rate since the derivatives are needed. In Assumption $\boldsymbol{C}1$, the finite 8th moment requirement is very conservative because of our own convenience. Like Assumption 3.6 of Hafner and Preminger (2009b), the finite 6th moment would work if the corresponding neighbourhood $\upsilon(\boldsymbol{\theta}_0) \in \boldsymbol{\Theta}$ around $\boldsymbol{\theta}_0$ can be found such that for all i_1, i_2 and i_3,

$$\mathbb{E} \sup_{\boldsymbol{\theta}\in\upsilon(\boldsymbol{\theta}_0)} \left|\frac{\partial^3 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_{i_1}\partial\boldsymbol{\theta}_{i_2}\partial\boldsymbol{\theta}_{i_3}}\right| < \infty.$$

The finite 8th moment leads to the largest neighbourhood since $\mathbb{E}\left|\frac{\partial^3 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_{i_1}\partial\boldsymbol{\theta}_{i_2}\partial\boldsymbol{\theta}_{i_3}}\right| < \infty$ is true for any $\boldsymbol{\theta}$ in $\boldsymbol{\Theta}$.

3.3.1 Proof of Theorem 3.3.1

Under Assumptions $\boldsymbol{B}1 - \boldsymbol{B}4$, we have the strong consistency result above. For a point in any compact set around $\boldsymbol{\theta}_0$, the results in Lemma 3.2.3- 3.2.8 will follow. Therefore, we can always select a proper set $\upsilon(\boldsymbol{\theta}_0)$ for a large n such that $\hat{\boldsymbol{\theta}}_n \in \upsilon(\boldsymbol{\theta}_0)$.

The mean-value expansion is applied to the score function around the true parameter $\boldsymbol{\theta}_0$. Hence,

$$\begin{aligned} 0 =& \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t(\hat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}} \\ =& \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\intercal}\right)\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \end{aligned} \tag{3.33}$$

where $\tilde{\boldsymbol{\theta}}_n$ is on the chord between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$ and $\tilde{\boldsymbol{\theta}}$ is between $\tilde{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$.

The proof of this theorem will be divided into the following intermediate steps:

1. $\mathbb{E}\left\|\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\frac{\partial l_t(\boldsymbol{\theta}_0)^\intercal}{\partial \boldsymbol{\theta}}\right\| < \infty$.
2. $\mathbb{E}\left\|\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\intercal}\right\|$ is finite.
3. $\frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{D} N(\mathbf{0}, V)$.
4. $\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right| = o_p(1)$.
5. There exists a neighbourhood $\upsilon(\boldsymbol{\theta}_0) \in \boldsymbol{\Theta}$ around $\boldsymbol{\theta}_0$ such that
$$\sup_{\boldsymbol{\theta}\in\upsilon(\boldsymbol{\theta}_0)}\left|\frac{1}{n}\sum_{t=1}^n \frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\intercal} - \frac{\partial^2 \tilde{l}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\intercal}\right| = o_p(1).$$
6. There exists a neighbourhood $\upsilon(\boldsymbol{\theta}_0) \in \boldsymbol{\Theta}$ around $\boldsymbol{\theta}_0$ such that for all i_1, i_2 and i_3,
$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\upsilon(\boldsymbol{\theta}_0)}\left|\frac{\partial^3 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{i_1}\partial \boldsymbol{\theta}_{i_2}\partial \boldsymbol{\theta}_{i_3}}\right| < \infty.$$

Once again, L_2 vector norm and the corresponding induced matrix norm are used throughout this section. (7) in Appendix A tells us that all the inequalities will remain the same subject to a scale when the norm is changed to some other L_p norms. We will prove these steps one by one. Some notations are used in the following proof, including ${}_i\dot{H}_t = \frac{\partial H_t(\boldsymbol{\theta})}{\partial \theta_i}$, ${}_{ij}\dot{H}_t = \frac{\partial^2 H_t(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}$ and ${}_{ijk}\dot{H}_t = \frac{\partial^2 H_t(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j \partial \theta_k}$.

1. First derivative criterion.

   We will calculate the components of the score function, which are given by

$$\begin{aligned}\frac{\partial l_t(\boldsymbol{\theta})}{\partial \theta_i} &= \frac{\partial}{\partial \theta_i}\log|H_t(\boldsymbol{\theta})| + \frac{\partial}{\partial \theta_i} tr(\boldsymbol{x_t}' H_t(\boldsymbol{\theta})^{-1}\boldsymbol{x_t}) \\ &= |H_t^{-1}(\boldsymbol{\theta})|\frac{\partial}{\partial \theta_i}|H_t(\boldsymbol{\theta})| + tr(\boldsymbol{x_t}\boldsymbol{x_t}^\intercal \frac{\partial}{\partial \theta_i} H_t(\boldsymbol{\theta})^{-1})\end{aligned}$$

$$
\begin{aligned}
&= |H_t^{-1}(\boldsymbol{\theta})|\frac{\partial}{\partial\theta_i}|H_t(\boldsymbol{\theta})| + tr(\boldsymbol{x_t}\boldsymbol{x_t}^\top\frac{\partial}{\partial\theta_i}H_t(\boldsymbol{\theta})^{-1}) \\
&= tr(H_t^{-1}(\boldsymbol{\theta})\,{}_i\dot{H}_t(\boldsymbol{\theta})) + tr(\boldsymbol{x_t}\boldsymbol{x_t}^\top H_t(\boldsymbol{\theta})^{-1}\,{}_i\dot{H}_t(\boldsymbol{\theta})H_t(\boldsymbol{\theta_0})^{-1}) \\
&= tr[(\mathbb{I}_m - \boldsymbol{x_t}\boldsymbol{x_t}^\top H_t(\boldsymbol{\theta})^{-1})\,{}_i\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})]. \qquad (3.34)
\end{aligned}
$$

Therefore, when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\Big|\mathcal{F}_{t-1}\right] \\
=&tr[(\mathbb{I}_m - \mathbb{E}\left(H_t^{1/2}(\boldsymbol{\theta}_0)\boldsymbol{\xi_t}\boldsymbol{\xi_t}^\top H_t^{1/2}(\boldsymbol{\theta}_0)\Big|\mathcal{F}_{t-1}\right)H_t(\boldsymbol{\theta}_0)^{-1})\,{}_i\dot{H}_t(\boldsymbol{\theta}_0)H_t^{-1}(\boldsymbol{\theta}_0)] \\
=&tr[(\mathbb{I}_m - \mathbb{I}_m)\,{}_i\dot{H}_t(\boldsymbol{\theta}_0)H_t^{-1}(\boldsymbol{\theta}_0)] = 0. \qquad (3.35)
\end{aligned}
$$

Now we move on to prove that $\mathbb{E}\left\|\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial l_t(\boldsymbol{\theta}_0)^\top}{\partial\boldsymbol{\theta}}\right\|$ is finite. The elements in the target matrix are

$$
\begin{aligned}
&\mathbb{E}\left|\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\theta_i}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\theta_j}\right| \\
\le& C_1\mathbb{E}(\left\|\mathbb{I}_m - \boldsymbol{x_t}\boldsymbol{x_t}^\top H_t(\boldsymbol{\theta}_0)^{-1}\right\|^2 \left\|{}_i\dot{H}_t(\boldsymbol{\theta}_0)\right\| \left\|{}_j\dot{H}_t(\boldsymbol{\theta}_0)\right\|) \\
\le& 2C_1\mathbb{E}\left[\left(1 + \left\|H_t(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{\xi_t}\boldsymbol{\xi_t}^\top H_t(\boldsymbol{\theta}_0)^{-1/2}\right\|^2\right)\left\|{}_i\dot{H}_t(\boldsymbol{\theta}_0)\right\| \left\|{}_j\dot{H}_t(\boldsymbol{\theta}_0)\right\|\right] \\
\le& 2C_1\left(1 + \mathbb{E}\left\|\boldsymbol{\xi_t}\boldsymbol{\xi_t}^\top\right\|^2\right)\mathbb{E}\left(\left\|{}_i\dot{H}_t(\boldsymbol{\theta}_0)\right\| \left\|{}_j\dot{H}_t(\boldsymbol{\theta}_0)\right\|\right) \\
\le& C_2\left[\mathbb{E}\left(\left\|{}_i\dot{H}_t(\boldsymbol{\theta}_0)\right\|\right)^2\right]^{1/2}\left[\mathbb{E}\left(\left\|{}_j\dot{H}_t(\boldsymbol{\theta}_0)\right\|\right)^2\right]^{1/2} \\
<& \infty
\end{aligned}
$$

Lemma 3.2.8 has proved that $\left\|H_t(\boldsymbol{\theta}_0)^{-1}\right\| < \infty$. This result along with (8) in Appendix A bring us the first inequality. The second inequality uses the C_r inequality and the fact that the model follows (3.2) at $\boldsymbol{\theta}_0$. We note that $\left\|H_t(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{\xi_t}\boldsymbol{\xi_t}^\top H_t(\boldsymbol{\theta}_0)^{-1/2}\right\|^2 \le C_1\left\|\boldsymbol{\xi_t}\boldsymbol{\xi_t}^\top\right\|^2 \le C_2\left\|\boldsymbol{\xi_t}\right\|^4$ a.s.. The third inequality is obtained by this result, the independence between $\boldsymbol{\xi_t}$ and H_t, and also the independence between $\boldsymbol{\xi_t}$ and H_t's derivatives. The second last line was implied by the Cauchy-Schwarz inequality and Assumption $\boldsymbol{C}1$. The last line results from Lemma 3.3.2.

2. By (19) and (20) in Appendix A, the second derivative with respect to the $\boldsymbol{i}$th and $\boldsymbol{j}$th parameter can be rewritten as following

$$
\begin{aligned}
&\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\\
=&\frac{\partial}{\partial\boldsymbol{\theta}_j}tr[(\mathbb{I}_m-\boldsymbol{x_t}\boldsymbol{x_t}^\intercal H_t(\boldsymbol{\theta})^{-1})\,{}_i\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})].\\
=&tr[\,{}_{ij}\ddot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})+{}_i\dot{H}_t(\boldsymbol{\theta})\frac{H_t^{-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}-\boldsymbol{x_t}\boldsymbol{x_t}^\intercal(\frac{H_t^{-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}\,{}_i\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})\\
&+H_t^{-1}(\boldsymbol{\theta})\,{}_{ij}\ddot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})+H_t^{-1}(\boldsymbol{\theta})\,{}_i\dot{H}_t(\boldsymbol{\theta})\frac{\partial H_t^{-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j})]\\
=&tr[\,{}_{ij}\ddot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})+{}_i\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})\,{}_j\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})+\boldsymbol{x_t}\boldsymbol{x_t}^\intercal H_t^{-1}(\boldsymbol{\theta})\\
&\left({}_j\dot{H}_t(\boldsymbol{\theta})\,{}_i\dot{H}_t(\boldsymbol{\theta})-{}_{ij}\ddot{H}_t(\boldsymbol{\theta})+{}_i\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})\,{}_j\dot{H}_t(\boldsymbol{\theta})\right)H_t^{-1}(\boldsymbol{\theta})]\\
=&tr[(\mathbb{I}_m-\boldsymbol{x_t}\boldsymbol{x_t}^\intercal H_t^{-1}(\boldsymbol{\theta}))({}_{ij}\ddot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})-{}_i\dot{H}_tH_t^{-1}(\boldsymbol{\theta})\,{}_j\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta}))\\
&+\boldsymbol{x_t}\boldsymbol{x_t}^\intercal H_t^{-1}(\boldsymbol{\theta})\,{}_i\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})\,{}_j\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})]
\end{aligned}
\tag{3.36}
$$

Using the fact that $\boldsymbol{\xi_t}$ and H_t are independent given the past, the expectation of the same element is

$$
\begin{aligned}
&\mathbb{E}\left|\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\right|\\
=&\mathbb{E}\Big\{tr[(\mathbb{I}_m-\boldsymbol{x_t}\boldsymbol{x_t}^\intercal H_t^{-1}(\boldsymbol{\theta}))({}_{ij}\ddot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})-{}_i\dot{H}_tH_t^{-1}(\boldsymbol{\theta})\,{}_j\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta}))\\
&+\boldsymbol{x_t}\boldsymbol{x_t}^\intercal\,{}_i\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})\,{}_j\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})]\Big\}\\
\le&C_1\mathbb{E}\left\|(\mathbb{I}_m-\boldsymbol{x_t}\boldsymbol{x_t}^\intercal H_t^{-1}(\boldsymbol{\theta}))\left({}_{ij}\ddot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})-{}_i\dot{H}_tH_t^{-1}(\boldsymbol{\theta})\,{}_j\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})\right)\right\|\\
&+C_1\mathbb{E}\left\|\boldsymbol{x_t}\boldsymbol{x_t}^\intercal\,{}_i\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})\,{}_j\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})\right\|\\
\le&C_2\mathbb{E}\Big[\,\|\boldsymbol{x_t}\boldsymbol{x_t}^\intercal\|\,(\left\|{}_{ij}\ddot{H}_t(\boldsymbol{\theta})\right\|+2\left\|{}_i\dot{H}_t(\boldsymbol{\theta})\right\|\left\|{}_j\dot{H}_t(\boldsymbol{\theta})\right\|)\Big]\\
&+C_3\mathbb{E}\left\|{}_{ij}\ddot{H}_t(\boldsymbol{\theta})\right\|+C_4\mathbb{E}\Big(\left\|{}_i\dot{H}_t(\boldsymbol{\theta})\right\|\left\|{}_j\dot{H}_t(\boldsymbol{\theta})\right\|\Big)\\
\le&2C_2\Big(\mathbb{E}\,\|\boldsymbol{x_t}\|^6\Big)^{1/3}\Big(\mathbb{E}\left\|{}_i\dot{H}_t(\boldsymbol{\theta})\right\|^3\Big)^{1/3}\Big(\mathbb{E}\left\|{}_j\dot{H}_t(\boldsymbol{\theta})\right\|^3\Big)^{1/3}
\end{aligned}
$$

$$
\begin{aligned}
&+ C_2(\mathbb{E}\,\|\boldsymbol{x_t}\|^4)^{1/2}(\mathbb{E}\left\|{}_{ij}\ddot{H}_t(\boldsymbol{\theta})\right\|^2)^{1/2} + C_3\mathbb{E}\left\|{}_{ij}\ddot{H}_t(\boldsymbol{\theta})\right\| \\
&+ C_4\left(\mathbb{E}\left\|{}_{j}\dot{H}_t(\boldsymbol{\theta})\right\|^2\right)^{1/2}\left(\mathbb{E}\left\|{}_{i}\dot{H}_t(\boldsymbol{\theta})\right\|^2\right)^{1/2} < \infty
\end{aligned}
\tag{3.37}
$$

The first inequality holds because of Minkowski's inequality and the second inequality results from that $\left\|H_t^{-1}(\boldsymbol{\theta})\right\|$ has an upper bound. By repeatedly using the Cauchy-Schwarz inequality, we can get the third inequality. Assumption $\boldsymbol{C}1$ and (ii) in Lemma 3.3.3 lead to the finiteness result in the end.

3. The results in the previous two parts imply the existence of the matrix V and the asymptotic normality of the score function. Similar to the argument in Section 3.2.1, an extension of the martingale central limit theorem in Billingsley (1961) can be applied here because that $\dfrac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ is stationary and ergodic, see page 61 of Jiang (2011) for more details. The desired result follows.

4. By the generalized Chebyshev inequality and the $\boldsymbol{C_r}$ inequality we have for $\epsilon > 0$ and $1/4 > v_2 > 0$,

$$
P(\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right| \geq \epsilon) \leq n^{-0.5v_2}\sum_{t=1}^{n}\mathbb{E}\left|\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right|^{v_2}. \tag{3.38}
$$

Hence, it is sufficient to show that for some $v_2 > 0$, the summation is finite.

Based on the score function in (3.34), the terms on the right hand side can be written as ($\boldsymbol{\theta}_0$ was dropped from the equations for simplicity)

$$
\begin{aligned}
&\mathbb{E}\left|\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right|^{v_2} \\
\leq & C_1\mathbb{E}\left\|\left(\mathbb{I}_m - \boldsymbol{x_t}\boldsymbol{x_t}^\intercal H_t^{-1}\right){}_i\dot{H}_t H_t - \left(\mathbb{I}_m - \boldsymbol{x_t}\boldsymbol{x_t}^\intercal \tilde{H}_t^{-1}\right){}_i\dot{\tilde{H}}_t\tilde{H}_t\right\|^{v_2} \\
= & C_1\mathbb{E}\Big\|\left(\mathbb{I}_m - \boldsymbol{x_t}\boldsymbol{x_t}^\intercal \tilde{H}_t^{-1}\right)\left({}_i\dot{\tilde{H}}_t\tilde{H}_t^{-1} - {}_i\dot{H}_t H_t^{-1}\right) \\
& \quad + \boldsymbol{x_t}\boldsymbol{x_t}^\intercal\left(H_t^{-1} - \tilde{H}_t^{-1}\right){}_i\dot{H}_t H_t^{-1}\Big\|^{v_2} \\
= & C_1\mathbb{E}\Big\|\left(\mathbb{I}_m - \boldsymbol{x_t}\boldsymbol{x_t}^\intercal \tilde{H}_t^{-1}\right)\left[\left({}_i\dot{\tilde{H}}_t - {}_i\dot{H}_t\right)\tilde{H}_t^{-1} + {}_i\dot{H}_t\left(\tilde{H}_t^{-1} - H_t^{-1}\right)\right]
\end{aligned}
$$

$$+ \boldsymbol{x_t}\boldsymbol{x_t}^\intercal \tilde{H}_t^{-1}\left(\tilde{H}_t - H_t\right) H_t^{-1}\, {}_i\dot{H}_t H_t^{-1}\Big\|^{v_2}$$

From the C_r inequality, the Cauchy inequality and both $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\tilde{H}_t^{-1}(\boldsymbol{\theta})\right\|$ and $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\tilde{H}_t^{-1}(\boldsymbol{\theta})\right\|$ are bounded by constants by the proof in Lemma 3.2.8, we have

$$\begin{aligned}
&\mathbb{E}\left|\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} - \frac{\partial\tilde{l}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right|^{v_2}\\
\le& C_1\mathbb{E}\left\|(\mathbb{I}_m - \boldsymbol{x_t}\boldsymbol{x_t}^\intercal\tilde{H}_t^{-1})\left[\left({}_i\dot{\tilde{H}}_t - {}_i\dot{H}_t\right)\tilde{H}_t^{-1} + {}_i\dot{H}_t H_t^{-1}\left(H_t - \tilde{H}_t\right)\tilde{H}_t^{-1}\right]\right\|^{v_2}\\
&+ C_2\mathbb{E}\left\|\boldsymbol{x_t}\boldsymbol{x_t}^\intercal\tilde{H}_t^{-1}(\tilde{H}_t - H_t)H_t^{-1}\,{}_i\dot{H}_t H_t^{-1}\right\|^{v_2}\\
\le& C_1\mathbb{E}[\left(C_3 + C_4\|\boldsymbol{x_t}\|^{2v_2}\right)\left(\|{}_i\dot{\tilde{H}}_t - {}_i\dot{H}_t\|^{v_2} + \left\|{}_i\dot{H}_t\right\|^{v_2}\|H_t - \tilde{H}_t\|^{v_2}\right)]\\
&+ C_5\mathbb{E}\left[\left\|\boldsymbol{x_t}\right\|^{2v_2}\|\tilde{H}_t - H_t\right\|^{v_2}\left\|{}_i\dot{H}_t\right\|^{v_2}]\\
\le& C_1C_3\mathbb{E}\left\|{}_i\dot{\tilde{H}}_t - {}_i\dot{H}_t\right\|^{v_2} + C_1C_4\left(\mathbb{E}\left\|\boldsymbol{x_t}\right\|^{4v_2}\right)^{1/2}\left(\mathbb{E}\left\|{}_i\dot{H}_t\right\|^{4v_2}\right)^{1/4}(\mathbb{E}\left\|H_t - \tilde{H}_t\right\|^{4v_2})^{1/4}\\
&+ C_1C_3\left(\mathbb{E}\left\|{}_i\dot{H}_t\right\|^{2v_2}\right)^{1/2}\left(\mathbb{E}\left\|H_t - \tilde{H}_t\right\|^{2v_2}\right)^{1/2} + C_1C_4\left(\mathbb{E}\left\|\boldsymbol{x_t}\right\|^{4v_2}\right)^{1/2}\left(\mathbb{E}\left\|{}_i\dot{\tilde{H}}_t - {}_i\dot{H}_t\right\|^{2v_2}\right)^{1/2}\\
&+ C_5\mathbb{E}\left(\left\|\boldsymbol{x_t}\right\|^{4v_2}\right)^{1/2}\left(\mathbb{E}\left\|\tilde{H}_t - H_t\right\|^{4v_2}\right)^{1/4}\left(\mathbb{E}\left\|{}_i\dot{H}_t\right\|^{4v_2}\right)^{1/4}
\end{aligned}\tag{3.39}$$

Similar to the arguments when we prove Lemma 3.2.8, for some $v_2 > 0$, $\mathbb{E}\|\boldsymbol{x_t}\|^{4v_2} < \infty$. Using the results in Lemma 3.3.3 and Lemma 3.3.4 and the Markov inequality, we can tell, for the same $v_2 > 0$, that $\mathbb{E}\|{}_i\dot{H}_t\|^{4v_2} < \infty$, $\sum_{t=1}^{\infty}\mathbb{E}\|{}_i\dot{\tilde{H}}_t - {}_i\dot{H}_t\|^{v_2} < \infty$ and $\sum_{t=1}^{\infty}\left(\mathbb{E}\|{}_i\dot{\tilde{H}}_t - {}_i\dot{H}_t\|^{2v_2}\right)^{1/2} < \infty$.

Then, the finiteness of the summation in (3.38) has been verified.

$$\sum_{t=1}^{\infty}\mathbb{E}\left|\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} - \frac{\partial\tilde{l}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right|^{v_2} < \infty \tag{3.40}$$

This leads to that $\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} - \frac{\partial\tilde{l}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right|$ converges to 0 almost surely. Therefore, $\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} - \frac{\partial\tilde{l}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right| = o_p(1)$.

5. We can use the result in (3.37) to get the expectation of the difference between $l_t(\boldsymbol{\theta})$

and $\tilde{l}_t(\boldsymbol{\theta})$. If $0 < v_3 < 1/4$, then

$$
\begin{aligned}
&\mathbb{E}\Big|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j} - \frac{\partial^2 \tilde{l}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\Big|^{v_3} \\
=&\mathbb{E}tr\Big[\Big(\mathbb{I}_m - \boldsymbol{x}_t\boldsymbol{x}_t^\intercal H_t^{-1}\Big)\Big({}_{ij}\ddot{H}_t H_t^{-1} - {}_i\dot{H}_t H_t^{-1}\,{}_j\dot{H}_t H_t^{-1}\Big) \\
&- \Big(\mathbb{I}_m - \boldsymbol{x}_t\boldsymbol{x}_t^\intercal \tilde{H}_t^{-1}\Big)\Big({}_{ij}\ddot{\tilde{H}}_t \tilde{H}_t^{-1} - {}_i\dot{\tilde{H}}_t \tilde{H}_t^{-1}\,{}_j\dot{\tilde{H}}_t \tilde{H}_t^{-1}\Big) \\
&+ \boldsymbol{x}_t\boldsymbol{x}_t^\intercal H_t^{-1}\,{}_j\dot{H}_t H_t^{-1}\,{}_i\dot{H}_t H_t^{-1} - \boldsymbol{x}_t\boldsymbol{x}_t^\intercal \tilde{H}_t^{-1}\,{}_j\dot{\tilde{H}}_t \tilde{H}_t^{-1}\,{}_i\dot{\tilde{H}}_t \tilde{H}_t^{-1}\Big] \\
\leq& C_1\mathbb{E}\Big\|\Big[\Big(\mathbb{I}_m - \boldsymbol{x}_t\boldsymbol{x}_t^\intercal H_t^{-1}\Big)\Big({}_{ij}\ddot{H}_t H_t^{-1} - {}_i\dot{H}_t H_t^{-1}\,{}_j\dot{H}_t H_t^{-1}\Big) \\
&- \Big(\mathbb{I}_m - \boldsymbol{x}_t\boldsymbol{x}_t^\intercal \tilde{H}_t^{-1}\Big)\Big({}_{ij}\ddot{\tilde{H}}_t \tilde{H}_t^{-1} - {}_i\dot{\tilde{H}}_t \tilde{H}_t^{-1}\,{}_j\dot{\tilde{H}}_t \tilde{H}_t^{-1}\Big) \\
&+ \boldsymbol{x}_t\boldsymbol{x}_t^\intercal H_t^{-1}\,{}_j\dot{H}_t H_t^{-1}\,{}_i\dot{H}_t H_t^{-1} - \boldsymbol{x}_t\boldsymbol{x}_t^\intercal \tilde{H}_t^{-1}\,{}_j\dot{\tilde{H}}_t \tilde{H}_t^{-1}\,{}_i\dot{\tilde{H}}_t \tilde{H}_t^{-1}\Big]\Big\| \\
\leq& C_1\mathbb{E}\Big\{\Big\|\Big(\mathbb{I}_m - \boldsymbol{x}_t\boldsymbol{x}_t^\intercal H_t^{-1}\Big)\,{}_{ij}\ddot{H}_t H_t^{-1} - \Big(\mathbb{I}_m - \boldsymbol{x}_t\boldsymbol{x}_t^\intercal \tilde{H}_t^{-1}\Big)\,{}_{ij}\ddot{\tilde{H}}_t \tilde{H}_t^{-1}\Big\| \\
&+ \Big\|\Big(\mathbb{I}_m - \boldsymbol{x}_t\boldsymbol{x}_t^\intercal H_t^{-1}\Big)\Big({}_i\dot{H}_t H_t^{-1}\,{}_j\dot{H}_t H_t^{-1} - {}_i\dot{\tilde{H}}_t \tilde{H}_t^{-1}\,{}_j\dot{\tilde{H}}_t \tilde{H}_t^{-1}\Big) \\
&- \boldsymbol{x}_t\boldsymbol{x}_t^\intercal\Big(H_t^{-1} - \tilde{H}_t^{-1}\Big)\,{}_i\dot{\tilde{H}}_t \tilde{H}_t^{-1}\,{}_j\dot{\tilde{H}}_t \tilde{H}_t^{-1}\Big\| \\
&+ \Big\|\boldsymbol{x}_t\boldsymbol{x}_t^\intercal \tilde{H}_t^{-1}\Big({}_i\dot{\tilde{H}}_t \tilde{H}_t^{-1}\,{}_j\dot{\tilde{H}}_t \tilde{H}_t^{-1} - {}_i\dot{H}_t H_t^{-1}\,{}_j\dot{H}_t H_t^{-1}\Big) \\
&- \boldsymbol{x}_t\boldsymbol{x}_t^\intercal\Big(H_t^{-1} - \tilde{H}_t^{-1}\Big)\,{}_i\dot{H}_t H_t^{-1}\,{}_j\dot{H}_t H_t^{-1}\Big\|\Big\}^{v_3} \\
\leq& C_1\Big\{\mathbb{E}\Big\|\Big(\mathbb{I}_m - \boldsymbol{x}_t\boldsymbol{x}_t^\intercal \tilde{H}_t^{-1}\Big)\Big({}_{ij}\ddot{H}_t H_t^{-1} - {}_{ij}\ddot{\tilde{H}}_t \tilde{H}_t^{-1}\Big) + \boldsymbol{x}_t\boldsymbol{x}_t^\intercal\Big(\tilde{H}_t^{-1} - H_t^{-1}\Big)\,{}_{ij}\ddot{\tilde{H}}_t \tilde{H}_t^{-1}\Big\|\Big\}^{v_3} \\
&+ \mathbb{E}\Big\|\Big(\mathbb{I}_m - \boldsymbol{x}_t\boldsymbol{x}_t^\intercal \tilde{H}_t^{-1}\Big)\Big[\Big({}_i\dot{H}_t H_t^{-1} - {}_i\dot{\tilde{H}}_t \tilde{H}_t^{-1}\Big)\,{}_j\dot{\tilde{H}}_t \tilde{H}_t^{-1} + {}_i\dot{H}_t H_t^{-1}\Big({}_j\dot{\tilde{H}}_t \tilde{H}_t^{-1} - {}_j\dot{H}_t H_t^{-1}\Big)\Big]\Big\|^{v_3} \\
&+ \mathbb{E}\Big\|\boldsymbol{x}_t\boldsymbol{x}_t^\intercal \tilde{H}_t^{-1}\Big(H_t - \tilde{H}_t\Big)H_t^{-1}\,{}_i\dot{H}_t H_t^{-1}\,{}_j\dot{\tilde{H}}_t H_t^{-1}\Big\|^{v_3} \\
&+ \mathbb{E}\Big\|\boldsymbol{x}_t\boldsymbol{x}_t^\intercal H_t^{-1}({}_j\dot{H}_t H_t^{-1} - {}_j\dot{\tilde{H}}_t \tilde{H}_t^{-1})\,{}_i\dot{H}_t H_t^{-1} - {}_j\dot{\tilde{H}}_t \tilde{H}_t^{-1}({}_i\dot{\tilde{H}}_t \tilde{H}_t^{-1} - {}_i\dot{H}_t H_t^{-1})\Big\| \\
&+ \mathbb{E}\Big\|\boldsymbol{x}_t\boldsymbol{x}_t^\intercal \tilde{H}_t^{-1}\Big(H_t - \tilde{H}_t\Big)H_t^{-1}\,{}_j\dot{\tilde{H}}_t \tilde{H}_t^{-1}\,{}_i\dot{\tilde{H}}_t \tilde{H}_t^{-1}\Big\|^{v_3} \\
\leq& C_1\mathbb{E}\Big\|\Big(\mathbb{I}_m - \boldsymbol{x}_t\boldsymbol{x}_t^\intercal \tilde{H}_t^{-1}\Big)\Big[({}_{ij}\ddot{H}_t - {}_{ij}\ddot{\tilde{H}}_t\Big)H_t^{-1} + {}_{ij}\ddot{\tilde{H}}_t \tilde{H}_t^{-1}(H_t - \tilde{H}_t)H_t^{-1}\Big]\Big\|^{v_3} \\
&+ \mathbb{E}\Big\|\boldsymbol{x}_t\boldsymbol{x}_t^\intercal \tilde{H}_t^{-1}(H_t - \tilde{H}_t)H_t^{-1}\,{}_{ij}\ddot{\tilde{H}}_t \tilde{H}_t^{-1}\Big\|^{v_3} \\
&+ \mathbb{E}\Big\|\Big(\mathbb{I}_m - \boldsymbol{x}_t\boldsymbol{x}_t^\intercal \tilde{H}_t^{-1}\Big)\Big[({}_i\dot{H}_t - {}_i\dot{\tilde{H}}_t)H_t^{-1} + {}_i\dot{H}_t \tilde{H}_t^{-1}(H_t - \tilde{H}_t)H_t^{-1}\Big]\,{}_j\dot{\tilde{H}}_t \tilde{H}_t^{-1}\Big\|^{v_3} \\
&+ \mathbb{E}\Big\|\Big(\mathbb{I}_m - \boldsymbol{x}_t\boldsymbol{x}_t^\intercal H_t^{-1}\Big)\,{}_i\dot{H}_t H_t^{-1}[({}_j\dot{H}_t - {}_j\dot{\tilde{H}}_t)H_t^{-1} + {}_j\dot{\tilde{H}}_t \tilde{H}_t^{-1}(H_t - \tilde{H}_t)H_t^{-1}]\Big\|^{v_3} \\
&+ \mathbb{E}\Big\|\boldsymbol{x}_t\boldsymbol{x}_t^\intercal \tilde{H}_t^{-1}(H_t - \tilde{H}_t)H_t^{-1}\,{}_i\dot{H}_t H_t^{-1}\,{}_j\dot{\tilde{H}}_t H_t^{-1}\Big\|^{v_3}
\end{aligned}
$$

$$
\begin{aligned}
&+\mathbb{E}\Big\|\boldsymbol{x}_t\boldsymbol{x}_t^\intercal H_t^{-1}[({}_j\dot{H}_t - {}_i\dot{\tilde{H}}_t)H_t^{-1} + {}_j\dot{\tilde{H}}_t\tilde{H}_t^{-1}(\tilde{H}_t - H_t)H_t^{-1}]\,{}_i\dot{H}_tH_t^{-1}\Big\|^{v_3}\\
&+\mathbb{E}\Big\|\boldsymbol{x}_t\boldsymbol{x}_t^\intercal H_t^{-1}\,{}_j\dot{\tilde{H}}_t\tilde{H}_t^{-1}[({}_j\dot{\tilde{H}}_t - {}_j\dot{H}_t)H_t^{-1} + {}_i\dot{\tilde{H}}_t\tilde{H}_t^{-1}(H_t - \tilde{H}_t)H_t^{-1}]\Big\|^{v_3}\\
&+\mathbb{E}\Big\|\boldsymbol{x}_t\boldsymbol{x}_t^\intercal \tilde{H}_t^{-1}(H_t - \tilde{H}_t)H_t^{-1}\,{}_j\dot{\tilde{H}}_t\tilde{H}_t^{-1}\,{}_i\dot{\tilde{H}}_t\tilde{H}_t^{-1}\Big\|^{v_3}\\
\le& C_2\mathbb{E}\Big\|{}_{ij}\ddot{H}_t - {}_{ij}\ddot{\tilde{H}}_t\Big\|^{v_3} + C_3\mathbb{E}\Big\|\boldsymbol{x}_t\Big\|^{2v_3}\Big\|{}_{ij}\ddot{H}_t - {}_{ij}\ddot{\tilde{H}}_t\Big\|^{v_3} + C_4\mathbb{E}\Big\|\tilde{H}_t - H_t\Big\|^{v_3}\Big\|{}_{ij}\ddot{\tilde{H}}_t\Big\|^{v_3}\\
&+ C_5\mathbb{E}\Big\|\boldsymbol{x}_t\Big\|^{2v_3}\Big\|\tilde{H}_t - H_t\Big\|^{v_3}\Big\|{}_{ij}\ddot{\tilde{H}}_t\Big\|^{v_3} + C_6\mathbb{E}\Big\|\boldsymbol{x}_t\Big\|^{2v_3}\Big\|{}_i\dot{H}_t - {}_i\dot{\tilde{H}}_t\Big\|^{v_3}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{v_3}\\
&+ C_7\mathbb{E}\Big\|({}_i\dot{H}_t - {}_i\dot{\tilde{H}}_t)\Big\|^{v_3}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{v_3} + C_8\mathbb{E}\Big\|{}_i\dot{\tilde{H}}_t\Big\|^{v_3}\Big\|H_t - \tilde{H}_t\Big\|^{v_3}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{v_3}\\
&+ C_9\mathbb{E}\Big\|\boldsymbol{x}_t\Big\|^{2v_3}\Big\|{}_i\dot{\tilde{H}}_t\Big\|^{v_3}\Big\|H_t - \tilde{H}_t\Big\|^{v_3}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{v_3} + C_{10}\mathbb{E}\Big\|{}_i\dot{H}_t\Big\|^{v_3}\Big\|{}_j\dot{H}_t - {}_j\dot{\tilde{H}}_t\Big\|^{v_3}\\
&+ C_{11}\mathbb{E}\Big\|{}_i\dot{H}_t\Big\|^{v_3}\Big\|{}_j\dot{H}_t - {}_j\dot{\tilde{H}}_t\Big\|^{v_3}\Big\|\boldsymbol{x}_t\Big\|^{2v_3} + C_{12}\mathbb{E}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{v_3}\Big\|H_t - \tilde{H}_t\Big\|^{v_3}\\
&+ C_{13}\mathbb{E}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{v_3}\Big\|H_t - \tilde{H}_t\Big\|^{v_3}\Big\|\boldsymbol{x}_t\Big\|^{2v_3} + C_{14}\mathbb{E}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{v_3}\Big\|H_t - \tilde{H}_t\Big\|^{v_3}\Big\|\boldsymbol{x}_t\Big\|^{2v_3}\Big\|{}_i\dot{\tilde{H}}_t\Big\|^{v_3}\\
&+ C_{15}\mathbb{E}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{v_3}\Big\|H_t - \tilde{H}_t\Big\|^{v_3}\Big\|\boldsymbol{x}_t\Big\|^{2v_3}\Big\|{}_i\dot{H}_t\Big\|^{v_3}.
\end{aligned}
\tag{3.41}
$$

By using Assumption $\boldsymbol{C}1$ and Cauchy-Schwarz inequality repeatedly, the summation can be rewritten as

$$
\begin{aligned}
&\sum_{t=1}^{\infty}\mathbb{E}\Big|\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j} - \frac{\partial^2 \tilde{l}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\Big|^{v_3}\\
\le&\sum_{t=1}^{\infty} C_2\mathbb{E}\Big\|{}_{ij}\ddot{H}_t - {}_{ij}\ddot{\tilde{H}}_t\Big\|^{v_3} + C_3\Big(\mathbb{E}\Big\|\boldsymbol{x}_t\Big\|^{4v_3}\mathbb{E}\Big\|{}_{ij}\ddot{H}_t - {}_{ij}\ddot{\tilde{H}}_t\Big\|^{2v_3}\Big)^{1/2}\\
&+ C_4\Big(\mathbb{E}\Big\|\tilde{H}_t - H_t\Big\|^{2v_3}\mathbb{E}\Big\|{}_{ij}\ddot{\tilde{H}}_t\Big\|^{2v_3}\Big)^{1/2} + C_5\Big(\mathbb{E}\Big\|\boldsymbol{x}_t\Big\|^{6v_3}\mathbb{E}\Big\|\tilde{H}_t - H_t\Big\|^{3v_3}\mathbb{E}\Big\|{}_{ij}\ddot{\tilde{H}}_t\Big\|^{3v_3}\Big)^{1/3}\\
&+ C_6\Big(\mathbb{E}\Big\|\boldsymbol{x}_t\Big\|^{6v_3}\mathbb{E}\Big\|{}_i\dot{H}_t - {}_i\dot{\tilde{H}}_t\Big\|^{3v_3}\mathbb{E}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{3v_3}\Big)^{1/3} + C_7\Big(\mathbb{E}\Big\|({}_i\dot{H}_t - {}_i\dot{\tilde{H}}_t)\Big\|^{2v_3}\mathbb{E}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{2v_3}\Big)^{1/2}\\
&+ C_8\Big(\mathbb{E}\Big\|{}_i\dot{\tilde{H}}_t\Big\|^{3v_3}\mathbb{E}\Big\|H_t - \tilde{H}_t\Big\|^{3v_3}\mathbb{E}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{3v_3}\Big)^{1/3} + C_{10}\Big(\mathbb{E}\Big\|{}_i\dot{H}_t\Big\|^{2v_3}\mathbb{E}\Big\|{}_j\dot{H}_t - {}_j\dot{\tilde{H}}_t\Big\|^{2v_3}\Big)^{1/2}\\
&+ C_9\Big(\mathbb{E}\Big\|\boldsymbol{x}_t\Big\|^{8v_3}\mathbb{E}\Big\|{}_i\dot{\tilde{H}}_t\Big\|^{4v_3}\mathbb{E}\Big\|H_t - \tilde{H}_t\Big\|^{4v_3}\mathbb{E}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{4v_3}\Big)^{1/4}\\
&+ C_{11}\Big(\mathbb{E}\Big\|{}_i\dot{H}_t\Big\|^{3v_3}\mathbb{E}\Big\|{}_j\dot{H}_t - {}_j\dot{\tilde{H}}_t\Big\|^{3v_3}\mathbb{E}\Big\|\boldsymbol{x}_t\Big\|^{6v_3}\Big)^{1/3} + C_{12}\Big(\mathbb{E}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{2v_3}\mathbb{E}\Big\|H_t - \tilde{H}_t\Big\|^{2v_3}\Big)^{1/2}\\
&+ C_{13}\Big(\mathbb{E}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{3v_3}\mathbb{E}\Big\|H_t - \tilde{H}_t\Big\|^{3v_3}\mathbb{E}\Big\|\boldsymbol{x}_t\Big\|^{6v_3}\Big)^{1/3}\\
&+ C_{14}\Big(\mathbb{E}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{4v_3}\mathbb{E}\Big\|H_t - \tilde{H}_t\Big\|^{4v_3}\mathbb{E}\Big\|\boldsymbol{x}_t\Big\|^{8v_3}\mathbb{E}\Big\|{}_i\dot{\tilde{H}}_t\Big\|^{4v_3}\Big)^{1/4}\\
&+ C_{15}\Big(\mathbb{E}\Big\|{}_j\dot{\tilde{H}}_t\Big\|^{4v_3}\mathbb{E}\Big\|H_t - \tilde{H}_t\Big\|^{4v_3}\mathbb{E}\Big\|\boldsymbol{x}_t\Big\|^{8v_3}\mathbb{E}\Big\|{}_i\dot{H}_t\Big\|^{4v_3}\Big)^{1/4}
\end{aligned}
$$

$$\leq \sum_{t=1}^{\infty} k_1 \mathbb{E}\left\| {}_{ij}\ddot{H}_t - {}_{ij}\ddot{\tilde{H}}_t \right\|^{v_3} + k_2\Big(\mathbb{E}\left\| {}_{ij}\ddot{H}_t - {}_{ij}\ddot{\tilde{H}}_t \right\|^{2v_3}\Big)^{1/2} + k_3\Big(\mathbb{E}\left\| \tilde{H}_t - H_t \right\|^{2v_3}\Big)^{1/2}$$
$$+ k_4\Big(\mathbb{E}\left\| {}_{i}\dot{H}_t - {}_{i}\dot{\tilde{H}}_t \right\|^{3v_3}\Big)^{1/3} + k_5\Big(\mathbb{E}\left\| \tilde{H}_t - H_t \right\|^{3v_3}\Big)^{1/3} + k_6\Big(\mathbb{E}\left\| {}_{i}\dot{H}_t - {}_{i}\dot{\tilde{H}}_t \right\|^{2v_3}\Big)^{1/2}$$
$$+ k_7\Big(\mathbb{E}\left\| {}_{j}\dot{H}_t - {}_{j}\dot{\tilde{H}}_t \right\|^{2v_3}\Big)^{1/2} + k_8\Big(\mathbb{E}\left\| \tilde{H}_t - H_t \right\|^{4v_3}\Big)^{1/4} + k_9\Big(\mathbb{E}\left\| {}_{j}\dot{H}_t - {}_{j}\dot{\tilde{H}}_t \right\|^{3v_3}\Big)^{1/3} \tag{3.42}$$

In Lemma 3.3.3 and 3.3.4, each of these summations has been proved to be finite. Hence,

$$\sum_{t=1}^{\infty} \mathbb{E} \sup_{\boldsymbol{\theta}\in\upsilon(\boldsymbol{\Theta})} \left| \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j} - \frac{\partial^2 \tilde{l}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j} \right|^{v_3} < \infty \tag{3.43}$$

By Borel-Cantelli lemma, $\sup_{\boldsymbol{\theta}\in\upsilon(\boldsymbol{\Theta})} \left| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\intercal} - \frac{\partial^2 \tilde{l}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\intercal} \right| = o(1) \quad a.s..$

6. We will apply Results (19) and (18) in Appendix A on the third derivatives. Based on the terms and steps in (3.36),

$$\begin{aligned}
&\frac{\partial^3 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_{i_1}\partial\boldsymbol{\theta}_{i_2}\partial\boldsymbol{\theta}_{i_3}} \\
=&\frac{\partial}{\partial\boldsymbol{\theta}_j} tr\Big[\Big(\mathbb{I}_m - \boldsymbol{x_t}\boldsymbol{x_t}^\intercal H_t^{-1}(\boldsymbol{\theta})\Big)\Big({}_{ij}\ddot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta}) - {}_{i}\dot{H}_t H_t^{-1}(\boldsymbol{\theta})\ {}_{j}\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})) \\
&+ \boldsymbol{x_t}\boldsymbol{x_t}^\intercal\ {}_{i}\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})\ {}_{j}\dot{H}_t(\boldsymbol{\theta})H_t^{-1}(\boldsymbol{\theta})\Big] \\
=&tr\Big[\Big(\mathbb{I}_m - \boldsymbol{x_t}\boldsymbol{x_t}^\intercal H_t^{-1}\Big)\Big\{ {}_{ijk}\dddot{H}_t H_t^{-1} - {}_{ij}\ddot{H}_t H_t^{-1}\ {}_{k}\dot{H}_t H_t^{-1} - {}_{ik}\ddot{H}_t H_t^{-1}\ {}_{k}\dot{H}_t H_t^{-1} \\
&+ {}_{i}\dot{H}_t H_t^{-1}\ {}_{k}\dot{H}_t H_t^{-1}\ {}_{j}\dot{H}_t H_t^{-1} - {}_{i}\dot{H}_t H_t^{-1}\ {}_{jk}\ddot{H}_t H_t^{-1} + {}_{i}\dot{H}_t H_t^{-1}\ {}_{j}\dot{H}_t H_t^{-1}\ {}_{k}\dot{H}_t H_t^{-1}\Big\} \\
&+ \boldsymbol{x_t}\boldsymbol{x_t}^\intercal H_t^{-1}\Big\{ {}_{k}\dot{H}_t H_t^{-1}\ {}_{ij}\ddot{H}_t H_t^{-1} - {}_{k}\dot{H}_t H_t^{-1}\ {}_{i}\dot{H}_t H_t^{-1}\ {}_{j}\dot{H}_t H_t^{-1} + {}_{ik}\ddot{H}_t H_t^{-1}\ {}_{j}\dot{H}_t H_t^{-1} \\
&- {}_{k}\dot{H}_t H_t^{-1}\ {}_{i}\dot{H}_t H_t^{-1}\ {}_{j}\dot{H}_t H_t^{-1} - {}_{i}\dot{H}_t H_t^{-1}\ {}_{k}\dot{H}_t H_t^{-1}\ {}_{j}\dot{H}_t H_t^{-1} + {}_{i}\dot{H}_t H_t^{-1}\ {}_{jk}\ddot{H}_t H_t^{-1} H_t^{-1} \\
&- {}_{i}\dot{H}_t H_t^{-1}\ {}_{j}\dot{H}_t H_t^{-1}\ {}_{k}\dot{H}_t H_t^{-1}\Big\}\Big] \\
\leq& c_1\Big[\mathbb{E}\,\|\boldsymbol{x_t}\boldsymbol{x_t}^\intercal\|^2\, \mathbb{E}\left\|\Big({}_{ijk}\dddot{H}_t\Big)\right\|^2\Big]^{1/2} + c_2\Big[\mathbb{E}\,\|\boldsymbol{x_t}\boldsymbol{x_t}^\intercal\|^3\, \mathbb{E}\left\|{}_{ij}\ddot{H}_t\right\|^3 \mathbb{E}\left\|{}_{k}\dot{H}_t\right\|^3\Big]^{1/3} \\
&+ c_3\Big[\mathbb{E}\,\|\boldsymbol{x_t}\boldsymbol{x_t}^\intercal\|^3\, \mathbb{E}\left\|{}_{ik}\ddot{H}_t\right\|^3 \mathbb{E}\left\|{}_{k}\dot{H}_t\right\|^3\Big]^{1/3}
\end{aligned}$$

$$
\begin{aligned}
&+ c_4\Big[\mathbb{E}\,\|\boldsymbol{x_t}\boldsymbol{x_t}^\intercal\|^4\,\mathbb{E}\,\big\|{}_i\dot{H}_t\big\|^4\,\mathbb{E}\,\big\|{}_k\dot{H}_t\big\|^4\,\mathbb{E}\,\big\|{}_j\dot{H}_t\big\|^4\,\Big]^{1/4}\\
&+ c_5\Big[\mathbb{E}\,\|\boldsymbol{x_t}\boldsymbol{x_t}^\intercal\|^3\,\mathbb{E}\,\big\|{}_i\dot{H}_t\big\|^3\,\mathbb{E}\,\big\|{}_{jk}\ddot{H}_t\big\|^3\,\Big]^{1/3}\\
&+ c_6\Big[\mathbb{E}\,\|\boldsymbol{x_t}\boldsymbol{x_t}^\intercal\|^4\,\mathbb{E}\,\big\|{}_i\dot{H}_t\big\|^4\,\mathbb{E}\,\big\|{}_j\dot{H}_t\big\|^4\,\mathbb{E}\,\big\|{}_k\dot{H}_t\big\|^4\,\Big]^{1/4}\\
&+ c_7\mathbb{E}\,\big\|{}_{ijk}\dddot{H}_t\big\| + c_8\Big[\mathbb{E}\,\big\|{}_{ij}\ddot{H}_t\big\|^2\,\mathbb{E}\,\big\|{}_k\dot{H}_t\big\|^2\,\Big]^{1/2}\\
&+ c_9\Big[\mathbb{E}\,\big\|{}_{ik}\ddot{H}_t\big\|^2\,\mathbb{E}\,\big\|{}_k\dot{H}_t\big\|^2\,\Big]^{1/2}\\
&+ c_{10}\Big[\mathbb{E}\,\big\|{}_i\dot{H}_t\big\|^3\,\mathbb{E}\,\big\|{}_k\dot{H}_t\big\|^3\,\mathbb{E}\,\big\|{}_j\dot{H}_t\big\|^3\,\Big]^{1/3}\\
&+ c_{11}\Big[\mathbb{E}\,\big\|{}_i\dot{H}_t\big\|^2\,\mathbb{E}\,\big\|{}_{jk}\ddot{H}_t\big\|^2\,\Big]^{1/2}\\
&+ c_{12}\Big[\mathbb{E}\,\big\|{}_i\dot{H}_t\big\|^3\,\mathbb{E}\,\big\|{}_j\dot{H}_t\big\|^3\,\mathbb{E}\,\big\|{}_k\dot{H}_t\big\|^3\,\Big]^{1/3}
\end{aligned}
\tag{3.44}
$$

Lemma 3.3.3 and Assumption $C1$ tell us that every term in the last inequality is finite. The result is true for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, so we can definitely find such a set $\upsilon(\boldsymbol{\theta}_0)$ around $\boldsymbol{\theta}_0$ to satisfy this finite result.

After proving all the intermediate steps above, it brings us back to (3.33). By Steps 3 and 4 above, we can get that

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\\
=&\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right) + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \xrightarrow{D} N(\mathbf{0}, V)
\end{aligned}
$$

since the term in the bracket converges to 0 in probability and the second term converges to a normal distribution with $\mathbf{0}$ mean and variance V.

Apply the Taylor expansion on the stationary second derivative term around $\boldsymbol{\theta}_0$,

$$
\left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\intercal}\right)_{i_1 i_2} = \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\intercal}\right)_{i_1 i_2} + \frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}^\intercal}\left(\frac{\partial^2 l_t(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\intercal}\right)_{i_1 i_2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \tag{3.45}
$$

where $\boldsymbol{\theta}^*$ is between $\tilde{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$.

$\tilde{\boldsymbol{\theta}}_n$ is within the neighbourhood of $\boldsymbol{\theta}_0$ because of the strong consistency. Moreover, within the same compact set $\upsilon(\boldsymbol{\theta}_0)$, when the sample size n is sufficiently large,

$$\begin{aligned}
&\limsup_{n\to\infty}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}^{\top}}\Big(\frac{\partial^2 l_t(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}\Big)_{i_1 i_2}\right\| \\
\leq&\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in\upsilon(\boldsymbol{\theta}_0)}\left\|\frac{\partial}{\partial\boldsymbol{\theta}^{\top}}\Big(\frac{\partial^2 l_t(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}\Big)_{i_1 i_2}\right\| \\
\leq&\mathbb{E}\sup_{\boldsymbol{\theta}\in\upsilon(\boldsymbol{\theta}_0)}\left\|\frac{\partial}{\partial\boldsymbol{\theta}^{\top}}\Big(\frac{\partial^2 l_t(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}\Big)_{i_1 i_2}\right\| < \infty
\end{aligned} \tag{3.46}$$

by Step 6 above. This leads to another convergence result,

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}^{\top}}\left(\frac{\partial^2 l_t(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}\right)_{i_1 i_2}(\tilde{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\xrightarrow{D}0.$$

The first term in (3.45) converges to $\boldsymbol{J}$ in probability by applying the ergodic theorem,

$$\left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}\right)\xrightarrow{P}J.$$

Then, the left side of (3.45) converges,

$$\left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}\right)\xrightarrow{P}J.$$

By Step 5, the same results apply to the term within the bracket in (3.33),

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \tilde{l}_t(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}=\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}+\left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \tilde{l}_t(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}-\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}}\right)\xrightarrow{P}J \tag{3.47}$$

The final step in this proof is to use the Slutsky's theorem, the desired result can be obtained.

3.3.2 Lemmas

Lemma 3.3.2 *Under Assumption* $\boldsymbol{C}_1-\boldsymbol{C}_2$*,* *(i)* $\mathbb{E}\sigma_{i,t}^8 < \infty$ *and* $\mathbb{E}\tilde{\sigma}_{i,t}^8 < \infty$. *In addition,* $\mathbb{E}\left(\sigma_{i,t}\sigma_{j,t}\right)^4 < \infty$ *and* $\mathbb{E}\left(\tilde{\sigma}_{i,t}\tilde{\sigma}_{j,t}\right)^4 < \infty$*for* $i,j = 1,\ldots,m$, $i \neq j$. *Therefore,* $\mathbb{E}\left\|H_t\right\|^4 < \infty$ *and* $\mathbb{E}\left\|\tilde{H}_t\right\|^4 < \infty$.

(ii) $\mathbb{E}\left(\sum_{l=0}^{\infty}\beta_i^l x_{i,t-1-l}^2\right)^4 < \infty$

(iii) $\mathbb{E}\left(\sum_{l=1}^{\infty} l\beta_i^{l-1} x_{i,t-1-l}^2\right)^4 < \infty$

(iv) $\mathbb{E}\left(\sum_{l=2}^{\infty} l(l-1)\beta_i^{l-2} x_{i,t-1-l}^2\right)^4 < \infty$

(v) $\mathbb{E}\left[\frac{\sum_{l=0}^{\infty}\beta_i^l x_{i,t-1-l}^2}{\sigma_{i,t}}\right]^8 < \infty$ *and* $\mathbb{E}\left[\frac{(\sum_{l=0}^{\infty}\beta_i^l x_{i,t-1-l}^2)^2}{\sigma_{i,t}^3}\right]^8 < \infty$.

(vi) $\mathbb{E}\left[\frac{\sum_{l=1}^{\infty} l\beta_i^{l-1} x_{i,t-1-l}^2}{\sigma_{i,t}}\right]^8 < \infty$, $\mathbb{E}\left[\frac{(\sum_{l=1}^{\infty} l\beta_i^{l-1} x_{i,t-1-l}^2)^2}{\sigma_{i,t}^3}\right]^8 < \infty$ *and* $\mathbb{E}\left[\frac{(\sum_{l=1}^{\infty} l\beta_i^{l-1} x_{i,t-1-l}^2)^3}{\sigma_{i,t}^5}\right]^8 < \infty$.

(vii) $\mathbb{E}\left[\frac{(\sum_{l=0}^{\infty}\beta_i^l x_{i,t-1-l}^2)(\sum_{l=1}^{\infty} l\beta_i^{l-1} x_{i,t-1-l}^2)}{\sigma_{i,t}^3}\right]^8 < \infty$

(viii) $\mathbb{E}\left[\frac{\sum_{l=2}^{\infty} l(l-1)\beta_i^{l-2} x_{i,t-1-l}^2}{\sigma_{i,t}}\right]^8 < \infty$ *and* $\mathbb{E}\left[\frac{\sum_{l=3}^{\infty} l(l-1)(l-2)\beta_i^{l-3} x_{i,t-1-l}^2}{\sigma_{i,t}}\right]^8 < \infty$.

(ix) $\mathbb{E}\left[\frac{(\sum_{l=2}^{\infty} l(l-1)\beta_i^{l-2} x_{i,t-1-l}^2)(\sum_{l=1}^{\infty} l\beta_i^{l-1} x_{i,t-1-l}^2)}{\sigma_{i,t}^3}\right]^8 < \infty$.

Proof $a, b, c, d, a_1, b_1, c_1, d_1 \ldots$ will be used to represent some finite constants and they may have different values in lines in the following proof.

(i) Apply Holder's and Minkowski's inequality ((14) and (15) in Appendix A), for $i = 1, 2, \ldots, m$,

$$\begin{aligned}\mathbb{E}\sigma_{i,t}^8 =& \mathbb{E}\left(\frac{\omega_i}{1-\beta_i} + \sum_{j=0}^{\infty}\alpha_i\beta_i^j x_{i,t-1-j}^2\right)^4 \\ \leq& \left(\frac{\omega_i}{1-\beta_i} + \sum_{j=0}^{\infty}\alpha_i\beta_i^j \left[\mathbb{E}(x_{i,t-1-j}^2)^4\right]^{1/4}\right)^4 < \infty\end{aligned}$$

The last line is from Assumption $\boldsymbol{C}_2$, $\mathbb{E}x_{i,t}^8$ is finite. Then by (3) in Appendix A, the second result follows.

For the practical term,

$$\begin{aligned}
\mathbb{E}\tilde{\sigma}_{i,t}^8 =&\mathbb{E}\left(\tilde{\sigma}_{i,0}^2\beta_i^t + \sum_{j=0}^{t-2}\alpha_i\beta_i^j x_{i,t-1-j}^2 + \sum_{j=0}^{t-1}\beta_i^j\omega_i + \alpha_i\beta_i^{t-1}\tilde{x}_{i,0}^2\right)^4 \\
\leq& a_1\left[(\tilde{\sigma}_{i,0}^2\beta_i^t)^4 + \mathbb{E}\left(\sum_{j=0}^{\infty}\alpha_i\beta_i^j x_{i,t-1-j}^2\right)^4 + a_2 + a_3\beta_i^{(t-1)4}\right] < \infty
\end{aligned}$$

The second part of the result follows by Cauchy-Schwarz inequality.

Move to the last part of this result, both expectations will be proved following the same logic. The diagonal elements of $\boldsymbol{H}_t$ are finite summations of $\sigma_{i,t}^2$ and the non-diagonal elements are finite summations of both $\sigma_{i,t}^2$ and $\sigma_{i,t}\sigma_{j,t}$. By (4) in Appendix A, $\mathbb{E}\left\|\boldsymbol{H}_t\right\|^4$ is finite . So is $\mathbb{E}\left\|\tilde{\boldsymbol{H}}_t\right\|^4$.

(ii) From (i) above and (4) in Appendix A, the term on the left hand side becomes

$$\begin{aligned}
\mathbb{E}\left(\sum_{l=0}^{\infty}\beta_i^l x_{i,t-1-l}^2\right)^4 =&\mathbb{E}\left[(\sigma_{i,t}^2 - \frac{\omega_i}{1-\beta_i})/\alpha_i\right]^4 \\
\leq& a\left[\mathbb{E}(\sigma_{i,t}^2/\alpha_i)^4 + \left(\frac{\omega_i}{(1-\beta_i)\alpha_i}\right)^4\right] < \infty.
\end{aligned}$$

(iii) The inequality follows by applying (14) and (15) in Appendix A.

$$\mathbb{E}\left(\sum_{l=1}^{\infty} l\beta_i^{l-1}x_{i,t-1-l}^2\right)^4 = \left\{\sum_{l=1}^{\infty} l\beta_i^{l-1}\left[\mathbb{E}\left(x_{i,t-1-l}^2\right)^4\right]^{1/4}\right\}^4 < \infty.$$

(iv) The results (10), (11) and (12) in Appendix A lead to

$$\mathbb{E}\left(\sum_{l=2}^{\infty} l(l-1)\beta_i^{l-2}x_{i,t-1-l}^2\right)^4 \leq \left\{\sum_{l=2}^{\infty} l(l-1)\beta_i^{l-2}\mathbb{E}\left(x_{i,t-1-l}^8\right)^{1/4}\right\}^4 < \infty.$$

(v) The denominator $\sigma_{i,t}$ can be reduced according to the terms in the numerator.

$$\begin{aligned}
\mathbb{E}\left[\frac{\sum_{l=0}^{\infty}\beta_i^l x_{i,t-1-l}^2}{\sigma_{i,t}}\right]^8 =& \mathbb{E}\left[\frac{\sum_{l=0}^{\infty}\beta_i^l x_{i,t-1-l}^2}{\sqrt{\frac{\omega_i}{1-\beta_i}+\alpha_i\sum_{k=0}^{\infty}\beta_i^k x_{i,t-1-k}^2}}\right]^8 \\
\leq& \mathbb{E}\left[\sum_{l=0}^{\infty}\frac{\beta_i^l x_{i,t-1-l}^2}{\sqrt{\frac{\omega_i}{1-\beta_i}+\alpha_i\beta_i^l x_{i,t-1-l}^2}}\right]^8 \\
\leq& \mathbb{E}\left[\sum_{l=0}^{\infty}\sqrt{\beta_i^l x_{i,t-1-l}^2}/\sqrt{\alpha_i}\right]^8 \\
\leq& \left[\sum_{l=0}^{\infty}\left[\mathbb{E}(\sqrt{\beta_i^l x_{i,t-1-l}^2}/\sqrt{\alpha_i})^8\right]^{1/8}\right]^8 \\
\leq& \left[\sum_{l=0}^{\infty}\frac{\beta_i^{l/2}}{\alpha_i^{1/2}}\left(\mathbb{E}x_{i,t-1-l}^8\right)^{1/8}\right]^8 < \infty
\end{aligned}$$

The Holder's and Minkowski's inequality can be applied afterwards to get an upper bound of the summation in the second last line. The first inequality holds, then we can apply the same technique and transform the terms on the left hand side of the second inequality.

Therefore,

$$\begin{aligned}
\mathbb{E}\left[\frac{(\sum_{l=0}^{\infty}\beta_i^l x_{i,t-1-l}^2)^2}{\sigma_{i,t}^3}\right]^8 =& \mathbb{E}\left\{\left[\frac{\sigma_{i,t}^2-\omega_i/(1-\beta_i)}{\alpha_i}\right]^2/\sigma_{i,t}^3\right\}^8 \\
\leq& a\left[\mathbb{E}\left(\frac{\sigma_{i,t}^4}{\alpha_i^2\sigma_{i,t}^3}\right)^8+\left(\frac{\omega_i}{(1-\beta_i)\alpha_i}\right)^8\right] \\
=& a\frac{1}{\alpha_i^2}\mathbb{E}\left(\sigma_{i,t}^8\right)+a\left(\frac{\omega_i}{(1-\beta_i)\alpha_i}\right)^8 < \infty.
\end{aligned}$$

(vi) Using (10) and (11) in Appendix A, and the similar arguments in both (v) and (iv),

$$\mathbb{E}\left[\frac{\sum_{l=1}^{\infty} l\beta_i^{l-1}x_{i,t-1-l}^2}{\sigma_{i,t}}\right]^8 \leq a\mathbb{E}\left[\sum_{l=0}^{\infty} l\sqrt{\beta_i^{l-1}x_{i,t-1-l}^2}\right]^8$$

$$\leq a\left[\sum_{l=0}^{\infty} l\beta_i^{(l-1)/2}\left[\mathbb{E}(x_{i,t-1-l}^8)\right]^{1/8}\right]^8 < \infty.$$

The other two inequalities can be proved as well using the same method. Then,

$$\mathbb{E}\left[\frac{(\sum_{l=1}^{\infty} l\beta_i^{l-1}x_{i,t-1-l}^2)^2}{\sigma_{i,t}^3}\right]^8 \leq b_1\mathbb{E}\left[\sum_{l=1}^{\infty}\frac{l\beta_i^{l-1}x_{i,t-1-l}^2}{(\alpha_i\beta_i^l x_{i,t-1-l}^2)^{3/4}}\right]^{16}$$
$$\leq b_2\left[\sum_{l=1}^{\infty} l\beta_i^{(l-1)/4}[\mathbb{E}(x_{i,t-1-l}^{1/2})^{16}]^{1/16}\right]^{16} < \infty$$

and

$$\mathbb{E}\left[\frac{\left(\sum_{l=1}^{\infty} l\beta_i^{l-1}x_{i,t-1-l}^2\right)^3}{\sigma_{i,t}^5}\right]^8 \leq c_1\mathbb{E}\left[\sum_{l=1}^{\infty}\frac{l\beta_i^{l-1}x_{i,t-1-l}^2}{(\alpha_i\beta_i^l x_{i,t-1-l}^2)^{5/6}}\right]^{24}$$
$$\leq c_2\left[\sum_{l=1}^{\infty} l\beta_i^{(l-1)/6}\left[\mathbb{E}\left(x_{i,t-1-l}^{1/3}\right)^{24}\right]^{1/24}\right]^{24} < \infty.$$

(vii) By transforming the term inside the first pair of brackets in the numerator and applying the result in (vi),

$$\mathbb{E}\left[\frac{\left(\sum_{l=0}^{\infty}\beta_i^l x_{i,t-1-l}^2\right)\left(\sum_{l=1}^{\infty} l\beta_i^{l-1}x_{i,t-1-l}^2\right)}{\sigma_{i,t}^3}\right]^8$$
$$\leq\mathbb{E}\left[\frac{\sigma_{i,t}^2/\alpha_i(\sum_{l=1}^{\infty} l\beta_i^{l-1}x_{i,t-1-l}^2)}{\sigma_{i,t}^3}\right]^8$$
$$\leq a\mathbb{E}\left[\frac{\sum_{l=1}^{\infty} l\beta_i^{l-1}x_{i,t-1-l}^2}{\sigma_{i,t}}\right]^8 < \infty.$$

(viii) Using (10) and (11) in Appendix A, and the result (vi) above in this lemma, we can get

$$\mathbb{E}\left[\frac{\sum_{l=2}^{\infty} l(l-1)\beta_i^{l-2}x_{i,t-1-l}^2}{\sigma_{i,t}}\right]^8 \leq a\mathbb{E}\left[\sum_{l=2}^{\infty} l(l-1)\sqrt{\beta_i^{l-2}x_{i,t-1-l}^2}\right]^8$$

$$\leq a\left[\sum_{l=1}^{\infty} l(l-1)\beta_i^{(l-2)/2}\left[\mathbb{E}(x_{i,t-1-l}^8)\right]^{1/8}\right]^8 < \infty.$$

Similarly, the second inequality becomes

$$\begin{aligned}\mathbb{E}\left[\frac{\sum_{l=3}^{\infty} l(l-1)(l-2)\beta_i^{l-3}x_{i,t-1-l}^2}{\sigma_{i,t}}\right]^8 &\leq a\mathbb{E}\left[\sum_{l=3}^{\infty} l(l-1)(l-2)\sqrt{\beta_i^{l-3}x_{i,t-1-l}^2}\right]^8\\ &\leq b\left[\sum_{l=3}^{\infty} l(l-1)(l-2)\beta_i^{(l-3)/2}\left[\mathbb{E}(x_{i,t-1-l}^8)\right]^{1/8}\right]^8\\ &<\infty.\end{aligned}$$

(ix) We can use (vi) in the second line of the inequality below in order to get the desired finiteness.

$$\begin{aligned}&\mathbb{E}\left[\frac{\left(\sum_{l=2}^{\infty} l(l-1)\beta_i^{l-2}x_{i,t-1-l}^2\right)\left(\sum_{l=1}^{\infty} l\beta_i^{l-1}x_{i,t-1-l}^2\right)}{\sigma_{i,t}^3}\right]^8\\ =&\mathbb{E}\left[\frac{(\sum_{l=2}^{\infty} l(l-1)\beta_i^{l-2}x_{i,t-1-l}^2)}{\sigma_{i,t}^{3/2}}\frac{(\sum_{l=1}^{\infty} l\beta_i^{l-1}x_{i,t-1-l}^2)}{\sigma_{i,t}^{3/2}}\right]^8\\ \leq&\sqrt{\mathbb{E}\left[\frac{\sum_{l=2}^{\infty} l(l-1)\beta_i^{l-2}x_{i,t-1-l}^2}{\sigma_{i,t}^{3/2}}\right]^{12}\mathbb{E}\left[\frac{\sum_{l=1}^{\infty} l\beta_i^{l-1}x_{i,t-1-l}^2}{\sigma_{i,t}^{3/2}}\right]^{16}}\\ \leq&a\sqrt{\mathbb{E}\left[\frac{\sum_{l=2}^{\infty} l(l-1)\beta_i^{l-2}x_{i,t-1-l}^2}{(\beta_i x_{i,t-1-l}^2)^{3/2}}\right]^{16}}\\ \leq&b\sqrt{\mathbb{E}\left[\sum_{l=2}^{\infty} l(l-1)\beta_i^{(l-2)/2}x_{i,t-1-l}^{1/2}\right]^{16}}\\ \leq&c\sqrt{\left\{\sum_{l=2}^{\infty} l(l-1)\beta_i^{(l-2)/2}\left[\mathbb{E}(x_{i,t-1-l}^{1/2})^{16}\right]^{1/16}\right\}^{16}} < \infty.\end{aligned}$$

□

Lemma 3.3.3 *Under Assumptions $C_1 - C_2$,*

(i) $\max(\mathbb{E}\left|\frac{\partial\sigma_{i,t}^2}{\partial\omega_i}\right|^{z_1}, \mathbb{E}\left|\frac{\partial\sigma_{i,t}^2}{\partial\alpha_i}\right|^{z_1}, \mathbb{E}\left|\frac{\partial\sigma_{i,t}^2}{\partial\beta_i}\right|^{z_1}) \leq k_{z_1} < \infty$ *for* $i = 1, \ldots, m$ *and any positive number*

$z_1 \leq 4$. *Therefore,* $\mathbb{E}\left\|{}_i\dot{H}_t\right\|^4 < \infty$, *where* ${}_i\dot{H}_t = \dfrac{\partial H_t}{\partial \theta_i}$, *represents the first derivative of matrix* H_t *with respect to the* i*-th parameter.*

(ii) $\max(\mathbb{E}\left|\frac{\partial \tilde{\sigma}^2_{i,t}}{\partial \omega_i}\right|^{z_2}, \mathbb{E}\left|\frac{\partial \tilde{\sigma}^2_{i,t}}{\partial \alpha_i}\right|^{z_2}, \mathbb{E}\left|\frac{\partial \tilde{\sigma}^2_{i,t}}{\partial \beta_i}\right|^{z_2}) \leq \tilde{k}_{z_2} < \infty$ *for* $i = 1, \ldots, m$ *and any positive number* $0 < z_2 \leq 4$. *In addition,* $\mathbb{E}\left\|{}_i\dot{\tilde{H}}_t\right\|^4 < \infty$.

(iii) $\mathbb{E}\left\|{}_{ij}\ddot{H}_t\right\|^4 < \infty$, *where* ${}_{ij}\ddot{H}_t = \dfrac{\partial^2 H_t}{\partial \theta_i \partial \theta_j}$.

(iv) $\mathbb{E}\left\|{}_{ijl}\dddot{H}_t\right\|^4 < \infty$, *where* ${}_{ijl}\dddot{H}_t = \dfrac{\partial^3 H_t}{\partial \theta_i \partial \theta_j \partial \theta_l}$

(v) $\mathbb{E}\left\|{}_{ij}\ddot{\tilde{H}}_t\right\|^{z_3} < \infty$, *where* ${}_{ij}\ddot{\tilde{H}}_t = \dfrac{\partial^2 \tilde{H}_t}{\partial \theta_i \partial \theta_j}$ *for any positive number* $0 < z_3 < 1$.

Proof We can see from (7) in Appendix A and the proof of Lemma 3.2.8, the norm constant p is irrelevant here when we work on (iii), (iv) and (v). Therefore, it is equivalent to prove that the partial derivative of each element of the matrix H_t with respect to θ_i has a finite absolute third moment.

Since $H_t(\boldsymbol{\theta})$ is a symmetric matrix, we could denote the element in the ith row and jth column of $H_t(\boldsymbol{\theta})$ by $H_{ij,t}$, and assume $i < j$ and $l \leq k$ in $\rho_{l,k}$ without loss of generality, $i, j, l, k = 1, \ldots, m$. The lower case letters with or without a subscript, $a, b, a_1, b_1, \ldots$, have been used as symbols to represent some finite constants, they may have distinct values in different lines below.

(i) The first derivatives of $\sigma^2_{i,t}$ with respect to the parameters can be easily calculated. By using the inequalities in Lemma 3.3.2 and (4) in Appendix A, the following inequalities can be obtained.

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial \sigma^2_{i,t}}{\partial \omega_i}\right|^4 &= \left(\frac{1}{1-\beta_i}\right)^4 < \infty \\
\mathbb{E}\left|\frac{\partial \sigma^2_{i,t}}{\partial \alpha_i}\right|^4 &= \mathbb{E}(\sum_{l=0}^{\infty} \beta_j^l x^2_{j,t-1-l}{}^4 < \infty \\
\mathbb{E}\left|\frac{\partial \sigma^2_{i,t}}{\partial \beta_i}\right|^4 &= \mathbb{E}\left(\frac{\omega_i}{(1-\beta_i)} + \sum_{l=1}^{\infty} l\alpha_i \beta_i^{l-1} x^2_{i,t-1-l}\right)^4 < \infty
\end{aligned}
$$

$$\leq a\left(\mathbb{1}_{i=j}+\beta_{0j}\right)^4\left[\left(\frac{\omega_j}{(1-\beta_j)}\right)^4+\mathbb{E}\left(\sum_{l=1}^{\infty} l\alpha_j\beta_j^{l-1}x_{j,t-1-l}^2\right)^4\right]<\infty$$

It is easier to consider the diagonal terms and the other terms of $\boldsymbol{H}_t(\boldsymbol{\theta})$ separately.

$$\mathbb{E}\left(\frac{\partial H_{ii,t}}{\partial\omega_j}\right)^4=\mathbb{E}\left(\frac{\partial H_{ii,t}}{\partial\sigma_{j,t}^2}\frac{\partial\sigma_{j,t}^2}{\partial\omega_j}\right)^4=\left(\mathbb{1}_{i=j}+\beta_{0j}\right)^4\mathbb{E}\left|\frac{\partial\sigma_{i,t}^2}{\partial\omega_j}\right.^4<\infty$$

$$\mathbb{E}\left(\frac{\partial H_{ii,t}}{\partial\alpha_j}\right)^4=\left(\mathbb{1}_{i=j}+\beta_{0j}\right)^3\mathbb{E}\left|\frac{\partial\sigma_{i,t}^2}{\partial\alpha_j}\right|^4<\infty$$

$$\mathbb{E}\left(\frac{\partial H_{ii,t}}{\partial\beta_j}\right)^3=\left(\mathbb{1}_{i=j}+\beta_{0j}\right)^4\mathbb{E}\left|\frac{\partial\sigma_{i,t}^2}{\partial\beta_j}\right|^4<\infty$$

$$\mathbb{E}\left(\frac{\partial H_{ii,t}}{\partial\beta_{0j}}\right)^4=\mathbb{E}\left(\sigma_{j,t}^2\right)^4<\infty$$

$$\mathbb{E}\left(\frac{\partial H_{ii,t}}{\partial\rho_{l,k}}\right)^4=0$$

The first derivatives of the diagonal elements have been proved. Then, for the non-diagonal terms, we have the similar finiteness result.

$$\mathbb{E}\left(\frac{\partial H_{ij,t}}{\partial\omega_l}\right)^4=\mathbb{E}\left(\mathbb{1}_{i=l}\rho_{i,j}\frac{\sigma_{j,t}}{2\sigma_{l,t}}+\mathbb{1}_{j=l}\rho_{i,j}\frac{\sigma_{i,t}}{2\sigma_{l,t}}+\beta_{0l}\right)^4\left(\frac{1}{1-\beta_l}\right)^4<\infty$$

$$\begin{aligned}\mathbb{E}\left(\frac{\partial H_{ij,t}}{\partial\alpha_l}\right)^4=&\mathbb{E}\left(\mathbb{1}_{i=l}\rho_{i,j}\frac{\sigma_{j,t}}{2\sigma_{l,t}}+\mathbb{1}_{j=l}\rho_{i,j}\frac{\sigma_{i,t}}{2\sigma_{l,t}}+\beta_{0l}\right)^4\left(\sum_{k=0}^{\infty}\beta_j^k x_{l,t-1-k}^2\right)^4\\
\leq&\mathbb{E}\left(\mathbb{1}_{i=l}\rho_{i,j}\sigma_{j,t}+\mathbb{1}_{j=l}\rho_{i,j}\sigma_{i,t}+2\beta_{0l}\sigma_{l,t}\right)^4\left(\frac{\sum_{k=0}^{\infty}\beta_j^k x_{l,t-1-k}^2}{2\sigma_{l,t}}\right)^4\\
\leq&\left\{\mathbb{E}\left(\mathbb{1}_{i=l}\rho_{i,j}\sigma_{j,t}+\mathbb{1}_{j=l}\rho_{i,j}\sigma_{i,t}+2\beta_{0l}\sigma_{l,t}\right)^8\mathbb{E}\left(\frac{\sum_{k=0}^{\infty}\beta_j^k x_{l,t-1-k}^2}{2\sigma_{l,t}}\right)^8\right\}^{1/2}\\
<&\infty\end{aligned}$$

Based on (4) in Appendix A and $\rho_{i,j}\in[-1,1]$,

$$\mathbb{E}\left(\frac{\partial H_{ij,t}}{\partial\beta_l}\right)^4=\mathbb{E}\left[(\mathbb{1}_{i=l}\rho_{i,j}\frac{\sigma_{j,t}}{2\sigma_{l,t}}+\mathbb{1}_{j=l}\rho_{i,j}\frac{\sigma_{i,t}}{2\sigma_{l,t}}+\beta_{0l})^4\left(\frac{\omega_l}{(1-\beta_l)^2}+\sum_{k=1}^{\infty}k\alpha_l\beta_l^{k-1}x_{l,t-1-k}^2\right)^4\right]$$

$$
\begin{aligned}
\leq&\mathbb{E}\left[(\frac{\sigma_{j,t}}{\sigma_{i,t}}+\frac{\sigma_{i,t}}{\sigma_{j,t}}+\beta_{0l})\frac{\omega_l}{(1-\beta_l)^2}+\left(\frac{\sigma_{j,t}}{\sigma_{i,t}}+\frac{\sigma_{i,t}}{\sigma_{j,t}}+\beta_{0l}\right)(\sum_{k=1}^{\infty}k\alpha_l\beta_l^{k-1}x_{l,t-1-k}^2)\right]^4\\
\leq&4\mathbb{E}\left[\left(\frac{\sigma_{j,t}}{\sigma_{i,t}}+\frac{\sigma_{i,t}}{\sigma_{j,t}}+\beta_{0l}\right)\frac{\omega_l}{(1-\beta_l)^2}\right]^4+4\mathbb{E}\left[\left(\frac{\sigma_{j,t}}{\sigma_{i,t}}+\frac{\sigma_{i,t}}{\sigma_{j,t}}+\beta_{0l}\right)\left(\sum_{k=1}^{\infty}k\alpha_l\beta_l^{k-1}x_{l,t-1-k}^2\right)\right]^4\\
\leq&a\mathbb{E}\left[\left(\frac{\sigma_{j,t}}{\sqrt{\omega_i}}+\frac{\sigma_{i,t}}{\sqrt{\omega_j}}+\beta_{0l}\right)\frac{\omega_l}{(1-\beta_l)^2}\right]^4+b\mathbb{E}\left[\frac{\sigma_{j,t}}{\sigma_{i,t}}\left(\sum_{k=1}^{\infty}k\alpha_i\beta_i^{k-1}x_{i,t-1-k}^2\right)\right]^4\\
&+b\mathbb{E}\left[\frac{\sigma_{i,t}}{\sigma_{j,t}}\left(\sum_{k=1}^{\infty}k\alpha_j\beta_j^{k-1}x_{j,t-1-k}^2\right)\right]^4+b\mathbb{E}\left[\beta_{0l}\left(\sum_{k=1}^{\infty}k\alpha_l\beta_l^{k-1}x_{l,t-1-k}^2\right)\right]^4\\
\leq&a\mathbb{E}\left[\left(\frac{\sigma_{j,t}}{\sqrt{\omega_i}}+\frac{\sigma_{i,t}}{\sqrt{\omega_j}}+\beta_{0l}\right)\frac{\omega_l}{(1-\beta_l)^2}\right]^4\\
\leq&c+d\left\{\mathbb{E}\sigma_{j,t}^8\mathbb{E}\left[\frac{\sum_{k=1}^{\infty}k\alpha_i\beta_i^{k-1}x_{i,t-1-k}^2}{\sigma_{i,t}}\right]^8\right\}^{1/2}\\
&+d\left\{\mathbb{E}\sigma_{i,t}^8\mathbb{E}\left[\frac{\sum_{k=1}^{\infty}k\alpha_j\beta_j^{k-1}x_{j,t-1-k}^2}{\sigma_{j,t}}\right]^8\right\}^{1/2}+b\mathbb{E}\left[\beta_{0l}\left(\sum_{k=1}^{\infty}k\alpha_l\beta_l^{k-1}x_{l,t-1-k}^2\right)\right]^4<\infty.
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left(\frac{\partial H_{ij,t}}{\partial\beta_{0l}}\right)^4=&\mathbb{E}\left(\sigma_{l,t}^2\right)^3<\infty\\
\mathbb{E}\left(\frac{\partial H_{ij,t}}{\partial\rho_{l,k}}\right)^4=&\mathbb{E}\left[\mathbb{1}_{i=l}\mathbb{1}_{j=k}\sigma_{l,t}\sigma_{k,t}\right]^4\leq\mathbb{E}\left(\sigma_{l,t}\sigma_{k,t}\right)^4<\infty
\end{aligned}
$$

The desired result follows.

(ii) We can easily write down the first derivatives of $\tilde{\sigma}_{i,t}^2$ in (3.22).

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial\tilde{\sigma}_{i,t}^2}{\partial\omega_i}\right|^4=&\left(\frac{1-\beta_i^t}{1-\beta_i}\right)^4<\infty\\
\mathbb{E}\left|\frac{\partial\tilde{\sigma}_{i,t}^2}{\partial\alpha_i}\right|^4=&\mathbb{E}\left(\sum_{l=0}^{t-2}\beta_j^l x_{j,t-1-l}^2+\beta_i^{t-1}\tilde{x}_{1,0}^2\right)^4\\
\leq&a\mathbb{E}\left(\sum_{l=0}^{\infty}\beta_j^l\mathbb{E}x_{j,t-1-l}^2\right)^4+a\beta_i^{(t-1)4}\tilde{x}_{1,0}^8\leq b+c\beta_i^{tv_3}<\infty
\end{aligned}
$$

It is worth to notice that the first part in the second inequality is a continuous function with respect to any fixed β_i. It is easy to see that this term has a maximum a on $t\in\mathbb{N}^+$ which leads to the result in the last line.

The last term, the derivative of $\tilde{\sigma}^2_{i,t}$ with respect to β_i is

$$\begin{aligned}
\mathbb{E}\left|\frac{\partial\tilde{\sigma}^2_{i,t}}{\partial\beta_i}\right|^4 =&\mathbb{E}\left(t\beta_i^{t-1}\tilde{\sigma}^2_{i,0}+\sum_{l=1}^{t-2}l\beta_i^{l-1}\left(\alpha_i x^2_{i,t-1-l}+\omega_i\right)+(t-1)\beta_i^{t-2}\left(\omega_i+\alpha_i\tilde{x}^2_{i,0}\right)\right)^4\\
\leq& a\left[t^4\beta_i^{(t-1)4}\tilde{\sigma}^8_{i,0}+\mathbb{E}\left(\sum_{l=1}^{\infty}l\beta_i^{(l-1)}(\alpha_i x^2_{i,t-1-l}+\omega_i)\right)^4+(t-1)^4\beta_i^{(t-2)4}\alpha_i^4\tilde{x}^8_{i,0}\right]\\
\leq& c_1t^4\beta_i^{t4}+c_2+c_3(t-1)^v\beta_i^{t4}.
\end{aligned}$$

The last line of the inequality above is a continuous function of t, $h_1(t)$. Since $\beta_i<1$, we can prove that $h_1(t)$ has a maximum. The term $\mathbb{E}\left|\frac{\partial\tilde{\sigma}^2_{i,t}}{\partial\beta_i}\right|^4$ is finite for any $t>0$. We move on to the first derivative of the practical conditional covariance matrix $\dot{\tilde{H}}_t$. We study all the elements in our target and start with the diagonal elements. By the results above in this lemma,

$$\begin{aligned}
\mathbb{E}\left|\frac{\partial\tilde{H}_{ii,t}}{\partial\omega_j}\right|^4 =&(\mathbb{1}_{i=j}+\beta_{0j})^4\mathbb{E}\left|\frac{\partial\tilde{\sigma}^2_{j,t}}{\partial\omega_j}\right|^4<\infty,\\
\mathbb{E}\left|\frac{\partial\tilde{H}_{ii,t}}{\partial\alpha_j}\right|^4 =&\left(\mathbb{1}_{i=j}+\beta_{0j}\right)^4\mathbb{E}\left|\frac{\partial\tilde{\sigma}^2_{j,t}}{\partial\alpha_j}\right|^4<\infty,\\
\mathbb{E}\left|\frac{\partial\tilde{H}_{ii,t}}{\partial\beta_j}\right|^4 =&\Big(\mathbb{1}_{i=j}+\beta_{0j})^4\mathbb{E}\left|\frac{\partial\tilde{\sigma}^2_{j,t}}{\partial\beta_j}\right|^4<\infty,\\
\mathbb{E}\left|\frac{\partial\tilde{H}_{ii,t}}{\partial\beta_{0j}}\right|^4 =&\mathbb{E}\left|\tilde{\sigma}^2_{j,t}\right|^4<\infty.
\end{aligned}$$

Similar results are true for the $4th$ moment of the non-diagonal terms as following. Suppose that $i<j$ and $l<k$,

$$\begin{aligned}
\mathbb{E}\left|\frac{\partial\tilde{H}_{ij,t}}{\partial\rho_{lk}}\right|^4 =&\mathbb{1}_{i=l}\mathbb{1}_{j=k}\mathbb{E}\left|\tilde{\sigma}_{i,t}\tilde{\sigma}_{j,t}\right|^4<\infty,\\
\mathbb{E}\left|\frac{\partial\tilde{H}_{ij,t}}{\partial\omega_l}\right|^4 =&\mathbb{E}\left|\left(\frac{\mathbb{1}_{i=l}\sigma_{j,t}+\mathbb{1}_{j=l}\sigma_{i,t}}{\sigma_{l,t}}\rho_{ij}+\beta_{0l}\right)\frac{\sigma^2_{l,t}}{\omega_l}\right|^4<\infty,\\
\mathbb{E}\left|\frac{\partial\tilde{H}_{ij,t}}{\partial\alpha_l}\right|^4 =&\mathbb{E}\left|\left(\frac{\mathbb{1}_{i=l}\sigma_{j,t}+\mathbb{1}_{j=l}\sigma_{i,t}}{\sigma_{l,t}}\rho_{ij}+\beta_{0l}\right)\frac{\sigma^2_{l,t}}{\alpha_l}\right|^4<\infty,
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial \tilde{H}_{ij,t}}{\partial \beta_l}\right|^4 =& \mathbb{E}\left|\left(\frac{\mathbb{1}_{i=l}\sigma_{j,t} + \mathbb{1}_{j=l}\sigma_{i,t}}{\sigma_{l,t}}\rho_{ij} + \beta_{0l}\right)\frac{\sigma^2_{l,t}}{\beta_l}\right|^4 < \infty, \\
\mathbb{E}\left|\frac{\partial \tilde{H}_{ij,t}}{\partial \beta_{0l}}\right|^4 =& \mathbb{E}\left|\tilde{\sigma}^2_{l,t}\right|^4 < \infty.
\end{aligned}
$$

(iii) A sufficient condition is that all absolute elements of the second partial derivative of $\boldsymbol{H}_t$ are finite in terms of the third moment. Only the non-zero terms are listed below because the number of all second derivative elements is large.

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^2 H_{ii,t}}{\partial \omega_j \partial \beta_{0j}}\right|^4 =& \left(\frac{1}{1-\beta_j}\right)^4 < \infty \\
\mathbb{E}\left|\frac{\partial^2 H_{ii,t}}{\partial \omega_j \partial \beta_j}\right|^4 =& (\mathbb{1}_{i=j} + \beta_{0j})^4 \left(\frac{1}{1-\beta_j}\right)^8 < \infty \\
\mathbb{E}\left|\frac{\partial^2 H_{ii,t}}{\partial \alpha_j \partial \beta_{0j}}\right|^4 =& \mathbb{E}\left(\sum_{l=0}^{\infty} \beta_j^l x^2_{j,t-1-l}\right)^4 < \infty \\
\mathbb{E}\left|\frac{\partial^2 H_{ii,t}}{\partial \alpha_j \partial \beta_j}\right|^4 =& \left(\mathbb{1}_{i=j} + \beta_{0j}\right)^4 \mathbb{E}\left(\sum_{l=1}^{\infty} l\beta_j^{l-1} x^2_{j,t-1-l}\right)^4 < \infty \\
\mathbb{E}\left|\frac{\partial^2 H_{ii,t}}{\partial \beta_j \partial \beta_{0j}}\right|^4 =& \mathbb{E}\left(\frac{\omega_j}{(1-\beta_j)^3} + \sum_{l=1}^{\infty} l\alpha_j\beta_j^{l-1} x^2_{j,t-1-l}\right)^4 \\
\leq& a\mathbb{E}\left(\frac{\omega_j}{(1-\beta_j)^3}\right)^4 + \mathbb{E}\left(\sum_{l=1}^{\infty} l\alpha_j\beta_j^{l-1} x^2_{j,t-1-l}\right)^4 < \infty \\
\mathbb{E}\left|\frac{\partial H_{ii,t}}{\partial \beta_j \partial \beta_j}\right|^4 =& (\mathbb{1}_{i=j} + \beta_{0j})^4 \mathbb{E}\left(\frac{2\omega_j}{(1-\beta_j)^3} + \sum_{l=2}^{\infty} l(l-1)\alpha_j\beta_j^{l-2} x^2_{j,t-1-l}\right)^4 \\
\leq& b\mathbb{E}\left(\sum_{l=2}^{\infty} l(l-1)\alpha_j\beta_j^{l-2} x^2_{j,t-1-l}\right)^4 + b\frac{(2\omega_j)^4}{(1-\beta_j)^{12}} \\
\leq& c\left(\sum_{l=2}^{\infty} l(l-1)\alpha_j\beta_j^{l-2}\left[\mathbb{E}(x^2_{j,t-1-l})^4\right]^{1/3}\right)^4 + b\frac{(2\omega_j)^4}{(1-\beta_j)^{12}} < \infty
\end{aligned}
$$

Above proved the finiteness of the absolute diagonal terms of the second derivative. The non-diagonal elements will be shown below.

$$
\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{\partial \rho_{l,k} \partial \omega_{i_1}}\right|^4 = \mathbb{E}\left[\mathbb{1}_{i=l}\mathbb{1}_{j=k}\left(\frac{\mathbb{1}_{i=i_1}\sigma_{j,t}}{2\sigma_{i,t}(1-\beta_i)} + \frac{\mathbb{1}_{j=i_1}\sigma_{i,t}}{2\sigma_{j,t}(1-\beta_j)}\right)\right]^4
$$

$$
\begin{aligned}
&\leq a\mathbb{E}\left(\sigma_{i,t}\right)^4 + a\mathbb{E}(\sigma_{j,t})^4 < \infty && (3.48)\\
\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{\partial\rho_{l,k}\partial\alpha_{i_1}}\right|^4 &= \mathbb{E}\left[\mathbb{1}_{i=l}\mathbb{1}_{j=k}\Big(\frac{\mathbb{1}_{i=i_1}\sigma_{j,t}}{2\sigma_{i,t}} + \frac{\mathbb{1}_{j=i_1}\sigma_{i,t}}{2\sigma_{j,t}}\Big)\sum_{h=0}^{\infty}\beta_{i_1}^h x_{i_1,t-1-h}^2\right]^4\\
&\leq \mathbb{E}\left[\mathbb{1}_{i=l}\mathbb{1}_{j=k}(\sigma_{i,t}+\sigma_{j,t})\frac{\sum_{h=0}^{\infty}\beta_{i_1}^h x_{i_1,t-1-h}^2}{2\sigma_{i_1,t}}\right]^4\\
&\leq \left\{\mathbb{E}\sigma_{k,t}^8\mathbb{E}\left[\frac{\sum_{h=0}^{\infty}\beta_{i_1}^h x_{i_1,t-1-h}^2}{2\sigma_{l,t}}\right]^8\right\}^{1/2} < \infty && (3.49)\\
\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{\partial\rho_{l,k}\partial\beta_{i_1}}\right|^4 &= \mathbb{E}\left[\mathbb{1}_{i=l}\mathbb{1}_{j=k}\left(\frac{\mathbb{1}_{i=i_1}\sigma_{j,t}}{2\sigma_{i,t}} + \frac{\mathbb{1}_{j=i_1}\sigma_{i,t}}{2\sigma_{j,t}}\right)\Big(\sum_{h=1}^{\infty}\alpha_{i_1}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2 + \frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\Big)\right]^4\\
&\leq \mathbb{E}\left[\frac{\sum_{h=1}^{\infty}\alpha_{i_1}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{2\sigma_{i_1,t}}\left(\sigma_{i,t}+\sigma_{j,t}\right) + \frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\frac{\sigma_{i,t}+\sigma_{j,t}}{2\sigma_{i_1,t}}\right]^4\\
&\leq a\mathbb{E}\left[\frac{\sum_{h=1}^{\infty}\alpha_{i_1}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{2\sigma_{i_1,t}}\sigma_{i,t}\right]^4 + b\mathbb{E}\left[\frac{\sigma_{i,t}}{2\sigma_{i_1,t}}\right]^4\\
&\quad + c\mathbb{E}\left[\frac{\sum_{h=1}^{\infty}\alpha_{i_1}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{2\sigma_{i_1,t}}\sigma_{j,t}\right]^4 + d\mathbb{E}\left[\frac{\sigma_{j,t}}{2\sigma_{i_1,t}}\right]^4 < \infty\\
\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{\partial^2\omega_{i_1}}\right|^4 &= \mathbb{E}\left|\mathbb{1}_{i=i_1}\rho_{i,j}\frac{-\sigma_{j,t}}{4\sigma_{i_1,t}^3} + \mathbb{1}_{j=i_1}\rho_{i,j}\frac{-\sigma_{i,t}}{4\sigma_{i_1,t}^3}\right|^4\left(\frac{1}{(1-\beta_{i_1})^2}\right)^4 < \infty\\
\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{\partial\omega_l\partial\omega_k}\right|^4 &= \mathbb{E}\left(\frac{\mathbb{1}_{i=l}\mathbb{1}_{j=k}\rho_{i,j}}{4\sigma_{l,t}\sigma_{k,t}(1-\beta_l)(1-\beta_k)}\right)^4 < \infty\\
\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{\partial\omega_l\partial\alpha_k}\right|^4 &= \mathbb{E}\left|\frac{(\mathbb{1}_{i=l}\mathbb{1}_{j=k}+\mathbb{1}_{i=k}\mathbb{1}_{j=l})\rho_{i,j}\mathbb{1}_{k\neq l}}{4\sigma_{l,t}\sigma_{k,t}(1-\beta_l)}\sum_{h=0}^{\infty}\beta_k^h x_{k,t-1-h}^2\right.\\
&\quad \left. + \frac{-(\mathbb{1}_{i=l}\sigma_{i,t}+\mathbb{1}_{j=l}\sigma_{j,t})\rho_{i,j}\mathbb{1}_{k=l}}{4\sigma_{l,t}^3(1-\beta_l)}\sum_{h=0}^{\infty}\beta_l^h x_{l,t-1-h}^2\right|^4\\
&\leq a\mathbb{E}\left|\sum_{h=0}^{\infty}\beta_k^h x_{k,t-1-h}^2\right|^4 + b\mathbb{E}\left|\sum_{h=0}^{\infty}\beta_l^h x_{l,t-1-h}^2\right|^4 < \infty\\
\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{\partial\omega_l\partial\beta_{i_1}}\right|^4 &= \mathbb{E}\left|\frac{(\mathbb{1}_{i=l}\sigma_{j,t}+\mathbb{1}_{j=l}\sigma_{i,t})\mathbb{1}_{l=i_1}\rho_{i,j}}{2\sigma_{l,t}(1-\beta_l)^2} + \frac{\beta_{0l}}{(1-\beta_l)^2}\right.\\
&\quad \left. - \frac{(\mathbb{1}_{i=l}\sigma_{j,t}+\mathbb{1}_{j=l}\sigma_{i,t})\mathbb{1}_{l=i_1}\rho_{i,j}}{4\sigma_{l,t}^3(1-\beta_l)}\left[\sum_{h=0}^{\infty}h\alpha_l\beta_l^{h-1}x_{l,t-1-h}^2 + \frac{\omega_l}{(1-\beta_l)^2}\right]\right|^4\\
&= a\mathbb{E}\left|\frac{(\sigma_{j,t}+\sigma_{i,t})\rho_{i,j}}{2\sigma_{l,t}}\right|^4 + b\frac{\beta_{0l}^4}{(1-\beta_l)^8}
\end{aligned}
$$

$$
\begin{aligned}
&+ c\mathbb{E}\left|\frac{(\sigma_{j,t}+\sigma_{i,t})\rho_{i,j}}{4\sigma_{l,t}^3(1-\beta_l)}\left[\sum_{h=0}^{\infty} h\alpha_l\beta_l^{h-1}x_{l,t-1-h}^2+\frac{\omega_l}{(1-\beta_l)^2}\right]\right|^4 \\
&<\infty
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{\partial^2\alpha_l}\right|^4 =&\mathbb{E}\left[\left|-\frac{\mathbb{1}_{i=l}\rho_{i,j}\sigma_{j,t}}{4\sigma_{l,t}^3}-\frac{\mathbb{1}_{j=l}\rho_{i,j}\sigma_{i,t}}{4\sigma_{l,t}^3}\right|^4\left|\sum_{k=0}^{\infty}\beta_l^k x_{l,t-1-k}^2\right|^4\right] \\
\leq&\mathbb{E}\left(\frac{\rho_{i,j}\sigma_{j,t}}{4\sigma_{l,t}^3}+\frac{\rho_{i,j}\sigma_{i,t}}{4\sigma_{l,t}^3}\right)^4(\sum_{k=0}^{\infty}\beta_l^k x_{l,t-1-k}^2)^4 \\
\leq& a\left\{\mathbb{E}\left(\rho_{i,j}\sigma_{j,t}+\rho_{i,j}\sigma_{i,t}\right)^8\mathbb{E}\left[\frac{(\sum_{k=0}^{\infty}\beta_l^k x_{l,t-1-k}^2)}{\sigma_{l,t}}\right]^8\right\}^{1/2} < \infty \\
\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{\partial\alpha_l\partial\alpha_k}\right|^4 =&\mathbb{E}\left[(\mathbb{1}_{i=l}\mathbb{1}_{j=k}+\mathbb{1}_{i=k}\mathbb{1}_{j=l})\frac{\rho_{i,j}}{4\sigma_{l,t}\sigma_{k,t}}\left(\sum_{h=0}^{\infty}\beta_l^h x_{l,t-1-h}^2\right)\left(\sum_{g=0}^{\infty}\beta_k^g x_{k,t-1-g}^2\right)\right]^4 \\
\leq&\frac{\rho_{i,j}^4}{4^4}\left\{\mathbb{E}\frac{\left(\sum_{h=0}^{\infty}\beta_l^h x_{l,t-1-h}^2\right)^8}{\sigma_{l,t}^8}\mathbb{E}\frac{\left(\sum_{g=0}^{\infty}\beta_k^g x_{k,t-1-g}^2\right)^8}{\sigma_{k,t}^8}\right\}^{1/2} < \infty \\
\mathbb{E}\left|\frac{\partial H_{ij,t}}{\partial\alpha_l\partial\beta_l}\right|^4 =&\rho_{i,j}^4\mathbb{E}\left|\left(\sum_{k=1}^{\infty}k\beta_j^k x_{l,t-1-k}^2\right)\left(\frac{\mathbb{1}_{i=l}\sigma_{j,t}}{2\sigma_{l,t}}+\frac{\mathbb{1}_{j=l}\sigma_{i,t}}{2\sigma_{l,t}}\right)\right. \\
&\left.-\left(\frac{\mathbb{1}_{i=l}\sigma_{j,t}}{4\sigma_{l,t}^3}+\frac{\mathbb{1}_{j=l}\sigma_{i,t}}{4\sigma_{l,t}^3}\right)\left(\sum_{k=0}^{\infty}\beta_j^k x_{l,t-1-k}^2\right)\left(\sum_{k=1}^{\infty}k\beta_j^{k-1}\alpha_l x_{l,t-1-k}^2+\frac{\omega_l}{(1-\beta_l)^2}\right)\right|^4 \\
\leq& a\mathbb{E}\left[\left(\frac{\sum_{k=1}^{\infty}k\beta_j^k x_{l,t-1-k}^2}{\sigma_{l,t}}\right)\left(\sigma_{i,t}+\sigma_{j,t}\right)\right]^4+b\mathbb{E}\left[\left(\sum_{k=0}^{\infty}\beta_j^k x_{l,t-1-k}^2\right)\left(\sigma_{i,t}+\sigma_{j,t}\right)\right]^4 \\
&+c\mathbb{E}\left[\left(\sigma_{i,t}+\sigma_{j,t}\right)\frac{\left(\sum_{k=0}^{\infty}\beta_j^k x_{l,t-1-k}^2\right)\left(\sum_{k=1}^{\infty}k\beta_j^{k-1}\alpha_l x_{l,t-1-k}^2\right)}{4\sigma_{l,t}^3}\right]^4 \\
<&\infty
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial H_{ij,t}}{\partial\alpha_l\partial\beta_k}\right|^4 =&\mathbb{E}\left|\frac{(\mathbb{1}_{i=l}+\mathbb{1}_{j=l})\mathbb{1}_{k\neq l}\rho_{i,j}}{4\sigma_{l,t}\sigma_{k,t}}\left(\sum_{g=0}^{\infty}\beta_j^g x_{l,t-1-g}^2\right)\left(\sum_{h=0}^{\infty}\alpha_k\beta_j^{h-1}x_{k,t-1-h}^2+\frac{\omega_k}{(1-\beta_k)^2}\right)\right|^4 \\
\leq& a\left[\mathbb{E}\left|\frac{\sum_{g=0}^{\infty}\beta_j^g x_{l,t-1-g}^2}{\sigma_{l,t}}\right|^8\mathbb{E}\left|\frac{\partial H_{ij,t}}{\partial\beta_l}\right|^8\right]^{1/2} < \infty \\
\mathbb{E}\left|\frac{\partial H_{ij,t}}{\partial\alpha_l\partial\beta_{0k}}\right|^4 =&\mathbb{E}\left|(\mathbb{1}_{i=l}+\mathbb{1}_{j=l})\mathbb{1}_{l=k}\sum_{h=0}^{\infty}\beta_l^h x_{l,t-1-h}^2\right|^4 < \infty \\
\mathbb{E}\left|\frac{\partial H_{ij,t}}{\partial^2\beta_l}\right|^4 =&\mathbb{E}\left|(\mathbb{1}_{i=l}\sigma_{j,t}+\mathbb{1}_{j=l}\sigma_{i,t})\frac{-\rho_{i,j}}{4\sigma_{l,t}^3}\left(\frac{\omega_l}{(1-\beta_l)^2}+\alpha_l\sum_{h=1}^{\infty}h\beta_l^{j-1}x_{l,t-1-j}^2\right)^2\right.
\end{aligned}
$$

$$
\begin{aligned}
&+\left(\frac{(\mathbb{1}_{i=l}\sigma_{j,t}+\mathbb{1}_{j=l}\sigma_{i,t})}{2\sigma_{l,t}}+\beta_{0l}\right)\left(\frac{2\omega_l}{(1-\beta_l)^4}+\alpha_l\sum_{h=2}^{\infty}h(h-1)\beta_l^{h-2}x^2_{l,t-1-h}\right)\Bigg|^4\\
\leq &a\mathbb{E}(\sigma_{i,t}+\sigma_{j,t})^4+b\mathbb{E}\left\{\left(\frac{(\alpha_l\sum_{h=1}^{\infty}h\beta_l^{h-1}x^2_{l,t-1-h})^2}{\sigma^3_{l,t}}\right)^8\mathbb{E}(\mathbb{1}_{i=l}\sigma_{j,t}+\mathbb{1}_{j=l}\sigma_{i,t})^8\right\}^{1/2}\\
&+c\mathbb{E}\left\{\left(\frac{\alpha_l\sum_{h=2}^{\infty}h(h-1)\beta_l^{h-2}x^2_{l,t-1-h}}{\sigma_{l,t}}\right)^8\mathbb{E}(\mathbb{1}_{i=l}\sigma_{j,t}+\mathbb{1}_{j=l}\sigma_{i,t})^8\right\}+d\\
<&\infty
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial H_{ij,t}}{\partial\beta_l\partial\beta_k}\right|^4=&\mathbb{E}\Bigg|\left(\mathbb{1}_{i=l}\mathbb{1}_{j=k}+\mathbb{1}_{j=l}\mathbb{1}_{i=k}\right)\frac{\rho_{i,j}}{4\sigma_{l,t}\sigma_{k,t}}\left(\frac{\omega_l}{(1-\beta_l)^2}+\alpha_l\sum_{h=1}^{\infty}h\beta_l^{j-1}x^2_{l,t-1-j}\right)\\
&\left(\frac{\omega_k}{(1-\beta_k)^2}+\alpha_k\sum_{h=1}^{\infty}h\beta_k^{j-1}x^2_{k,t-1-j}\right)\Bigg|^4\\
\leq &a\left\{\mathbb{E}\left[\frac{\alpha_k\sum_{h=1}^{\infty}h\beta_k^{j-1}x^2_{k,t-1-j}}{\sigma_{k,t}}\right]^8\mathbb{E}\left[\frac{\alpha_l\sum_{h=1}^{\infty}h\beta_l^{j-1}x^2_{l,t-1-j}}{\sigma_{l,t}}\right]^8\right\}^{1/2}+d\\
&+b\mathbb{E}\left[\frac{\alpha_l\sum_{h=1}^{\infty}h\beta_l^{j-1}x^2_{l,t-1-j}}{\sigma_{l,t}}\right]^4+c\mathbb{E}\left[\frac{\alpha_k\sum_{h=1}^{\infty}h\beta_k^{j-1}x^2_{k,t-1-j}}{\sigma_{k,t}}\right]^4\\
<&\infty
\end{aligned}
$$

$$
\mathbb{E}\left|\frac{\partial H_{ij,t}}{\partial\beta_l\partial\beta_{0k}}\right|^4=\mathbb{E}\left|(\mathbb{1}_{i=l}\mathbb{1}_{l=k}+\mathbb{1}_{j=l}\mathbb{1}_{l=k})\left(\frac{\omega_l}{(1-\beta_l)^2}+\alpha_l\sum_{h=1}^{\infty}\beta_l^{h-1}x^2_{l,t-1-h}\right)\right|^4<\infty
$$

All the absolute third moments of the terms have been proved to be finite. The elements which are not listed above have the value 0. The desired result, $\mathbb{E}\left\|{}_{ij}\ddot{H}_t\right\|_p^4<\infty$, follows.

(iv) Among all the third derivatives of the diagonal elements, only 6 terms are not zero and they are listed below.

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^3 H_{ii,t}}{\partial\omega_j\partial\beta_{0j}\partial\beta_j}\right|^4=&(\frac{1}{1-\beta_j})^8<\infty\\
\mathbb{E}\left|\frac{\partial^3 H_{ii,t}}{\partial\omega_j\partial\beta_j\partial\beta_j}\right|^4=&(\mathbb{1}_{i=j}+\beta_{0j})^4(\frac{2}{1-\beta_j})^{12}<\infty\\
\mathbb{E}\left|\frac{\partial^3 H_{ii,t}}{\partial\alpha_j\partial\beta_{0j}\partial\beta_j}\right|^4=&\mathbb{E}\left|\sum_{l=1}^{\infty}l\beta_j^{l-1}x^2_{j,t-1-j}\right|^4<\infty
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^3 H_{ii,t}}{\partial\alpha_j\partial\beta_j\partial\beta_j}\right|^4 &=(\mathbb{1}_{i=j}+\beta_{0j})^4\mathbb{E}\left|\sum_{l=1}^{\infty} l(l-1)\beta_j^{l-2}x_{j,t-1-j}^2\right|^4 < \infty \\
\mathbb{E}\left|\frac{\partial^3 H_{ii,t}}{\partial\beta_j\partial\beta_{0j}\partial\beta_j}\right|^4 &=\mathbb{E}\left|\frac{3\omega_j}{(1-\beta_j)^4}+\sum_{l=1}^{\infty} l(l-1)\beta_j^{l-2}x_{j,t-1-j}^2\right|^4 < \infty \\
\mathbb{E}\left|\frac{\partial^3 H_{ii,t}}{\partial\beta_j\partial\beta_j\partial\beta_j}\right|^4 &=(\mathbb{1}_{i=j}+\beta_{0j})^4\mathbb{E}\left|\frac{6\omega_j}{(1-\beta_j)^4}+\sum_{l=1}^{\infty} l(l-1)(l-2)\beta_j^{l-3}x_{j,t-1-j}^2\right|^4 < \infty
\end{aligned}
$$

The non-diagonal elements have more non-zero terms. By Lemma 3.3.2 and (3), (4) in Appendix A, we could prove the following results.

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\rho_{l,k}\partial\omega_{i_1}\partial\omega_{i_1}}\right|^4 &=\mathbb{E}\left[\mathbb{1}_{i=l}\mathbb{1}_{j=k}\left(\frac{\mathbb{1}_{i=i_1}\sigma_{j,t}}{4\sigma_{i,t}^3(1-\beta_i)^2}+\frac{\mathbb{1}_{j=i_1}\sigma_{i,t}}{4\sigma_{j,t}^3(1-\beta_j)^2}\right)\right]^4 < \infty \\
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\rho_{l,k}\partial\omega_{i_1}\partial\omega_{i_2}}\right|^4 &=\mathbb{E}\left[\mathbb{1}_{i=l}\mathbb{1}_{j=k}\left(\frac{\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2}}{4\sigma_{j,t}\sigma_{i,t}(1-\beta_i)(1-\beta_j)}\right)\right]^4 < \infty \\
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\rho_{l,k}\partial\omega_{i_1}\partial\alpha_{i_1}}\right|^4 &=\mathbb{E}\left|-\mathbb{1}_{i=l}\mathbb{1}_{j=k}\left(\frac{\mathbb{1}_{i=i_1}\sigma_{j,t}+\mathbb{1}_{j=i_1}\sigma_{i,t}}{4\sigma_{i_1,t}^3(1-\beta_{i_1})}\right)\sum_{h=0}^{\infty}\beta_{i_1}^h x_{i_1,t-1-h}^2\right|^4 < \infty \\
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\rho_{l,k}\partial\omega_{i_1}\partial\alpha_{i_2}}\right|^4 &=\mathbb{E}\left[\mathbb{1}_{i=l}\mathbb{1}_{j=k}\left(\frac{\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2}}{4\sigma_{j,t}\sigma_{i,t}(1-\beta_{i_1})}\right)\sum_{h=0}^{\infty}\beta_{i_2}^h x_{i_2,t-1-h}^2\right]^4 < \infty \\
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\rho_{l,k}\partial\omega_{i_1}\partial\beta_{i_1}}\right|^4 &=\mathbb{E}\Bigg|-\left(\frac{\mathbb{1}_{i=i_1}\sigma_{j,t}+\mathbb{1}_{j=i_1}\sigma_{i,t}}{4\sigma_{i_1,t}^3(1-\beta_{i_1})}\right)\Bigg(\alpha_{i_1}\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2 \\
&\quad+\frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\Bigg)+\frac{\mathbb{1}_{i=i_1}\sigma_{j,t}+\mathbb{1}_{j=i_1}\sigma_{i,t}}{2\sigma_{i_1,t}(1-\beta_{i_1})^2}\Bigg|^4\mathbb{1}_{i=l}\mathbb{1}_{j=k} \\
&\le b\sqrt{(\mathbb{E}\sigma_i^8+\mathbb{E}\sigma_j^8)\mathbb{E}\left(\frac{\alpha_{i_1}\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{\sigma_{i_1,t}^3}\right)^8} \\
&\quad+a\mathbb{E}(\sigma_i+\sigma_j)^4 < \infty
\end{aligned} \tag{3.50}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\rho_{l,k}\partial\omega_{i_1}\partial\beta_{i_2}}\right|^4 &=\mathbb{E}\Bigg[\mathbb{1}_{i=l}\mathbb{1}_{j=k}\left(\frac{\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2}}{4\sigma_{j,t}\sigma_{i,t}(1-\beta_{i_1})}\right)\Bigg(\sum_{h=1}^{\infty}h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2 \\
&\quad+\frac{\omega_{i_2}}{(1-\beta_{i_2})^2}\Bigg)\Bigg]^4 \\
&\le c\mathbb{E}\left(\sum_{h=1}^{\infty}h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2+\frac{\omega_{i_2}}{(1-\beta_{i_2})^2}\right)^4 < \infty \\
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\rho_{l,k}\partial\alpha_{i_1}\partial\alpha_{i_1}}\right|^4 &=\mathbb{E}\left|-\mathbb{1}_{i=l}\mathbb{1}_{j=k}\Big(\frac{\mathbb{1}_{i=i_1}\sigma_{j,t}}{4\sigma_{i_1,t}^3}+\frac{\mathbb{1}_{j=i_1}\sigma_{i,t}}{4\sigma_{i_1,t}^3}\Big)\Big(\sum_{h=0}^{\infty}\beta_{i_1}^h x_{i_1,t-1-h}^2\Big)^2\right|^4
\end{aligned}
$$

$$\leq a\sqrt{\mathbb{E}(\sigma_{i,t}^8+\sigma_{j,t}^8)\mathbb{E}\bigg(\frac{(\sum_{h=0}^\infty \beta_{i_1}^h x_{i_1,t-1-h}^2)^2}{\sigma_{i_1,t}^3}\bigg)^8} < \infty \tag{3.51}$$

$$\begin{aligned}
\mathbb{E}\bigg|\frac{\partial^3 H_{ij,t}}{\partial\rho_{l,k}\partial\alpha_{i_1}\partial\alpha_{i_2}}\bigg|^4 =&\mathbb{E}\bigg[\mathbb{1}_{i=l}\mathbb{1}_{j=k}\bigg(\frac{\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2}}{4\sigma_{i_1,t}\sigma_{j,t}}\bigg)\\
&(\sum_{h=0}^\infty \beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2)(\sum_{h_2=0}^\infty \beta_{i_2}^{h_2}x_{i_2,t-1-h_2}^2)\bigg]^4\\
\leq& b\sqrt{\mathbb{E}\bigg(\frac{\sum_{h_1=0}^\infty \beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2}{\sigma_{i_1,t}}\bigg)^8\mathbb{E}\bigg(\frac{\sum_{h_2=0}^\infty \beta_{i_2}^{h_2}x_{i_2,t-1-h_2}^2}{\sigma_{i_2,t}}\bigg)^8} < \infty
\end{aligned}\tag{3.52}$$

$$\begin{aligned}
\mathbb{E}\bigg|\frac{\partial^2 H_{ij,t}}{\partial\rho_{l,k}\partial\alpha_{i_1}\partial\beta_{i_1}}\bigg|^4 =&\mathbb{E}\bigg|-\mathbb{1}_{i=l}\mathbb{1}_{j=k}\bigg(\frac{\mathbb{1}_{i=i_1}\sigma_{j,t}}{2\sigma_{i_1,t}}+\frac{\mathbb{1}_{j=i_1}\sigma_{i,t}}{2\sigma_{i_1,t}}\bigg)\bigg[\frac{\sum_{h=0}^\infty \beta_{i_1}^h x_{i_1,t-1-h}^2}{2\sigma_{i_1,t}^2}\\
&(\sum_{h=1}^\infty \alpha_{i_1}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2+\frac{\omega_{i_1}}{(1-\beta_{i_1})^2})+\sum_{h=0}^\infty \beta_{i_1}^{h-1}hx_{i_1,t-1-h}^2\bigg]\bigg|^4\\
\leq&\bigg\{a\bigg[\mathbb{E}\bigg(\frac{(\sum_{h=0}^\infty \alpha_{i_1}\beta_{i_1}^h x_{i_1,t-1-h}^2)(\sum_{h=1}^\infty \alpha_{i_1}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2)}{\sigma_{i_1,t}^3}\bigg)^8\bigg]^{1/2}\\
&+b\bigg[\mathbb{E}\bigg(\frac{\sum_{h=1}^\infty \alpha_{i_1}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{\sigma_{i_1,t}^3}\bigg)^8\bigg]^{1/2}\\
&+c\bigg[\mathbb{E}\bigg(\frac{\sum_{h=1}^\infty \alpha_{i_1}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{\sigma_{i_1,t}}\bigg)^8\bigg]^{1/2}\bigg\}\sqrt{\mathbb{E}\sigma_{i,t}^8+\mathbb{E}\sigma_{j,t}^8}\\
<&\infty
\end{aligned}\tag{3.53}$$

$$\begin{aligned}
\mathbb{E}\bigg|\frac{\partial^3 H_{ij,t}}{\partial\rho_{l,k}\partial\alpha_{i_1}\partial\beta_{i_2}}\bigg|^4 =&\mathbb{E}\bigg[\mathbb{1}_{i=l}\mathbb{1}_{j=k}\bigg(\frac{\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2}}{4\sigma_{i_1,t}\sigma_{j,t}}\bigg)\bigg(\sum_{h=0}^\infty \beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2\bigg)\\
&\bigg(\sum_{h_2=0}^\infty \alpha_{i_2}\beta_{i_2}^{h_2}x_{i_2,t-1-h_2}^2+\frac{\omega_{i_2}}{(1-\beta_{i_2})^2}\bigg)\bigg]^4\\
\leq& a\sqrt{\mathbb{E}\bigg(\frac{\sum_{h_1=0}^\infty \beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2}{\sigma_{i_1,t}}\bigg)^8\mathbb{E}\bigg(\frac{\sum_{h_2=0}^\infty \alpha_{i_2}\beta_{i_2}^{h_2}x_{i_2,t-1-h_2}^2}{\sigma_{i_2,t}}\bigg)^8}\\
&+b\mathbb{E}\bigg(\frac{\sum_{h_1=0}^\infty \beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2}{\sigma_{i_1,t}}\bigg)^4 < \infty
\end{aligned}\tag{3.54}$$

$$\begin{aligned}
\mathbb{E}\bigg|\frac{\partial^3 H_{ij,t}}{\partial\rho_{l,k}\partial\beta_{i_1}\partial\beta_{i_1}}\bigg|^4 =&\mathbb{E}\bigg|-\mathbb{1}_{i=l}\mathbb{1}_{j=k}\bigg(\frac{\mathbb{1}_{i=i_1}\sigma_{j,t}}{2\sigma_{i_1,t}}+\frac{\mathbb{1}_{j=i_1}\sigma_{i,t}}{2\sigma_{i_1,t}}\bigg)\\
&\bigg[\bigg(\frac{\sum_{h=1}^\infty \alpha_{i_1}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{2\sigma_{i_1,t}^2}+\frac{\omega_{i_1}}{2\sigma_{i_1,t}^2(1-\beta_{i_1})^2}\bigg)^2+\\
&\frac{2\omega_{i_1}}{(1-\beta_{i_1})^3}+\sum_{h=2}^\infty \alpha_{i_1}h(h-1)\beta_{i_1}^{h-2}x_{i_1,t-1-h}^2\bigg]\bigg|^4
\end{aligned}$$

$$
\begin{aligned}
&\leq a\Big\{\mathbb{E}\Big[\frac{(\sum_{h=1}^{\infty}\alpha_{i_1}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2)^2}{\sigma_{i_1,t}^3}\Big]^8\Big\}^{1/2}\sqrt{\mathbb{E}(\sigma_{i,t}^8+\sigma_{j,t}^8)}\\
&\quad + c\sqrt{\mathbb{E}\Big[\frac{\sum_{h=2}^{\infty}\alpha_{i_1}h(h-1)\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{\sigma_{i_1,t}}\Big]^8}\sqrt{\mathbb{E}(\sigma_{i,t}^8+\sigma_{j,t}^8)}\\
&\quad + b\mathbb{E}(\sigma_{i,t}^4+\sigma_{j,t}^4)<\infty
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{\partial\rho_{l,k}\partial\beta_{i_1}\partial\beta_{i_2}}\Big|^4 =&\mathbb{E}\Big[\mathbb{1}_{i=l}\mathbb{1}_{j=k}\Big(\frac{\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2}}{4\sigma_{i_1,t}\sigma_{j,t}}\Big)\Big(\sum_{h=1}^{\infty}\alpha_{i_1}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2\\
&+\frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\Big)\Big(\sum_{h_2=0}^{\infty}\alpha_{i_2}\beta_{i_2}^{h_2}x_{i_2,t-1-h_2}^2+\frac{\omega_{i_2}}{(1-\beta_{i_2})^2}\Big)\Big]^4\\
\leq& a\sqrt{\mathbb{E}\Big(\frac{\sum_{h_1=0}^{\infty}\beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2}{\sigma_{i_1,t}}\Big)^8\mathbb{E}\Big(\frac{\sum_{h_2=0}^{\infty}\alpha_{i_2}\beta_{i_2}^{h_2}x_{i_2,t-1-h_2}^2}{\sigma_{i_2,t}}\Big)^8}\\
&+b\mathbb{E}\Big(\frac{\sum_{h_1=0}^{\infty}\beta_{i_1}^{h_1}\alpha_{i_1}x_{i_1,t-1-h_1}^2}{\sigma_{i_1,t}}\Big)^4\\
&+c\mathbb{E}\Big(\frac{\sum_{h_2=0}^{\infty}\beta_{i_2}^{h_2}\alpha_{i_2}x_{i_2,t-1-h_1}^2}{\sigma_{i_2,t}}\Big)^4+d<\infty
\end{aligned}
\tag{3.55}
$$

$$
\begin{aligned}
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{(\partial\omega_{i_1})^3}\Big|^4 =&\mathbb{E}\Big|\mathbb{1}_{i=i_1}\rho_{i,j}\frac{3\sigma_{j,t}}{8\sigma_{i_1,t}^5}+\mathbb{1}_{j=i_1}\rho_{i,j}\frac{3\sigma_{i,t}}{8\sigma_{i_1,t}^5}\Big|^4\Big(\frac{1}{(1-\beta_{i_1})^3}\Big)^4<\infty\\
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{(\partial\omega_{i_1})^2\partial\omega_{i_2}}\Big|^4 =&\mathbb{E}\Big|\Big(\frac{\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}\rho_{i,j}}{8\sigma_{i_1,t}^3\sigma_{i_2,t}(1-\beta_{i_1})^2(1-\beta_{i_2})}\Big)\Big|^4<\infty\\
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{(\partial\omega_{i_1})^2\partial\alpha_{i_1}}\Big|^4 =&\mathbb{E}\Big(\Big|\mathbb{1}_{i=i_1}\sigma_{j,t}+\mathbb{1}_{j=i_1}\sigma_{i,t}\Big|\frac{3\rho_{i,j}\sum_{h_1=0}^{\infty}\beta_{i_1}^{h_1}\alpha_{i_1}x_{i_1,t-1-h_1}^2}{8\sigma_{i_1,t}^5(1-\beta_{i_1})^2}\Big)^4\\
\leq& a\sqrt{\mathbb{E}(\rho_{i,j}\frac{\sum_{h_1=0}^{\infty}\beta_{i_1}^{h_1}\alpha_{i_1}x_{i_1,t-1-h_1}^2}{\sigma_{i_1,t}^5})^8}\sqrt{\mathbb{E}(\sigma_i^8+\sigma_j^8)}<\infty\\
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{(\partial\omega_{i_1})^2\partial\alpha_{i_2}}\Big|^4 =&\mathbb{E}\Big|-\frac{\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}\rho_{i,j}}{8\sigma_{i_1,t}^3\sigma_{i_2,t}(1-\beta_{i_1})^2}\sum_{h_1=0}^{\infty}\beta_{i_2}^{h_1}x_{i_2,t-1-h_1}^2\Big|^4<\infty\\
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{(\partial\omega_{i_1})^2\partial\beta_{i_1}}\Big|^4 =&\mathbb{E}\Big|\Big(\frac{\mathbb{1}_{i=i_1}\sigma_{j,t}+\mathbb{1}_{j=i_1}\sigma_{i,t}}{4\sigma_{i_1,t}^3}\Big)\Big[-\frac{2\rho_{i,j}}{(1-\beta_{i_1})^3}+\\
&\Big(\sum_{h=1}^{\infty}\alpha_{i_1}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2+\frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\Big)^2\frac{3\rho_{i,j}}{2\sigma_{i_1,t}^2(1-\beta_{i_1})^2}\Big]\Big|^4\\
\leq& a\Big\{\mathbb{E}\Big[\frac{\sum_{h=1}^{\infty}\alpha_{i_1}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{\sigma_{i_1,t}^5}\Big]^8\Big\}^{1/2}\sqrt{\mathbb{E}(\sigma_{i,t}^8+\sigma_{j,t}^8)}\\
&+c\sqrt{\mathbb{E}\Big[\frac{\sum_{h=2}^{\infty}\alpha_{i_1}h(h-1)\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{\sigma_{i_1,t}}\Big]^8}\sqrt{\mathbb{E}(\sigma_{i,t}^8+\sigma_{j,t}^8)}
\end{aligned}
$$

$$
\begin{aligned}
&+ b\mathbb{E}(\sigma_{i,t}^4 + \sigma_{j,t}^4) < \infty \\
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{(\partial\omega_{i_1})^2\partial\beta_{i_2}}\right|^4 =&\mathbb{E}\left|\left(\rho_{i,j}\frac{\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2} + \mathbb{1}_{i=i_2}\mathbb{1}_{j=i_1}}{8\sigma_{i,t}\sigma_{j,t}(1-\beta_{i_1})^2}\right)\left[\sum_{h=1}^{\infty}\alpha_{i_1} h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2\right.\right. \\
&\left.\left. + \frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\right]\right|^4 < \infty \\
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\omega_{i_1}\partial\omega_{i_2}\partial\alpha_{i_1}}\right|^4 =&(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2} + \mathbb{1}_{i=i_2}\mathbb{1}_{j=i_1})\mathbb{E}\left| - \rho_{i,j}\frac{\sum_{h_1=0}^{\infty}\beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2}{8\sigma_{i_1,t}^3\sigma_{i_2,t}(1-\beta_{i_1})(1-\beta_{i_2})}\right|^4 \\
\leq& a\mathbb{E}\Big(\sum_{h_1=0}^{\infty}\beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2\Big)^4 < \infty \\
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\omega_{i_1}\partial\omega_{i_2}\partial\beta_{i_1}}\right|^4 =&\mathbb{E}\left|\frac{\rho_{i,j}}{4\sigma_{i_1,t}\sigma_{i_2,t}(1-\beta_{i_1})(1-\beta_{i_2})}\right[- \frac{\sum_{h_1=0}^{\infty}\beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2}{2\sigma_{i_1,t}^2} - \\
&\left.\frac{\omega_{i_1}}{(1-\beta_{i_1})^2 2\sigma_{i_1,t}^2} + \frac{1}{1-\beta_{i_1}}\right]\Bigg|^4 (\mathbb{1}_{i=i_1} + \mathbb{1}_{j=i_1}) < \infty \\
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\omega_{i_1}(\partial\alpha_{i_1})^2}\right|^4 =&\mathbb{E}(\mathbb{1}_{i=i_1}\sigma_{j,t} + \mathbb{1}_{j=i_1}\sigma_{i,t})\left|\rho_{i,j}\frac{3(\sum_{h_1=0}^{\infty}\beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2)^2}{8\sigma_{i_1,t}^5(1-\beta_{i_1})}\right|^4 \\
\leq& a\mathbb{E}\Big(\frac{\sum_{h_1=0}^{\infty}\beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2}{\sigma_{i_1,t}^5}\Big)^4 < \infty \\
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\omega_{i_1}\partial\alpha_{i_1}\partial\alpha_{i_2}}\right|^4 =&\mathbb{E}\left|\frac{-\rho_{i,j}\sum_{h_1=0}^{\infty}\beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2\sum_{h_2=0}^{\infty}\beta_{i_2}^{h_2}x_{i_2,t-1-h_2}^2}{8\sigma_{i_1,t}^3\sigma_{i_2,t}(1-\beta_{i_1})}\right|^4 \\
&(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2} + \mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2}) < \infty \\
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\omega_{i_1}\partial\alpha_{i_1}\partial\beta_{i_1}}\right|^4 =&\mathbb{E}(\mathbb{1}_{i=i_1}\sigma_{j,t} + \mathbb{1}_{j=i_1}\sigma_{i,t})\left|\frac{\rho_{i,j}}{4\sigma_{i_1,t}^3(1-\beta_{i_1})}\right|^4 \\
&\Bigg| - \sum_{h_2=0}^{\infty}h_2\beta_{i_1}^{h_2-1}x_{i_1,t-1-h_2}^2 - \frac{\sum_{h_1=0}^{\infty}\beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2}{1-\beta_{i_1}} \\
&+ \frac{3\sum_{h_1=0}^{\infty}\beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2}{2\sigma_{i_1,t}^2}\Big(\alpha_{i_1}\sum_{h_2=0}^{\infty}h_2\beta_{i_1}^{h_2-1}x_{i_1,t-1-h_2}^2 \\
&+ \frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\Big)\Bigg|^4 \\
\leq& a\Bigg[\sqrt{\mathbb{E}\Big(\frac{(\sum_{h=0}^{\infty}\beta_{i_1}^{h}x_{i_1,t-1-h}^2)(\alpha_{i_1}\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2)}{\sigma_{i_1,t}^3}\Big)^8} \\
&+ d\sqrt{\mathbb{E}\Big(\frac{(\sum_{h=0}^{\infty}\beta_{i_1}^{h}x_{i_1,t-1-h}^2)(\alpha_{i_1}\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2)}{\sigma_{i_1,t}^3}\Big)^6}
\end{aligned}
$$

$$
\begin{aligned}
&+ c\sqrt{\mathbb{E}(\frac{\sum_{h=0}^{\infty}\beta_{i_1}^{h}x_{i_1,t-1-h}^2}{\sigma_{i_1,t}})^8}\Big]\sqrt{\mathbb{E}(\sigma_i^8+\sigma_j^8)}\\
&+ b\mathbb{E}(\sum_{h=0}^{\infty}\beta_{i_1}^{h}x_{i_1,t-1-h}^2)^4<\infty \qquad (3.56)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{\partial\omega_{i_1}\partial\alpha_{i_1}\partial\beta_{i_2}}\Big|^4 =&\mathbb{E}\Big|\frac{-\rho_{i,j}\sum_{h_1=0}^{\infty}\beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2}{8\sigma_{i_1,t}^3\sigma_{i_2,t}(1-\beta_{i_1})}\Big(\sum_{h_2=0}^{\infty}\beta_{i_2}^{h_2-1}\alpha_{i_2}h_2x_{i_2,t-1-h_2}^2\\
&+\frac{\omega_{i_2}}{(1-\beta_{i_2})^2}\Big)\Big|^4(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2})\\
\le& c\sqrt{\mathbb{E}\Big(\frac{\sum_{h_1=0}^{\infty}\beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2}{\sigma_{i_1,t}}\Big)^8\mathbb{E}\Big(\frac{\sum_{h_2=0}^{\infty}\alpha_{i_2}h_2\beta_{i_2}^{h_2-1}x_{i_2,t-1-h_2}^2}{\sigma_{i_2,t}}\Big)^8}\\
&+ d\mathbb{E}(\sum_{h_1=0}^{\infty}\beta_{i_1}^{h_1}x_{i_1,t-1-h_1}^2)^4<\infty \qquad (3.57)
\end{aligned}
$$

$$
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{\partial\omega_{i_1}(\partial\alpha_{i_2})^2}\Big|^4 =\mathbb{E}\Big|-\frac{(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2})\rho_{i,j}(\sum_{h=0}^{\infty}\beta_{i_2}^{h}x_{i_2,t-1-h}^2)^2}{8\sigma_{i_1,t}\sigma_{i_2,t}^3(1-\beta_{i_1})}\Big|^4<\infty \qquad (3.58)
$$

$$
\begin{aligned}
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{\partial\omega_{i_1}\partial\alpha_{i_2}\partial\beta_{i_1}}\Big|^4 =&\mathbb{E}(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{i=i_2}\mathbb{1}_{j=i_1})\Big|\frac{\rho_{i,j}}{4\sigma_{i_1,t}\sigma_{i_2,t}(1-\beta_{i_1})}\Big|^4\\
&\Big[\frac{\sum_{h=0}^{\infty}\beta_{i_2}^{h}x_{i,t-1-h}^2}{1-\beta_{i_1}}-\frac{\sum_{h=0}^{\infty}\beta_{i_2}^{h}x_{i_2,t-1-h_1}^2}{2\sigma_{i_1,t}^2}\\
&\Big(\alpha_{i_1}\sum_{h=1}^{\infty}h\alpha_{i_1}\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2+\frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\Big)\Big]^4\\
\le& a\sqrt{\mathbb{E}\Big(\frac{\sum_{h=0}^{\infty}\beta_{i_2}^{h}x_{i_2,t-1-h}^2}{\sigma_{i_2,t}}\Big)^8}\sqrt{\mathbb{E}\Big(\frac{\alpha_{i_1}\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{\sigma_{i_1}}\Big)^8}\\
&+ b\mathbb{E}(\sum_{h=0}^{\infty}\beta_{i_2}^{h}x_{i_2,t-1-h}^2)^4+c\mathbb{E}\Big(\frac{(\sum_{h=1}^{\infty}h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2)}{\sigma_{i_2,t}}\Big)^4<\infty \qquad (3.59)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{\partial\omega_{i_1}\partial\alpha_{i_2}\partial\beta_{i_2}}\Big|^4 =&\mathbb{E}(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{i=i_2}\mathbb{1}_{j=i_1})\Big|\frac{\rho_{i,j}}{4\sigma_{i_1,t}\sigma_{i_2,t}(1-\beta_{i_1})}\Big|^4\\
&\Big|\sum_{h=1}^{\infty}h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2-\frac{\sum_{h=0}^{\infty}\beta_{i_2}^{h}x_{i_2,t-1-h_1}^2}{2\sigma_{i_1,t}^2}\\
&\Big(\alpha_{i_2}\sum_{h=1}^{\infty}h\alpha_{i_2}\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2+\frac{\omega_{i_2}}{(1-\beta_{i_2})^2}\Big)\Big|^4\\
\le& a\mathbb{E}\Big(\frac{(\sum_{h=0}^{\infty}\beta_{i_2}^{h}x_{i_2,t-1-h}^2)(\alpha_{i_2}\sum_{h=1}^{\infty}h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2)}{\sigma_{i_2,t}^3}\Big)^4\\
&+ b\mathbb{E}\Big(\sum_{h=1}^{\infty}h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2\Big)^4+c\mathbb{E}(\sum_{h=0}^{\infty}h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2)^4<\infty \qquad (3.60)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\omega_{i_1}(\partial\beta_{i_1})^2}\right|^4 =&\mathbb{E}\Big| -\frac{(\mathbb{1}_{i=i_1}\sigma_{j,t}+\mathbb{1}_{j=i_1}\sigma_{i,t})\rho_{i,j}}{4\sigma^3_{i_1,t}(1-\beta_{i_1})^2}2\Big[\alpha_{i_1}\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h}\\
&+\frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\Big]-\frac{(\mathbb{1}_{i=i_1}\sigma_{j,t}+\mathbb{1}_{j=i_1}\sigma_{i,t})\rho_{i,j}}{4\sigma^3_{i_1,t}(1-\beta_{i_1})}\Big[\frac{2\omega_{i_1}}{(1-\beta_{i_1})^3}\\
&+\alpha_{i_1}\sum_{h=2}^{\infty}h(h-1)\beta_{i_1}^{h-2}x^2_{i_1,t-1-h}\Big]\\
&+\Big[\frac{(\mathbb{1}_{i=i_1}\sigma_{j,t}+\mathbb{1}_{j=i_1}\sigma_{i,t})\rho_{i,j}}{2\sigma_{i_1,t}}+\beta_{0i_1}\Big]\frac{2}{(1-\beta_{i_1})^3}\\
&+\Big[\frac{(\mathbb{1}_{i=i_1}\sigma_{j,t}+\mathbb{1}_{j=i_1}\sigma_{i,t})3\rho_{i,j}}{8\sigma^5_{i_1,t}(1-\beta_{i_1})}\Big]\Big[\frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\\
&+\alpha_{i_1}\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h}\Big]^2\Big|^4\\
\leq& a\sqrt{\mathbb{E}\sigma^8_{i,t}+\mathbb{E}\sigma^8_{j,t}}\sqrt{\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h}}{\sigma_{i_1,t}}\Big)^8}\\
&+b(\mathbb{E}\sigma^4_{i,t}+\mathbb{E}\sigma^4_{j,t})+c\\
&+d\sqrt{\mathbb{E}\sigma^8_{i,t}+\mathbb{E}\sigma^8_{j,t}}\sqrt{\mathbb{E}\Big(\frac{\sum_{h=2}^{\infty}h(h-1)\beta_{i_1}^{h-2}x^2_{i_1,t-1-h}}{\sigma_{i_1,t}}\Big)^8}\\
&+e\sqrt{\mathbb{E}\sigma^8_{i,t}+\mathbb{E}\sigma^8_{j,t}}\sqrt{\mathbb{E}\Big(\frac{(\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h})^2}{\sigma^3_{i_1,t}}\Big)^8}<\infty
\end{aligned}\tag{3.61}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial^2\alpha_{i_1}\beta_{i_1}\beta_{i_2}}\right|^4 =&\mathbb{E}\Big|\frac{(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{i=i_2}\mathbb{1}_{j=i_1})\rho_{i,j}}{4\sigma_{i_1,t}\sigma_{i_2,t}(1-\beta_{i_1})^2}\Big[\frac{\omega_{i_1}}{(1-\beta_{i_1})^2}+\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h}\Big]\\
&-\frac{(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{i=i_2}\mathbb{1}_{j=i_1})\rho_{i,j}}{8\sigma^3_{i_1}\sigma_{i_2}(1-\beta_{i_1})}\Big[\frac{\omega_{i_1}}{(1-\beta_{i_1})^2}+\alpha_{i_1}\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h}\Big]\\
&\Big[\frac{\omega_{i_2}}{(1-\beta_{i_2})^2}+\sum_{h=1}^{\infty}h\beta_{i_2}^{h-1}x^2_{i_2,t-1-h}\Big]\Big|^4\\
\leq& a\sqrt{\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h}}{\sigma_{i_1}}\Big)^8}\sqrt{\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty}h\beta_{i_2}^{h-1}x^2_{i_2,t-1-h}}{\sigma_{i_2}}\Big)^8}\\
&+b\mathbb{E}\Big(\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h}\Big)^4+c<\infty
\end{aligned}\tag{3.62}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{\partial\alpha_{i_1}\partial\beta_{i_2}\partial\beta_{i_2}}\right|^4 =&\mathbb{E}\Big|\frac{(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{i=i_2}\mathbb{1}_{j=i_1})\rho_{i,j}}{4\sigma_{i_1,t}\sigma_{i_2,t}(1-\beta_{i_1})^2}\Big[\alpha_{i_2}\sum_{h=2}^{\infty}h(h-1)\beta_{i_2}^{h-2}x^2_{i_2,t-1-h}\\
&+\frac{2\omega_{i_2}}{(1-\beta_{i_2})^3}\Big]-\frac{(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{i=i_2}\mathbb{1}_{j=i_1})\rho_{i,j}}{8\sigma^3_{i_2}\sigma_{i_1,t}(1-\beta_{i_1})}
\end{aligned}
$$

$$
\begin{aligned}
&\Big[\frac{\omega_{i_2}}{(1-\beta_{i_2})^2}+\alpha_{i_2}\sum_{h=1}^{\infty}h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2\Big]^2\Big|^4\\
\leq& a\mathbb{E}(\sum_{h=2}^{\infty}h(h-1)\beta_{i_2}^{h-2}x_{i_2,t-1-h}^2)^4+b\\
&+c\mathbb{E}\Big(\frac{(\sum_{h=1}^{\infty}h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2)^2}{\sigma_{i_2,t}^3}\Big)^4<\infty
\end{aligned}\tag{3.63}
$$

$$
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{\partial\omega_{i_1}\partial\beta_{i_1}\partial\beta_{0i_1}}\Big|^4=(\mathbb{1}_{i=i_1}+\mathbb{1}_{j=i_1})\Big[\frac{1}{(1-\beta_{i_1})^2}\Big]^4<\infty
$$

$$
\begin{aligned}
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{(\partial\alpha_{i_1})^3}\Big|^4=&\mathbb{E}\Big|\frac{(\mathbb{1}_{i=i_1}\sigma_{j,t}+\mathbb{1}_{j=i_1}\sigma_{i,t})\rho_{i,j}}{8\sigma_{i_1,t}^5}(\sum_{h=0}^{\infty}\beta_{i_1}^{h}x_{i_1,t-1-h}^2)^3\Big|^4\\
\leq& a\sqrt{\mathbb{E}\sigma_{i,t}^6+\mathbb{E}\sigma_{j,t}^8}\sqrt{\mathbb{E}\Big(\frac{(\sum_{h=0}^{\infty}\beta_{i_1}^{h}x_{i_1,t-1-h}^2)^3}{\sigma_{i_1,t}^5}\Big)^8}<\infty
\end{aligned}\tag{3.64}
$$

$$
\begin{aligned}
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{(\partial\alpha_{i_1})^2\partial\alpha_{i_2}}\Big|^4=&\mathbb{E}\Big|-\frac{(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2})\rho_{i,j}}{8\sigma_{i_1,t}^3\sigma_{i_2,t}}(\sum_{h=0}^{\infty}\beta_{i_1}^{h}x_{i_1,t-1-h}^2)^2\\
&(\sum_{h=0}^{\infty}\beta_{i_2}^{h}x_{i_2,t-1-h}^2)\Big|^4\\
\leq& a\sqrt{\mathbb{E}\Big(\frac{(\sum_{h=0}^{\infty}\beta_{i_1}^{h}x_{i_1,t-1-h}^2)^2}{\sigma_{i_1}^3}\Big)^8}\sqrt{\mathbb{E}\Big(\frac{(\sum_{h=0}^{\infty}\beta_{i_2}^{h}x_{i_2,t-1-h}^2)}{\sigma_{i_2,t}}\Big)^8}\\
<&\infty
\end{aligned}\tag{3.65}
$$

$$
\begin{aligned}
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{(\partial\alpha_{i_1})^2\partial\beta_{i_1}}\Big|^4=&\mathbb{E}\Big|\frac{3(\mathbb{1}_{i=i_1}\sigma_{j,t}+\mathbb{1}_{j=i_1}\sigma_{i,t})\rho_{i,j}}{8\sigma_{i_1,t}^5}(\sum_{h=0}^{\infty}\beta_{i_1}^{h}x_{i_1,t-1-h}^2)^2\Big[\frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\\
&+\sum_{h=0}^{\infty}\beta_{i_1}^{h}x_{i_1,t-1-h}^2\Big]-(\sum_{h=0}^{\infty}\beta_{i_1}^{h}x_{i_1,t-1-h}^2)(\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2)\\
&\frac{(\mathbb{1}_{i=i_1}\sigma_{j,t}+\mathbb{1}_{j=i_1}\sigma_{i,t})\rho_{i,j}}{2\sigma_{i_1,t}^3}\Big|^4\\
\leq&\mathbb{E}\Big|\Big|\frac{\rho_{i,j}(\sigma_{i,t}+\sigma_{j,t})}{\sigma_{i_1,t}^5}\Big(\frac{\sigma_{i_1,t}^2}{\alpha_{i_1}}\Big)^2\Big[\frac{\omega_{i_1}}{(1-\beta_{i_1})^2}+\sum_{h=0}^{\infty}\beta_{i_1}^{h}x_{i_1,t-1-h}^2\Big]\\
&-\frac{\sigma_{i_1,t}^2}{\alpha_{i_1}}(\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2)\frac{\sigma_{j,t}+\mathbb{1}_{j=i_1})\rho_{i,j}}{2\sigma_{i_1,t}^3}\Big|^4\\
\leq& a\sqrt{\mathbb{E}(\sigma_i^8+\sigma_j^8)}\sqrt{\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty}h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{\sigma_{i_1}}\Big)^8}\\
&+b\mathbb{E}(\sigma_i^4+\sigma_j^4)<\infty
\end{aligned}\tag{3.66}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{(\partial\alpha_{i_1})^2\partial\beta_{i_2}}\right|^4 =&\mathbb{E}\Big| - \frac{(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2} + \mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2})\rho_{i,j}}{8\sigma^3_{i_1,t}\sigma_{i_2,t}}(\sum_{h=0}^{\infty}\beta_{i_1}^h x^2_{i_1,t-1-h})\\
&\Big[\frac{\omega_{i_2}}{(1-\beta_{i_2})^2} + \alpha_{i_2}\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1}x^2_{i_2,t-1-h}\Big]\Big|^4\\
\leq&\mathbb{E}\Big| - \frac{\rho_{i,j}}{8\sigma^3_{i_1,t}\sigma_{i_2,t}}\frac{\sigma^2_{i_1,t}}{\alpha_{i_1}}\Big[\frac{\omega_{i_2}}{(1-\beta_{i_2})^2} + \alpha_{i_2}\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1}x^2_{i_2,t-1-h}\Big]\Big|^4\\
\leq& a\mathbb{E}\Big[\frac{\omega_{i_2}}{(1-\beta_{i_2})^2} + \alpha_{i_2}\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1}x^2_{i_2,t-1-h}\Big]^4\\
\leq& b + c\mathbb{E}(\alpha_{i_2}\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1}x^2_{i_2,t-1-h})^4 < \infty
\end{aligned}
\tag{3.67}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\alpha_{i_1}(\partial\beta_{i_1})^2}\right|^4 =&\mathbb{E}\Big|\frac{3(\mathbb{1}_{i=i_1}\sigma_{j,t} + \mathbb{1}_{j=i_1}\sigma_{i,t})\rho_{i,j}}{8\sigma^5_{i_1,t}}(\sum_{h=0}^{\infty}\beta_{i_1}^h x^2_{i_1,t-1-h})\Big[\frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\\
&+ \alpha_{i_1}\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h}\Big] + (\sum_{h=2}^{\infty} h(h-1)\beta_{i_1}^{h-2}x^2_{i_1,t-1-h})\\
&\frac{(\mathbb{1}_{i=i_1} + \mathbb{1}_{j=i_1})\rho_{i,j}}{\sigma_{i_1,t}} - \frac{(\mathbb{1}_{i=i_1}\sigma_{j,t} + \mathbb{1}_{j=i_1}\sigma_{i,t})\rho_{i,j}}{2\sigma^3_{i_1,t}}\Big[\frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\\
&+ \alpha_{i_1}\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h}\Big](\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h})\\
&- \frac{(\mathbb{1}_{i=i_1}\sigma_{j,t} + \mathbb{1}_{j=i_1}\sigma_{i,t})\rho_{i,j}}{4\sigma^3_{i_1,t}}\Big[\frac{2\omega_{i_1}}{(1-\beta_{i_1})^3}\\
&+ \alpha_{i_1}\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h}\Big](\sum_{h=2}^{\infty} h(h-1)\beta_{i_1}^{h-2}x^2_{i_1,t-1-h})\Big|^4\\
\leq&\mathbb{E}\Big|\frac{3(\sigma_{j,t} + \sigma_{i,t})\rho_{i,j}}{8\sigma^5_{i_1,t}}\frac{\sigma^2_{i_1,t}}{\alpha_{i_1}}\Big[\frac{\omega_{i_1}}{(1-\beta_{i_1})^2} + \alpha_{i_1}\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h}\Big]\\
&+ (\sum_{h=2}^{\infty} h(h-1)\beta_{i_1}^{h-2}x^2_{i_1,t-1-h})\frac{\rho_{i,j}}{\sigma_{i_1,t}} - \frac{(\sigma_{j,t} + \sigma_{i,t})\rho_{i,j}}{2\sigma^3_{i_1,t}}\\
&\Big[\frac{\omega_{i_1}}{(1-\beta_{i_1})^2} + \alpha_{i_1}\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h}\Big](\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h})\\
&- \frac{(\sigma_{j,t} + \sigma_{i,t})\rho_{i,j}}{4\sigma^3_{i_1,t}}\Big[\frac{2\omega_{i_1}}{(1-\beta_{i_1})^3} + \alpha_{i_1}\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h}\Big]\frac{\sigma^2_{i_1,t}}{\alpha_{i_1}}\Big|^4\\
\leq& a\mathbb{E}(\sigma_i^4 + \sigma_j^4) + b\sqrt{\mathbb{E}(\sigma_i^8 + \sigma_j^8)}\sqrt{\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x^2_{i_1,t-1-h}}{\sigma_{i_1,t}}\Big)^8}
\end{aligned}
\tag{3.68}
$$

$$
+ c\sqrt{\mathbb{E}(\sigma_i^8 + \sigma_j^8)}\sqrt{\mathbb{E}\Big(\frac{\sum_{h=2}^{\infty} h(h-1)\beta_{i_1}^{h-2}x^2_{i_1,t-1-h}}{\sigma_{i_1,t}}\Big)^8}
\tag{3.69}
$$

$$
\begin{aligned}
&+ d\sqrt{\mathbb{E}(\sigma_i^8+\sigma_j^8)}\Big[\sqrt{\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{\sigma_{i_1,t}}\Big)^8} \\
&+ \sqrt{\mathbb{E}\Big(\frac{(\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2)^2}{\sigma_{i_1,t}^3}\Big)^8} \qquad (3.70)\\
&+ \sqrt{\mathbb{E}\Big(\frac{(\sum_{h=2}^{\infty} h(h-1)\beta_{i_1}^{h-2}x_{i_1,t-1-h}^2)^2}{\sigma_{i_1,t}}\Big)^8}\Big] < \infty \qquad (3.71)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\alpha_{i_1}\partial\beta_{i_1}\partial\beta_{i_2}}\right|^4 =&\mathbb{E}\Big|\frac{(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2})\rho_{i,j}}{4\sigma_{i_1,t}\sigma_{i_2,t}}(\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2) \\
&\Big[\frac{\omega_{i_2}}{(1-\beta_{i_2})^2}+\alpha_{i_2}\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2\Big] - (\sum_{h=0}^{\infty}\beta_{i_1}^{h}x_{i_1,t-1-h}^2) \\
&\frac{(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2})\rho_{i,j}}{8\sigma_{i_1,t}^3\sigma_{i_2,t}}\Big[\alpha_{i_2}\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2 \\
&+\frac{\omega_{i_2}}{(1-\beta_{i_2})^2}\Big]\Big[\frac{\omega_{i_1}}{(1-\beta_{i_1})^2}+\alpha_{i_1}\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2\Big]\Big|^4 \\
\le& a\sqrt{\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{\sigma_{i_1,t}}\Big)^8}\sqrt{\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2}{\sigma_{i_2,t}}\Big)^8} \\
&+ c\mathbb{E}\Big|\frac{\rho_{i,j}\sigma_{i_1,t}^2/\alpha_{i_1}}{\sigma_{i_1,t}^3\sigma_{i_2,t}}\Big[\frac{\omega_{i_2}}{(1-\beta_{i_2})^2}+\alpha_{i_2}\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2\Big] \\
&\Big[\frac{\omega_{i_1}}{(1-\beta_{i_1})^2}+\alpha_{i_1}\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2\Big]\Big|^4 \\
&+ b\mathbb{E}(\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2)^4 \\
\le& d_1 + d_2\mathbb{E}(\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2)^4 + d_3\mathbb{E}(\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2)^4 \\
&+ d_4\sqrt{\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{\sigma_{i_1,t}}\Big)^8\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2}{\sigma_{i_2,t}}\Big)^8} < \infty \qquad (3.72)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{\partial\alpha_{i_1}(\partial\beta_{i_2})^2}\right|^4 =&\mathbb{E}\Big|\frac{(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2})\rho_{i,j}}{4\sigma_{i_1,t}\sigma_{i_2,t}}(\sum_{h=0}^{\infty}\beta_{i_1}^{h}x_{i_1,t-1-h}^2) \\
&\Big[\frac{2\omega_{i_2}}{(1-\beta_{i_2})^3}+\alpha_{i_2}\sum_{h=2}^{\infty} h(h-1)\beta_{i_2}^{h-2}x_{i_2,t-1-h}^2\Big] \\
&-\frac{(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2}+\mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2})\rho_{i,j}}{8\sigma_{i_1,t}\sigma_{i_2,t}^3}\Big[\alpha_{i_2}\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2
\end{aligned}
$$

$$
\begin{aligned}
&+ \frac{\omega_{i_2}}{(1-\beta_{i_2})^2}\Big]^2 \Big(\sum_{h=0}^{\infty} \beta_{i_1}^{h} x_{i_1,t-1-h}^2\Big)\Big|^4 \\
\leq & a\mathbb{E}\Big|\frac{\sigma_{i_1,t}^2/\alpha_{i_1}}{\sigma_{i_1,t}\sigma_{i_2,t}}\Big[\alpha_{i_2}\sum_{h=2}^{\infty} h(h-1)\beta_{i_2}^{h-2}x_{i_2,t-1-h}^2 + \frac{2\omega_{i_2}}{(1-\beta_{i_2})^3}\Big] \\
&- \frac{\sigma_{i_1,t}^2/\alpha_{i_1}}{\sigma_{i_1,t}\sigma_{i_2,t}^3}\Big[\alpha_{i_2}\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1}x_{i_2,t-1-h}^2 + \frac{\omega_{i_2}}{(1-\beta_{i_2})^2}\Big]^2\Big|^4 \\
\leq & b + c\sqrt{\mathbb{E}\sigma_{i_1,t}^8}\sqrt{\mathbb{E}\Big(\frac{\sum_{h=2}^{\infty} h(h-1)\beta_{i_2}^{h-2}x_{i_2,t-1-h}^2}{\sigma_{i_2,t}}\Big)^8} && (3.73) \\
&+ d\sqrt{\mathbb{E}\sigma_{i_1,t}^8}\sqrt{\mathbb{E}\Big(\frac{(\sum_{h=2}^{\infty} h(h-1)\beta_{i_2}^{h-2}x_{i_2,t-1-h}^2)^2}{\sigma_{i_2,t}^3}\Big)^8} && (3.74) \\
< & \infty
\end{aligned}
$$

$$
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{\partial\alpha_{i_1}\partial\beta_{i_1}\partial\beta_{0i_1}}\Big|^4 = \mathbb{E}\Big(\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2\Big)^3 < \infty
$$

$$
\begin{aligned}
\mathbb{E}\Big|\frac{\partial^3 H_{ij,t}}{(\partial\beta_{i_1})^3}\Big|^4 =& \mathbb{E}\Big|\frac{3(\mathbb{1}_{i=i_1}\sigma_{j,t} + \mathbb{1}_{j=i_1}\sigma_{i,t})\rho_{i,j}}{8\sigma_{i_1,t}^5}\Big[\frac{\omega_{i_1}}{(1-\beta_{i_1})^2} \\
&+ \alpha_{i_1}\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2\Big]^3 \\
&- \frac{(\mathbb{1}_{i=i_1}\sigma_{j,t} + \mathbb{1}_{j=i_1}\sigma_{i,t})\rho_{i,j}}{4\sigma_{i_1,t}^3}\Big[\alpha_{i_1}\sum_{h=2}^{\infty} h(h-1)\beta_{i_1}^{h-2}x_{i_1,t-1-h}^2 \\
&+ \frac{2\omega_{i_1}}{(1-\beta_{i_1})^3}\Big]\Big[\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2 + \frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\Big] \\
&- \frac{(\mathbb{1}_{i=i_1}\sigma_{j,t} + \mathbb{1}_{j=i_1}\sigma_{i,t})\rho_{i,j}}{2\sigma_{i_1,t}^3}\Big[\alpha_{i_1}\sum_{h=2}^{\infty} h(h-1)\beta_{i_1}^{h-2}x_{i_1,t-1-h}^2 \\
&+ \frac{2\omega_{i_1}}{(1-\beta_{i_1})^3}\Big] + \Big[\frac{(\mathbb{1}_{i=i_1}\sigma_{j,t} + \mathbb{1}_{j=i_1}\sigma_{i,t})\rho_{i,j}}{2\sigma_{i_1,t}} + \beta_{0i_1}(\mathbb{1}_{i=i_1} + \mathbb{1}_{j=i_1})\Big] \\
&\Big[\alpha_{i_1}\sum_{h=3}^{\infty} h(h-1)(h-2)\beta_{i_1}^{h-3}x_{i_1,t-1-h}^2 + \frac{6\omega_{i_1}}{(1-\beta_{i_1})^4}\Big]\Big| \\
\leq & a\mathbb{E}(\sigma_{j,t}^4 + \sigma_{i,t}^4) + \sqrt{\mathbb{E}(\sigma_{i,t}^8 + \sigma_{j,t}^8)}\Big[b_1\sqrt{\mathbb{E}\Big(\frac{(\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2)^3}{\sigma_{i_1,t}^5}\Big)^8} \\
&+ b_2\sqrt{\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2}{\sigma_{i_1,t}}\Big)^8} + b_3 \\
&+ b_4\sqrt{\mathbb{E}\Big(\frac{(\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1}x_{i_1,t-1-h}^2)(\sum_{h=2}^{\infty} h(h-1)\beta_{i_1}^{h-2}x_{i_1,t-1-h}^2)}{\sigma_{i_1,t}^3}\Big)^8}
\end{aligned}
$$

$$
\begin{aligned}
&+ b_5 \sqrt{\mathbb{E}\Big(\frac{\sum_{h=2}^{\infty} h(h-1)\beta_{i_1}^{h-2} x_{i_1,t-1-h}^2}{\sigma_{i_1,t}}\Big)^8} \\
&+ b_6 \sqrt{\mathbb{E}\Big(\frac{\sum_{h=3}^{\infty} h(h-1)(h-2)\beta_{i_1}^{h-3} x_{i_1,t-1-h}^2}{\sigma_{i_1,t}}\Big)^8}\Bigg]\Bigg| < \infty
\end{aligned}
\tag{3.75}
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{(\partial\beta_{i_1})^2\partial\beta_{i_2}}\right|^4 =&\mathbb{E}\Bigg| - \frac{(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2} + \mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2})\rho_{i,j}}{8\sigma_{i_1,t}^3\sigma_{i_2,t}}\Big[\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1} x_{i_1,t-1-h}^2 \\
&+ \frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\Big]^2\Big[\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1} x_{i_2,t-1-h}^2 + \frac{\omega_{i_2}}{(1-\beta_{i_2})^2}\Big] \\
&+ \frac{(\mathbb{1}_{i=i_1}\mathbb{1}_{j=i_2} + \mathbb{1}_{j=i_1}\mathbb{1}_{i=i_2})\rho_{i,j}}{4\sigma_{i_1,t}\sigma_{i_2,t}}\Big[\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1} x_{i_1,t-1-h}^2 \\
&+ \frac{\omega_{i_1}}{(1-\beta_{i_1})^2}\Big]\Big[\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1} x_{i_2,t-1-h}^2 + \frac{\omega_{i_2}}{(1-\beta_{i_2})^2}\Big]\Bigg|^4 \\
\le& a + b_1 \sqrt{\mathbb{E}\Big(\frac{(\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1} x_{i_1,t-1-h}^2)^2}{\sigma_{i_1,t}^3}\Big)^8}\sqrt{\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1} x_{i_2,t-1-h}^2}{\sigma_{i_2,t}}\Big)^8} \\
&+ b_2\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1} x_{i_1,t-1-h}^2}{\sigma_{i_1,t}}\Big)^4 + b_3\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1} x_{i_2,t-1-h}^2}{\sigma_{i_2,t}}\Big)^4 + b_4 \\
&+ b_5 \sqrt{\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty} h\beta_{i_1}^{h-1} x_{i_1,t-1-h}^2}{\sigma_{i_1,t}}\Big)^8}\sqrt{\mathbb{E}\Big(\frac{\sum_{h=1}^{\infty} h\beta_{i_2}^{h-1} x_{i_2,t-1-h}^2}{\sigma_{i_2,t}}\Big)^8} \\
<&\infty
\end{aligned}
\tag{3.76}
$$

$$
\mathbb{E}\left|\frac{\partial^3 H_{ij,t}}{(\partial\beta_{i_1})^2\partial\beta_{0i_1}}\right|^4 =\mathbb{E}\left|\frac{2\omega_{i_1}}{(1-\beta_{i_1})^3} + \alpha_{i_1}\sum_{h=2}^{\infty} h(h-1)\beta_{i_1}^{h-2} x_{i_1,t-1-h}^2\right|^3 < \infty
\tag{3.77}
$$

As of now, we have proofed that all the elements of $\mathbb{E}\left\|_{ijl}\dddot{H}_t\right\|^4$ are finite.

(v) Similar to the proof of (iii) above, only the non-zero terms are listed below because the number of all second derivative elements is large. The only difference is that these elements have an initial value $(\tilde{\boldsymbol{x}}_0, \tilde{\boldsymbol{\sigma}}_0)$ while the ones in (iii) have infinite past.

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^2 \tilde{H}_{ii,t}}{\partial\omega_j\partial\beta_{0j}}\right|^{z_3} =&\Big(\frac{1-\beta_j^t}{1-\beta_j}\Big)^{z_3} < \infty \\
\mathbb{E}\left|\frac{\partial^2 \tilde{H}_{ii,t}}{\partial\omega_j\partial\beta_j}\right|^{z_3} =&(\mathbb{1}_{i=j} + \beta_{0j})^{z_3}\Big(\frac{t\beta_j^{t-1}}{1-\beta_j}\Big)^{2z_3}
\end{aligned}
$$

$$\mathbb{E}\left|\frac{\partial^2 \tilde{H}_{ii,t}}{\partial \alpha_j \partial \beta_{0j}}\right|^{z_3} = \mathbb{E}(\sum_{l=0}^{t-2} \beta_j^l x_{j,t-1-l}^2 + \beta_j^{t-1} \tilde{x}_{j,0})^{z_3}$$

$$\leq \sum_{l=0}^{t-2} \mathbb{E}\beta_j^{lz_3} x_{j,t-1-l}^{2z_3} + \beta_j^{(t-1)z_3} \tilde{x}_{j,0}^{2z_3} \tag{3.78}$$

$$\mathbb{E}\left|\frac{\partial^2 \tilde{H}_{ii,t}}{\partial \alpha_j \partial \beta_j}\right|^{z_3} = (\mathbb{1}_{i=j} + \beta_{0j})^{z_3} \mathbb{E}(\sum_{l=1}^{t-2} l\beta_j^{l-1} x_{j,t-1-l}^2 + (t-1)\beta_j^{t-2} \tilde{x}_{j,0}^2)^{z_3}$$

$$\leq a + (t-1)^{z_3} \beta_j^{(t-2)z_3} \tilde{x}_{j,0}^{z_3}) \tag{3.79}$$

$$\mathbb{E}\left|\frac{\partial^2 \tilde{H}_{ii,t}}{\partial \beta_j \partial \beta_{0j}}\right|^{z_3} = \mathbb{E}\left(t\beta_j^{t-1} \tilde{\sigma}_{j,0}^2 + \sum_{l=1}^{t-1} l\beta_j^{l-1} \omega_j + \sum_{l=1}^{t-2} l\beta_j^{l-1} \alpha_j x_{j,t-1-l}^2 + (t-1)\alpha_j \beta_j^{t-2} \tilde{x}_{j,0}^2\right)^{z_3}$$

$$\leq a + t^{z_3} \beta_j^{(t-1)z_3} \tilde{\sigma}_{j,0}^{2z_3} + (t-1)^{z_3} \alpha_j^{z_3} \beta_j^{(t-2)z_3} \tilde{x}_{j,0}^{2z_3} \tag{3.80}$$

$$\mathbb{E}\left|\frac{\partial \tilde{H}_{ii,t}}{\partial \beta_j \partial \beta_j}\right|^{z_3} = (\mathbb{1}_{i=j} + \beta_{0j})^{z_3} \mathbb{E}\Bigg[t(t-1)\beta_j^{t-2} \tilde{\sigma}_{j,0}^2 + \sum_{l=2}^{t-1} l(l-1)\beta_j^{l-2} \omega_j$$

$$+ \sum_{l=2}^{t-2} l(l-1)\alpha_j \beta_j^{l-2} x_{j,t-1-l}^2 + (t-1)(t-2)\alpha_j \beta_j^{t-3} \tilde{x}_{j,0}^2\Bigg]^{z_3}$$

$$\leq a + \left[t(t-1)\beta_j^{t-2} \tilde{\sigma}_{j,0}^2\right]^{z_3} + b\left[(t-1)(t-2)\alpha_j \beta_j^{t-3} \tilde{x}_{j,0}^2\right]^{z_3} \tag{3.81}$$

Since $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{\sigma}}_0$ are fixed, the terms in (3.78) - (3.81) are functions of t, all have maximum on $t \in \mathbb{N}^+$. Then, above proved the finiteness of the absolute diagonal terms of the second derivative.

The non-diagonal elements will be shown below.

$$\mathbb{E}\left|\frac{\partial^2 \tilde{H}_{ij,t}}{\partial \rho_{l,k} \partial \omega_{i_1}}\right|^{z_3} = \mathbb{E}\left[\mathbb{1}_{i=l}\mathbb{1}_{j=k}\left(\frac{\mathbb{1}_{i=i_1}\tilde{\sigma}_{j,t}(1-\beta_i^t)}{2\tilde{\sigma}_{i,t}(1-\beta_i)} + \frac{\mathbb{1}_{j=i_1}\tilde{\sigma}_{i,t}(1-\beta_j^t)}{2\tilde{\sigma}_{j,t}(1-\beta_j)}\right)\right]^{z_3}$$

$$\leq a\mathbb{E}(\tilde{\sigma}_{i,t})^{z_3} + b\mathbb{E}(\tilde{\sigma}_{j,t})^{z_3} < \infty$$

$$\mathbb{E}\left|\frac{\partial^2 \tilde{H}_{ij,t}}{\partial \rho_{l,k} \partial \alpha_{i_1}}\right|^{z_3} = \mathbb{E}\left[\mathbb{1}_{i=l}\mathbb{1}_{j=k}\left(\frac{\mathbb{1}_{i=i_1}\tilde{\sigma}_{j,t}}{2\tilde{\sigma}_{i,t}} + \frac{\mathbb{1}_{j=i_1}\tilde{\sigma}_{i,t}}{2\tilde{\sigma}_{j,t}}\right)\frac{\partial \tilde{\sigma}_{i_1,t}^2}{\partial \alpha_{i_1}}\right]^{z_3}$$

$$\leq a\left[\mathbb{E}\tilde{\sigma}_{i,t}^{2z_3}\mathbb{E}\left(\frac{\partial \tilde{\sigma}_{j,t}^2}{\partial \alpha_j}\right)^{2z_3}\right]^{1/2} + b\left[\mathbb{E}\tilde{\sigma}_{j,t}^{2z_3}\mathbb{E}\left(\frac{\partial \tilde{\sigma}_{i,t}^2}{\partial \alpha_i}\right)^{2z_3}\right]^{1/2} < \infty$$

$$\mathbb{E}\left|\frac{\partial^2 \tilde{H}_{ij,t}}{\partial \rho_{l,k} \partial \beta_{i_1}}\right|^{z_3} = \mathbb{E}\left[\mathbb{1}_{i=l}\mathbb{1}_{j=k}\left(\frac{\mathbb{1}_{i=i_1}\tilde{\sigma}_{j,t}}{2\tilde{\sigma}_{i,t}} + \frac{\mathbb{1}_{j=i_1}\tilde{\sigma}_{i,t}}{2\tilde{\sigma}_{j,t}}\right)\frac{\partial \tilde{\sigma}_{i_1,t}^2}{\partial \beta_{i_1}}\right]^{z_3}$$

$$
\begin{aligned}
&\leq a\left[\mathbb{E}\tilde{\sigma}_{i,t}^{2z_3}\mathbb{E}\left(\frac{\partial\tilde{\sigma}_{j,t}^2}{\partial\beta_j}\right)^{2z_3}\right]^{1/2}+b\left[\mathbb{E}\tilde{\sigma}_{j,t}^{2z_3}\mathbb{E}\left(\frac{\partial\tilde{\sigma}_{i,t}^2}{\partial\beta_i}\right)^{2z_3}\right]^{1/2}<\infty\\
\mathbb{E}\left|\frac{\partial^2\tilde{H}_{ij,t}}{\partial\omega_l\partial\omega_k}\right|^{z_3}&=\mathbb{E}\left|\frac{\mathbb{1}_{i=l}\mathbb{1}_{j=k}\rho_{i,j}}{4\tilde{\sigma}_{l,t}\tilde{\sigma}_{k,t}}\frac{\partial\tilde{\sigma}_{l,t}^2}{\partial\omega_l}\frac{\partial\tilde{\sigma}_{k,t}^2}{\partial\omega_k}\mathbb{1}_{l\neq k}-\mathbb{1}_{l=k}\rho_{i,j}\left(\frac{\mathbb{1}_{i=l}\tilde{\sigma}_{j,t}}{4\tilde{\sigma}_{l,t}^3}+\frac{\mathbb{1}_{j=l}\tilde{\sigma}_{i,t}}{4\tilde{\sigma}_{l,t}^3}\right)\left(\frac{\partial\tilde{\sigma}_{l,t}^2}{\partial\omega_l}\right)^2\right|^{z_3}\\
&\leq a\left[\mathbb{E}\left(\frac{\partial\tilde{\sigma}_{l,t}^2}{\partial\omega_l}\right)^{2z_3}\mathbb{E}\left(\frac{\partial\tilde{\sigma}_{k,t}^2}{\partial\omega_k}\right)^{2z_3}\right]^{1/2}+b\left[\mathbb{E}\tilde{\sigma}_{i,t}^{2z_3}\mathbb{E}\left(\frac{\partial\tilde{\sigma}_{j,t}^2}{\partial\omega_j}\right)^{4z_3}\right]^{1/2}\\
&\quad+c\left[\mathbb{E}\tilde{\sigma}_{j,t}^{2z_3}\mathbb{E}\left(\frac{\partial\tilde{\sigma}_{i,t}^2}{\partial\omega_i}\right)^{4z_3}\right]^{1/2}<\infty\\
\mathbb{E}\left|\frac{\partial^2\tilde{H}_{ij,t}}{\partial\omega_l\partial\alpha_k}\right|^{z_3}&=\mathbb{E}\left|\frac{\mathbb{1}_{i=l}\mathbb{1}_{j=k}\rho_{i,j}\mathbb{1}_{l\neq k}}{4\tilde{\sigma}_{i_1,t}\tilde{\sigma}_{i_2,t}}\frac{\partial\tilde{\sigma}_{i_1,t}^2}{\partial\omega_{i_1}}\frac{\partial\tilde{\sigma}_{i_2,t}^2}{\partial\alpha_{i_2}}\right.\\
&\quad\left.+\frac{-(\mathbb{1}_{i=l}\tilde{\sigma}_{i,t}+\mathbb{1}_{j=l}\tilde{\sigma}_{j,t})\rho_{i,j}\mathbb{1}_{l=k}}{4\tilde{\sigma}_{l,t}^3}\frac{\partial\tilde{\sigma}_{l,t}^2}{\partial\omega_l}\frac{\partial\tilde{\sigma}_{k,t}^2}{\partial\alpha_k}\right|^{z_3}\\
&\leq a\left[\mathbb{E}\left|\frac{\partial\tilde{\sigma}_{l,t}^2}{\partial\omega_l}\right|^{2z_3}\mathbb{E}\left|\frac{\partial\tilde{\sigma}_{k,t}^2}{\partial\alpha_k}\right|^{2z_3}\right]^{1/2}+b\left[\mathbb{E}\tilde{\sigma}_{j,t}^{3z_3}\mathbb{E}\left|\frac{\partial\tilde{\sigma}_{i,t}^2}{\partial\omega_i}\right|^{3z_3}\mathbb{E}\left|\frac{\partial\tilde{\sigma}_{i,t}^2}{\partial\alpha_i}\right|^{3z_3}\right]^{1/3}\\
&\quad+c\left[\mathbb{E}\tilde{\sigma}_{i,t}^{3z_3}\mathbb{E}\left|\frac{\partial\tilde{\sigma}_{j,t}^2}{\partial\omega_j}\right|^{3z_3}\mathbb{E}\left|\frac{\partial\tilde{\sigma}_{j,t}^2}{\partial\alpha_j}\right|^{3z_3}\right]^{1/3}<\infty\\
\mathbb{E}\left|\frac{\partial^2\tilde{H}_{ij,t}}{\partial\omega_l\partial\beta_k}\right|^{z_3}&=\mathbb{E}\left|\frac{(\mathbb{1}_{i=l}\mathbb{1}_{j=k}+\mathbb{1}_{i=k}\mathbb{1}_{j=l})\rho_{i,j}\mathbb{1}_{k\neq l}}{4\tilde{\sigma}_{l,t}\tilde{\sigma}_{k,t}}\frac{\partial\tilde{\sigma}_{l,t}^2}{\partial\omega_l}\frac{\partial\tilde{\sigma}_{k,t}^2}{\partial\beta_k}\right.\\
&\quad\left.-\frac{(\mathbb{1}_{i=l}\tilde{\sigma}_{j,t}+\mathbb{1}_{j=k}\tilde{\sigma}_{i,t})\mathbb{1}_{k=l}\rho_{i,j}}{4\tilde{\sigma}_{l,t}^3}\frac{\partial\tilde{\sigma}_{l,t}^2}{\partial\omega_l}\frac{\partial\tilde{\sigma}_{l,t}^2}{\partial\beta_l}\right|^{z_3}\\
&\leq a\left[\mathbb{E}\left|\frac{\partial\tilde{\sigma}_{l,t}^2}{\partial\omega_l}\right|^{2z_3}\mathbb{E}\left|\frac{\partial\tilde{\sigma}_{k,t}^2}{\partial\beta_k}\right|^{2z_3}\right]^{1/2}+b\left[\mathbb{E}\tilde{\sigma}_{j,t}^{3z_3}\mathbb{E}\left|\frac{\partial\tilde{\sigma}_{i,t}^2}{\partial\omega_i}\right|^{3z_3}\mathbb{E}\left|\frac{\partial\tilde{\sigma}_{i,t}^2}{\partial\beta_i}\right|^{3z_3}\right]^{1/3}\\
&\quad+c\left[\mathbb{E}\tilde{\sigma}_{i,t}^{3z_3}\mathbb{E}\left|\frac{\partial\tilde{\sigma}_{j,t}^2}{\partial\omega_j}\right|^{3z_3}\mathbb{E}\left|\frac{\partial\tilde{\sigma}_{j,t}^2}{\partial\beta_j}\right|^{3z_3}\right]^{1/3}<\infty\\
\mathbb{E}\left|\frac{\partial^2\tilde{H}_{ij,t}}{\partial\alpha_l\partial\alpha_k}\right|^{z_3}&=\mathbb{E}\left[\left|-\frac{(\mathbb{1}_{i=l}\tilde{\sigma}_{j,t}+\mathbb{1}_{j=l}\tilde{\sigma}_{i,t})}{4\tilde{\sigma}_{l,t}^3}\mathbb{1}_{k=l}\rho_{i,j}\right|\left(\frac{\partial\tilde{\sigma}_{l,t}^2}{\partial\alpha_l}\right)^2\right.\\
&\quad\left.+(\mathbb{1}_{i=l}\mathbb{1}_{j=k})\frac{\rho_{i,j}}{4\tilde{\sigma}_{l,t}\tilde{\sigma}_{k,t}}\frac{\partial\tilde{\sigma}_{l,t}^2}{\partial\alpha_l}\frac{\partial\tilde{\sigma}_{k,t}^2}{\partial\alpha_k}\right]^{z_3}\\
&\leq a\left[\mathbb{E}\tilde{\sigma}_{i,t}^{2z_3}\mathbb{E}\left(\frac{\partial\tilde{\sigma}_{j,t}^2}{\partial\alpha_j}\right)^{4z_3}\right]^{1/2}+b\left[\mathbb{E}\tilde{\sigma}_{j,t}^{2z_3}\mathbb{E}\left(\frac{\partial\tilde{\sigma}_{i,t}^2}{\partial\alpha_i}\right)^{4z_3}\right]^{1/2}\\
&\quad+c\left[\mathbb{E}\left(\frac{\partial\tilde{\sigma}_{j,t}^2}{\partial\alpha_j}\right)^{2z_3}\mathbb{E}\left(\frac{\partial\tilde{\sigma}_{k,t}^2}{\partial\alpha_k}\right)^{2z_3}\right]^{1/2}<\infty\\
\mathbb{E}\left|\frac{\partial\tilde{H}_{ij,t}}{\partial\alpha_l\partial\beta_k}\right|^{z_3}&=\rho_{i,j}^{z_3}\mathbb{E}\left|\left[\left(\sum_{k=1}^{t-2}k\beta_j^k x_{l,t-1-k}^2+(t-1)\beta_j^{t-2}\tilde{x}_{j,0}^2\right)\left(\frac{\mathbb{1}_{i=l}\tilde{\sigma}_{j,t}}{2\tilde{\sigma}_{l,t}}+\frac{\mathbb{1}_{j=l}\tilde{\sigma}_{i,t}}{2\tilde{\sigma}_{l,t}}\right)\right.\right.
\end{aligned}
$$

$$
\begin{aligned}
&- \Big(\frac{\mathbb{1}_{i=l}\tilde{\sigma}_{j,t}}{4\tilde{\sigma}^3_{l,t}} + \frac{\mathbb{1}_{j=l}\tilde{\sigma}_{i,t}}{4\tilde{\sigma}^3_{l,t}}\Big)\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\alpha_l}\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\beta_l}\Bigg]\mathbb{1}_{l=k} + \frac{(\mathbb{1}_{i=l}+\mathbb{1}_{j=l})\mathbb{1}_{k\neq l}}{4\tilde{\sigma}_{l,t}\tilde{\sigma}_{k,t}}\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\alpha_l}\frac{\partial\tilde{\sigma}^2_{k,t}}{\partial\beta_k}\Bigg|^{z_3} \\
&\leq a\left[\mathbb{E}\left(\sum_{k=1}^{\infty} k\beta_j^k x^2_{l,t-1-k}\right)^{2z_3}(\mathbb{E}\tilde{\sigma}^{2z_3}_{i,t} + \mathbb{E}\tilde{\sigma}^{2z_3}_{j,t})\right]^{1/2} \\
&\quad + b\left[(t-1)\beta_j^{t-2}\tilde{x}^2_{j,0}\right]^{z_3}(\mathbb{E}\tilde{\sigma}^{z_3}_{i,t} + \mathbb{E}\tilde{\sigma}^{z_3}_{j,t}) \\
&\quad + c\left[(\mathbb{E}\tilde{\sigma}^{3z_3}_{i,t} + \mathbb{E}\tilde{\sigma}^{3z_3}_{j,t})\mathbb{E}\left(\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\alpha_l}\right)^{3z_3}\mathbb{E}\left(\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\beta_l}\right)^{3z_3}\right]^{1/3} \\
&\quad + d\left[\mathbb{E}\left(\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\alpha_l}\right)^{2z_3}\mathbb{E}\left(\frac{\partial\tilde{\sigma}^2_{k,t}}{\partial\beta_k}\right)^{2z_3}\right]^{1/2} < \infty
\end{aligned}
$$

$$
\mathbb{E}\left|\frac{\partial\tilde{H}_{ij,t}}{\partial\alpha_l\partial\beta_{0k}}\right|^{z_3} = \mathbb{E}\left|(\mathbb{1}_{i=l}+\mathbb{1}_{j=l})\mathbb{1}_{l=k}\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\alpha_l}\right|^{z_3} < \infty
$$

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial\tilde{H}_{ij,t}}{\partial\beta_l\partial\beta_k}\right|^{z_3} &= \mathbb{E}\Bigg|\frac{\rho_{i,j}\mathbb{1}_{i=l}\mathbb{1}_{j=k}}{4\tilde{\sigma}_{l,t}\tilde{\sigma}_{k,t}}\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\beta_l}\frac{\partial\tilde{\sigma}^2_{k,t}}{\partial\beta_k} - \Bigg\{\rho_{i,j}\frac{\mathbb{1}_{i=l}\tilde{\sigma}_{j,t}+\mathbb{1}_{j=l}\tilde{\sigma}_{i,t}}{4\tilde{\sigma}^3_{l,t}}\left(\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\beta_l}\right)^2 \\
&\quad + \Bigg[t(t-1)\tilde{\sigma}^2_{l,0} + \sum_{h=2}^{t-2} h(h-1)(\alpha_l\beta_l^{h-2}x^2_{l,t-1-h}+\omega_l) \\
&\quad + (t-1)(t-2)\beta_l^{t-3}(\omega_l+\alpha_l\tilde{x}^2_{l,0})\Bigg]\Bigg[\frac{\rho_{i,j}(\mathbb{1}_{i=l}\tilde{\sigma}_{j,t}+\mathbb{1}_{j=l}\tilde{\sigma}_{i,t})}{2\tilde{\sigma}_{l,t}} + \beta_{0l}\Bigg]\Bigg\}\mathbb{1}_{k=l}\Bigg|^{z_3} \\
&\leq a_1\left[\mathbb{E}\left(\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\beta_l}\right)^{2z_3}\mathbb{E}\left(\frac{\partial\tilde{\sigma}^2_{k,t}}{\partial\beta_k}\right)^{2z_3}\right]^{1/2} + a_2\left[\mathbb{E}\left(\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\beta_l}\right)^{4z_3}(\mathbb{E}\tilde{\sigma}^{2z_3}_{i,t}+\mathbb{E}\tilde{\sigma}^{2z_3}_{j,t})\right]^{1/2} \\
&\quad + a_3\left[t(t-1)\tilde{\sigma}^2_{l,0}\right]^{z_3}(\mathbb{E}\tilde{\sigma}^{z_3}_{i,t}+\mathbb{E}\tilde{\sigma}^{z_3}_{j,t}+a_4) \\
&\quad + a_5\mathbb{E}\left[(\mathbb{1}_{i=l}\tilde{\sigma}_{j,t}+\mathbb{1}_{j=l}\tilde{\sigma}_{i,t}+a_6)\right\}\sum_{h=2}^{\infty} h(h-1)(\alpha_l\beta_l^{h-2}x^2_{l,t-1-h}+\omega_l)\Bigg]^{z_3} \\
&\quad + a_7\left[(t-1)(t-2)\beta_l^{t-3}\alpha_l\tilde{x}^2_{l,0}\right]^{z_3}(\mathbb{E}\tilde{\sigma}^{z_3}_{i,t}+\mathbb{E}\tilde{\sigma}^{z_3}_{j,t}) \\
&\leq b_1\left[t(t-1)\right]^{z_3} + b_2\left\{(\mathbb{E}\tilde{\sigma}^{2z_3}_{j,t}+\mathbb{E}\tilde{\sigma}^{2z_3}_{i,t})\mathbb{E}\left[\sum_{h=2}^{\infty} h(h-1)\beta_l^{h-2}x^2_{l,t-1-h}\right]^{2z_3}\right\}^{1/2} \\
&\quad + b_3 + b_4\left[(t-1)(t-2)\beta_l^{t-3}\right]^{z_3} < \infty
\end{aligned}
$$

$$
\mathbb{E}\left|\frac{\partial\tilde{H}_{ij,t}}{\partial\beta_l\partial\beta_{0k}}\right|^{z_3} = \mathbb{E}\left|(\mathbb{1}_{i=l}\mathbb{1}_{l=k}+\mathbb{1}_{j=l}\mathbb{1}_{l=k})\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\beta_l}\right|^{z_3} < \infty
$$

All the absolute third moments of the terms have been proved to be finite. The elements which are not listed above have the value 0. The desired result, $\mathbb{E}\left\|{}_{ij}\ddot{H}_t\right\|^{z_3} <$

∞, follows.

□

Lemma 3.3.4 *Under Assumptions $C_1 - C_2$,*

(i) $\sum_{t=1}^{\infty} \mathbb{E}\left|\dfrac{\partial(\sigma_{i,t}^2 - \tilde{\sigma}_{i,t}^2)}{\partial \boldsymbol{\theta}_j}\right|^4 < \infty$ *where θ_j means the jth parameter.*

(ii) $\mathbb{E}|\sigma_{i,t}\sigma_{j,t} - \tilde{\sigma}_{i,t}\tilde{\sigma}_{j,t}|^2 \leq O(\beta_i^{2t}) + O(\beta_j^{2t})$ *and* $\mathbb{E}|\sigma_{i,t}\tilde{\sigma}_{j,t} - \tilde{\sigma}_{i,t}\sigma_{j,t}|^2 \leq O(\beta_i^{2t}) + O(\beta_j^{2t})$. *Therefore,* $\mathbb{E}\left|\dfrac{\partial H_{ij,t}}{\partial \sigma_{i,t}^2} - \dfrac{\partial \tilde{H}_{ij,t}}{\partial \tilde{\sigma}_{i,t}^2}\right|^2 = \mathbb{E}\left|\dfrac{\rho_{ij}\sigma_{j,t}}{\sigma_{i,t}} - \dfrac{\rho_{ij}\tilde{\sigma}_{j,t}}{\tilde{\sigma}_{i,t}}\right| \leq O(\beta_i^{2t}) + O(\beta_j^{2t})$.

(iii) $\mathbb{E}|\sigma_{i,t}^3 - \tilde{\sigma}_{i,t}^3|^2 \leq O(\beta_i^{t2})$.

(iv) $\mathbb{E}|\frac{\partial^2 H_{i_1 j_1,t}}{(\partial \sigma_{i_1,t}^2)^2} - \frac{\partial^2 \tilde{H}_{i_1 j_1,t}}{(\partial \tilde{\sigma}_{i_1,t}^2)^2}|^{z_7} \leq O(\beta_{i_1}^{tz_7}) + O(\beta_{j_1}^{tz_7})$ *for any* $0 \leq z_7 \leq 1$.

(v) $\mathbb{E}|\frac{\partial^2 H_{i_1 j_1,t}}{\partial \sigma_{i_1,t}^2 \partial \sigma_{j_1,t}^2} - \frac{\partial^2 \tilde{H}_{i_1 j_1,t}}{\partial \tilde{\sigma}_{i_1,t}^2 \partial \tilde{\sigma}_{j_1,t}^2}|^2 \leq O(\beta_{i_1}^{t2}) + O(\beta_{j_1}^{t2})$

(vi) $\sum_{t=1}^{n} \mathbb{E}\left\|{}_i\dot{\tilde{H}}_t - {}_i\dot{H}_t\right\|^{z_9} < \infty$ *for any* $0 < z_9 < 1$, $\sum_{t=1}^{n}\left(\mathbb{E}\left\|{}_i\dot{\tilde{H}}_t - {}_i\dot{H}_t\right\|^{2z_{10}}\right)^{1/2} < \infty$ *for any* $0 < z_{10} < 1/2$, *and* $\sum_{t=1}^{n}\left(\mathbb{E}\left\|{}_i\dot{\tilde{H}}_t - {}_i\dot{H}_t\right\|^{3z_{11}}\right)^{1/3} < \infty$ *for any* $0 \leq z_{11} < 1/3$ *and* $\sum_{t=1}^{n}\left(\mathbb{E}\left\|{}_i\dot{\tilde{H}}_t - {}_i\dot{H}_t\right\|^{4z_{12}}\right)^{1/4} < \infty$ *for any* $0 < z_{12} < 1/4$.

(vii) $\sum_{t=1}^{n} \mathbb{E}\left\|{}_{ij}\ddot{\tilde{H}}_t - {}_{ij}\ddot{H}_t\right\|^{z_{13}} < \infty$ *for any* $0 \leq z_{13} < 1/2$, $\sum_{t=1}^{n}\left(\mathbb{E}\left\|{}_{ij}\ddot{\tilde{H}}_t - {}_{ij}\ddot{H}_t\right\|^{2z_{14}}\right)^{1/2} < \infty$ *for any* $0 \leq z_{14} < 1/4$, *and* $\sum_{t=1}^{n}\left(\mathbb{E}\left\|{}_{ij}\ddot{\tilde{H}}_t - {}_{ij}\ddot{H}_t\right\|^{4z_{15}}\right)^{1/4} < \infty$ *for any* $0 \leq z_{15} < 1/8$.

(viii) $\sum_{t=1}^{\infty} \mathbb{E}\left|\sigma_{i,t}^2 - \tilde{\sigma}_{i,t}^2\right|^4 < \infty$. *Then,* $\sum_{t=1}^{\infty} \mathbb{E}\left\|H_t - \tilde{H}_t\right\|^4 < \infty$, $\sum_{t=1}^{\infty}(\mathbb{E}\left\|H_t - \tilde{H}_t\right\|^{2v_{16}})^{1/2} < \infty$ *for any* $0 < v_{16} < 1/2$ *and* $\sum_{t=1}^{\infty}(\mathbb{E}\left\|H_t - \tilde{H}_t\right\|^{4v_{17}})^{1/4} < \infty$ *for any* $0 < v_{17} < 1$.

Proof After dropping the zero terms, the non-zero terms can be proved by the following inequalities. Still, the lower case letters $a, b, \ldots$ have been used to represent some finite constant, they may have distinct values in different lines.

(i) Since the term $\sigma_{i,t}$ only relates to $\omega_i, \alpha_i, \beta_i$ based on the expression. The derivative $\dfrac{\partial(\sigma_{i,t}^2 - \tilde{\sigma}_{i,t}^2)}{\partial \boldsymbol{\theta}_j}$ equals to 0 if θ_j is not one of ω_i, α_i or $,\beta_i$.

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial(\sigma_{i,t}^2 - \tilde{\sigma}_{i,t}^2)}{\partial\omega_i}\right|^4 =& \left|\frac{\beta_i^t\beta_{0i}}{1-\beta_i}\right|^4 = O(\beta_i^{4t}) \\
\mathbb{E}\left|\frac{\partial(\sigma_{i,t}^2 - \tilde{\sigma}_{i,t}^2)}{\partial\alpha_i}\right|^4 \leq& \beta_i^{t4}\frac{1}{(1-\beta_i)^4}\mathbb{E}(x_{i,t}^8) + \beta_i^{(t-1)4}\mathbb{E}|x_{i,0}^2 - \tilde{x}_{i,0}^2|^4 = O(\beta_i^{t4}) \\
\mathbb{E}\left|\frac{\partial(\sigma_{i,t}^2 - \tilde{\sigma}_{i,t}^2)}{\partial\beta_i}\right|^4 \leq& t^4\beta_i^{(t-1)4}\mathbb{E}|\sigma_{i,0}^2 - \tilde{\sigma}_{i,0}^2|^4 \\
&+ \beta_i^{t4}\left(\frac{\omega_i^4}{(1-\beta_i)^{24}} + \alpha_i\sum_{j=0}^{\infty}\beta_i^{j-1}x_{i,0-j}^2\right) \\
&+ \alpha_i^4(t-1)^4\beta_i^{(t-2)4}\mathbb{E}|x_{i,0}^2 - \tilde{x}_{i,0}^2|^4 = O(\beta_i^{t4}) + O(t^4\beta_i^{t4})
\end{aligned}
$$

Since each of these summations is finite, so the result is true.

(ii) We can replace $v_1/4$ by 2 in (3.30), then the first part has been proved. Following the similar steps, the second part becomes

$$
\begin{aligned}
\mathbb{E}|\sigma_{i,t}\tilde{\sigma}_{j,t} - \tilde{\sigma}_{i,t}\sigma_{j,t}|^2 \leq& (\mathbb{E}\sigma_{i,t}^4\mathbb{E}|\sigma_{j,t} - \tilde{\sigma}_{j,t}|^4)^{1/2} + (\mathbb{E}\sigma_{j,t}^4\mathbb{E}|\sigma_{i,t} - \tilde{\sigma}_{i,t}|^4)^{1/2} \\
\leq& \left(\frac{\mathbb{E}\sigma_{i,t}^4}{(4\omega_j)^2}\mathbb{E}|\sigma_{j,t}^2 - \tilde{\sigma}_{j,t}^2|^4\right)^{1/2} + \left(\frac{\mathbb{E}\sigma_{j,t}^4}{(4\omega_i)^2}\mathbb{E}|\sigma_{i,t}^2 - \tilde{\sigma}_{i,t}^2|^4\right)^{1/2} \\
\leq& O(\beta_j^{2t}) + O(\beta_i^{2t}).
\end{aligned}
$$

(iii) By (3.26), (i) in Lemma 3.3.2 and Cauchy-Schwarz inequality,

$$
\begin{aligned}
\mathbb{E}|\sigma_{i,t}^3 - \tilde{\sigma}_{i,t}^3|^2 =& \mathbb{E}|\sigma_{i,t}^2\sigma_{i,t} - \tilde{\sigma}_{i,t}^2\tilde{\sigma}_{i,t}|^2 \\
\leq& (\mathbb{E}|\sigma_{i,t}^2|^4)^{1/2}\left(\frac{\mathbb{E}|\sigma_{i,t}^2 - \tilde{\sigma}_{i,t}^2|^4}{(2\sqrt{\omega_i})^4}\right)^{1/2} + (\mathbb{E}|\sigma_{i,t}|^4)^{1/2}(\mathbb{E}|\sigma_{i,t}^2 - \tilde{\sigma}_{i,t}^2|^4)^{1/2} \\
=& O(\beta_i^{t2}).
\end{aligned}
$$

(iv) $0 < 2z_7 \leq 2$, we can use (iii) in this same lemma to prove this result. Hence,

$$
\begin{aligned}
&\mathbb{E}\left|\frac{\partial^2 H_{i_1j_1,t}}{(\partial\sigma^2_{i_1,t})^2} - \frac{\partial^2 \tilde{H}_{i_1j_1,t}}{(\partial\tilde{\sigma}^2_{i_1,t})^2}\right|^{z_7}\\
=&\mathbb{E}\left|\frac{\rho_{i_1j_1}\sigma_{j_1,t}}{4\sigma^3_{i_1,t}} - \frac{\rho_{i_1j_1}\tilde{\sigma}_{j_1,t}}{4\tilde{\sigma}^3_{i_1,t}}\right|^{z_7}\\
=&\mathbb{E}\left|\rho_{i_1j_1}\frac{\sigma_{j_1,t}\tilde{\sigma}^3_{i_1,t} - \sigma^3_{i_1,t}\tilde{\sigma}_{j_1,t}}{4\sigma^3_{i_1,t}\tilde{\sigma}^3_{i_1,t}}\right|^{z_7}\\
\leq&\frac{\rho^{z_7}_{i_1j_1}}{4^{z_7}\omega^{3z_7}_{i_1}}\mathbb{E}\left|\frac{\sigma_{j_1,t}\tilde{\sigma}^3_{i_1,t} - \sigma^3_{i_1,t}\tilde{\sigma}_{j_1,t}}{4\sigma^3_{i_1,t}\tilde{\sigma}^3_{i_1,t}}\right|^{z_7}\\
\leq&a[(\mathbb{E}|\sigma_{j_1,t}|^{2z_7}\mathbb{E}|\tilde{\sigma}^3_{i_1,t} - \sigma^3_{i_1,t}|^{2z_7})^{1/2} + (\mathbb{E}|\tilde{\sigma}^3_{i_1,t}|^{2z_7}\mathbb{E}|\tilde{\sigma}_{j_1,t} - \sigma_{j_1,t}|^{2z_7})^{1/2}]\\
\leq&O(\beta^{tz_7}_{i_1}) + O(\beta^{tz_7}_{j_1}).
\end{aligned}
$$

(v) We can work on the derivative directly,

$$
\begin{aligned}
\mathbb{E}\left|\frac{\partial^2 H_{i_1j_1,t}}{\partial\sigma^2_{i_1,t}\partial\sigma^2_{j_1,t}} - \frac{\partial^2 \tilde{H}_{i_1j_1,t}}{\partial\tilde{\sigma}^2_{i_1,t}\partial\tilde{\sigma}^2_{j_1,t}}\right|^2 =&E\left|\frac{\rho_{i_1j_1}}{4\sigma_{j_1,t}\sigma_{i_1,t}} - \frac{\rho_{i_1j_1}}{4\tilde{\sigma}_{j_1,t}\tilde{\sigma}_{i_1,t}}\right|^2\\
\leq&a\mathbb{E}\left|\tilde{\sigma}_{j_1,t}\tilde{\sigma}_{i_1,t} - \sigma_{j_1,t}\sigma_{i_1,t}\right|^2\\
\leq&a[O(\beta^{t2}_i) + O(\beta^{t2}_j)] < \infty
\end{aligned}
$$

(vi) The proof will be based on the elements of the difference matrix as well. If $\boldsymbol{\theta}_i \in \{\omega_k, \alpha_k, \beta_k\}$, it is easy to see the diagonal terms have finite summations,

$$
\begin{aligned}
\sum_{t=1}^{n}\mathbb{E}\left|\frac{\partial\tilde{H}_{jj,t} - H_{jj,t}}{\partial\boldsymbol{\theta}_i}\right|^{z_9} =&\mathbb{E}\left|\frac{\partial H_{jj,t}}{\partial\sigma^2_{k,t}}\frac{\partial\sigma^2_{k,t}}{\partial\boldsymbol{\theta}_i} - \frac{\partial\tilde{H}_{jj,t}}{\partial\tilde{\sigma}^2_{k,t}}\frac{\partial\tilde{\sigma}^2_{k,t}}{\partial\boldsymbol{\theta}_i}\right|^{z_9}\\
\leq&\sum_{t=1}^{n}(\mathbb{1}_{k=j} + \beta_{0i})^{z_9}\mathbb{E}\left|\frac{\partial(\sigma^2_{k,t} - \tilde{\sigma}^2_{k,t})}{\partial\boldsymbol{\theta}_i}\right|^{z_9}\\
\leq&\sum_{t=1}^{n}O(\beta^{tz_9}_i) + O(t^{z_9}\beta^{tz_9}_i) < \infty
\end{aligned}
$$

because of (i) in this lemma.

The next step is to check the non-diagonal terms. Assume $i < j$ and $l < k$ in $\rho_{l,k}$ without loss of generality, $i, j, l, k = 1, \ldots, m$. The total summation of any term that has not been included below is 0.

$$\begin{aligned}
\sum_{t=1}^{n}\mathbb{E}\left|\frac{\partial(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\rho_{ij}}\right|^{z_9} &= \sum_{t=1}^{n}\mathbb{E}\left|\sigma_{i,t}\sigma_{j,t}-\tilde{\sigma}_{i,t}\tilde{\sigma}_{j,t}\right|^{z_9} < \infty \\
\sum_{t=1}^{n}\mathbb{E}\left|\frac{\partial(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\omega_l}\right|^{z_9} &= \sum_{t=1}^{n}\mathbb{E}\left|\left(\frac{\rho_{ij}(\mathbb{1}_{j=l}\sigma_{i,t}+\mathbb{1}_{i=l}\sigma_{j,t})}{\sigma_{l,t}}+\beta_{0l}\mathbb{1}_{j=l}\mathbb{1}_{i=l}\right)\frac{\partial\sigma^2_{l,t}}{\partial\omega_l}\right. \\
&\quad\left.-\left(\frac{\rho_{ij}(\mathbb{1}_{j=l}\tilde{\sigma}_{i,t}+\mathbb{1}_{i=l}\tilde{\sigma}_{j,t})}{\tilde{\sigma}_{l,t}}+\beta_{0l}\mathbb{1}_{j=l}\mathbb{1}_{i=l}\right)\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\omega_l}\right|^{z_9} \\
&\le \sum_{t=1}^{n} a\mathbb{1}_{i=l}\mathbb{1}_{j=l}\left(\frac{\rho_{ij}}{(1-\beta_l)}\right)^{z_9}\mathbb{E}\left|\frac{\sigma_{i,t}\tilde{\sigma}_{j,t}-\tilde{\sigma}_{i,t}\sigma_{j,t}}{\sigma_{l,t}\tilde{\sigma}_{l,t}}\right|^{z_9} \\
&\quad+\sum_{t=1}^{n}\frac{\beta_l^{tz_9}\mathbb{1}_{i=l}\mathbb{1}_{j=l}}{(1-\beta_l)^{z_9}}\mathbb{E}\left|\rho_{ij}\frac{\sigma_{i,t}\mathbb{1}_{j=l}+\sigma_{j,t}\mathbb{1}_{i=l}}{\tilde{\sigma}_{l,t}}+\beta_{0l}\right|^{z_9} \\
&\le \sum_{t=1}^{n}\left(O(\beta_i^{tz_9})+O(\beta_j^{tz_9})+O(\beta_i^{tz_9}\beta_j^{tz_9})\right) < \infty
\end{aligned}$$

$$\begin{aligned}
\sum_{t=1}^{n}\mathbb{E}\left|\frac{\partial(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\alpha_l}\right|^{z_9} &= \sum_{t=1}^{n}\mathbb{E}\left|(\rho_{ij}\frac{\sigma_{i,t}\mathbb{1}_{j=l}+\sigma_{j,t}\mathbb{1}_{i=l}}{2\sigma_{l,t}}+\beta_{0l}\mathbb{1}_{j=l}\mathbb{1}_{i=l})\frac{\partial\sigma^2_{l,t}}{\partial\alpha_l}\right. \\
&\quad\left.-\left(\frac{\rho_{ij}(\mathbb{1}_{j=l}\tilde{\sigma}_{i,t}+\mathbb{1}_{i=l}\tilde{\sigma}_{j,t})}{2\tilde{\sigma}_{l,t}}+\beta_{0l}\mathbb{1}_{j=l}\mathbb{1}_{i=l}\right)\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\alpha_l}\right|^{z_9} \\
&\le a\sum_{t=1}^{n}\mathbb{E}\left|\frac{\partial(\sigma^2_{l,t}-\tilde{\sigma}^2_{l,t})}{\partial\alpha_l}\right|^{z_9} \\
&\quad+b\sum_{t=1}^{n}\mathbb{E}\left|\rho_{ij}\frac{\sigma_{i,t}\mathbb{1}_{j=l}+\sigma_{j,t}\mathbb{1}_{i=l}}{\sigma_{l,t}}\frac{\partial\sigma^2_{l,t}}{\partial\alpha_l}\right. \\
&\quad\left.-\rho_{ij}\frac{\tilde{\sigma}_{i,t}\mathbb{1}_{j=l}+\tilde{\sigma}_{j,t}\mathbb{1}_{i=l}}{\tilde{\sigma}_{l,t}}\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\alpha_l}\right|^{z_9}
\end{aligned}$$

Use (16) in Appendix A and (ii) in this lemma since $2z_9 < 2$, the second expectation term becomes

$$\begin{aligned}
&\sum_{t=1}^{n}\mathbb{E}\left|\frac{\sigma_{i,t}\mathbb{1}_{j=l}+\sigma_{j,t}\mathbb{1}_{i=l}}{\sigma_{l,t}}\frac{\partial\sigma^2_{l,t}}{\partial\alpha_l}-\frac{\tilde{\sigma}_{i,t}\mathbb{1}_{j=l}+\tilde{\sigma}_{j,t}\mathbb{1}_{i=l}}{\tilde{\sigma}_{l,t}}\frac{\partial\tilde{\sigma}^2_{l,t}}{\partial\alpha_l}\right|^{2z_9} \\
&\le \sum_{t=1}^{n} c\sqrt{\mathbb{E}\left(\frac{\sigma_{j,t}}{\sigma_{i,t}}\right)^{2z_9}\mathbb{E}\left|\frac{\partial(\sigma^2_{i,t}-\tilde{\sigma}^2_{i,t})}{\partial\alpha_l}\right|^{2z_9}}+d\sqrt{\mathbb{E}\left(\frac{\partial\tilde{\sigma}_{i,t}}{\partial\alpha_i}\right)^{2z_9}\mathbb{E}\left|\frac{\sigma_{j,t}}{\sigma_{i,t}}-\frac{\tilde{\sigma}_{j,t}}{\tilde{\sigma}_{i,t}}\right|^{2z_9}}
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{t=1}^{n} c\sqrt{\mathbb{E}\left(\frac{\sigma_{i,t}}{\sigma_{j,t}}\right)^{2z_9}\mathbb{E}\left|\frac{\partial(\sigma_{j,t}^2-\tilde{\sigma}_{j,t}^2)}{\partial\alpha_l}\right|^{2z_9}}+d\sqrt{\mathbb{E}\left(\frac{\partial\tilde{\sigma}_{j,t}}{\partial\alpha_j}\right)^{2z_9}\mathbb{E}\left|\frac{\sigma_{i,t}}{\sigma_{j,t}}-\frac{\tilde{\sigma}_{i,t}}{\tilde{\sigma}_{j,t}}\right|^{2z_9}} \\
&\le\sum_{t=1}^{n} O(\beta_i^{tz_9})+O(\beta_j^{tz_9})+O(\beta_j^{tz_9}\beta_i^{tz_9})<\infty
\end{aligned}
$$

Then, we can get $\sum_{t=1}^{n}\mathbb{E}\left|\frac{\partial(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\alpha_l}\right|^{v_2}<\infty$

Similarly, the non-diagonal terms with respect to β_l can be proved.

$$
\begin{aligned}
\sum_{t=1}^{n}\mathbb{E}\left|\frac{\partial(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\beta_l}\right|^{z_9}=&\sum_{t=1}^{n}\mathbb{E}\Bigg|\Big(\rho_{ij}\frac{\sigma_{i,t}\mathbb{1}_{j=l}+\sigma_{j,t}\mathbb{1}_{i=l}}{2\sigma_{l,t}}+\beta_{0l}\mathbb{1}_{j=l}\mathbb{1}_{i=l}\Big)\frac{\partial\sigma_{l,t}^2}{\partial\beta_l} \\
&-\left(\frac{\rho_{ij}(\mathbb{1}_{j=l}\tilde{\sigma}_{i,t}+\mathbb{1}_{i=l}\tilde{\sigma}_{j,t})}{2\tilde{\sigma}_{l,t}}+\beta_{0l}\mathbb{1}_{j=l}\mathbb{1}_{i=l}\right)\frac{\partial\tilde{\sigma}_{l,t}^2}{\partial\beta_l}\Bigg|^{z_9} \\
\le& a\sum_{t=1}^{n}\mathbb{E}\left|\frac{\partial(\sigma_{l,t}^2-\tilde{\sigma}_{l,t}^2)}{\partial\beta_l}\right|^{z_9} \\
&+b\sum_{t=1}^{n}\mathbb{E}\Bigg|\rho_{ij}\frac{\sigma_{i,t}\mathbb{1}_{j=l}+\sigma_{j,t}\mathbb{1}_{i=l}}{\sigma_{l,t}}\frac{\partial\sigma_{l,t}^2}{\partial\beta_l} \\
&-\rho_{ij}\frac{\tilde{\sigma}_{i,t}\mathbb{1}_{j=l}+\tilde{\sigma}_{j,t}\mathbb{1}_{i=l}}{\tilde{\sigma}_{l,t}}\frac{\partial\tilde{\sigma}_{l,t}^2}{\partial\beta_l}\Bigg|^{z_9} \\
\le&\sum_{t=1}^{n} O(\beta_i^{tz_9})+O(\beta_j^{tz_9})+O(t^{z_9}\beta_l^{tz_9})<\infty
\end{aligned}
$$

After checking all the elements of the difference matrix, we can conclude that

$$
\begin{aligned}
&\sum_{t=1}^{n}\mathbb{E}\left\|{}_i\dot{\tilde{H}}_t-{}_i\dot{H}_t\right\|^{z_9} \\
=&\sum_{t=1}^{n}\mathbb{E}\left\|\frac{\partial(H_t-\tilde{H}_t)}{\partial\theta_i}\right\|^{z_9} \\
\le&\sum_{t=1}^{n}\Bigg[\sum_{\substack{j=1\\j\neq k}}^{m}\sum_{k=1}^{m} O(\beta_j^{tz_9})+O(t^{z_9}\beta_j^{tz_9})+O(\beta_k^{tz_9})+O(t^{z_9}\beta_k^{tz_9})+O(\beta_k^{tz_9}\beta_j^{tz_9})\Bigg] \\
&+\sum_{t=1}^{n}\left[O(\beta_i^{tz_9})+O(t^{z_9}\beta_i^{tz_9})+\sum_{j=1}^{m}O(\beta_j^{tz_9})+O(t^{z_9}\beta_j^{tz_9})\right] \\
\le&\sum_{t=1}^{n} O(b_1^{tz_9})+O(b_1^{2tz_9})+O(t^{z_9}b_1^{tz_9})
\end{aligned}
$$

The desired results follow, e.g.

$$\lim_{n\to\infty}\sum_{t=1}^{n}\mathbb{E}\left\|{}_i\dot{\tilde{H}}_t - {}_i\dot{H}_t\right\|^{z_9} < \infty$$

Simply apply (5) in Appendix A on the above inequalities, we can get

$$\lim_{n\to\infty}\sum_{t=1}^{n}\left(\mathbb{E}\left\|{}_i\dot{\tilde{H}}_t - {}_i\dot{H}_t\right\|^{2z_{10}}\right)^{1/2} < \infty$$

$$\lim_{n\to\infty}\sum_{t=1}^{n}\left(\mathbb{E}\left\|{}_i\dot{\tilde{H}}_t - {}_i\dot{H}_t\right\|^{3z_{11}}\right)^{1/3} < \infty$$

and

$$\lim_{n\to\infty}\sum_{t=1}^{n}\left(\mathbb{E}\left\|{}_i\dot{\tilde{H}}_t - {}_i\dot{H}_t\right\|^{4z_{12}}\right)^{1/4} < \infty$$

since $0 \le 2z_{10} < 1$, $0 \le 3z_{11} < 1$ and $0 \le 4z_{12} < 1$.

(vii) We need some preliminary inequalities before working on the non-diagonal elements. One is

$$\begin{aligned}
&\mathbb{E}\left|\frac{\partial^2 H_{i_1 j_1,t}}{\partial\rho_{i_1 j_1}\partial\sigma^2_{i_1,t}} - \frac{\partial^2 \tilde{H}_{i_1 j_1,t}}{\partial\rho_{i_1 j_1}\partial\tilde{\sigma}^2_{i_1,t}}\right|^{z_{13}} = \mathbb{E}\left|\frac{\sigma_{j_1,t}}{2\sigma_{i_1,t}} - \frac{\tilde{\sigma}_{j_1,t}}{2\tilde{\sigma}_{i_1,t}}\right|^{z_{13}} \\
=&\mathbb{E}\left|\frac{\sigma_{j_1,t}\tilde{\sigma}_{i_1,t} - \tilde{\sigma}_{j_1,t}\sigma_{i_1,t}}{2\sigma_{i_1,t}\tilde{\sigma}_{i_1,t}}\right|^{z_{13}} < O(\beta_{i_1}^{tz_{13}}) + O(\beta_{j_1}^{tz_{13}})
\end{aligned} \tag{3.82}$$

for any $0 < z_{13} < 1$ because of (ii) above.

The other set of inequalities is, for any $z \le 4$, we have

$$\begin{aligned}
\mathbb{E}\left|\frac{\partial^2\sigma^2_{i,t} - \tilde{\sigma}^2_{i,t}}{\partial\omega_i\partial\beta_i}\right|^z =&\left(\frac{t\beta_i^t}{(1-\beta_i)^2}\right)^z = O(t^z\beta_i^{tz}) \\
\mathbb{E}\left|\frac{\partial^2\sigma^2_{i,t} - \tilde{\sigma}^2_{i,t}}{\partial\alpha_i\partial\beta_i}\right|^z =&\mathbb{E}\left\{\left[\frac{t\beta_i^t}{\beta_i(1-\beta_i)} + \frac{\beta_i^t}{(1-\beta_i)^2}\right]x^2_{i,t} + (t-1)\beta_i^{t-2}(x^2_{i,0} - \tilde{x}^2_{i,0})\right\}^z \\
\le& a[(t\beta_i^t)^z + \beta_i^{tz}]\mathbb{E}x^{2z}_{i,t} + a(t-1)^z\beta_i^{(t-2)z}\mathbb{E}x^{2z}_{i,0} + a(t-1)^z\beta_i^{(t-2)z}\tilde{x}^{2z}_{i,0} \\
=&O(t^z\beta_i^{tz}) + O(\beta_i^{tz})
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left|\frac{\partial^2\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2}{\partial\beta_i\partial\beta_i}\right|^z =&\mathbb{E}\left|\frac{t\beta_i^t}{\beta_i(1-\beta_i)^2}+\frac{2\beta_i^t}{(1-\beta_i)^3}+t\beta_i^{t-1}\alpha_i\sum_{l=1}^{\infty}l\beta_i^{l-1}x_{i,t-1-l}^2+\beta_i^t\alpha_i\sum_{l=2}^{\infty}l(l-1)\beta_i^{l-2}x_{i,t-1-l}^2\right.\\
&+\frac{t\beta_i^{t-1}}{\beta_i(1-\beta_i)}+\frac{t\beta_i^{t-1}}{(1-\beta_i)^2}+t(t-1)\beta_i^{t-2}\alpha_i\sum_{l=0}^{\infty}\beta_i^l x_{i,t-1-l}^2\\
&\left.+t\beta_i^{t-1}\alpha_i\sum_{l=1}^{\infty}l\beta_i^{l-1}x_{i,t-1-l}^2-t(t-1)\beta_i^{t-2}\tilde{\sigma}_{i,0}^2+\alpha_i(t-1)(t-2)\beta_i^{t-3}(x_{i,0}^2-\tilde{x}_{i,0}^2)\right|^z\\
\le& a_1(t^z\beta_i^{tz}+\beta_i^{tz})+a_2t^z\beta_i^{(t-1)z}+a_3\beta_i^{tz}+a_4t^z(t-1)^z\beta_i^{(t-2)z}+a_5t^z\beta_i^{(t-1)z}\\
&+(t-1)^z(t-2)^z\beta_i^{(t-3)z}\\
=&O(t^z\beta_i^{tz})+O(\beta_i^{tz})+O(t^z(t-1)^z\beta_i^{tz}) \qquad (3.83)
\end{aligned}$$

and $\mathbb{E}\left|\frac{\partial^2\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2}{\partial\omega_i\partial\omega_i}\right|=\mathbb{E}\left|\frac{\partial^2\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2}{\partial\omega_i\partial\alpha_i}\right|=\mathbb{E}\left|\frac{\partial^2\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2}{\partial\omega_i\partial\omega_i}\right|=\mathbb{E}\left|\frac{\partial^2\sigma_{i,t}^2-\tilde{\sigma}_{i,t}^2}{\partial\alpha_i\partial\alpha_i}\right|=0.$
Sufficiently, we will prove the finiteness of each element and start with the diagonal terms. If $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_l$ are both in the set $\{\omega_k,\alpha_k,\beta_k\}$,

$$\begin{aligned}
\sum_{t=1}^{n}\mathbb{E}\left|\frac{\partial^2\tilde{H}_{jj,t}-H_{jj,t}}{\boldsymbol{\theta}_i\boldsymbol{\theta}_l}\right|^{z_{13}} =&\mathbb{E}\left|\frac{\partial^2H_{jj,t}}{(\partial\sigma_{k,t}^2)^2}\frac{\partial\sigma_{k,t}^2}{\partial\boldsymbol{\theta}_j}\frac{\partial\sigma_{k,t}^2}{\partial\boldsymbol{\theta}_l}-\frac{\partial^2\tilde{H}_{jj,t}}{(\partial\tilde{\sigma}_{k,t}^2)^2}\frac{\partial\tilde{\sigma}_{k,t}^2}{\partial\boldsymbol{\theta}_j}\frac{\partial\tilde{\sigma}_{k,t}^2}{\partial\boldsymbol{\theta}_l}\right.\\
&\left.+\frac{\partial H_{jj,t}}{\partial\sigma_{k,t}^2}\frac{\partial^2\sigma_{k,t}^2}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_l}-\frac{\partial\tilde{H}_{jj,t}}{\partial\tilde{\sigma}_{k,t}^2}\frac{\partial^2\tilde{\sigma}_{k,t}^2}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_l}\right|^{z_{13}}\\
\le&\sum_{t=1}^{n}(1+\beta_{0k})^{z_{13}}\mathbb{E}\left|\frac{\partial^2\sigma_{k,t}^2}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_l}-\frac{\partial^2\tilde{\sigma}_{k,t}^2}{\partial\boldsymbol{\theta}_j\partial\boldsymbol{\theta}_l}\right|^{z_{13}}\\
\le&\sum_{t=1}^{n}O(\beta_k^{tz_{13}})+\sum_{t=1}^{n}O(t^{z_{13}}\beta_k^{tz_{13}})\sum_{t=1}^{n}O(t^{z_{13}}(t-1)^{z_{13}}\beta_k^{tz_{13}})<\infty.
\end{aligned}$$

If one of $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_l$ is $\beta_{0k},k=1,2,\ldots,m$ and the other one is in the set $\{\omega_k,\alpha_k,\beta_k\}$. Assume $\boldsymbol{\theta}_i=\beta_{0k}$ and $\boldsymbol{\theta}_l$ is one element in $\{\omega_k,\alpha_k,\beta_k\}$, then we can conclude from (i) in Lemma 3.3.4 that

$$\sum_{t=1}^{n}\mathbb{E}\left|\frac{\partial^2\tilde{H}_{jj,t}-H_{jj,t}}{\boldsymbol{\theta}_i\boldsymbol{\theta}_l}\right|^{z_{13}} =\mathbb{E}\left|\frac{\partial\sigma_{k,t}^2}{\partial\boldsymbol{\theta}_l}-\frac{\partial\tilde{\sigma}_{k,t}^2}{\partial\boldsymbol{\theta}_l}\right|^{z_{13}}<\infty$$

The next step is to check the non-diagonal terms. Assume $i<j$ and $l<k$ in $\rho_{l,k}$ without loss of generality, $i,j,l,k=1,\ldots,m$.

It is easier to start working with the non-diagonal elements in the matrix.

$$
\begin{aligned}
&\mathbb{E}|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\rho_{lk}\partial\omega_{i_1}}|^{z_{13}}\\
=&\mathbb{E}|\frac{\partial^2 H_{ij,t}}{\partial\rho_{ij}\partial\sigma^2_{i_1,t}}\frac{\partial\sigma^2_{i_1,t}}{\partial\omega_{i_1}}-\frac{\partial^2\tilde{H}_{ij,t}}{\partial\rho_{ij}\partial\tilde{\sigma}^2_{i_1,t}}\frac{\partial\tilde{\sigma}^2_{i_1,t}}{\partial\omega_{i_1}}|^{z_{13}}(\mathbb{1}_{j=k}\mathbb{1}_{i=l}+\mathbb{1}_{j=l}\mathbb{1}_{i=k})(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
\leq&(\mathbb{E}|\frac{\partial^2 H_{ij,t}}{\partial\rho_{ij}\partial\sigma^2_{i_1,t}}-\frac{\partial^2\tilde{H}_{ij,t}}{\partial\rho_{ij}\partial\tilde{\sigma}^2_{i_1,t}}|^{2z_{13}}\mathbb{E}|\frac{\partial\sigma^2_{i_1,t}}{\partial\omega_{i_1}}|^{2z_{13}})^{1/2}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
&+(\mathbb{E}|\frac{\partial^2\tilde{H}_{ij,t}}{\partial\rho_{ij}\partial\tilde{\sigma}^2_{i_1,t}}|^{2z_{13}}\mathbb{E}|\frac{\partial\sigma^2_{i_1,t}}{\partial\omega_{i_1}}-\frac{\partial\tilde{\sigma}^2_{i_1,t}}{\partial\omega_{i_1}}|^{2z_{13}})^{1/2}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
\leq&O(\beta_i^{tz_{13}})+O(\beta_j^{tz_{13}})\\
&\mathbb{E}\left|\frac{\partial^2(H_{ij,t}-\tilde{H}_{iji,t})}{\partial\rho_{lk}\partial\alpha_{i_1}}\right|^{z_{13}}\\
=&\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{\partial\rho_{ij}\partial\sigma^2_{i_1,t}}\frac{\partial\sigma^2_{i_1,t}}{\partial\alpha_{i_1}}-\frac{\partial^2\tilde{H}_{ij,t}}{\partial\rho_{ij}\partial\tilde{\sigma}^2_{i_1,t}}\frac{\partial\tilde{\sigma}^2_{i_1,t}}{\partial\alpha_{i_1}}\right|^{z_{13}}(\mathbb{1}_{j=k}\mathbb{1}_{i=l}+\mathbb{1}_{j=l}\mathbb{1}_{i=k})(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
\leq&\left(\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{\partial\rho_{ij}\partial\sigma^2_{i_1,t}}-\frac{\partial^2\tilde{H}_{ij,t}}{\partial\rho_{ij}\partial\tilde{\sigma}^2_{i_1,t}}\right|^{2z_{13}}\mathbb{E}\left|\frac{\partial\sigma^2_{i_1,t}}{\partial\alpha_{i_1}}\right|^{2z_{13}}\right)^{1/2}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
&+\left(\mathbb{E}\left|\frac{\partial^2\tilde{H}_{ij,t}}{\partial\rho_{ij}\partial\tilde{\sigma}^2_{i_1,t}}\right|^{2z_{13}}\mathbb{E}\left|\frac{\partial\sigma^2_{i_1,t}}{\partial\alpha_{i_1}}-\frac{\partial\tilde{\sigma}^2_{i_1,t}}{\partial\alpha_{i_1}}\right|^{2z_{13}}\right)^{1/2}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
\leq&O(\beta_i^{tz_{13}})+O(\beta_j^{tz_{13}})\\
&\mathbb{E}\left|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\rho_{lk}\partial\beta_{i_1}}\right|^{z_{13}}\\
=&\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{\partial\rho_{lk}\partial\sigma^2_{i_1,t}}\frac{\partial\sigma^2_{i_1,t}}{\partial\beta_{i_1}}-\frac{\partial^2\tilde{H}_{ij,t}}{\partial\rho_{lk}\partial\tilde{\sigma}^2_{i_1,t}}\frac{\partial\tilde{\sigma}^2_{i_1,t}}{\partial\beta_{i_1}}\right|^{z_{13}}(\mathbb{1}_{j=k}\mathbb{1}_{i=l}+\mathbb{1}_{j=l}\mathbb{1}_{i=k})(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
\leq&(\mathbb{E}|\frac{\partial^2 H_{ij,t}}{\partial\rho_{ij}\partial\sigma^2_{i_1,t}}-\frac{\partial^2\tilde{H}_{ij,t}}{\partial\rho_{ij}\partial\tilde{\sigma}^2_{i_1,t}}|^{2z_{13}}\mathbb{E}|\frac{\partial\sigma^2_{i_1,t}}{\partial\beta_{i_1}}|^{2z_{13}})^{1/2}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
&+(\mathbb{E}|\frac{\partial^2\tilde{H}_{ij,t}}{\partial\rho_{ij}\partial\tilde{\sigma}^2_{i_1,t}}|^{2z_{13}}\mathbb{E}\left|\frac{\partial\sigma^2_{i_1,t}}{\partial\beta_{i_1}}-\frac{\partial\tilde{\sigma}^2_{i_1,t}}{\partial\beta_{i_1}}|^{2z_{13}}\right)^{1/2}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
\leq&O(\beta_i^{tz_{13}})+O(\beta_j^{tz_{13}})+O(t^{z_{13}}\beta_{i_1}^{tz_{13}})\\
&\mathbb{E}\left|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{(\partial\omega_{i_1})^2}\right|^{z_{13}}\\
=&\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{(\partial\sigma^2_{i_1})^2}(\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}})^2-\frac{\partial^2\tilde{H}_{ij,t}}{(\partial\tilde{\sigma}^2_{i_1})^2}(\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}})^2\right|^{z_{13}}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})
\end{aligned}
$$

$$
\begin{aligned}
&\le(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\Big(\mathbb{E}\Big|\frac{\partial^2 H_{ij,t}}{(\partial\sigma^2_{i_1})^2}-\frac{\partial^2 \tilde{H}_{ij,t}}{(\partial\tilde{\sigma}^2_{i_1})^2}\Big|^{2z_{13}}\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}}\Big|^{4z_{13}}\Big)^{1/2}\\
&\quad+(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\Big(\mathbb{E}\Big|\frac{\partial^2 \tilde{H}_{ij,t}}{(\partial\tilde{\sigma}^2_{i_1})^2}\Big|^{2z_{13}}\mathbb{E}\Big|(\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}})^2-(\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}})^2\Big|^{2z_{13}}\Big)^{1/2}\\
&\le[O(\beta_i^{tz_{13}})+O(\beta_j^{tz_{13}})+O(\beta_{i_1}^{tz_{13}})](\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})=O(\beta_i^{tz_{13}})+O(\beta_j^{tz_{13}})
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\Big|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\omega_{i_1}\partial\omega_{i_2}}\Big|^{z_{13}}\\
=&\mathbb{E}\Big|\frac{\partial^2 H_{ij,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{i_2}}\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}}\frac{\partial\sigma^2_{i_2}}{\partial\omega_{i_2}}-\frac{\partial^2 \tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{i_2}}\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}}\frac{\partial\tilde{\sigma}^2_{i_2}}{\partial\omega_{i_2}}\Big|^{z_{13}}(\mathbb{1}_{i_1=i}\mathbb{1}_{i_2=j}+\mathbb{1}_{i_2=i}\mathbb{1}_{i_1=j})\\
\le&\Big(\mathbb{E}|\frac{\partial^2 H_{ij,t}}{\partial\sigma^2_{i}\partial\sigma^2_{j}}-\frac{\partial^2 \tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i}\partial\tilde{\sigma}^2_{j}}|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}|\frac{\partial\sigma^2_{i}}{\partial\omega_{i}}|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{j}}{\partial\omega_{j}}\Big|^{4z_{13}}\Big)^{1/4}\\
&+\Big(\mathbb{E}\Big|\frac{\partial^2 H_{ij,t}}{\partial\sigma^2_{i}\partial\sigma^2_{j}}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i}}{\partial\omega_{i}}-\frac{\partial\tilde{\sigma}^2_{i}}{\partial\omega_{i}}\Big|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\frac{\partial\tilde{\sigma}^2_{j}}{\partial\omega_{j}}|^{4z_{13}}\Big)^{1/4}\\
&+\Big(\mathbb{E}\Big|\frac{\partial^2 \tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i}\partial\tilde{\sigma}^2_{j}}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\tilde{\sigma}^2_{i}}{\partial\omega_{i}}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{j}}{\partial\omega_{j}}-\frac{\partial\tilde{\sigma}^2_{j}}{\partial\omega_{j}}\Big|^{2z_{13}}\Big)^{1/2}\\
\le&O(\beta_i^{tz_{13}})+O(\beta_j^{tz_{13}})+O(\beta_i^{tz_{13}})+O(\beta_j^{tz_{13}})=O(\beta_i^{tz_{13}})+O(\beta_j^{tz_{13}})
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\Big|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\omega_{i_1}\partial\alpha_{i_1}}\Big|^{z_{13}}\\
=&\mathbb{E}\Big|\frac{\partial^2 H_{ij,t}}{(\partial\sigma^2_{i_1})^2}\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}}\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}-\frac{\partial^2 \tilde{H}_{ij,t}}{(\partial\tilde{\sigma}^2_{i_1})^2}\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}}\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}}\Big|^{z_{13}}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
\le&\Big(\mathbb{E}\Big|\frac{\partial^2 H_{ij,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{i_1}}-\frac{\partial^2 \tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{i_1}}\Big|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}\Big|^{4z_{13}}\Big)^{1/4}\\
&+\Big(\mathbb{E}\Big|\frac{\partial^2 H_{ij,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{i_1}}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}}-\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}}\Big|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\Big|\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}}\Big|^{4z_{13}}\Big)^{1/4}\\
&+\Big(\mathbb{E}\Big|\frac{\partial^2 \tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{i_1}}|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}|\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}}|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}-\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}}\Big|^{2z_{13}}\Big)^{1/2}\\
\le&O(\beta_i^{tz_{13}})+O(\beta_j^{tz_{13}})
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\Big|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\omega_{i_1}\partial\alpha_{i_2}}\Big|^{z_{13}}\\
=&\mathbb{E}\Big|\frac{\partial^2 H_{ij,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{i_2}}\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}}\frac{\partial\sigma^2_{i_2}}{\partial\alpha_{i_2}}-\frac{\partial^2 \tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{i_2}}\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}}\frac{\partial\tilde{\sigma}^2_{i_2}}{\partial\alpha_{i_2}}\Big|^{z_{13}}(\mathbb{1}_{i_1=i}\mathbb{1}_{i_2=j}+\mathbb{1}_{i_1=j}\mathbb{1}_{i_2=i})\\
\le&\Big(\mathbb{E}\Big|\frac{\partial^2 H_{i_1i_2,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{i_2}}-\frac{\partial^2 \tilde{H}_{i_1i_2,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{i_2}}\Big|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_2}}{\partial\alpha_{i_2}}\Big|^{4z_{13}}\Big)^{1/4}
\end{aligned}
$$

$$
\begin{aligned}
&+\Big(\mathbb{E}\Big|\frac{\partial^2 H_{i_1 i_2,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{i_2}}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}}-\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}}\Big|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\frac{\partial\tilde{\sigma}^2_{i_2}}{\partial\alpha_{i_2}}\Big|^{4z_{13}}\Big)^{1/4}\\
&+\Big(\mathbb{E}\Big|\frac{\partial^2 \tilde{H}_{i_1 i_2,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{i_2}}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_2}}{\partial\alpha_{i_2}}-\frac{\partial\tilde{\sigma}^2_{i_2}}{\partial\alpha_{i_2}}\Big|^{2z_{13}}\Big)^{1/2}\\
\le& O(\beta_{i_1}^{tz_{13}})+O(\beta_{i_2}^{tz_{13}})+O(\beta_{i_1}^{tz_{13}})+O(\beta_{i_2}^{tz_{13}})=O(\beta_{i_1}^{tz_{13}})+O(\beta_{i_2}^{tz_{13}})\\
&\mathbb{E}\Big|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\omega_{i_1}\partial\beta_{j_1}}\Big|^{z_{13}}\\
=&\mathbb{E}\Big|\frac{\partial^2 H_{ij,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{j_1}}\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}}\frac{\partial\sigma^2_{j_1}}{\partial\beta_{j_1}}-\frac{\partial^2 \tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{j_1}}\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}}\frac{\partial\tilde{\sigma}^2_{j_1}}{\partial\beta_{j_1}}\Big|^{z_{13}}(\mathbb{1}_{i_1=i}\mathbb{1}_{j_1=j}+\mathbb{1}_{i_1=j}\mathbb{1}_{j_1=i})\\
\le&\Big(\mathbb{E}\Big|\frac{\partial^2 H_{i_1j_1,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{j_1}}-\frac{\partial^2 \tilde{H}_{i_1j_1,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{j_1}}\Big|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}|\frac{\partial\sigma^2_{j_1}}{\partial\beta_{j_1}}\Big|^{4z_{13}}\Big)^{1/4}\\
&+\Big(\mathbb{E}\Big|\frac{\partial^2 \tilde{H}_{i_1j_1,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{j_1}}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}}-\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}}\Big|^{2z_{13}})^{1/2}(\mathbb{E}\frac{\partial\sigma^2_{j_1}}{\partial\beta_{j_1}}\Big|^{4z_{13}}\Big)^{1/4}\\
&+\Big(\mathbb{E}\Big|\frac{\partial^2 \tilde{H}_{i_1j_1,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{j_1}}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{j_1}}{\partial\beta_{j_1}}-\frac{\partial\tilde{\sigma}^2_{j_1}}{\partial\beta_{j_1}}\Big|^{2z_{13}}\Big)^{1/2}\\
\le& O(\beta_{i_1}^{tz_{13}})+O(\beta_{j_1}^{tz_{13}})+O(\beta_{i_1}^{tz_{13}})+O(\beta_{j_1}^{tz_{13}})+O(t^{z_{13}}\beta_{j_1}^{tz_{13}})\\
=& O(\beta_{i_1}^{tz_{13}})+O(\beta_{j_1}^{tz_{13}})+O(t^{z_{13}}\beta_{j_1}^{tz_{13}})\\
&\mathbb{E}\Big|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\omega_{i_1}\partial\beta_{0i_1}}\Big|^{z_{13}}=\mathbb{E}\Big|\frac{\partial^2(\sigma^2_{i_1,t}-\tilde{\sigma}^2_{i_1,t})}{\partial\omega_{i_1}}\Big|^{z_{13}}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\le O(\beta_{i_1}^{tz_{13}})
\end{aligned}
$$

By using (i) above and Cauchy-Schwarz inequality to enlarge the terms repeatedly,

$$
\begin{aligned}
&\mathbb{E}\Big|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\alpha_{i_1}\partial\alpha_{i_1}}\Big|^{z_{13}}\\
=&\mathbb{E}\Big|\frac{\partial^2 H_{ij,t}}{(\partial\sigma^2_{i_1})^2}\Big(\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}\Big)^2-\frac{\partial^2 \tilde{H}_{ij,t}}{(\partial\tilde{\sigma}^2_{i_1})^2}(\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}})^2\Big|^{z_{13}}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
\le&\Big(\mathbb{E}\Big|\frac{\partial^2 H_{ij,t}}{(\partial\sigma^2_{i_1})^2}-\frac{\partial^2 \tilde{H}_{ij,t}}{(\partial\tilde{\sigma}^2_{i_1})^2}|^{2z_{13}}\Big)^{1/2}\Big(\Big(\mathbb{E}|\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}|^{4z_{13}}\Big)^{1/4}\Big)^2(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
&+\Big(\mathbb{E}\Big|\frac{\partial^2 \tilde{H}_{i_1j_1,t}}{(\partial\sigma^2_{i_1})^2}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}-\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}}\Big|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}\Big|^{4z_{13}}\Big)^{1/4}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
&+\Big(\mathbb{E}\Big|\frac{\partial^2 \tilde{H}_{i_1j_1,t}}{(\partial\tilde{\sigma}^2_{i_1})^2}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}|\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}}|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}-\frac{\partial\tilde{\sigma}^2_{j_1}}{\partial\alpha_{i_1}}\Big|^{2z_{13}}\Big)^{1/2}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
\le& O(\beta_i^{tz_{13}})+O(\beta_j^{tz_{13}}).
\end{aligned}
$$

By using Cauchy-Schwarz inequality to enlarge the terms repeatedly,

$$
\begin{aligned}
&\mathbb{E}\left|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\alpha_{i_1}\partial\alpha_{j_1}}\right|^{z_{13}}\\
\leq&\mathbb{E}\left|\frac{\partial^2 H_{i_1j_1,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{j_1}}\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}\frac{\partial\sigma^2_{j_1}}{\partial\alpha_{j_1}}-\frac{\partial^2 \tilde{H}_{i_1j_1,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{j_1}}\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}}\frac{\partial\tilde{\sigma}^2_{j_1}}{\partial\alpha_{j_1}}\right|^{z_{13}}\\
\leq&\Big(\mathbb{E}\left|\frac{\partial^2 H_{i_1j_1,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{j_1}}-\frac{\partial^2 \tilde{H}_{i_1j_1,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{j_1}}\right|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}\right|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{j_1}}{\partial\alpha_{j_1}}\right|^{4z_{13}}\Big)^{1/4}\\
&+\Big(\mathbb{E}\left|\frac{\partial^2 \tilde{H}_{i_1j_1,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{j_1}}\right|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}-\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}}\right|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{j_1}}{\partial\alpha_{j_1}}\right|^{4z_{13}}\Big)^{1/4}\\
&+\Big(\mathbb{E}\left|\frac{\partial^2 \tilde{H}_{i_1j_1,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{j_1}}\right|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\left|\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}}\right|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{j_1}}{\partial\alpha_{j_1}}-\frac{\partial\tilde{\sigma}^2_{j_1}}{\partial\alpha_{j_1}}\right|^{2z_{13}}\Big)^{1/2}\\
=&O(\beta_{i_1}^{tz_{13}})+O(\beta_{j_1}^{tz_{13}})\\
&\mathbb{E}\left|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\alpha_{i_1}\partial\beta_{j_1}}\right|^{z_{13}}\\
\leq&\mathbb{E}\left|\frac{\partial^2 H_{i_1j_1,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{j_1}}\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}\frac{\partial\sigma^2_{j_1}}{\partial\beta_{j_1}}-\frac{\partial^2 \tilde{H}_{i_1j_1,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{j_1}}\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}}\frac{\partial\tilde{\sigma}^2_{j_1}}{\partial\beta_{j_1}}\right|^{z_{13}}\\
\leq&\Big(\mathbb{E}\left|\frac{\partial^2 H_{i_1j_1,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{j_1}}-\frac{\partial^2 \tilde{H}_{i_1j_1,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{j_1}}\right|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}\right|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{j_1}}{\partial\beta_{j_1}}\right|^{4z_{13}}\Big)^{1/4}\\
&+\Big(\mathbb{E}\left|\frac{\partial^2 \tilde{H}_{i_1j_1,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{j_1}}\right|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}-\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}}\right|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{j_1}}{\partial\beta_{j_1}}\right|^{4z_{13}}\Big)^{1/4}\\
&+\Big(\mathbb{E}\left|\frac{\partial^2 \tilde{H}_{i_1j_1,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{j_1}}\right|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\left|\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}}\right|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{j_1}}{\partial\beta_{j_1}}-\frac{\partial\tilde{\sigma}^2_{j_1}}{\partial\beta_{j_1}}\right|^{2z_{13}}\Big)^{1/2}\\
=&O(\beta_{i_1}^{tz_{13}})+O(\beta_{j_1}^{tz_{13}})+O(\beta_{i_1}^{tz_{13}})+O(\beta_{j_1}^{tz_{13}})+O(t^{z_{13}}\beta_{j_1}^{tz_{13}})\\
=&O(\beta_{i_1}^{tz_{13}})+O(\beta_{j_1}^{tz_{13}})+O(t^{z_{13}}\beta_{j_1}^{tz_{13}})\\
&\mathbb{E}\left|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\alpha_{i_1}\partial\beta_{0i_1}}\right|^{z_{13}}=\mathbb{E}\left|\frac{\partial^2(\sigma^2_{i_1,t}-\tilde{\sigma}^2_{i_1,t})}{\partial\alpha_{i_1}}\right|^{z_{13}}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\leq O(\beta_{i_1}^{tz_{13}})\\
&\mathbb{E}\left|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\beta_{i_1}\partial\beta_{j_1}}\right|^{z_{13}}\\
\leq&\mathbb{E}\left|\frac{\partial^2 H_{i_1j_1,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{j_1}}\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}}\frac{\partial\sigma^2_{j_1}}{\partial\beta_{j_1}}-\frac{\partial^2 \tilde{H}_{i_1j_1,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{j_1}}\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\beta_{i_1}}\frac{\partial\tilde{\sigma}^2_{j_1}}{\partial\beta_{j_1}}\right|^{z_{13}}\\
\leq&O(\beta_{i_1}^{tz_{13}})+O(\beta_{j_1}^{tz_{13}})+O(t^{z_{13}}\beta_{j_1}^{tz_{13}})+O(t^{p}\beta_{i_1}^{tz_{13}})
\end{aligned}
$$

$$\mathbb{E}\left|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\beta_{i_1}\partial\beta_{0i_1}}\right|^{z_{13}} = \mathbb{E}\left|\frac{\partial^2(\sigma^2_{i_1,t}-\tilde{\sigma}^2_{i_1,t})}{\partial\beta_{i_1}}\right|^{z_{13}}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j}) \leq O(\beta_{i_1}^{tz_{13}}) + O(t^{z_{13}}\beta_{i_1}^{tz_{13}})$$

Based on the preliminary results at the beginning and Cauchy-Schwarz inequality,

$$\begin{aligned}
&\mathbb{E}\left|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\omega_{i_1}\partial\beta_{i_1}}\right|^{z_{13}}\\
=&\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{(\partial\sigma^2_{i_1})^2}\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}}\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}} - \frac{\partial^2\tilde{H}_{ij,t}}{(\partial\tilde{\sigma}^2_{i_1})^2}\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}}\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\beta_{i_1}}\right.\\
&\left.+\frac{\partial H_{ij,t}}{\partial\sigma^2_{i_1}}\frac{\partial^2\sigma^2_{i_1}}{\partial\omega_{i_1}\partial\beta_{i_1}} - \frac{\partial\tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}}\frac{\partial^2\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}\partial\beta_{i_1}}\right|^{z_{13}}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
\leq&\Big[\Big(\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{i_1}} - \frac{\partial^2\tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{i_1}}\right|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}}\right|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}}\right|^{4z_{13}}\Big)^{1/4}\\
&+\Big(\mathbb{E}\left|\frac{\partial^2\tilde{H}_{ij,t}}{\partial\sigma^2_{i_1}\partial\sigma^2_{i_1}}\right|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{i_1}}{\partial\omega_{i_1}} - \frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}}\right|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{j_1}}{\partial\beta_{i_1}}\right|^{4z_{13}}\Big)^{1/4}\\
&+\Big(\mathbb{E}\left|\frac{\partial^2\tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}\partial\tilde{\sigma}^2_{j_1}}\right|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\left|\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}}\right|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}} - \frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\beta_{i_1}}\right|^{2z_{13}}\Big)^{1/2}\\
&+\Big(\mathbb{E}\left|\frac{\partial H_{ij,t}}{\partial\sigma^2_{i_1}} - \frac{\partial\tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}}\right|^{2z_13}\Big)^{1/2}\Big(\mathbb{E}\left|\frac{\partial^2\sigma^2_{i_1}}{\partial\omega_{i_1}\partial\beta_{i_1}}\right|^{2z_{13}}\Big)^{1/2}\\
&+\Big(\mathbb{E}\left|\frac{\partial^2\sigma^2_{i_1}}{\partial\omega_{i_1}\partial\beta_{i_1}} - \frac{\partial^2\tilde{\sigma}^2_{i_1}}{\partial\omega_{i_1}\partial\beta_{i_1}}\right|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\left|\frac{\partial\tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}}\right|^{2z_{13}}\Big)^{1/2}\Big](\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
\leq&O(\beta_i^{tz_{13}}) + O(\beta_j^{tz_{13}}) + O(\beta_{i_1}^{tz_{13}}) + O(\beta_i^{tz_{13}}) + O(t^{z_{13}}\beta_{i_1}^{tz_{13}})\\
=&O(\beta_i^{tz_{13}}) + O(\beta_j^{tz_{13}}) + O(t^{z_{13}}\beta_{i_1}^{tz_{13}})
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}\left|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\alpha_{i_1}\partial\beta_{i_1}}\right|^{z_{13}}\\
=&\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{(\partial\sigma^2_{i_1})^2}\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}} - \frac{\partial^2\tilde{H}_{ij,t}}{(\partial\tilde{\sigma}^2_{i_1})^2}\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}}\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}}\right.\\
&\left.+\frac{\partial H_{ij,t}}{\partial\sigma^2_{i_1}}\frac{\partial^2\sigma^2_{i_1}}{\partial\alpha_{i_1}\partial\beta_{i_1}} - \frac{\partial\tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}}\frac{\partial^2\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}\partial\beta_{i_1}}\right|^{z_{13}}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
\leq&\Big[\Big(\mathbb{E}\left|\frac{\partial^2 H_{ij,t}}{(\partial\sigma^2_{i_1})^2} - \frac{\partial^2\tilde{H}_{ij,t}}{(\partial\tilde{\sigma}^2_{i_1})^2}\right|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}}\right|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}}\right|^{4z_{13}}\Big)^{1/4}\\
&+\Big(\mathbb{E}\left|\frac{\partial^2\tilde{H}_{ij,t}}{(\partial\sigma^2_{i_1})^2}\right|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{i_1}}{\partial\alpha_{i_1}} - \frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}}\right|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\left|\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}}\right|^{4z_{13}}\Big)^{1/4}
\end{aligned}$$

$$
\begin{aligned}
&+\Big(\mathbb{E}\Big|\frac{\partial^2\tilde{H}_{ij,t}}{(\partial\tilde{\sigma}^2_{i_1})^2}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\alpha_{i_1}}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}}-\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\beta_{i_1}}\Big|^{2z_{13}}\Big)^{1/2}\\
&+\Big(\mathbb{E}\Big|\frac{\partial H_{ij,t}}{\partial\sigma^2_{i_1}}-\frac{\partial\tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}}\Big|^{2z_13}\Big)^{1/2}\Big(\mathbb{E}\Big|\frac{\partial^2\sigma^2_{i_1}}{\partial\beta_{i_1}\partial\beta_{i_1}}\Big|^{2z_{13}}\Big)^{1/2}\\
&+\Big(\mathbb{E}\Big|\frac{\partial^2\sigma^2_{i_1}}{\partial\beta_{i_1}\partial\beta_{i_1}}-\frac{\partial^2\tilde{\sigma}^2_{i_1}}{\partial\beta_{i_1}\partial\beta_{i_1}}\Big|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\Big|\frac{\partial\tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}}\Big|^{2z_{13}}\Big)^{1/2}\Big]\\
\le& O(\beta_i^{tz_{13}})+O(\beta_j^{tz_{13}})+O(\beta_i^{tz_{13}})+O(t^p\beta_{i_1}^{tz_{13}})\\
&\mathbb{E}\Big|\frac{\partial^2(H_{ij,t}-\tilde{H}_{ij,t})}{\partial\beta_{i_1}\partial\beta_{i_1}}\Big|^{z_{13}}\\
=&\mathbb{E}\Big|\frac{\partial^2 H_{ij,t}}{(\partial\sigma^2_{i_1})^2}\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}}\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}}-\frac{\partial^2\tilde{H}_{ij,t}}{(\partial\tilde{\sigma}^2_{i_1})^2}\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\beta_{i_1}}\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}}\\
&+\frac{\partial H_{ij,t}}{\partial\sigma^2_{i_1}}\frac{\partial^2\sigma^2_{i_1}}{\partial\beta_{i_1}\partial\beta_{i_1}}-\frac{\partial\tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}}\frac{\partial^2\tilde{\sigma}^2_{i_1}}{\partial\beta_{i_1}\partial\beta_{i_1}}\Big|^{z_{13}}(\mathbb{1}_{i_1=i}+\mathbb{1}_{i_1=j})\\
\le&\Big[\Big(\mathbb{E}\Big|\frac{\partial^2 H_{ij,t}}{(\partial\sigma^2_{i_1})^2}-\frac{\partial^2\tilde{H}_{ij,t}}{(\partial\tilde{\sigma}^2_{i_1})^2}\Big|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}}\Big|^{4z_{13}}\Big)^{1/2}\\
&+\Big(\mathbb{E}\Big|\frac{\partial^2\tilde{H}_{ij,t}}{(\partial\sigma^2_{i_1})^2}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}}-\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\beta_{i_1}}\Big|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}}\Big|^{4z_{13}}\Big)^{1/4}\\
&+\Big(\mathbb{E}\Big|\frac{\partial^2\tilde{H}_{ij,t}}{(\partial\tilde{\sigma}^2_{i_1})^2}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\beta_{i_1}}\Big|^{4z_{13}}\Big)^{1/4}\Big(\mathbb{E}\Big|\frac{\partial\sigma^2_{i_1}}{\partial\beta_{i_1}}-\frac{\partial\tilde{\sigma}^2_{i_1}}{\partial\beta_{i_1}}\Big|^{2z_{13}}\Big)^{1/2}\\
&+\Big(\mathbb{E}\Big|\frac{\partial H_{ij,t}}{\partial\sigma^2_{i_1}}-\frac{\partial\tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}}\Big|^{2z_13}\Big)^{1/2}\Big(\mathbb{E}\Big|\frac{\partial^2\sigma^2_{i_1}}{\partial\beta_{i_1}\partial\beta_{i_1}}\Big|^{2z_{13}}\Big)^{1/2}\\
&+\Big(\mathbb{E}\Big|\frac{\partial^2\sigma^2_{i_1}}{\partial\beta_{i_1}\partial\beta_{i_1}}-\frac{\partial^2\tilde{\sigma}^2_{i_1}}{\partial\beta_{i_1}\partial\beta_{i_1}}\Big|^{2z_{13}}\Big)^{1/2}\Big(\mathbb{E}\Big|\frac{\partial\tilde{H}_{ij,t}}{\partial\tilde{\sigma}^2_{i_1}}\Big|^{2z_{13}}\Big)^{1/2}\Big]\\
\le& O(\beta_i^{tz_{13}})+O(\beta_j^{tz_{13}})+O(t^{z_{13}}\beta_{i_1}^{tz_{13}})+O(t^{z_{13}}\beta_{i_1}^{tz_{13}})+O(t^{z_{13}}(t-1)^{z_{13}}\beta_i^{tz_{13}})
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\sum_{t=1}^n\mathbb{E}\Big\|{}_{ij}\ddot{\tilde{H}}_t-{}_{ij}\ddot{H}_t\Big\|^{z_{13}}\\
=&\sum_{t=1}^n\mathbb{E}\Big\|\frac{\partial^2(H_t-\tilde{H}_t)}{\partial\theta_i\partial\theta_j}\Big\|^{z_{13}}\\
\le&\sum_{t=1}^n\Big[O(\beta_k^{tz_{13}})+O(t^{z_{13}}\beta_k^{tz_{13}})+O(\beta_l^{tz_{13}})+O(t^{z_{13}}\beta_l^{tz_{13}})+O(\beta_l^{tz_{13}})
\end{aligned}
$$

$$+ O(t^{z_{13}}(t-1)^{z_{13}}\beta_l^{tz_{13}}) + O(t^{z_{13}}(t-1)^{z_{13}}\beta_k^{tz_{13}})\Big]$$
$$\leq \sum_{t=1}^{n} O(b_1^{tz_{13}}) + O(b_1^{2tz_{13}}) + O(t^{z_{13}} b_1^{tz_{13}}) + O(t^{z_{13}}(t-1)^{z_{13}} b_1^{tz_{13}})$$

We use the similar argument used above in (vi) and get the rest of the desired results,

$$\lim_{n\to\infty} \sum_{t=1}^{n} \left(\mathbb{E}\left\| {}_{ij}\ddot{\tilde{H}}_t - {}_{ij}\ddot{H}_t \right\|^{2z_{14}}\right)^{1/2} < \infty$$

$$\lim_{n\to\infty} \sum_{t=1}^{n} \left(\mathbb{E}\left\| {}_{ij}\ddot{\tilde{H}}_t - {}_{ij}\ddot{H}_t \right\|^{4z_{15}}\right)^{1/4} < \infty$$

since $0 \leq 2z_{14} < 1/2$ and $0 \leq 4z_{15} < 1/2$

(viii) Recall the difference between $\sigma_{i,t}^2$ and $\tilde{\sigma}_{i,t}^2$ in (3.25),

$$\mathbb{E}\left|\sigma_{i,t}^2 - \tilde{\sigma}_{i,t}^2\right|^4 = \mathbb{E}\left|\beta_i^t(\sigma_{0,t}^2 - \tilde{\sigma}_{0,t}^2) + \alpha_i\beta_i^{t-1}(x_{0,t}^2 - \tilde{x}_{0,t}^2)\right|^4.$$

By Assumption $\boldsymbol{C}1$ and (i) in Lemma 3.3.2, $\mathbb{E}\left|\sigma_{i,t}^2 - \tilde{\sigma}_{i,t}^2\right|^4 < \infty$. The result along with (ii) leads to $\sum_{t=1}^{\infty} \mathbb{E}\left\|H_t - \tilde{H}_t\right\|^4 < \infty$. The other inequities are true by applying the same arguments used in (vi)and (vii).

□

Chapter 4

Simulation Study

4.1 Introduction

A Monte Carlo simulation will be performed to study the asymptotic properties of the QMLE for a given true value $\boldsymbol{\theta}_0$, which satisfies all the assumptions regarding the parameter, including $A2, A3, A4$ in Chapter 2 and $B1$ in Chapter 3. When the innovations follow the Gaussian distribution, then Assumptions $A1$ and $A5$ are satisfied. Assumptions $A2$ and $A3$ are explicit expressions which are easy to verify if we know the value of $\boldsymbol{\theta}_0$, while the other two assumptions $A4$ and $B1$ are complicated. The main problem we are facing is that the stationary and ergodic parameter space is not explicitly known, so the search area cannot be defined in the algorithm when we estimate the parameters.

In this chapter, we study a simplified version of the model with two-dimensional data. Prior to the simulation, we want to find a proper value in the parameter space. For Assumption $B1$, a closed interval is defined in the next section as a searching area so that the algorithm is looking for the estimate in a compact space. With the general consensus in financial economics, the log return series of stocks is stationary. We start with the log returns of two stocks. After applying numeric optimization with the conditions in Assumptions $A2$ and $A3$, we can get an estimate $\hat{\boldsymbol{\theta}}_d$ from the real data which maximizes the likelihood function with some initial values. The Monte Carlo method is used in the next section to verify whether the drift condition, Assumption $A4$, is satisfied. If the estimate $\hat{\boldsymbol{\theta}}_d$ passes the test, it will be used as "true" parameter $\boldsymbol{\theta}_0$ when we use the simulation method to study the statistical properties. Otherwise, a modification will be implemented on $\hat{\boldsymbol{\theta}}_d$ to create such a $\boldsymbol{\theta}_0$ which meets Assumptions $A1 - A5$ and $B1, B3, B4$.

It is worth to notice that even if $\hat{\boldsymbol{\theta}}_d$ does not pass the drift condition in Assumption $A4$, it does not mean that it cannot produce a stationary and ergodic observable sequence. As a sufficient condition, Assumption $A4$ leads to a smaller space than the true stationary and ergodic parameter space. Often the estimates obtained from real data analysis fail to pass the drift condition test. We included the results obtained from the Monte Carlo simulation on an artificial $\boldsymbol{\theta}_0$ in this chapter which passed the drift condition. The same simulation was also performed on $\hat{\boldsymbol{\theta}}_d$, and the convergence results are similar to we got for $\boldsymbol{\theta}_0$. This suggests that the sufficient drift condition $A4$ might be too strong.

4.2 Monte Carlo Study Preparation

To reduce the number of parameters and simplify the model, the contributions from each individual stock to the common risk indicator $\sigma^2_{0,t}$ can be set to be equal, which means $\beta_{01} = \beta_{02} = \cdots = \beta_{0m} = \beta_0$. This setup not only serves the convenience purpose but also saves the computation time. The number of parameters in $\sigma^2_{0,t}$ will be reduced to 2 from $m + 1$, which we have before the simplification.

A bivariate example is shown in this section. The bivariate realization of the model becomes

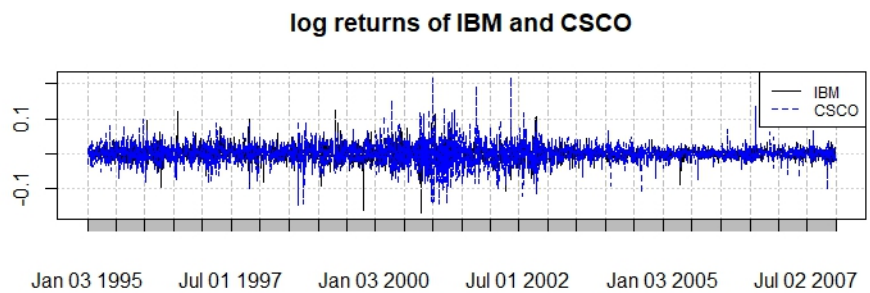

Figure 4.1: Log return of IBM and CSCO from 1995-01-01 to 2007-12-31

$$\begin{cases} x_{1,t} = \epsilon_{1,t}\sigma_{1,t} + \epsilon_{0,t}\sigma_{0,t} \\ x_{2,t} = \epsilon_{2,t}\sigma_{2,t} + \epsilon_{0,t}\sigma_{0,t} \\ \sigma^2_{1,t} = \omega_1 + \alpha_1 x^2_{1,t-1} + \beta_1\sigma^2_{1,t-1} \\ \sigma^2_{2,t} = \omega_2 + \alpha_2 x^2_{2,t-1} + \beta_2\sigma^2_{2,t-1} \\ \sigma^2_{0,t} = \beta_0(\sigma^2_{1,t} + \sigma^2_{2,t}). \end{cases} \tag{4.1}$$

and the number of parameters is 8. In order to choose a 'true' parameter $\boldsymbol{\theta}_0$ in this simulation study, we estimate the parameters based on the centered log returns of two equity series (two stocks in American stock market): International Business Machines Corporation (IBM) and Cisco Systems, Inc. (CSCO) from 1995 to 2007 with 3274 trading days in total. As we can see from Figure 4.1, more than 90% of the log returns lie within the range between -0.05 and 0.05.

The default searching box needs to be chosen by considering not only the assumptions but also some additional constraints. One of the constraints is that the meaning of these parameters. On the one hand, we want at least one of β_{0i}'s to be larger than 0 so that the common term $\sigma^2_{0,t}$ does not disappear. Therefore, in this case, a lower bound is needed for β_0 other than 0. On the other hand, we are not expecting the contribution from one term to the common term $\sigma^2_{0,t}$ higher than itself, which means that the upper bound of β_{0i} is set to be 1 for β_{0i}'s.

The final searching region is chosen to be $\boldsymbol{\Theta} = [-1, 1] \times [\iota, 1]^2 \times [\iota, 1]^2 \times [0, 1-\iota]^2 \times [\iota, 1]$ where ι is set to be 10^{-40}. Several numeric checks are added to verify the positive definite constraints on both $\boldsymbol{H}_t$ and $\boldsymbol{R}$ matrices through the eigen decomposition. Further numeric issues will be discussed in Section 4.4.

In Theorem 3.3.1, the observed $\boldsymbol{x}_t$ have a finite 8th moment which can lead to the weak stationarity. Let $\mathbb{E}\sigma_{1,t}^2 = \dfrac{\omega_1}{1-\beta_1} + \dfrac{\alpha_1 \mathbb{E}x_{1,t}^2}{1-\beta_1} = w_1$, $\mathbb{E}\sigma_{2,t}^2 = \dfrac{\omega_2}{1-\beta_2} + \dfrac{\alpha_2 \mathbb{E}x_{2,t}^2}{1-\beta_2} = w_2$ and $\mathbb{E}\sigma_{0,t}^2 = w_0$, then w_1, w_2 can be solved by the last two equations in (4.1). Therefore,

$$w_1 = \frac{w_0\alpha_1 + \omega_1}{1-\alpha_1-\beta_1},$$

$$w_2 = \frac{w_0\alpha_2 + \omega_2}{1-\alpha_2-\beta_2}.$$

To ensure these terms are positive, Assumption $A3$ is needed.

A non-linear optimization function `nlminb` in R with box constraints is used to estimate the parameter. While this function is extremely helpful, it has a few numeric problems when we use it in such a high dimensional case. These problems will be discussed in Section 4.4. The output from `nlminb` gives a locally optimal solution but without specifying whether it is the global ones. Varies methods in Section 4.4 are used to increase the likelihood of being the global optimum.

The estimated values are shown in the first row of Table 4.1 as $\hat{\boldsymbol{\theta}}_d$. The next step is to test whether this estimate satisfies Assumption $A4$.

A Monte Carlo study has been conducted to study the properties of the estimates in a numeric way. In this verification, the integer p_1 is set to be 2, then the corresponding matrix induced norm is also called spectral norm. The positive number s is chosen to be 1. Theoretically, the conditional standard deviation $(\sigma_{1,t}, \sigma_{2,t})$ can reach positive infinity without any upper bound. In theory, the expectation listed in Assumption $A4$ needs to be verified for all possible $\boldsymbol{y}$ in the state space. Nonetheless, we only need to know that the expectation is smaller than 1 within the sensible range from a practical point of view.

The time series sequence, $\boldsymbol{x_t}^\intercal = (x_{1,t}, x_{2,t})$, we are targeting is the daily stock return which never has any value beyond the interval $(-0.5, 0.5)$. From the expression in (4.1) and the constraint $\alpha_i + \beta_i < 1$, it is sensible to verify the conditional standard deviation using the upper bound $(1, 1)^\intercal$. In the meantime. the lower bound of this region is set to be $(\sqrt{\omega_1}, \sqrt{\omega_2})^\intercal$.

In this bivariate case, the steps to verify the drift assumption for a given value $\boldsymbol{\theta}$,

1. Choose a sample size n, simulate n independent and identically distributed 3 dimensional innovations $\boldsymbol{\epsilon_t}$ where $\boldsymbol{\epsilon_t}$'s are $N(\boldsymbol{0}, \Sigma)$.

2. Set up the region with $\boldsymbol{y}_{min} = (y_{min,1}, y_{min,2})^\intercal = (\sqrt{\omega_1}, \sqrt{\omega_2})^\intercal$ and $\boldsymbol{y}_{max} = (y_{max,1}, y_{max,2})^\intercal = (1, 1)^\intercal$. We can discretize the interval between $\boldsymbol{y}_{min}$ and $\boldsymbol{y}_{max}$ by creating grids using the weighted averages. The domain of each element in the 2 dimensional weight variable $\boldsymbol{s} = (s_1, s_2)$, [0,1], can be divided into K_3 equally spaced points $(0, 1/K_3, \cdots, (K_3 - 1)/K_3, 1)$. The grid points are set to be $\boldsymbol{y}_{\boldsymbol{s},t} = \boldsymbol{y}_{s_1,s_2,t} = (s_1 y_{min,1} + (1 - s_1) y_{max,1}, s_2 y_{min,2} + (1 - s_2) y_{max,2})^\intercal$, so the number of grids is $(K_3 + 1) \times (K_3 + 1)$.

3. For each grid point, the partial derivative matrix $B(\boldsymbol{y}_{s,t}, \epsilon_t)$ is calculated for these innovations $\boldsymbol{\epsilon_t}$. Then the average of these n values is obtained as the term $\mathbb{E}\,||B(\bar{y}, \epsilon_t)||_2$.

4. Finally, we can get the range of $\mathbb{E}\,||B(\bar{y}, \epsilon_t)||_2$ over the region defined. If the upper bound of the range is smaller than 1, then we can conclude that Assumption $A4$ is satisfied.

The steps are similar to the ones in Hafner and Preminger (2009b) and Jiang (2011), but more complicated and computationally intensive. The sample n is set to be 200 and the number of grids, K_3, is 1001 in this example.

```
> range(EDriftOrg)

[1] 0.7148196 1.0071080
```

Table 4.1: The numeric estimate from IBM and CSCO centered log return and the 'True' value used for the bootstrap simulations (rounded to two decimal digits)

	$\rho_{1,2}$	$10^6\omega_1$	$10^6\omega_2$	$100\alpha_1$	$100\alpha_2$	$10\beta_1$	$10\beta_2$	$10\beta_0$
$\hat{\boldsymbol{\theta}}_d$	4.18	2.93	8.03	7.65	5.54	9.09	9.30	0.44
$\boldsymbol{\theta}_0$	4.18	2.93	8.03	7.65	5.54	8.50	8.50	4.42

Unfortunately, the estimate $\hat{\boldsymbol{\theta}}_d$ does not pass the test. So we need to modify the values to create a $\boldsymbol{\theta}_0$ which satisfies the conditions stated above.

Some testing runs have been performed to check which element has a larger effect on this result. Only one value is changed at each run while the values of other elements of the parameter remain the same as $\hat{\boldsymbol{\theta}}_d$ in Table 4.1. As in the univariate or multivariate GARCH models, the changes in β_i have a significant impact on the partial derivative. The "true" values of parameters in the second row are set to be the modified values based on the first row by modifying the values of 3 parameters. The individual parameters β_1 and β_2 are set to be 0.85, and β_0 to be 10 times the estimate around 0.42. All the numbers in the table are rounded to 2 decimal places for a better display where the more accurate values are used in the simulation.

Once again, the same test is applied to this modified parameter value $\boldsymbol{\theta}_0$.

```
> range(EDriftMod)

[1] 0.6990809 0.9907809
```

The upper bound of that expectation is smaller than 1, this $\boldsymbol{\theta}_0$ satisfies all the assumptions in Chapter 2 and 3 regarding the parameter. Thus, we can use this $\boldsymbol{\theta}_0$ as the true value in the following examples.

4.3 Simulated Results

In order to investigate the consistency and asymptotic normality, we need to know the exact value of the true parameter. Then the simulated results can be compared with the theoretical values, which are calculated based on the theorems in the previous chapters.

The initial values are fixed at $(\boldsymbol{x}_0, \boldsymbol{\sigma}_0) = (0.008, 0.008, \sigma_{0,1}, \sigma_{0,2})$ where $\sigma_{0,1}, \sigma_{0,2}$, are the default values obtained from the log returns of IBM and CSCO by the method described in Section 4.4.3. Given the true value $\boldsymbol{\theta}_0$ satisfying the assumptions, we could simulate a path with i.i.d. normally distributed innovation and then estimate the parameters by maximizing the Gaussian likelihood function based on the simulated path. If the simulating and estimating processes were repeated for K times with K large enough, we could numerically verify the asymptotic properties of the ML estimators with the knowledge about the true parameter value.

The initial values $(\boldsymbol{x}_0, \boldsymbol{\sigma}_0)$ and the true parameter $\boldsymbol{\theta}_0$ are fixed as the values stated above. For the ith path, the detailed steps to generate the estimates are shown as following:

1. Generate $K_1 + K_2$ i.i.d multivariate normally distributed innovation $\boldsymbol{\epsilon}_1$ with mean $\mathbf{0}$ and known covariance Σ, where K_2 is the number of observations we desired and K_1 is the size of the burn-in period.

2. Calculate the simulated observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{K_1+K_2}$ and the corresponding $\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_{K_1+K_2}$ iteratively.

3. Drop the first K_1 points from the sequences, the sequences left are $\boldsymbol{x}_{K_1+1}, \ldots, \boldsymbol{x}_{K_1+K_2}$ and $\boldsymbol{\sigma}_{K_1+1}, \ldots, \boldsymbol{\sigma}_{K_1+K_2}$.

4. Get the estimate $\hat{\boldsymbol{\theta}}_{K_2}^{(i)}$ by using the method in Section 4.4.

We repeat this process for N_2 times, then examine the behavior of the estimates.

Some typical paths of $\boldsymbol{x}_t$ and $\boldsymbol{\sigma}_t$ are shown in Figure 4.2 which represent the paths $\boldsymbol{x}_1^{(i)}, \ldots, \boldsymbol{x}_{K_1+K_2}^{(i)}$ in Step 1 and 2. Though the true parameter we used to generate the path is an artificial $\boldsymbol{\theta}_0$, the simulated observed paths $\boldsymbol{x}_t$ still show a similar characteristic as the original stock returns. They have similar ranges, the same volatility clustering feature and both are heavy tail distributed.

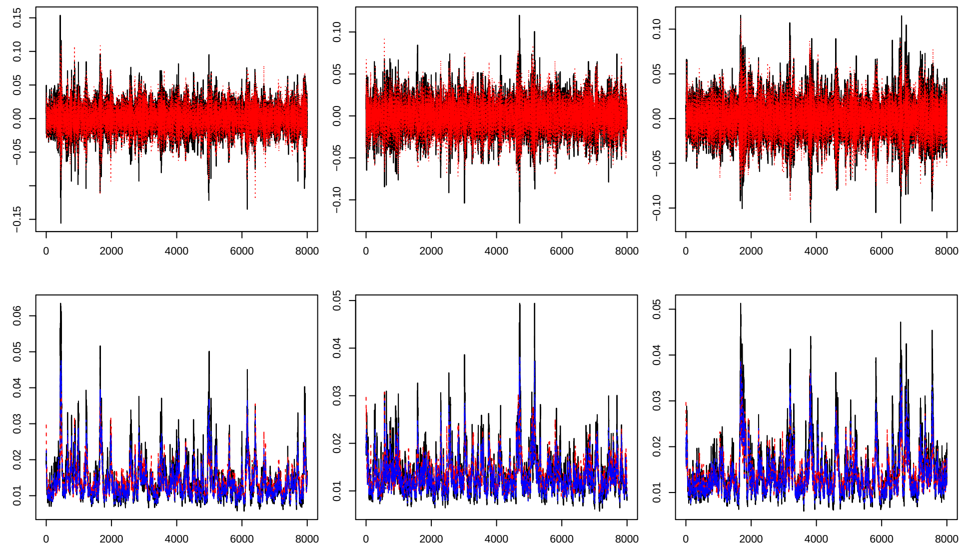

Figure 4.2: The simulated paths: the top three are the simulated $\boldsymbol{x}_t$ (the black solid line represents $x_{1,t}$ and the red dashed line represents $x_{2,t}$) and the three below are the corresponding $\boldsymbol{\sigma}_t$ (the black line represents $\sigma^2_{1,t}$, the red line represents $\sigma^2_{2,t}$ and the blue line represents $\sigma^2_{0,t}$)

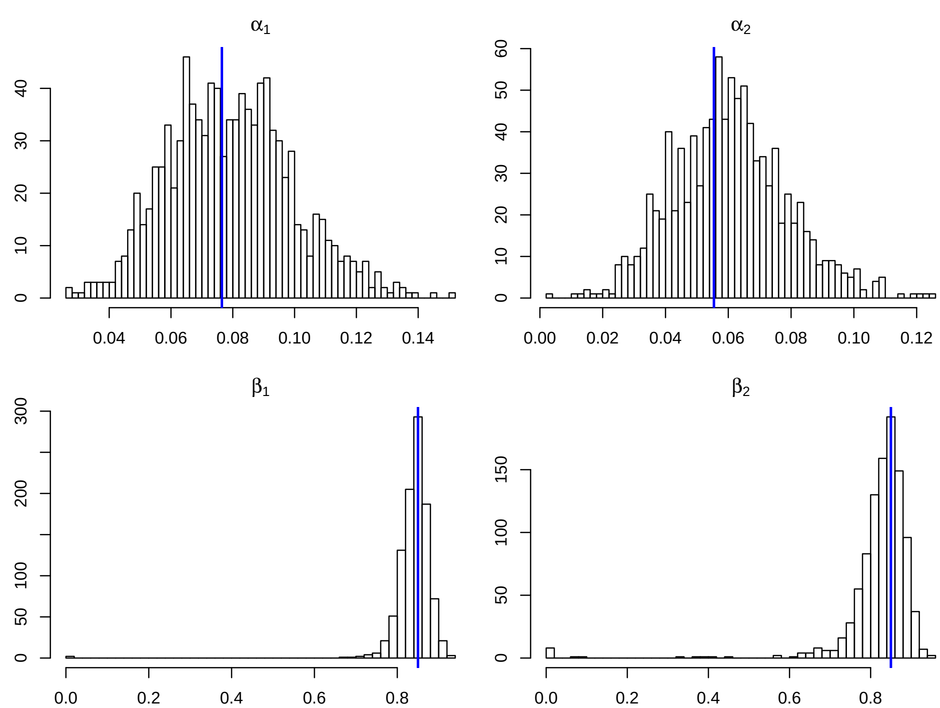

Figure 4.3: The histogram of $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2$ when $K_2 = 1000$. The blue lines represent the true values.

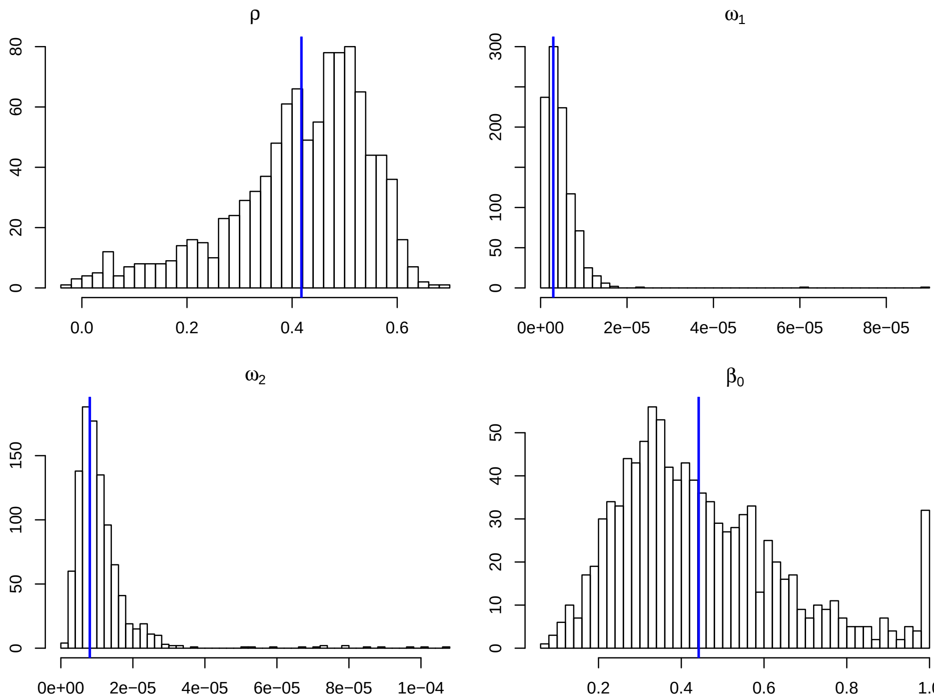

Figure 4.4: The histogram of $\hat{\rho}, \hat{\omega}_1, \hat{\omega}_2, \hat{\beta}_0$ when $K_2 = 1000$. The blue lines represent the true values.

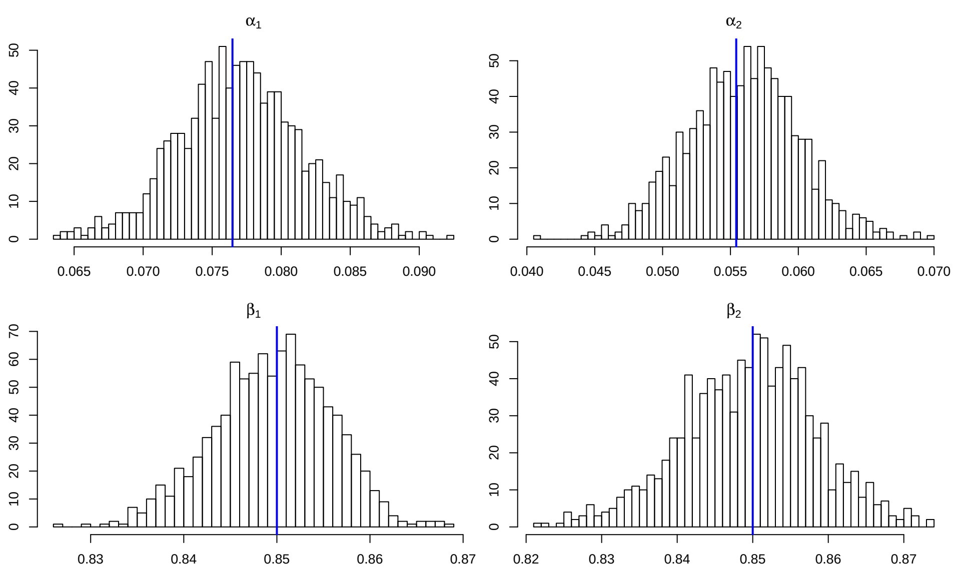

Figure 4.5: The histogram of $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2$ when $K_2 = 20000$. The blue lines represent the true values.

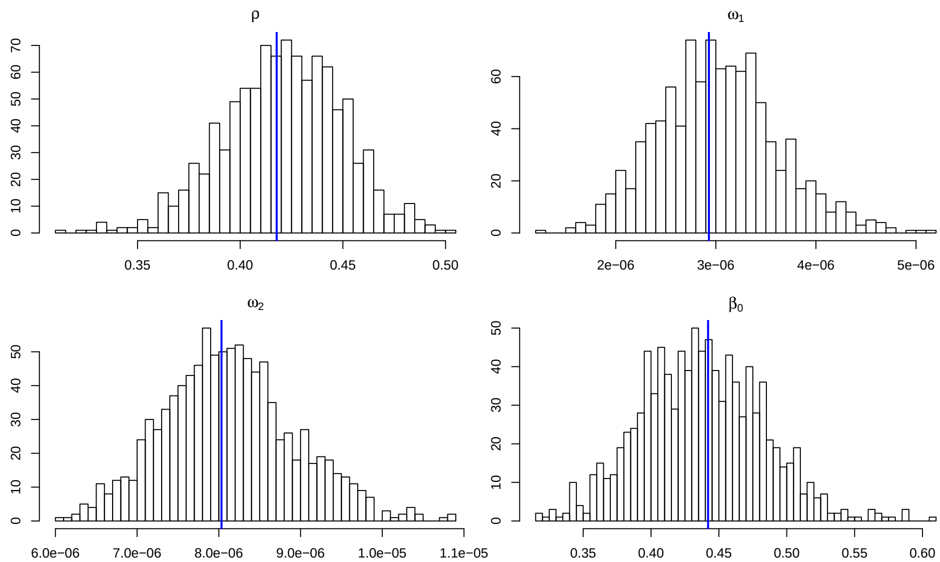

Figure 4.6: The histogram of $\hat{\rho}, \hat{\omega_1}, \hat{\omega_2}, \hat{\beta_0}$ when $K_2 = 20000$. The blue lines represent the true values.

We simulate 1000 paths ($N_2 = 1000$) with different sample sizes K_2 from 1000 to 20000 and set the burn-in size K_1 to be 7000. Figure 4.3 and Figure 4.4 show the histogram of the parameters. From these two figures, we can see that the estimate of ρ has a positive bias while the estimate of β_0 has a negative bias. Among the 1000 estimates of β_0, 26 of them hit the upper bound of the searching region, 1. Though 1000 is about 4 years of data in the stock market, it is still very small when we want to study the asymptotic properties. Most of the histograms are skewed. When the sample size is increased to 20000, the histograms of the estimates in Figure 4.5 and 4.6 become more symmetric. The centers of the bell shapes are much closer to the true values. Between these two sample sizes, 1000 and 20000, 4 sample sizes are chosen for K_2 and they are 3000, 5000, 7000 and 10000.

Section 4.4 uses a numeric optimizer, an element of the output gives an indicator whether the numeric iteration converges based on the some numerical criteria. The proportion of the convergenced estimates is 99.5% when the sample size equals to 1000. Then the proportion gradually decays to 91.6% when we increase the sample size to 20000(98.3% for $K_2 = 3000$, 96.1% for $K_2 = 5000$, 94.7% for $K_2 = 7000$ and 93.1% for

$K_2 = 10000$). Given the fact that the target function is summing up a large number of terms, the rounding error may become overwhelming when the sample size increases. The change in the convergence rate relates to the numerical stability which we have no control. All the results below are based on the converged estimates.

Ignore the sign of the bias, the absolute bias of a parameter with true value θ_0 is defined as

$$\left|\frac{1}{N_2}\sum_{i=1}^{N_2}\hat{\theta}_{K_2}^{(i)} - \theta_0\right|.$$

Then, the absolute biases of each parameter with different sample sizes are shown in Table 4.2. The absolute biases generally decrease as the sample size increases with few exceptions. Since the estimate of one parameter is related to the estimate of other parameters, the increasing of the absolute bias of one parameter may be because of the reduction in the absolute bias of another parameter. We cannot compare the overall effect for different sample sizes. It is important to realize that it is impossible to achieve 0 in any numeric study and the accuracy is limited by several factors. We define another measure called relative absolute bias as

$$\text{relative absolute bias (RAB) of } \theta = \frac{\text{absolute bias of } \theta}{\theta_0}$$

which takes the scale of the true parameter into account. The RABs for different parameters are comparable in terms of different K_2. Table 4.2 is converted to Table 4.3 by using this new measure. In the last column, the vanishing trend of the total RAB is clear as the sample size increases, which means the bigger the sample size, the better the estimate will be.

In addition to the absolute bias, the root mean square error (RMSE) is used to evaluate the performance of the estimate. The RMSE we used in this thesis is

$$\text{RMSE of } \theta = \sqrt{\frac{1}{N_2}\sum_{i=1}^{N_2}(\hat{\theta}_{K_2}^{(i)} - \theta_0)^2}$$

Table 4.2: The absolute biases with different sample sizes (rounded to two decimal digits)

K_2	$\rho_{1,2}(10^{-3})$	$\omega_1(10^{-7})$	$\omega_2(10^{-7})$	$\alpha_1(10^{-3})$	$\alpha_2(10^{-3})$	$\beta_1(10^{-3})$	$\beta_2(10^{-3})$	$\beta_0(10^{-3})$
1000	1.82	14.80	31.31	2.79	5.19	9.22	28.64	5.45
3000	12.86	6.59	8.24	3.12	2.78	3.77	4.93	19.55
5000	6.23	2.68	4.95	1.28	1.54	1.21	3.31	9.15
7000	1.26	1.73	2.85	0.56	0.55	0.99	1.71	0.65
10000	1.35	1.25	3.01	0.45	0.66	0.44	2.25	1.14
20000	3.25	0.91	0.94	0.59	0.50	0.45	0.50	4.62

Table 4.3: The relative absolute biases with different sample size (%)

K_2	$\rho_{1,2}$	ω_1	ω_2	α_1	α_2	β_1	β_2	β_0	total RAB
1000	0.43	50.51	38.98	3.65	9.36	1.08	3.37	1.23	108.63
3000	3.08	22.49	10.25	4.08	5.02	0.44	0.58	4.42	50.36
5000	1.49	9.14	6.16	1.68	2.77	0.14	0.39	2.07	23.84
7000	0.30	5.89	3.55	0.73	0.99	0.12	0.20	0.15	11.93
10000	0.32	4.28	3.74	0.58	1.19	0.05	0.26	0.26	10.69
20000	0.78	3.11	1.18	0.77	0.91	0.05	0.06	1.04	7.90

Unlike the one defined in Liu (2011), the mean of the estimates $\bar{\hat{\theta}}$ is replaced by the true value θ_0. The numeric results are shown in Table 4.4. In this table, the reduction in RMSE is dramatic when a larger sample is used to estimate the parameters. In terms of the RMSE, it also implies that a larger sample size will lead to a more accurate point estimate in most cases.

The histograms in Figure 4.5 and 4.6 both have nice bell shapes, but there are not enough evidence to confirm the normal distribution proved in Chapter 3. More convincing methods are used to check the normality in both graphical and statistical way. Instead of the raw estimate $\hat{\boldsymbol{\theta}}^{(i)}$, the distribution of the rescaled estimates in Section 3.3 is studied because $\sqrt{K_2}(\hat{\boldsymbol{\theta}}_{K_2} - \boldsymbol{\theta}_0)$ is more informative.

The graphical tests include quantile-quantile(Q-Q) plots and the density plots for each

Table 4.4: RMSE of the estimates (rounded to two decimal digits)

K_2	$\rho_{1,2}(10^{-3})$	$\omega_1(10^{-12})$	$\omega_2(10^{-12})$	$\alpha_1(10^{-4})$	$\alpha_2(10^{-4})$	$\beta_1(10^{-4})$	$\beta_2(10^{-4})$	$\beta_0(10^{-2})$
1000	17.78	21.74	104.13	4.16	3.54	25.03	109.18	4.11
3000	5.72	3.11	6.46	1.44	1.33	3.04	6.62	1.39
5000	3.21	1.41	3.56	0.73	0.72	1.51	3.81	0.76
7000	2.49	0.92	2.25	0.54	0.50	1.14	2.62	0.59
10000	1.87	0.71	1.52	0.40	0.37	0.73	1.64	0.44
20000	0.86	0.37	0.65	0.21	0.17	0.40	0.78	0.21

parameter. To save space, only the results for 4 parameters are included here despite the fact that the number of parameters is 8. The change in the skewness and bias of $\rho_{1,2}$ and β_0, which was shown in Figures 4.3 and 4.4, needs to be addressed. Other than these two parameters, ω_1 and β_1 are chosen to represent the estimates in both extremely small and relatively large scales. Figures 4.7 to 4.10 are the normal Q-Q plot of $\rho_{1,2}, \omega_1, \beta_1$ and β_0. The qqnorm plot of ω_1 has few extreme values at the tail when $K_2 = 1000$ but this gradually changes when K_2 increases to 20000. For the same 4 parameters, the densities are plotted in Figure 4.11 to 4.14. A black solid line represents the kernel density estimation and the blue dashed line shows the referenced normal density with the same mean and variance. The kernel densities are far from the referenced normal densities with a higher peak when $K_2 = 1000$. With the largest sample size 20000, the kernel estimations are close to the reference lines both in the centers and the tails.

From the numeric point of view, the distribution of $\sqrt{K_2}(\hat{\boldsymbol{\theta}}_{K_2} - \boldsymbol{\theta}_0)$ stabilized to the reference normal distribution and the kurtosis of kernel estimation is shown in Table 4.7 along with the skewness. The kurtosis of all parameters becomes stable around 3 which is the standard for normal distribution. There are several normality tests implemented in R, some of them are very sensitive to the tails or the size of the data. Though the distributions of the rescaled estimates are approaching normal as the sample size increases, there is no reason to expect these estimates to pass the normality tests even with the largest sample size 20000.

The assessments in the previous paragraphs show us that the distribution of $\sqrt{K_2}(\hat{\boldsymbol{\theta}}_{K_2} - \boldsymbol{\theta}_0)$ is approaching a normal distribution with mean $\mathbf{0}$. The first 6 rows in Table 4.5 show the rescaled standard deviations calculated from the Monte Carlo simulation. The rescaled standard deviations of the Monte Carlo simulation are relatively stable except for the case when $K_2 = 1000$. The last question we are interested in is that whether the limiting standard deviation is determined by $J^{-1}VJ^{-1}$ as stated in Section 3.3. Both the first and second derivatives of $l_t(\boldsymbol{\theta})$ are extremely hard to evaluate since they are given by very complex iterative formulas in the proof of Theorem 3.3.1.

By the definition of J in Section 3.3, it is a symmetric matrix. The component in row i and column j of this matrix can be computed by the limit of $\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j}$ when $n \to \infty$, which can be further simplified as $\frac{\partial^2 L_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j}$. The numeric methods can be used to estimate the second derivatives of the negative target function for a simulated path. Therefore, a numeric estimation of the matrix J is easily obtained. In the meantime, the elements in V can be approximated by the expression $-\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_i}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_j}$. However, this expression requires us to compute the numeric first derivatives of l_t at all time points, which means it can not be further simplified to the numeric derivatives of the overall target function L_n. After getting the numeric estimate of J and V, the diagonal elements of $J^{-1}VJ^{-1}$ can be computed, and they can be treated as the approximation of the asymptotic variance. It is worth to note that the numeric estimates of J and V are calculated at $\hat{\boldsymbol{\theta}}$ instead of $\boldsymbol{\theta}_0$ since the likelihood function reaches its maximum at $\hat{\boldsymbol{\theta}}$ given a path, not at $\boldsymbol{\theta}_0$. The estimated estimated theoretical asymptotic standard deviations are included in the last row of Table 4.5.

It is hard to conclude anything from the rescaled standard deviations of the Monte Carlo simulation. As explained in the table of the absolute bias, the value of one parameter may get closer to the asymptotic standard deviation when the value of another parameter is further away from the asymptotic standard deviation as the sample size

Table 4.5: SD of $\sqrt{K_2}(\hat{\boldsymbol{\theta}}_{K_2} - \boldsymbol{\theta}_0)$ and estimated asymptotic SD (rounded to two decimal digits)

K_2	$\rho_{1,2}$	$\omega_1(10^{-5})$	$\omega_2(10^{-4})$	$\alpha_1(10^{-1})$	$\alpha_2(10^{-1})$	$\beta_1(10^{-1})$	β_2	β_0
1000	4.22	13.99	3.07	6.40	5.72	15.56	3.18	6.42
3000	4.09	8.97	1.32	6.36	6.14	9.33	1.38	6.36
5000	3.98	8.17	1.29	6.00	5.92	8.66	1.36	6.15
7000	4.18	7.91	1.23	6.16	5.90	8.90	1.35	6.43
10000	4.33	8.31	1.19	6.28	6.02	8.55	1.26	6.66
20000	4.12	8.48	1.14	6.45	5.80	8.89	1.25	6.39
Asy	4.15	8.36	1.20	6.39	5.95	8.90	1.27	6.45

Table 4.6: RASY with different sample sizes (%)

K_2	$\rho_{1,2}$	ω_1	ω_2	α_1	α_2	β_1	β_2	β_0	total RASY
1000	1.56	67.33	155.62	0.15	3.87	74.82	150.37	0.46	454.17
3000	1.62	7.25	9.60	0.48	3.25	4.90	9.02	1.27	37.40
5000	4.12	2.23	7.19	6.12	0.42	2.69	7.21	4.55	34.54
7000	0.59	5.39	2.54	3.61	0.81	0.01	6.16	0.19	19.31
10000	4.18	0.55	0.61	1.71	1.16	3.97	0.78	3.26	16.23
20000	0.77	1.42	5.47	1.04	2.51	0.09	1.60	0.85	13.76

increases. Thus, a new measure called the relative absolute difference with respect to the asymptotic standard deviation can be defined as

$$\text{RASY of } \theta = \frac{|\text{ SD of } \sqrt{K_2}(\hat{\theta}_{K_2} - \theta_0) - \text{ asymptotic standard deviation}|}{\text{asymptotic standard deviation}}.$$

This measure uses the asymptotic standard deviation as a standard to compare the performance of different sample sizes. Using this measure, Table 4.5 is converted to Table 4.6. From the values in the last column, the total RASY is decreasing when the sample size becomes larger. Overall, the standard deviations obtained from the Monte Carlo simulation is getting closer to the asymptotic ones as the sample size increases.

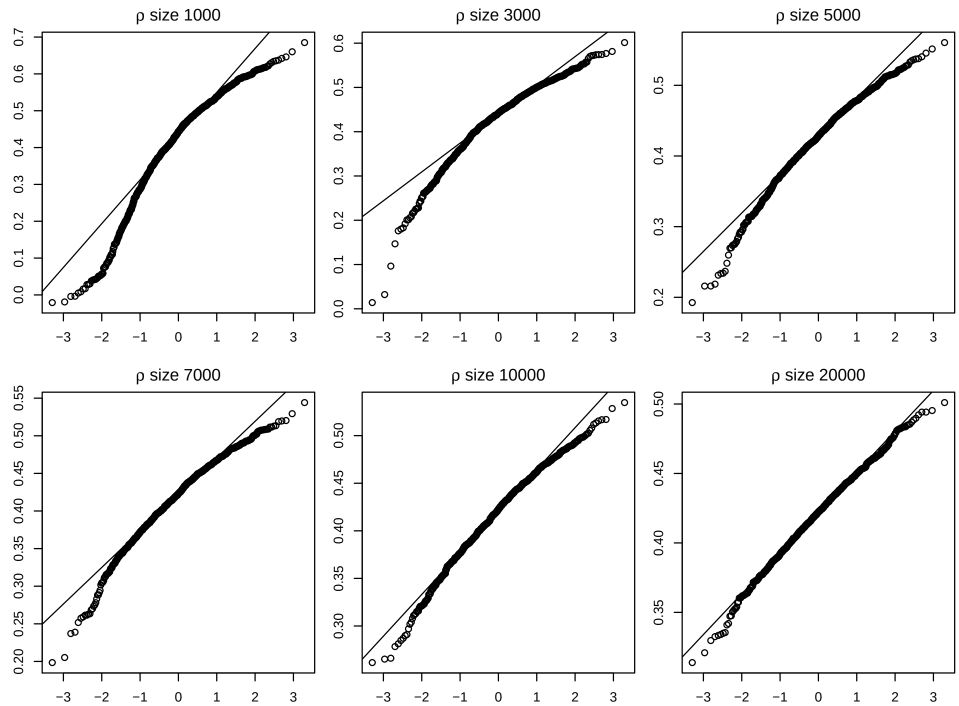

Figure 4.7: qqnorm of rescaled $\hat{\rho_{1,2}}$ with different sample sizes K_2

The histograms, the relative absolute error and the root mean square error show that the point estimator has a higher probability to fall into a close neighbourhood of the true value when the sample size becomes larger. It is a rectification of the consistency in finite samples. In the meantime, the distribution of $\sqrt{K_2}(\hat{\boldsymbol{\theta}}_{K_2} - \boldsymbol{\theta}_0)$ is gradually moving towards a stable normal distribution with the mean and variance specified in Theorem 3.3.1.

4.4 Numeric Issues with Solutions

4.4.1 The Scale Difference

Since the parameters are typical of quite different scales, the numerical algorithms to obtain the estimation can be sensitive to this. We could take a univariate GARCH(1,1) as an example since it is one of the most commonly used models for analyzing a single sequence in the stock log returns. When a GARCH(1,1) model is fitted on a stock return, the point estimates have considerable scale differences. The scale of $\hat{\omega}$ is in the order

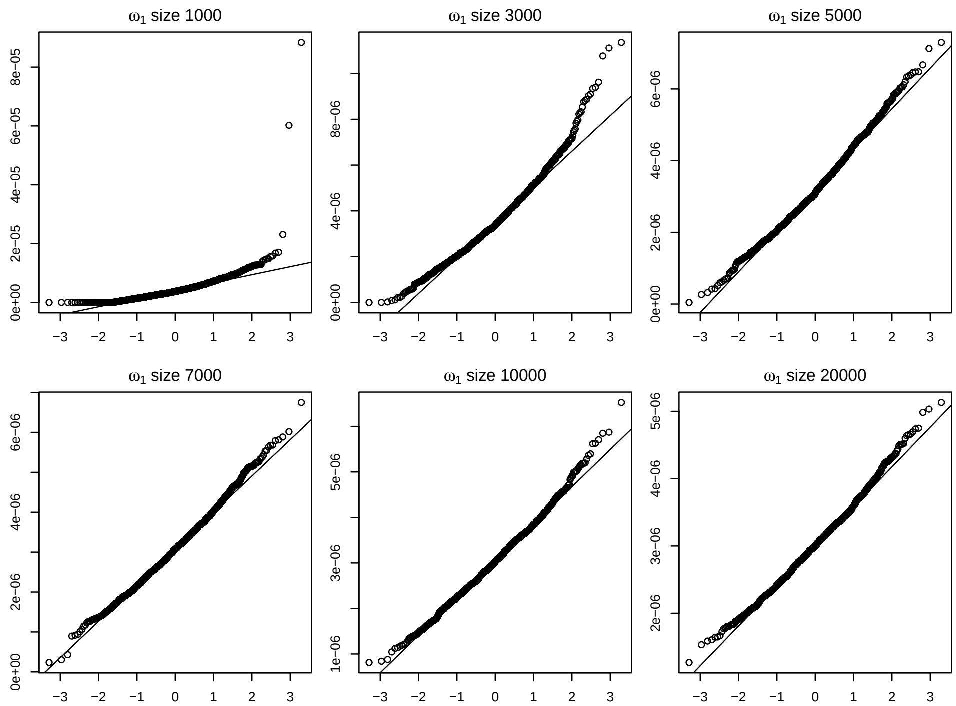

Figure 4.8: qqnorm of rescaled $\hat{\omega}_1$ with different sample sizes K_2

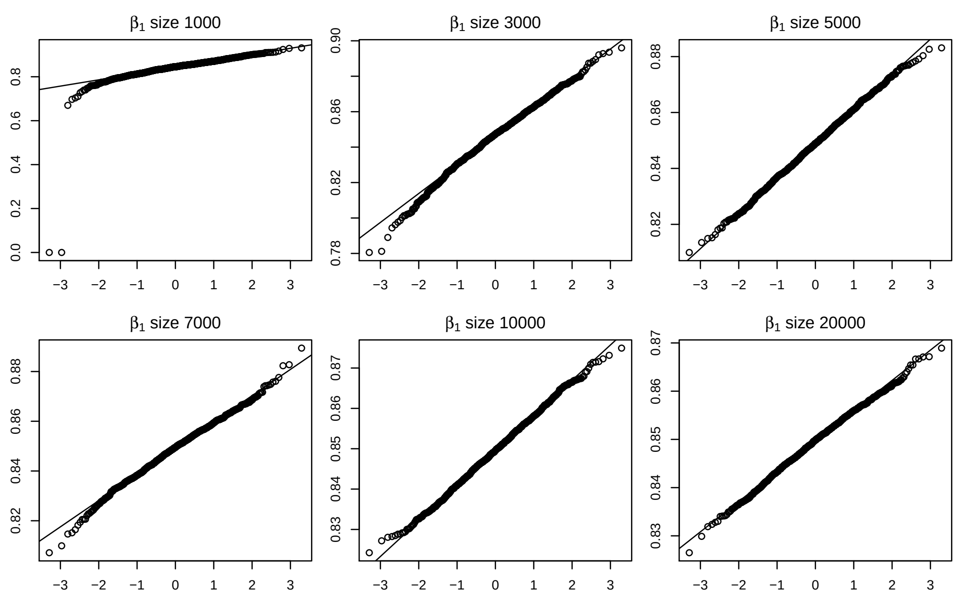

Figure 4.9: qqnorm of rescaled $\hat{\beta}_1$ with different sample sizes K_2

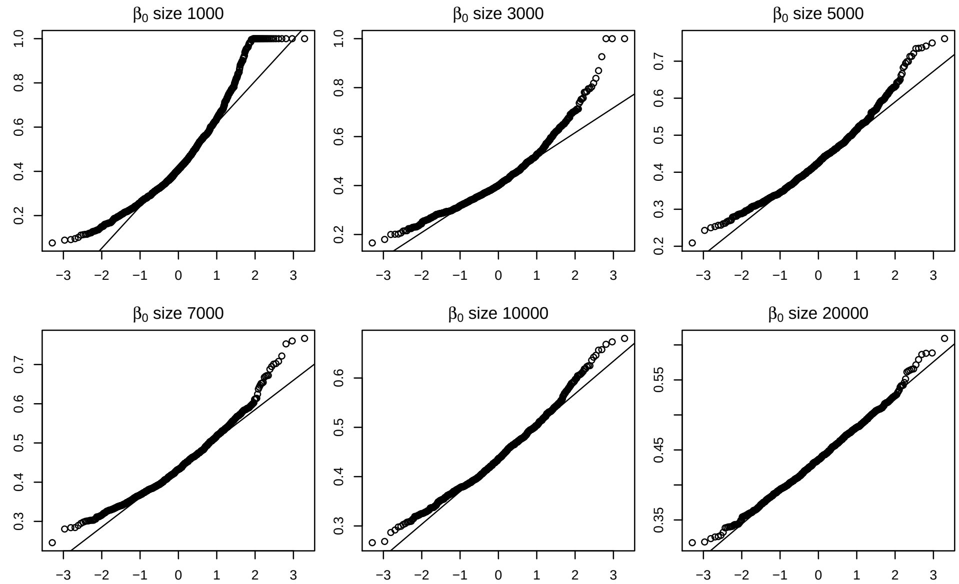

Figure 4.10: qqnorm of rescaled $\hat{\beta}_0$ with different sample sizes K_2

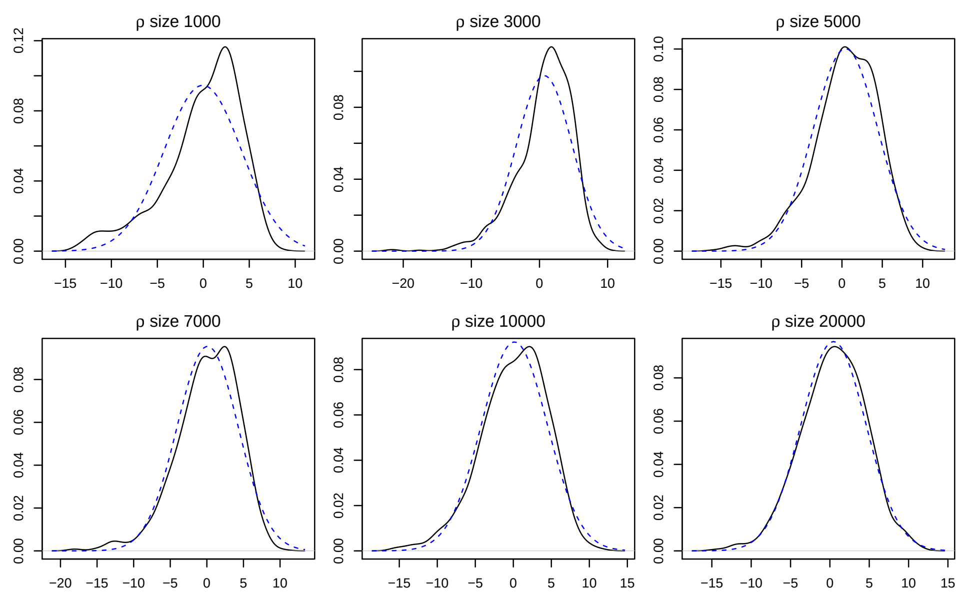

Figure 4.11: The black solid line: kernel density of rescaled $\hat{\rho}_{1,2}$ with different sample sizes K_2. The blue dashed line: standard normal using the same mean and sd from the estimates

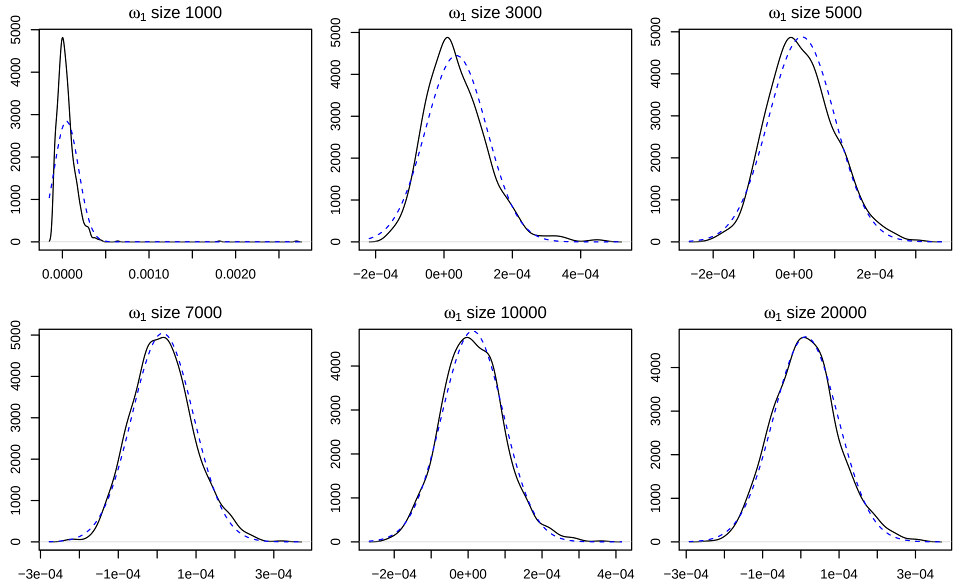

Figure 4.12: The black solid line: kernel density of rescaled $\hat{\omega}_1$ with different sample sizes K_2. The blue dashed line: standard normal using the same mean and sd from the estimates

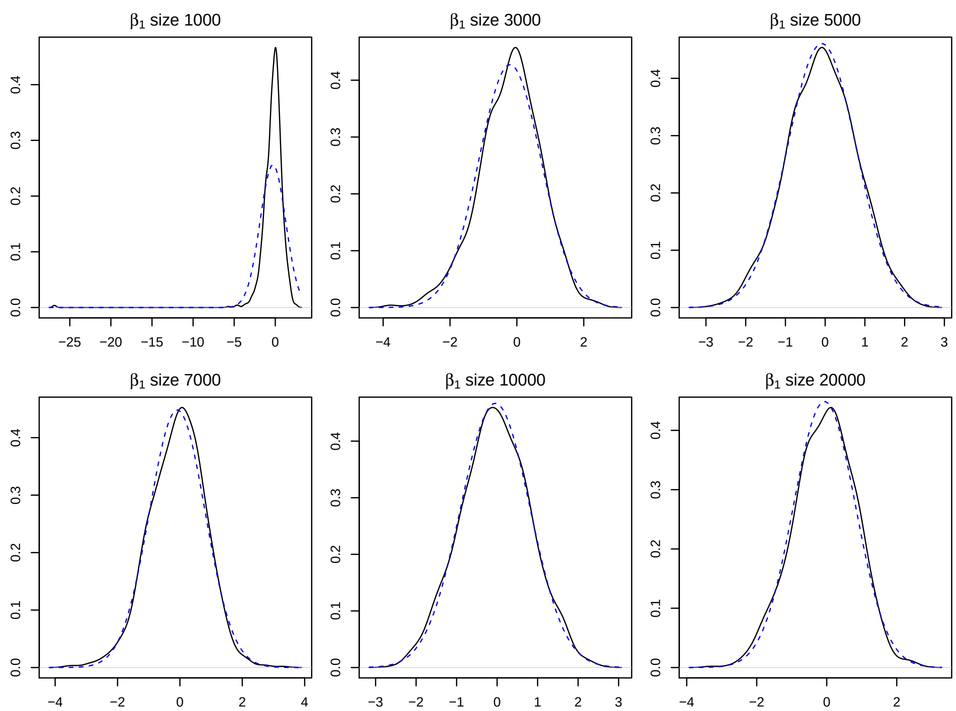

Figure 4.13: The black solid line: kernel density of rescaled $\hat{\beta}_1$ with different sample sizes K_2. The blue dashed line: standard normal using the same mean and sd from the estimates

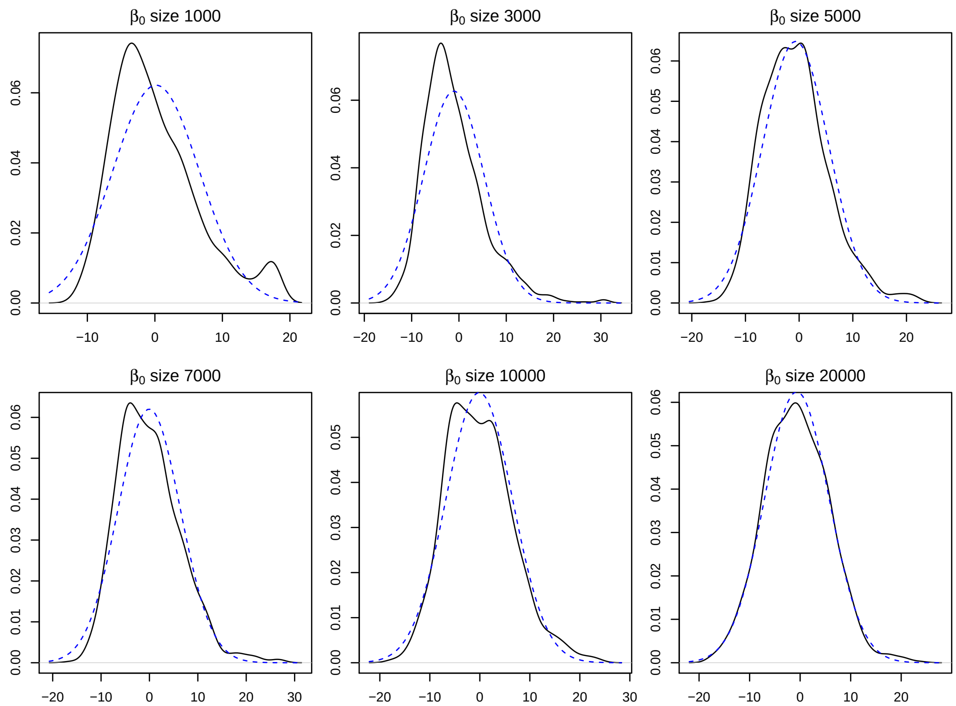

Figure 4.14: The black solid line: kernel density of rescaled $\hat{\beta}_0$ with different sample sizes K_2. The blue dashed line: standard normal using the same mean and sd from the estimates

Table 4.7: Kurtosis (skewness) of $\sqrt{K_2}(\hat{\boldsymbol{\theta}}_{K_2} - \boldsymbol{\theta}_0)$ (rounded to one decimal digits)

K_2	$\rho_{1,2}$	ω_1	ω_2	α_1	α_2	β_1	β_2	β_0
1000	3.7(−1.0)	156.7(9.2)	42.9(5.4)	3.0(0.3)	3.3(0.3)	171.8(−10.2)	43.9(−5.7)	3.5(0.9)
3000	5.4(−1.1)	4.5(0.8)	3.4(0.6)	3.0(0.1)	3.1(0.1)	3.4(−0.3)	3.8(−0.5)	5.4(1.2)
5000	3.7(−0.7)	3.0(0.3)	3.7(0.7)	3.1(0.0)	3.1(0.1)	2.8(−0.1)	3.9(−0.5)	3.7(0.7)
7000	4.0(−0.7)	3.1(0.3)	3.4(0.5)	3.4(−0.1)	3.0(0.1)	3.5(−0.2)	3.6(−0.4)	3.9(0.7)
10000	3.3(−0.5)	3.4(0.3)	3.5(0.3)	3.0(−0.0)	2.8(0.1)	2.8(−0.0)	3.1(−0.1)	3.3(0.5)
20000	3.2(−0.3)	3.1(0.3)	3.0(0.3)	3.1(0.1)	3.1(0.0)	3.0(−0.2)	3.0(−0.2)	3.2(0.2)

of 10^{-6} to 10^{-5} while $\hat{\alpha}$ and the scale of $\hat{\beta}$ is in the order of 10^{-1}. This leads one to imagine that the estimate from our model has a big difference on the scales of $\boldsymbol{\omega}_i$ and other parameters. In a more vivid picture, the optimization function begins the searching in the neighbourhood of the starting value and finds a direction with the deepest descent and moves in that direction with one step. Normally, the scale is related to the step size and the step size would not differ much if everything is kept as the default.

There is an argument in R function `nlminb` called `scale` which can be used to adjust the step size for each parameter and the default is 1. The larger the scale is, the smaller the step size is. The scale of ω_i's is changed to a larger value while the scales of other parameters remain as 1. Since no one knows what a proper scale is for ω_i's, the optimization function is fed with a vector of alternative scales. The default argument in the function is set to be $a1 = (0, 1, 2, 3, 4, 5, 6, 7)$, which means the possible scales for ω_i's are 10^{a1}. The function will start optimizing the negative log likelihood with the scale in the middle of $a1$ and if the number of elements in $a1$, n_{a1}, is even, it will start with the smaller one in the middle. Therefore, the scale will start from 10^3 in the default case and search into the two ends of the vector 10^{a1} until the target function convergence numerically.

4.4.2 Computational Speed

Since the likelihood is built based on the conditional distributions, the likelihood value needs to be calculated at each time point sequentially given a group of values for the parameters. The inverse matrix operation needs to be performed at each time point. The computation speed is a major problem when the maximum likelihood estimator is desired.

Better than most of other statistical and graphical software, R provides a completely programmable language for graphics, which makes the graphical capabilities of R extraordinary. As a scripting language, the computation speed is one of its main drawbacks. In the meantime, C++ is a relatively low-level compiled language. It is an object-oriented language with only some well-developed packages, which requires a higher cost in terms of coding. To combine the advantage of both languages, Eddelbuettel and Sanderson (2014) create a package `RcppArmadillo` which provides an interface to integrate C++ code within R. Since the optimization is a difficult job, it is not wise to code this part by ourselves in C++. Thus, only the negative log likelihood function is written in C++ and the rest part of the job is done in R code.

The computation speed of repeatedly computing the likelihood written in R and Rcpp was assessed. For simplicity, the assessment was running in a two-dimensional case. The computation is based on the centered log returns of the daily closing prices between 1995-1-1 to 2007-12-31, which is the same time range as Section 4.2. The closing price of IBM and BAC from the technology and the finance sector are used in this two-dimensional example. Both programs ran for 1000 times to compute the negative likelihood function with everything fixed. Figure 4.15 summarizes the time spent on 1000 computation as a violin plot. Note that the computation time of the Rcpp program is way faster than the pure R program. It takes about 690 milliseconds to run the R program on average while it only takes 9.5 milliseconds to do the same thing in Rcpp. It definitely has a huge improvement when the target function is written in Rcpp.

4.4.3 Initial values and Starting Point

The initial value is $(\tilde{\boldsymbol{x}}_0, \tilde{\boldsymbol{\sigma}}_0)$ which is needed in both the estimation and simulation processes and the starting point is an argument in the estimating function which is used to feed an initial value for the parameters to be optimized. The initial value is not important asymptotically because of the ergodicity, but it does have some impact on the estimate in the finite sample case. The longer the time series is, the smaller the difference will be in the estimates. There is no way we can find a "perfect" initial value. One can choose any value as long as it is a possible value. A good initial value means higher computation efficiency and converges more quickly and it is common to believe that a value within

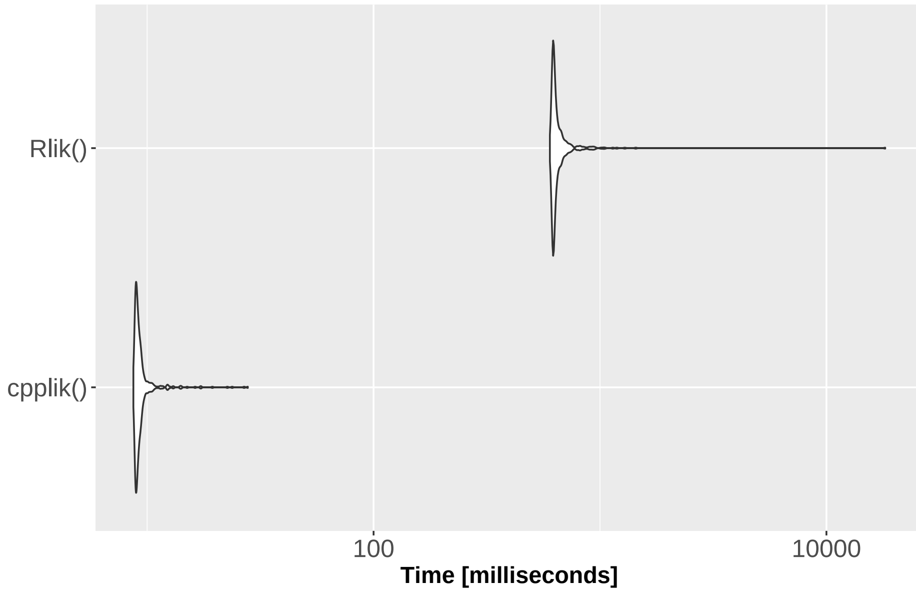

Figure 4.15: Violin plot for 1000 computation time using the same target function written in R and Rcpp

the high probability region is a good choice. Like the common choice in Zivot (2009), the default initial value is generated based on the data. For m dimensional time series with n observations, the default initial $\tilde{\boldsymbol{x}}_0$ is set to be $\boldsymbol{x}_1$ and the individual σ terms to be the mean square of each sequence. Therefore, $\tilde{\sigma}^2_{0,l} = \frac{1}{n}\sum_{t=1}^{n} x^2_{l,t}$ for $l = 1, \ldots, m$.

The choice of starting point is much more important than the initial value. Since this is a multivariate optimization problem and the surface of the likelihood function is really flat, the starting value determines the neighbourhood the optimization function looks into. Within the searching region, there are multiple local minimum points such that the function `nlminb` in R outputs them as the numeric convergences are reached.

In all the studies in this chapter, the value of $(\tilde{\boldsymbol{x}}_0, \tilde{\boldsymbol{\sigma}}_0)$ is set as the default mentioned a few lines above and $\boldsymbol{\theta}_{start}$ is chosen based on the estimated values from individual GARCH(1,1) model. The starting values of all $\rho_{i,j}$'s are set to be 0 since $\boldsymbol{\Sigma}$ is guaranteed to be positive definite. The starting values of ω_i, α_i and β_i are chosen as $\hat{\omega}_{ind,i}$, $\frac{\hat{\alpha}_{ind,i}}{2}$ and $\frac{\hat{\beta}_{ind,i}}{2}$ where $\hat{\omega}_{ind,i}, \hat{\alpha}_{ind,i}, \hat{\beta}_{ind,i}$ are the point estimates from univariate GARCH(1,1) models.

The importance of starting point can be seen in another highly related problem. The

Table 4.8: Estimates and corresponding values of the negative likelihood from Windows and Linux system: $\hat{\boldsymbol{\theta}}_1$ is the estimate from Windows system and $\hat{\boldsymbol{\theta}}_2$ is the estimate from Windows system.

	$\rho_{1,2} * 10$	$\omega_1(10^{-7})$	$\omega_2(10^{-7})$	$\alpha_1(10^{-2})$	$\alpha_2(10^{-2})$	$\beta_1(10^{-1})$	$\beta_2(10^{-1})$	$\beta_0(10^{-3})$	NLL(10^3)
$\hat{\boldsymbol{\theta}}_W$	−1.88	8.86	11.44	3.64	3.30	9.30	9.47	306.48	−58.64
$\hat{\boldsymbol{\theta}}_L$	2.73	11.85	17.33	3.86	4.68	9.58	9.51	0.12	−58.58

estimates under different operating systems are different. In Table 4.8, everything is the same including the observed data, initial value, starting value and the negative log likelihood function `NLL`. Surprisingly, the estimated values from Windows and Linux systems are different. The difference between the estimates may be considered as rounding error or precision problem except $\hat{\rho}$ and β_0. From the value of the target function, the estimate from Windows system outperforms the one from Linux system which one would never expect. This disagreement in the estimates starts from July 2017 and the estimates are the same prior to that. The function `nlminb` calls some low-level C functions, so an updated C library could cause such a difference. However, there is no clue what is changed inside the C library on Linux systems. This disagreement is also an evidence to show that the likelihood surface is flat.

A possible solution is proposed by adding another convergence criteria to the algorithm other than the default ones in `nlminb`. It does not make sense to do anything with the initial value since they are data oriented. The reasonable change needs to be done with the starting point. An iteration method is used to update the starting value. The convergence tag outputted from `nlminb` has two possible values, 1 means not converged and 0 means converged. The proposed criteria include the steps below.

1. Input the initial values $(\tilde{\boldsymbol{x}}_0, \tilde{\boldsymbol{\sigma}}_0)$ and the starting point $\boldsymbol{\theta}_{start}$ as well as the possible scales for ω_i's. Select a small tolerance *tol* as the break trigger and a maximum number of iteration $\boldsymbol{I}$. Set the iteration counter as 1.

2. The output using `nlminb` includes the estimated value $\hat{\boldsymbol{\theta}}_1$, the convergence tag

Table 4.9: Estimates and corresponding values of the negative likelihood from Windows and Linux system using $\boldsymbol{\theta}_{start,a}$ and $\boldsymbol{\theta}_{start,b}$.

	$\rho_{1,2} * 10$	$\omega_1(10^{-7})$	$\omega_2(10^{-7})$	$\alpha_1(10^{-2})$	$\alpha_2(10^{-2})$	$\beta_1(10^{-1})$	$\beta_2(10^{-1})$	$\beta_0(10^{-3})$	NLL(10^3)
$\boldsymbol{\theta}^{win}_{start,a}$	−1.88	8.86	11.44	3.64	3.30	9.30	9.47	306.48	−58.64
$\boldsymbol{\theta}^{win}_{start,b}$	−1.88	8.86	11.44	3.64	3.30	9.30	9.47	306.48	−58.64
$\boldsymbol{\theta}^{Lin}_{start,a}$	−1.88	8.86	11.44	3.64	3.30	9.30	9.47	306.48	−58.64
$\boldsymbol{\theta}^{Lin}_{start,b}$	−1.88	8.86	11.44	3.64	3.30	9.30	9.47	306.47	−58.64

Tag_1 and the likelihood value L_1. The selected scale power $Scale_1$ for ω_i's are also outputted.

3. Set the starting point $\boldsymbol{\theta}_{start} = \hat{\boldsymbol{\theta}}_1$, optimize the target function again. The group of values is outputted $\{\hat{\boldsymbol{\theta}}_2, Tag_2, L_2, Scale_2\}$. Update the counter by adding 1 to the current value.

4. Create a scale vector S such that it has the same length as the parameter vector, and initialize the elements as 1. Change the elements for ω_i's to 10^{Scale_2}.

5. If the summation of $|\hat{\boldsymbol{\theta}}_2 - \hat{\boldsymbol{\theta}}_1| * S$ is smaller than tol, the iteration is over. Otherwise, update $\hat{\boldsymbol{\theta}}_1 = \hat{\boldsymbol{\theta}}_2$, and go back to Step 3 until the counter reaches I.

6. If $Tag_2 = 0$ or $Tag_1 = 1$, output $\hat{\boldsymbol{\theta}}_2$, Tag_2, L_2, $Scale_2$. Otherwise output $\hat{\boldsymbol{\theta}}_1$, Tag_1, L_1, $Scale_1$.

The output is considered as a converged estimate if the maximum number of iteration is not reached. The same data, initial value that produced Table 4.8 are used to generate an updated example. The estimating results using the iteration method with two different starting values are shown in Table 4.9. The first starting point $\boldsymbol{\theta}_{start,a}$ uses what was described in the last paragraph and the second one $\boldsymbol{\theta}_{start,b}$ equals to $(0, 0.00001, 0.00001, 0.1, 0.1, 0.5, 0.5, 0.1)$.

From the table, two starting points lead to the same optimization result in both

Windows and Linux systems after applying the iteration method. Any reasonable starting point can be used in the algorithm since the effect of starting point becomes minimum.

Chapter 5

Concluding Remarks

In this thesis, a multivariate time series with a GARCH type structure has been proposed to address the common risk within multiple selected stock returns. Although it has been defined in an implicit form, the stationary and ergodic parameter space exists using the T-chain theory in Meyn and Tweedie (2009). The asymptotic theories of the quasi maximum likelihood estimator have been provided for the general model, including the consistency and asymptotic normality.

The geometrically ergodic theorem has been presented in Chapter 2. The initial values can be any possible state in the space since it has a positive chance to eventually get to all other states from the initial state. Jeantheau (1998) and Aue et al. (2009) show us a sufficient condition to control the stochastic process depending on the top Lyapunov exponent for the strict stationarity of CCC-GARCH models. The condition $A4$ used to control the drift in this model was much more complicated since the matrix norm of the partial derivative matrix $\boldsymbol{B}$ is not trackable in an implicit formula. It is possible to explore the stationary and ergodic theory further since the assumptions in this thesis are the sufficient conditions but not the necessary ones. The true space can be much larger than what has been studied in Chapter 2. The practical verification of Assumption $A4$ was done with a truncated state space within the sensible range. A deeper understanding of the process is needed to find a systematic method to check the assumption.

The Gaussian quasi likelihood function has been obtained by assuming that the innovations are i.i.d normally distributed. Under Assumptions $A1 - A5$ in Sections 2 and 3.2.1, the quasi maximum likelihood (QML) has become a sensible method to estimate the parameters in practice since the ergodicity and identifiability conditions were satisfied. The most desired statistical properties of the estimator (i.e. consistency and asymptotic normality) were wanted to ensure that the QMLE is approaching the true parameter value as the sample size increases no matter the actual innovation is consistent with Gaussian distribution or not. In addition to the assumptions for stationarity and ergodicity, the finite 8th moment plus some regular conditions have led to both desired asymptotic properties. It has a really high chance that the moment condition can be relaxed to the 6th moment instead of 8th as Hafner and Preminger (2009b) proved for multivariate GARCH models.

A parameter value satisfying all the conditions has been used to study the QMLE numerically. In Monte Carlo simulations, the simulated processes behaved similarly to a long memory time series. The numeric issues in R have been solved when the best nonlinear multivariate optimization function `nlminb` was used. The proposed solutions were useful, but the computation was time-consuming, there may exist certain ways to obtain the initial value and starting point such that a higher efficiency can be achieved. Despite our effort, we still do not have a good handle on the parameter space, or how widely the assumptions are met by the financial data. In the simulation example, the estimate typically converges with a sensible result when we apply the algorithm to a path with sample size 1000 (about 4 years daily data). With the largest sample size in our study, 20000 is unrealistic in practice. The algorithm needs a relatively large sample size in order to get a good estimate. Nevertheless, a long log return sequence in the stock market could have a structural change which will violate the stationarity assumption. A study on the sample size is needed in the future.

The financial application of this model needs to be addressed in the future with a comparison with the classical models. One possible application is to use this model to manage portfolios. This model will allow us to study different stocks which have comovements. The log returns might be highly correlated in certain periods, while they may be uncorrelated in other periods. Better insights into the dynamic correlation based on this model could affect the results of the portfolio optimization. If a portfolio only consists of fixed income securities and equities, the model we proposed here could be used to determine the allocation of the weights on them. The common risk term can reflect the shock within the series directly, which can be used as an indicator to provide guidance to adjust the weights on the equities and fixed income securities of the investment portfolio. More weights will be moved onto the fixed income securities when the common risks of the stocks invested rise sharply. This model setup could change the results in the portfolio optimization because of the covariance structure. The default risk model might benefit from this dynamic correlation setting as well.

Bibliography

Amemiya, T. (1985). *Advanced econometrics*. Harvard university press.

Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *Ann. Statist.*, 37(6B):4046–4087.

Bendat, J. S. and Piersol, A. G. (2010). *Random data: analysis and measurement procedures*. Wiley-Blackwell, Hoboken, N.J, 4th, new edition.

Berd, A. M., Engle, R. F., and Voronov, A. B. (2007). The underlying dynamics of credit correlations. *Journal of Credit Risk*, 3:27–62.

Berkes, I., Horvath, L., and Kokoszka, P. (2003). Garch processes: Structure and estimation. *Bernoulli*, 9(2):201–227.

Billingsley, P. (1961). Statistical methods in markov chains. *The Annals of Mathematical Statistics*, pages 12–40.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *J. Polit. Economy*, pages 637–654.

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

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *Rev. Econ. Statist.*, pages 498–505.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *J. Polit. Economy*, pages 116–131.

Bollerslev, T. and Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews*, 11(2):143–172.

Boussama, F., Fuchs, F., and Stelzer, R. (2011). Stationarity and geometric ergodicity of bekk multivariate garch models. *Stochastic Processes and their Applications*, 121(10):2331–2360.

Caporin, M. and McAleer, M. (2006). Dynamic asymmetric garch. *Journal of Financial Econometrics*, 4(3):385–412.

Carr, P. and Wu, L. (2009). Variance risk premiums. *The Review of Financial Studies*, 22(3):1311–1341.

Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory*, 18(1):17Ű39.

Christie, A. A. (1982). The stochastic behavior of common stock variances. value, leverage and interest rate effects. *J. Finan. Econ.*, 10(4):407–432.

Chu, J., Kulperger, R., and Yu, H. (2016). Modelling the common risk among equities: A multivariate time series model with an additive garch structure. In *Advanced Statistical Methods in Data Science*, pages 205–218. Springer.

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

Doob, J. L. (1990). *Stochastic processes.* Wiley, New York;Toronto;.

Eddelbuettel, D. and Sanderson, C. (2014). Rcpparmadillo: Accelerating r with high-performance c++ linear algebra. *Comput. Stat. Data Anal.*, 71:1054–1063.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, pages 987–1007.

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *J. Bus. Econ. Statist.*, 20(3):339–350.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric theory*, 11(1):122–150.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *J. Finan. Econ.*, 33(1):3–56.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *J. Finan. Econ.*, 116(1):1–22.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *The review of Economics and Statistics*, 82(4):540–554.

Francq, C. and Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli*, 10(4):605–637.

Francq, C. and Zakoian, J.-M. (2010). *GARCH models: structure, statistical inference and financial applications*. John Wiley and Sons, Chichester, West Sussex, U.K.

Hafner, C. M. and Preminger, A. (2009a). Asymptotic theory for a factor garch model. *Econometric Theory*, 25(2):336–363.

Hafner, C. M. and Preminger, A. (2009b). On asymptotic theory for multivariate garch models. *J. Multivariate Anal.*, 100(9):2044–2054.

Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617.

Jeantheau, T. (1998). Strong consistency of estimators for multivariate arch models. *Econometric Theory*, 14(01):70–86.

Jiang, W. (2011). *Asymptotic Theory of General Multivariate GARCH Models*. PhD thesis, The University of Western Ontario.

Klüppelberg, C., Lindner, A., and Maller, R. (2004). A continuous-time garch process driven by a lévy process: stationarity and second-order behaviour. *Journal of Applied Probability*, 41(3):601–622.

Kolmogorov, A. N. and Fomin, S. (1970). *Introductory Real Analysis*. New York: Dover Publications Inc.

Lee, S.-W. and Hansen, B. E. (1994). Asymptotic theory for the garch (1, 1) quasi-maximum likelihood estimator. *Econometric Theory*, 10(01):29–52.

Li, C. W. and Li, W. K. (1996). On a double-threshold autoregressive heteroskedastic time series model. *J. Appl. Econometrics*, 11(3):253.

Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector arma-garch model. *Econometric theory*, 19(2):280–310.

Liu, W. (2011). *Asymptotic Theory for GARCH-in-mean Models*. PhD thesis, The University of Western Ontario.

Lumsdaine, R. L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in igarch (1, 1) and covariance stationary garch (1, 1) models. *Econometrica: Journal of the Econometric Society*, pages 575–596.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, pages 867–887.

Meyn, S. P. and Tweedie, R. L. (2009). *Markov chains and stochastic stability*. Cambridge university press.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.

Nelson, D. B. and Cao, C. Q. (1992). Inequality constraints in the univariate garch model. *Journal of Business & Economic Statistics*, 10(2):229–235.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442.

Treynor, J. L. (1961). Market value, time, and risk. *Unpublished manuscript*, (95-209).

Treynor, J. L. (1962). Toward a theory of market value of risky assets. *Unpublished manuscript*, 40.

Tsai, H. and Chan, K.-S. (2008). A note on inequality constraints in the garch model. *Econometric Theory*, 24(3):823–828.

Tsay, R. (2010). *Analysis of Financial Time Series*. CourseSmart. Wiley.

Tse, Y. K. and Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *J. Bus. Econ. Statist.*, 20(3):351–362.

Van der Weide, R. (2002). Go-garch: a multivariate generalized orthogonal garch model. *Journal of Applied Econometrics*, 17(5):549–564.

Vasicek, O. A. (1987). *Probability of loss on loan portfolio*. Wiley Online Library.

Vrontos, I. D., Dellaportas, P., and Politis, D. N. (2003). A full-factor multivariate garch model. *The Econometrics Journal*, 6(2):312–334.

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *J. Econ. Dynam. Control*, 18(5):931–955.

Zivot, E. (2009). Practical issues in the analysis of univariate garch models. In *Handbook of financial time series*, pages 113–155. Springer.

Appendix A

Useful algebra results

1. For $x > -1$, $\log(1 + x) \leq x$.

2. Jensen's inequality. If X is a random variable and ψ is a convex function, then

$$\psi[\mathbb{E}(X)] \leq \mathbb{E}[\psi(X)].$$

3. Cauchy-Schwarz inequality

$$|E(XY)|^2 \leq EX^2 EY^2.$$

4. If $a_1, a_2, \cdots, a_n \geq 0$ and $p \geq 1$, then

$$\sum_{i=1}^{n} a_i^p \leq (\sum_{i=1}^{n} a_i)^p \leq n^{p-1} \sum_{i=1}^{n} a_i^p.$$

5. If $a_1, a_2, \cdots, a_n \geq 0$ and $0 < p < 1$, then

$$\sum_{i=1}^{n} a_i^p \geq (\sum_{i=1}^{n} a_i)^p \geq n^{p-1} \sum_{i=1}^{n} a_i^p.$$

6. If $a_1, a_2, \cdots \geq 0$ and $0 < p < 1$, then

$$\sum_{i=1}^{\infty} a_i^p \geq (\sum_{i=1}^{\infty} a_i)^p.$$

7. All matrix induced norms are equivalent. A is a m by m square matrix, p_1 and p_2 are a positive integers. $\|\cdot\|_{p_1}$ and $\|\cdot\|_{p_2}$ are two induced matrix norm, then there are positive constants l_1 and l_2 such that

$$l_1 \|A\|_{p_1} \leq \|A\|_{p_2} \leq l_2 \|A\|_{p_1}.$$

8. A and B are m by m square matrices, and p is a positive integer,

$$|tr(AB)| \leq m \|A\|_2 \|B\|_2.$$

9. A, B and C are m by m square matrices,

$$tr(ABC) = tr(BCA) = tr(CAB).$$

10. $\sum_{n=0}^{\infty} x^n = \dfrac{1}{1-x}$ if $|x| < 1$.

11. $\sum_{n=0}^{\infty} nx^n = \dfrac{x}{(1-x)^2}$ if $|x| < 1$.

12. $\sum_{n=0}^{\infty} n^2 x^n = \dfrac{x(1+x)}{(1-x)^3}$ if $|x| < 1$.

13. $\sum_{n=0}^{\infty} n^3 x^n = \dfrac{x(1+4x+x^2)}{(1-x)^4}$ if $|x| < 1$.

14. Holder's Inequality: Suppose that X and Y are two random variables, and $p, q > 1$ satisfy $1/p + 1/q = 1$

$$\mathbb{E}XY \leq \mathbb{E}|XY| \leq (\mathbb{E}|X|^p)^{1/p}(\mathbb{E}|Y|^q)^{1/q}.$$

15. Minkowski's Inequality: Suppose that X and Y are two random variables, and $1 \leq p < \infty$. Then

$$(\mathbb{E}|X+Y|^p)^{1/p} \leq (\mathbb{E}|X|^p)^{1/p} + (\mathbb{E}|Y|^p)^{1/p}.$$

16. If $0 < p < 1$, X and Y are two random variables, then

$$\mathbb{E}|XY - \tilde{X}\tilde{Y}|^p \leq \sqrt{\mathbb{E}|X|^{2p}\mathbb{E}|Y-\tilde{Y}|^{2p}} + \sqrt{\mathbb{E}|\tilde{Y}|^{2p}\mathbb{E}|X-\tilde{X}|^{2p}}.$$

17. If $0 < p < 1$, X, Y, Z are three random variables, then

$$\mathbb{E}|XYZ - \tilde{X}\tilde{Y}\tilde{Z}|^p \leq \sqrt{\mathbb{E}|XY|^{2r}\mathbb{E}|Z-\tilde{Z}|^{2r}} + \sqrt{\mathbb{E}|X\tilde{Z}|^{2p}\mathbb{E}|Y-\tilde{Y}|^{2p}} + \sqrt{\mathbb{E}|\tilde{Y}\tilde{Z}|^{2p}\mathbb{E}|X-\tilde{X}|^{2p}}$$

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

19. If $x \in \mathbb{R}$ and $A(x)$ is a matrix that the elements are functions of x, $\dfrac{dA(x)^{-1}}{dx} = -A(x)^{-1}\dfrac{dA(x)}{dx}A(x)^{-1}$.

20. If $|X| > 0$, $\dfrac{\partial \log |X|}{\partial X} = (X^{\intercal})^{-1}$.

Appendix B

Some Definitions in Markov Chain

Definition B.1 (T-Chains)

If $\mathbf{\Phi}$ is a Markov Chain for which there exists a sampling distribution a such that K_a possesses a continuous component T, with $T(x, X) > 0$ for all x, then $\mathbf{\Phi}$ is called a T-chain.

Definition B.2 (Minimal Sets)

A set M is called minimal for the deterministic control model CM(F), if it is (topological) closed, invariant, and does not contain any closed invariant set as a proper subset.

Definition B.3 (ψ-irreducible)

A chain $\mathbf{\Phi} = \{\Phi_t\}$ is called ψ-irreducible if there exists a measure ψ in $\mathcal{B}(X)$ such that, whenever $\psi(A) > 0$, we have $L(x, A) = P_x(\tau_A < \infty) > 0$ for all $x \in X$.

Definition B.4 (M-irreducible Control Models)

If CM(F) is indecomposable and also possesses a minimal set M, then CM(F) will be called M-irreducible.

Definition B.5 (Petite Set)

A set $C \in \mathcal{B}(X)$ is ν_a-petite if the sampled chain satisfies the bound

$$K_a(x, B) \geq \nu_a(B),$$

for all $x \in C, B \in \mathcal{B}(X)$*, where* ν_a *is a non-trivial measure on* $\mathcal{B}(X)$*.*

Definition B.6 (Period and Aperiod Chains)

Suppose that $\mathbf{\Phi}$ *is a* ψ*-irreducible Markov Chain. The largest* d *for which a* d*-cycle occurs for* $\mathbf{\Phi}$ *is called the period of* $\mathbf{\Phi}$*.*

When $d = 1$*, the chain* $\mathbf{\Phi}$ *is called aperiod.*

Definition B.7 (Harris Recurrence)

Define the occupation time random variable $\eta_A := \sum_{t=1}^{\infty} \mathbb{1}\{\mathbf{\Phi} \in A\}$*. For* $x \in X$*,* $A \in \mathcal{B}(X)$*, we consider the event that* $\mathbf{\Phi} \in A$ *infinitely often and define*

$$Q(x, A) := P_x(\mathbf{\Phi} \in A \quad i.o.).$$

The set A *is called Harris recurrent if*

$$Q(x, A) = P_x(\eta_A = \infty) = 1, \quad x \in A.$$

A chain $\mathbf{\Phi}$ *is called Harris recurrent if it is* ψ*-irreducible and every set in* $\mathcal{B}^+(X)$ *is Harris recurrent.*

Definition B.8 (Positive Recurrence)

Define the hitting time random variables $\tau_A := inf\{t \geq 1 : \Phi_t \in A\}$*. For* $x \in X$*,* $A \in \mathcal{B}(X)$*, we consider the expected hitting time* $\mathbb{E}_x \tau_A$*. If* $\mathbb{E}_x \tau_A < \infty$*, we say the set* A *is positive recurrent. A chain* $\mathbf{\Phi}$ *is called positive recurrent if it is* ψ*-irreducible and every set in* $\mathcal{B}^+(X)$ *is positive recurrent.*

Appendix C

Other Mathematical definitions

Definition C.1 (Converge in probability)

The sequence of random variables $X_1, X_2, \ldots$ *converges in probability to random variable* X*, denoted as* $X_n \xrightarrow{P} X$*, if*

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$$

for all $\epsilon > 0$.

Definition C.2 (Converge in distribution)

Consider a sequence of random variables $X_1, X_2, \ldots$ *and a corresponding sequence of cumulative distribution functions (cdfs),* F_{X_i}*. The sequence* $X_1, X_2, \ldots$ *is said to converge in distribution to a random variable* X *with cdf* F_X *if*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

for every $x \in \mathbb{R}$ *at which* F_X *is continuous.*

Definition C.3 (Almost sure convergence)

The sequence of random variables $X_1, X_2, \ldots$ *converges almost surely to random variable* X*, denoted as* $X_n \xrightarrow{a.s.} X$*, if*

$$P(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1.$$

Definition C.4 (Consistent estimator)

There are two kinds of consistent estimator, the weakly consistent estimator and the strongly consistent estimator.

An estimator $\hat{\theta}_n$ of parameter θ_0 is said to be weakly consistent, if it converges in probability to the true value of the parameter,

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

An estimator $\hat{\theta}_n$ of parameter θ_0 is said to be strongly consistent, if it converges almost surely to the true value of the parameter,

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

Definition C.5 (Stochastic equicontinuity)

A stochastic $\{X_n(\theta)\}$ is stochastically equicontinuous on Θ if $\forall\epsilon > 0, \forall\delta > 0, \exists\eta > 0$ such that

$$\limsup_{n\to\infty} P(\sup_{\theta\in\Theta} \sup_{\theta'\in B(\theta,\eta)} |X_n(\theta) - X_n(\theta')| > \epsilon) < \delta$$

where $B(\theta,\eta)$ is a open ball around the center θ with radius η.

Curriculum Vitae

Name:	Jingjia Chu
Post-Secondary Education and Degrees:	University of Western Ontario London, ON, Canada 2013 - 2018 Ph.D.
	University of Western Ontario London, ON, Canada 2012 - 2013 M.Sc.
	South China University of Technology Guangzhou, China 2008 - 2012 B.Sc.
Honours and Awards:	Student Research Presentation Award Statistical Society of Canada Annual Meeting, 2017 Winnipeg, Manitoba
	Department Teaching Assistant Award, 2016 University of Western Ontario

	Probability Section Student Poster Award Statistical Society of Canada Annual Meeting, 2015 Halifax, NS
Related Work Experience:	Teaching Assistant and Research Assistant The University of Western Ontario 2012 - 2017
	Statistical Consultant The University of Western Ontario 2014 - 2017