12-9-2013

Asymptotic Theory for GARCH-in-mean Models

Weiwei Liu
The University of Western Ontario

Supervisor
Reg Kulperger
The University of Western Ontario Joint Supervisor
Hao Yu
The University of Western Ontario

Graduate Program in Statistics and Actuarial Sciences
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

ASYMPTOTIC THEORY FOR GARCH-IN-MEAN MODELS

(Thesis format: Monograph)

by

Weiwei Liu

Graduate Program in Statistical & Actuarial Sciences

A thesis submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

The School of Graduate and Postdoctoral Studies

The University of Western Ontario

London, Ontario, Canada

Abstract

The GARCH-in-mean process is an important extension of the standard GARCH (generalized autoregressive conditional heteroscedastic) process and it has wide applications in economics and finance. The parameter estimation of GARCH type models usually involves the quasi-maximum likelihood (QML) technique as it produces consistent and asymptotically Gaussian distributed estimators under certain regularity conditions. For a pure GARCH model, such conditions were already found with asymptotic properties of its QML estimator well understood. However, when it comes to GARCH-in-mean models those properties are still largely unknown. The focus of this work is to establish a set of conditions under which the QML estimator of GARCH-in-mean models will have the desired asymptotic properties. Some general Markov model tools are applied to derive the result.

Keywords: GARCH, GARCH-in-mean, asymptotic theory, Markov model

To my parents and Xiami

Acknowledgements

This thesis could not be successfully completed without help from a lot of people. I would like to first express my deepest gratitude to my parents and Yanyan. Their encouragement and spiritual support have been my greatest strength throughout the full course of my study.

I would like to thank my academic supervisors Dr. Reg Kulperger and Dr. Hao Yu who have been extremely knowledgable, helpful and supportive. Special thanks are also due to my thesis examiners (ordered alphabetically) for their insightful comments and suggestions: Dr. John Knight, Dr. Adam Kolkiewicz, Dr. Xiaoming Liu and Dr. Mark Reesor.

I am also indebted to my instructors at Western University, including but not limited to Dr. Serge Provost, Dr. W. John Braun, Dr. Duncan Murdoch, Dr. Ricardas Zitikis, Dr. David Stanford. I also had the pleasure of working with a number of professors including Dr. Rogemar Mamon, Dr. Hristo Sendov as their teaching assistant. Their broad knowledge and strong work ethic are great lessons for myself.

During my stay at Western University I have received excellent support from the Department of Statistical and Actuarial Sciences. Many thanks to staff members including Ms. Jennifer Dungavell, Ms. Jane Bai, and Ms. Lisa Hunt. Lastly, I am thankful to all my fellow students and colleagues, with who the university feels like a big warm family to me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 GARCH Models

Understanding the trade-off between risk and return is essential for financial practices. Investors are not only concerned with the magnitude of asset returns but also want to know the size of the accompanying risk. Therefore, modeling financial volatilities is one of the core problems in financial econometrics.

Conventional discrete-time tools such as models of ARMA type usually find their limitations when dealing with certain financial time series, for example the log-return series. Empirical studies have confirmed a number of statistical regularities often observed on these series, also known as "stylized facts", which are hardly consistent with standard assumptions imposed on traditional models. For instance, the marginal distributions of some financial series were found to be leptokurtic: they have fatter tails and sharper peaks than normal distributions. Another example is the observation known as "volatility clustering", which was first documented by Fama (1965) and Mandelbrot (1963). The latter paper stated: "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes". Such observation suggests that financial volatilities are more likely to be dynamic instead of static, and the fluctuation depends on past information to some extent. For a detailed account of commonly-observed stylized facts one may refer to Taylor (2005) and Tsay (2010).

The existence of these stylized facts called for new tools that incorporate more flexibility in volatility structures. As a result, Engle (1982) proposed the famous autoregressive conditional heteroscedastic (ARCH) model which is defined by the following two equations:

$$\epsilon_t = \sigma_t \eta_t, \quad \eta_t \sim IID(0,1)$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2,$$

where $\omega > 0$, $\alpha_i \geq 0$, $i = 1, \ldots, q$ are constants. The stochastic process ϵ_t is known as the ARCH process of order q, denoted by $\epsilon_t \sim \text{ARCH}(q)$. It is not difficult to see that under the above specification

$$E(\epsilon_t|\mathcal{F}_{t-1}) = 0, \quad Var(\epsilon_t|\mathcal{F}_{t-1}) = \sigma_t^2,$$

where $\mathcal{F}_t = \sigma(\epsilon_s; s \leq t)$ is the information set available at time t.

From the definition equations above, we notice that ARCH model expresses the conditional variance term σ_t^2 as a linear function of the past observations of the squared process ϵ_t^2. Therefore it evolves over time as the most recent information becoming available. This dynamic volatility feature distinguishes ARCH models from conventional tools which operate under a constant conditional variance assumption. The algebraic structure of the conditional variance equation is also relatively simple yet has been found very powerful in capturing main stylized facts.

Early applications of ARCH were mainly focused on macroeconomic aspects, for example modeling inflation rates as in Engle (1982), Engle (1983) and Engle and Kraft (1983). As its popularity grew, practitioners soon started to notice an issue with the model: a long memory behavior was frequently found in empirical studies. In other words, a good fit of ARCH usually requires a considerable amount of parameters. For instance, the model constructed in Engle and Kraft (1983) was in the form of ARCH(8) which included 9 parameters in total. In order to optimize the structure and produce a more parsimonious fit, Bollerslev (1986) revised the ARCH model and proposed the generalized ARCH (GARCH) process with the following specification:

$$\epsilon_t = \sigma_t \eta_t, \quad \eta_t \sim IID(0,1)$$
$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,$$

where $\omega > 0$, $\alpha_i \geq 0$, $i = 1, \ldots, q$ and $\beta_j \geq 0$, $j = 1, \ldots, p$ are constants. The process ϵ_t is known as the GARCH process of orders p and q denoted by $\epsilon_t \sim$ GARCH(p, q). Notice that GARCH models have essentially inherited the spirit of ARCH: considering the conditional distribution of the process we have

$$\epsilon_t | \mathcal{F}_{t-1} \sim D(0, \sigma_t^2),$$

where D represents some generic distribution determined by the i.i.d innovation process $\{\eta_t\}$. Note that σ_t represents the volatility under the financial context, therefore is not directly observable. Nevertheless, under certain conditions it can be expressed by an infinite past representation of lagged values of ϵ_t therefore is well-defined and contained in $\mathcal{F}_{t-1}$. This representation will be discussed in more detail in Chapter 2.

Comparing to the original ARCH model, the only revision made by GARCH is the inclusion of past conditional variances in the volatility equation. The GARCH model assumes that the conditional variance σ_t^2 does not only depend on past observations of the squared process ϵ_t^2, but also on its lagged values σ_{t-j}^2, $1 \leq j \leq$

p. This may be interpreted as certain type of adaptive learning mechanism as Bollerslev (1986) pointed out. Bollerslev (1986) applied this GARCH model to the same problem studied by Engle and Kraft (1983) and showed that a GARCH(1,1) model provided a slightly better fit than the ARCH(8) model originally proposed.

Since its introduction GARCH models have been extremely popular and widely applied to various areas in financial modeling, for example, option pricing as discussed in Duan (1995). For literature reviews of empirical studies and financial applications see Bollerslev et al. (1992), Engle (2001) and Engle (2004). Also refer to Gouriéroux (1997), Francq and Zakoïan (2010) for comprehensive accounts of both theories and practices. In the meantime, numerous extensions to the original GARCH structure were also introduced by researchers to serve different purposes in applications. For example, the GJR-GARCH model of Glosten et al. (1993) focuses on modeling an asymmetric behavior by assuming the signs of past observations also have impact on the forecast. The Markov-switching GARCH model of Hamilton and Susmel (1994) aims to model different volatility dynamics within different sub-period of time. Other examples of GARCH extensions include the integrated GARCH (IGARCH), exponential GARCH (EGARCH), and threshold GARCH(TGARCH) just to name a few. For a survey of commonly encountered GARCH acronyms see Bollerslev (2010).

Among this large variety of GARCH extensions is the GARCH-in-mean model which is of great importance and will be the focus of our study. The motivation behind this particular model is to explain the excessive "risk premium" in the financial market. Engle et al. (1987) pointed out: "as the degree of uncertainty in asset returns varying over time, the compensation required by risk averse economic agents for holding these assets, must also be varying". Unfortunately, traditional GARCH models could not explain such excessive return since the condition expectation $E(\epsilon_t|\mathcal{F}_{t-1})$ remains to be zero throughout the time. Under other frameworks such as the GARCH regression model or ARMA-GARCH (where the GARCH process replaces the traditional i.i.d normal innovations), the conditional expectation of the process either depends on exogenous variables or past observations of the process as opposed to volatilities. The GARCH-in-mean model proposed by Engle et al. (1987) excels by directly establishing a risk-return relationship where the time-varying risk-premium is expressed as a linear function of the current size of risk. The model is defined by the following three equations:

$$
\begin{aligned}
y_t &= \lambda + \delta\sigma_t + \epsilon_t \\
\epsilon_t &= \sigma_t\eta_t, \quad \eta_t \sim IID(0,1) \\
\sigma_t^2 &= \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,
\end{aligned}
$$

where $\omega > 0$, $\alpha_i \geq 0$, $i = 1, \ldots, q$, $\beta_j \geq 0$, $j = 1, \ldots, p$ and λ, δ are constants. The process y_t is known as the GARCH-in-mean process of orders p and q denoted by $y_t \sim$ GARCH-M(p, q). Notice that based on the last two equations, ϵ_t is a well-defined pure GARCH process by itself, i.e. $\epsilon_t \sim$ GARCH(p,q). Therefore the GARCH-in-mean process is essentially a linear combination of a pure GARCH process and its underlying volatility process.

Time varying conditional expectation is the key feature of GARCH-in-mean models. Based on the definition equations above one could verify that

$$y_t|\mathcal{F}_{t-1} \sim D(\lambda + \delta\sigma_t, \sigma_t^2),$$

where D represents some generic distribution determined by the distribution of η_t. λ and δ are the two new parameters introduced by this model and under certain financial context, they may respectively represent the risk-free portion and risk-premium portion of the total excessive return. It also needs to be pointed out that the conditional mean specification can take different forms in practice. Besides the linear function $\lambda+\delta\sigma_t$, other popular choices include the squared form $\lambda+\delta\sigma_t^2$ and the log form $\lambda+\delta\log(\sigma_t^2)$. In this thesis, the consideration will be restricted to the linear form which is the most common one seen in literature. In other words, we assume the risk premium is proportional to the volatility which is on the same scale of the return, as opposed to the variance or the logarithm of variance.

When $\lambda = \delta = 0$ we have $y_t = \epsilon_t \sim$ GARCH(p, q). Therefore the GARCH process may be viewed as a special member of a more general GARCH-in-mean class. The GARCH-in-mean model plays an important role in financial modeling and econometric study, for examples cf. Grier and Perry (2000), Devaney (2001) and Brewer et al. (2007), just to list a few.

1.2 Estimation Theory

GARCH-type models are not only of great value to practitioners, but also contain rich theoretical contents which generate interesting problems. Considerable amount of research is available nowadays investigating various statistical properties of GARCH-type models: stationarity, moment structure, estimation, testing, etc. This thesis will focus on the aspect of parameter estimation. More specifically, we want to study large sample properties of the quasi-maximum likelihood estimator (QMLE) of the GARCH-in-mean process.

Early research such as Engle (1982), Bollerslev (1986), Engle et al. (1987) adopted the traditional maximum likelihood estimation (MLE) approach to estimate parameters of GARCH-type models. The innovation terms were assumed to be i.i.d Gaussian distributed therefore the likelihood function could be constructed based on the conditional distribution

$$\epsilon_t|\mathcal{F}_{t-1} \sim N(0, \sigma_t^2).$$

It needs to be pointed out that the estimation of GARCH-type models is based on the conditional likelihood function. Unlike conventional ARMA models where the likelihood function can be written explicitly, the marginal distribution of GARCH-type processes is usually unknown, and hence one need to work with the conditional distribution instead. Therefore, the QMLE is in fact based on the conditional quasi-likelihood. For convenience, we will just refer to it as the quasi-likelihood in the later chapters.

One concern around the above MLE approach is the Gaussian assumption: empirical studies of financial series often confirm a certain level of leptokurticity of the distribution of innovation process, which makes the distribution not likely to be Gaussian. In this light, the quasi-maximum likelihood estimation becomes quite popular as it does not rely on any particular distributional assumption of the innovation process. One can construct a QMLE without knowing the exact distribution of the process. It proceeds in a similar fashion of the standard ML estimation: we start by appointing a hypothetical distribution to η_t. This distribution does not necessarily coincide with the true one but simply serves as an ancillary tool to construct the quasi-likelihood function. Then the quasi-likelihood function can be calculated based on this postulated distribution. For instance, given observations ϵ_1, ..., ϵ_n of a GARCH(p, q) process, one may construct a Gaussian quasi-likelihood by deriving the log-likelihood function as if $\epsilon_t|\mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$, which yields:

$$I_n(\theta; \epsilon_1, \ldots, \epsilon_n) = \frac{1}{n} \sum_{t=1}^{n} l_t(\theta; \epsilon_t) = \frac{1}{n} \sum_{t=1}^{n} (\frac{\epsilon_t^2}{\sigma_t^2} + \log \sigma_t^2)$$

apart from some constants. Here θ is the parameter vector containing ω, $\alpha_1, \ldots, \alpha_q$ and $\beta_1, \ldots, \beta_p$. We want to point out that the above construction is based on Gaussian QMLE. For the purpose of estimation, other non-Gaussian density functions can also be used to construct the estimator and may have positive or negative impact on the efficiency of estimation. See Berkes and Horváth (2004) for discussion of non-Gaussian QMLEs for GARCH models. In the GARCH-in-mean context, since very limited knowledge is available about its estimators, we start by only considering the Gaussian QMLE in this work.

The term σ_t^2 from the above equation is calculated by the recursive relation

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

Since the volatility process is never observed, we need to estimate $\sigma_1^2, \ldots, \sigma_n^2$ to construct the quasi-likelihood. To do so one needs to iterate the above equation based on a set of initial values - we will explain the estimating procedure in more detail in Chapter 3. For now we just essentially view the QML estimator $\hat{\theta}_n$ as the minimizer of the quasi log-likelihood function $I_n(\theta)$ [1]. We want to emphasize that although this quasi-likelihood is derived based on the Gaussian (or other specific distributions) assumption, it does not imply that η_t is indeed normally distributed. Objects like $I_n(\theta)$ and $l_t(\theta)$ are simply functions to work with. For this reason $I_n(\theta)$ is known as the "quasi-likelihood" or "pseudo-likelihood" instead of just "likelihood".

Statisticians are curious to know whether asymptotic properties found in standard MLEs may somehow be extended to QML estimators. To be more specific,

[1] In fact $I(\theta)$ is obtained as the negative of quasi-likelihood apart from some other constants.

denoting the true value of the parameter by θ_0, we are interested in the following two properties:

Consistency: $\hat{\theta}_n \to \theta_0 \;\; a.s$ when $n \to \infty$.

Asymptotic Normality: $\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{D}{\to} N(0, \Sigma)$ when $n \to \infty$,

where $\overset{D}{\to}$ indicates the convergence in distribution.

It is well known that under a set of regularity conditions, a standard MLE will have the above two properties with the covariance matrix being the inverse of the Fisher information matrix, cf. Newey and McFadden (1994). Establishing similar results for the QMLEs of GARCH-type models has thus attracted a great amount of attention. Pioneering work includes Weiss (1986) which developed asymptotic results for ARCH models under the assumption that the process has finite moments up to the 4th order. Lee and Hansen (1994) further studied the QMLE of GARCH$(1,1)$ based on a re-scaled variable defined as the ratio of the disturbance to the conditional standard deviation. This variable was then assumed to have a bounded 4th moment and the asymptotic theory was established under such condition. However, their approach does not naturally extend to the GARCH(p,q) case.

The QMLE of the GARCH(p,q) model was first rigorously studied by Berkes et al. (2003) under the assumption that the process is strictly stationary and ergodic. Also refer to Berkes and Horváth (2003) and Berkes and Horváth (2004) for further studies of this estimator. Their approach requires the parameter space to be restricted in accordance with the stationarity theorem given by Bougerol and Picard (1992a). A moment condition involving the finiteness of the $(1+s)$th moment of the squared process ϵ_t^2 was also imposed (with s being an arbitrarily small positive number). Their result was improved by Francq and Zakoïan (2004) who removed this moment condition on the observed process and instead required a finite 4th moment of the innovation process. This is by far known as the weakest condition leading to asymptotic properties for QMLEs of GARCH(p, q) processes. Escanciano (2009) followed closely with this approach and extended the result to semi-strong GARCH models with non-i.i.d but martingale difference innovations. On the other hand, Jensen and Rahbek (2004a) and Jensen and Rahbek (2004b) considered the QMLEs of non-stationary ARCH(1)/GARCH(1,1) models and found that some of the parameters could be consistently estimated while fixing some other parameter. A few results are also available for ARMA-GARCH models: see Ling and Li (1998), Ling and McAleer (2003), Ling (2007) and Francq and Zakoïan (2004) for asymptotic theories established for both local and global QMLEs. Refer to Straumann (2005) for a monograph on the parameter estimation for general heteroscedastic models.

Compared to pure GARCH and ARMA-GARCH cases, the theoretical work focused on the GARCH-in-mean model is very limited, and its statistical properties are still largely unknown. Hong (1991) studied the autocorrelation structure of the GARCH-in-mean model and concluded that the autocorrelations behave similar to the autocorrelations of a pure GARCH model and are nonnegative under conventional parameter restrictions for GARCH. Arvanitis and Demos (2004) considered the autocovariances for both the GARCH-in-mean process and its squared process. Sufficient conditions for 4th-order stationarity of the process were proposed. Iglesias and Phillips (2012) investigated the finite sample properties of the QMLE of a restricted ARCH-M model and conducted numeric experiments on the estimator, but no asymptotic theory was established. Christensen et al. (2012) obtained asymptotic results for a modified version of the conditional heteroscedastic in mean model by combining both parametric and non-parametric approaches. However, the result is based on a few high-level assumptions that are difficult to verify, and it does not apply to the original GARCH-in-mean model specified by Engle et al. (1987).

Due to the lack of an asymptotic theory, empirical studies involving GARCH-in-mean models tend to either overlook the issue that the parameters may not be consistently estimated, or simply assume that the asymptotic result obtained under the pure GARCH setting also applies to GARCH-in-mean models automatically, for examples cf. Devaney (2001), Kontonikas (2004). The goal of this thesis is to fill in this gap by developing a proper asymptotic theory for the QMLE of the GARCH-in-mean process. Establishing such a result will not only help practitioners to better understand the validity of their estimates, but also serve as a cornerstone in developing further inference tools such as various goodness-of-fit tests of the innovation process.

1.3 Organization of Thesis

In practice, $p = 1$, $q = 1$ is the most popular specification for GARCH-type models as it provides both accurate and parsimonious fit. Therefore, in the following chapters we restrict our illustration to GARCH(1,1) and correspondingly GARCH-M(1,1) cases for a more straightforward and concise presentation. Note that apart from some added algebraic complexity, our approach applies to models of higher orders in the almost same fashion.

To establish asymptotic results for the QMLE we start by studying relevant statistical properties of the GARCH-M process. The next chapter will focus on its stochastic stability properties including stationarity and ergodicity. Those properties are essential for us to apply a specific type of ergodic theorem and central limit theorem later. Bougerol and Picard (1992a) established such properties for the pure GARCH models but their approach is not transferrable to GARCH-M models due to certain nonlinearity issues. Instead the general Markov model approach introduced by Meyn and Tweedie (2009) will be applied to obtain the desired result.

The procedure of quasi-maximum likelihood estimation is thoroughly discussed Chapter 3, with two main asymptotic results established including the consistency and asymptotic normality of the estimator. Those results are based on the geometric ergodicity theory obtained in Chapter 2, with addition of some other parameter restrictions that are common for GARCH-type models.

In Chapter 4 we conduct a number of simulation studies. By studying the QMLEs obtained under different sample sizes we investigate the overall convergence trend of the estimates. We also simulate t-distributed innovations to generate non-Gaussian GARCH observations and fit the model by Gaussian QMLE so the impact of the true distribution could be evaluated. Some final discussions and comments are included in Chapter 5.

Chapter 2

Stochastic Stability

Before diving into any specific estimation problems, the stochastic properties of the GARCH-in-mean process need to be well understood. Ideally we would like the process to be stochastically "stable" in some sense so that certain versions of limit theorems (law of large numbers, CLT) can be applied.

Strict stationarity and ergodicity are two key properties that are closely related to the asymptotic theories of GARCH-type models. Their definitions (cf. Appendix A, Definitions A.1 and A.2) may be easily found in a number of references, for examples Brockwell and Davis (1987), Francq and Zakoïan (2010). Broadly speaking, stationarity requires that the joint distribution of the process is unchanged when shifting the process over time. Ergodicity requires the process exhibiting the same behavior averaged over time as averaged over the state space. We start this chapter by reviewing some existing results for the pure GARCH model.

2.1 Stability of GARCH Models

A pure GARCH(1,1) process ϵ_t is defined by the following equations:

$$\epsilon_t = \sigma_t(\theta_0)\eta_t, \quad \eta_t \sim IID(0,1) \tag{2.1}$$

$$\sigma_t^2(\theta_0) = \omega_0 + \alpha_0\epsilon_{t-1}^2 + \beta_0\sigma_{t-1}^2(\theta_0) \tag{2.2}$$

with the parameter vector denoted by $\theta_0 = (\omega_0,\ \alpha_0,\ \beta_0)'$. Throughout the thesis we assume the following parameter restriction:

$$\omega_0 > 0, \quad \alpha_0 > 0, \quad \beta_0 > 0. \tag{2.3}$$

The above condition is quite standard in the GARCH literature. Some papers including Bollerslev (1986) may specify the condition as $\alpha_0 \geq 0$ and $\beta_0 \geq 0$. However, noticing that when $\beta_0 = 0$ the model reduces to an ARCH case and when $\alpha_0 = 0$ it becomes rather trivial, we want to exclude these senarios by using the strict inequality as in (2.3).

In equation (2.2) we use the notation $\sigma_t^2(\theta_0)$ instead of just σ_t^2 to emphasize the fact that it is the "true" conditional variance process driven by the true parameter θ_0. For the purpose of estimation we also need to introduce another "parametric form" of this process denoted by $\sigma_t^2(\theta)$, which is the solution of the following recursive relation

$$\sigma_t^2(\theta) = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2(\theta) \tag{2.4}$$

where ϵ_t is the GARCH observation and the parameter vector is $\theta = (\omega,\ \alpha,\ \beta)'$. This process is defined conditionally on a realization of the process $\{\epsilon_t\}$, and it directly relates to the construction of the QMLE which will be explained further in the next chapter. Here θ is just the argument of the quasi-likelihood function that needed to be minimized be minimized to obtain the QMLE, which is also sometimes known as the "dummy variable ". On the contrary, θ_0 denotes the true parameter value that defines the underlying model which generates observations. Notice that when $\theta = \theta_0$, equations (2.2) and (2.4) coincide and they both represent the true conditional variance process.

To better understand properties of this GARCH process we want to ask the following two questions:

- Do the equations (2.1) - (2.2) yield an unique strictly stationary and ergodic solution of ϵ_t and $\sigma_t^2(\theta_0)$?

- If the above is true, does equation (2.4) yields a strictly stationary and ergodic solution of $\sigma_t^2(\theta)$?

Consider equations (2.1) - (2.2) first. Substituting the ϵ_{t-1} term in the second equation by $\sigma_{t-1}(\theta_0)\eta_{t-1}$ gives us

$$\begin{aligned}\sigma_t^2(\theta_0) &= \omega_0 + \alpha_0\sigma_{t-1}^2(\theta_0)\eta_{t-1}^2 + \beta_0\sigma_{t-1}^2(\theta_0) \\ &= \omega_0 + (\alpha_0\eta_{t-1}^2 + \beta_0)\sigma_{t-1}^2(\theta_0).\end{aligned}$$

This is in the form of a random coefficient AR(1) process. The solution of $\sigma_t^2(\theta_0)$ could be found by repeatedly applying the recursive formula above. Nelson (1990) studied this process and obtained the following result.

Proposition 2.1. *If*

$$-\infty \leq E\log\{\alpha_0\eta_t^2 + \beta_0\} < 0 \tag{2.5}$$

then $\sigma_t^2(\theta_0)$ has an unique strictly stationary and ergodic solution

$$\sigma_t^2(\theta_0) = \{1 + \sum_{i=1}^{\infty} b(\eta_{t-1})\dots b(\eta_{t-i})\}\omega_0,$$

where $b(z) = \alpha_0 z^2 + \beta_0$. Moreover, $\epsilon_t = \sigma_t(\theta_0)\eta_t$ is the unique strictly stationary and ergodic solution of the GARCH$(1,1)$ model specified by equations (2.1) - (2.2). On the other hand, when

$$E\log\{\alpha_0\eta_t^2 + \beta_0\} \geq 0$$

there exists no strictly stationary solution.

Proof. See Nelson (1990). □

Taking a closer look at condition (2.5), by Jensen's inequality we know that

$$E \log \{\alpha_0 \eta_t^2 + \beta_0\} \leq \log E(\alpha_0 \eta_t^2 + \beta_0) = \log\{\alpha_0 + \beta_0\}.$$

Therefore, when

$$\alpha_0 + \beta_0 < 1 \tag{2.6}$$

we have $E \log \{\alpha_0 \eta_t^2 + \beta_0\} < \log 1 = 0$ in which case the GARCH process ϵ_t is strictly stationary and ergodic. In fact, condition (2.6) has its own important implication, cf. Bollerslev (1986):

Proposition 2.2. *The GARCH process defined by (2.1) - (2.2) is 2nd-order stationary with*

$$Var(\epsilon_t) = \frac{\omega_0}{1 - (\alpha_0 + \beta_0)}$$

if and only if condition (2.6) holds.

Proof. See Bollerslev (1986). □

From the above result we see that for a GARCH(1,1) process, 2nd-order stationarity actually requires more restrictive parameter conditions than the strictly stationarity. This is consistent with empirical studies which often found the existence of higher moments of financial series questionable. For the stationarity and ergodicity of a general GARCH(p, q) model, refer to results obtained by Bougerol and Picard (1992a), Bougerol and Picard (1992b) using techniques developed by

Brandt (1986).

As already mentioned, we also need to understand the stability properties of the parametric form process $\sigma_t^2(\theta)$ defined by (2.4) since it is directly connected to the quasi-likelihood function. Note that this equation resembles an ARMA(1,1) equation if viewing squared processes as stand-alone objects, although it is technically not true since ϵ_t^2 does not play the role of innovations. Nevertheless, we may still re-write equation (2.4) by introducing the lag operator L:

$$(1 - \beta L)\sigma_t^2(\theta) = \omega + \alpha\epsilon_{t-1}^2.$$

From the well-developed linear time series theory, we know that the lag polynomial on the left hand side of the equation could be inverted given the condition $\beta < 1$, cf. Brockwell and Davis (1987), which gives us the following infinite past representation:

$$\sigma_t^2(\theta) = \frac{\omega}{1-\beta} + \alpha \sum_{i=1}^{\infty} \beta^{i-1} \epsilon_{t-1-i}^2, \tag{2.7}$$

where these ϵ_{t-1-i} terms represent the strictly stationary and ergodic solution given by Proposition 2.1. This form is also known as the ARCH(∞) representation of the GARCH process as it greatly resembles an ARCH conditional variance equation with infinitely many parameters. Berkes et al. (2003) established an ARCH(∞) presentation for the general GARCH(p, q) process. Under such representation, $\sigma_t^2(\theta)$ is expressed as an measurable function of the strictly stationary and ergodic process $\{\epsilon_t\}$. Therefore $\sigma_t^2(\theta)$ is also strictly stationary and ergodic, cf. Billingsley (1995).

Based on the above discussion we see that under relatively simple conditions, both the conditional variance process $\sigma_t^2(\theta_0)$ and the parametric form $\sigma_t^2(\theta)$ of a pure GARCH process are strictly stationary and ergodic.

2.2 Stability of GARCH-M Models

Now we consider the same stability problem under the GARCH-in-mean context. Recall the definition of a GARCH-M$(1,1)$ model:

$$y_t = \lambda_0 + \delta_0 \sigma_t(\theta_0) + \epsilon_t \tag{2.8}$$

$$\epsilon_t = \sigma_t(\theta_0)\eta_t, \quad \eta_t \sim IID(0,1) \tag{2.9}$$

$$\sigma_t^2(\theta_0) = \omega_0 + \alpha_0 \epsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2(\theta_0), \tag{2.10}$$

where $\theta_0 = (\lambda_0,\ \delta_0,\ \omega_0,\ \alpha_0,\ \beta_0)'$ is the true parameter vector, with restrictions specified by (2.3). Same as in the pure GARCH case, we define the following parametric form $\sigma_t^2(\theta)$ to construct the quasi-likelihood later:

$$\sigma_t^2(\theta) = \omega + \alpha(y_{t-1} - \lambda - \delta\sigma_{t-1}(\theta))^2 + \beta\sigma_{t-1}^2(\theta). \tag{2.11}$$

For convenience we may denote

$$a_t(\theta) = y_t - \lambda - \delta\sigma_t(\theta). \tag{2.12}$$

Notice that when $\theta = \theta_0$, according to (2.8) - (2.10):

$$a_t(\theta_0) = y_t - \lambda_0 - \delta_0\sigma_t(\theta_0) = \epsilon_t.$$

Therefore equation (2.11) coincides with (2.10) at the true parameter value θ_0.

Consider the stability properties of $\sigma_t^2(\theta_0)$ and $\sigma_t^2(\theta)$. First of all notice that equations (2.9) - (2.10) independently define a pure GARCH process, i.e. $\epsilon_t \sim$ GARCH$(1,1)$. Hence existing results such as Proposition 2.1 will still apply to processes ϵ_t and $\sigma_t^2(\theta_0)$ here. Secondly, by equation (2.8) y_t is simply a measurable function of $\sigma_t^2(\theta_0)$ considering the fact that $\epsilon_t = \sigma_t(\theta_0)\eta_t$. In this case y_t is strictly stationary and ergodic if $\sigma_t^2(\theta_0)$ also has such property. To sum up we have the following corollary.

Corollary 2.3. *If condition (2.5) holds then the GARCH-M*$(1,1)$ *process defined by equations (2.8) - (2.10) admits an unique stationary and ergodic solution*

$$y_t = \lambda_0 + (\delta_0 + \eta_t)\sigma_t(\theta_0),$$

where $\sigma_t(\theta_0) = \sqrt{\sigma_t^2(\theta_0)}$ *and* $\sigma_t^2(\theta_0)$ *is the strictly stationary and ergodic solution* *defined in Proposition 2.1.*

Proof. By Proposition 2.1, equations (2.9) - (2.10) yield an unique stationary and ergodic solution $\sigma_t^2(\theta_0)$, which is a measurable function of lagged values of η_t. Therefore y_t is a measurable function of $\eta_t, \eta_{t-1}, \ldots$ which is also strictly stationary and ergodic, cf. Billingsley (1995).

The uniqueness of the solution is related to the identifiability issue of the GARCH-M parameters. Given y_t and $\sigma_t^2(\theta_0)$, the model is identifiable if there exists only one set of parameters such that equation (2.8) holds. This issue is discussed later in the proof of Lemma 3.3.

Carrasco and Chen (2002) also studied the stationarity and ergodicity of GARCH-M processes following a Markov modeling approach which we will discuss in more detail in the next section. However, the result only applies to the true processes y_t and $\sigma_t^2(\theta_0)$ but not the parametric form $\sigma_t^2(\theta)$. □

Now consider the parametric form $\sigma_t^2(\theta)$ defined by equation (2.11). One might want to try the same arguments that worked for the pure GARCH process. Assuming $\beta < 1$, by the invertibility property we can re-write the equation and obtain

$$\sigma_t^2(\theta) = \frac{\omega}{1-\beta} + \alpha \sum_{i=1}^{\infty} \beta^{i-1} a_{t-1-i}^2(\theta), \tag{2.13}$$

where the $a_t(\theta)$ process is defined by (2.12).

Although the above equation has the same form of (2.7), one can not conclude the stability property of $\sigma_t^2(\theta)$ by the same argument used in the GARCH case.

For equation (2.7), we have an infinite past representation in terms of the lagged values of ϵ_t, whose stability properties has been well-established via Proposition 2.1. However, it is not the same story for equation (2.13). The stability properties of $a_t(\theta)$ are largely unknown since $a_t(\theta)$ is still defined as a function of $\sigma_t(\theta)$. Only when $\theta = \theta_0$, we know that $a_t(\theta_0) = \epsilon_t$ is a stationary and ergodic sequence under condition (2.5), and in this case equation (2.13) will admit a stationary and ergodic solution since it defines a measurable transformation of a stationary and ergodic process. Unfortunately, this argument only works for this particular scenario. We are not able to conclude the stability properties of $\sigma_t^2(\theta)$ at any arbitrary θ within the parameter space.

For this reason we need to seek an alternative approach when studying the process $\sigma_t^2(\theta)$. Our main tool is the general Markov model technique introduced by Meyn and Tweedie (2009). The rest of this chapter will be heavily based on their theories. One may also refer to other related literatures such as Feigin and Tweedie (1985), Tjøstheim (1990), Doukhan (1994) as needed.

2.3 The Markov Model Approach

As discussed in the last section, the major problem we encountered when dealing with the GARCH-M structure is the nonlinear recursion given by equation (2.11). This nonlinear structure obstructed us from obtaining an infinite past representation as in (2.7). Alternatively, the nonlinear state space (NSS) model introduced by Meyn and Tweedie (2009) provides us a suitable framework to deal with this nonlinear process.

2.3.1 Nonlinear state space model

We begin with the definition of the NSS model.

Definition 2.4 (Nonlinear State Space Model). *Suppose a stochastic process* $\Phi = \{\Phi_k\}$. Φ *is called a nonlinear state space model if the following two conditions are satisfied:*

(NSS1) *for each* $k \geq 0$, Φ_k *and* W_k *are random variables on* $\mathbb{R}^n$ *and* $\mathbb{R}^p$ *respectively, satisfying inductively for* $k \geq 1$,

$$\Phi_k = F(\Phi_{k-1}, W_k),$$

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

(NSS2) *the random variables* $\{W_k\}$ *are an i.i.d. disturbance sequence on* $\mathbb{R}^p$, *whose marginal distribution* Γ *possesses a density* γ *which is supported on an open set* O.

This nonlinear state space model is Markovian since Φ_t only depends on the past information through Φ_{t-1}. Note that $\sigma_t^2(\theta)$ defined by equation (2.11) does not directly fit into this structure because y_t is not an i.i.d sequence thus can not be treated as the innovation process W_t. However, based on equations (2.8) - (2.10) we see that y_t and $\sigma_t^2(\theta_0)$ can be completely determined by their lagged values y_{t-1}, $\sigma_{t-1}^2(\theta_0)$ considering $\epsilon_t = y_t - \lambda_0 - \delta_0\sigma_t(\theta_0)$. We also know that according to (2.11), $\sigma_t^2(\theta)$ can be completely determined by y_{t-1} and $\sigma_{t-1}^2(\theta)$. Therefore if we set-up a three-dimensional process:

$$Y_t = (y_t,\ \sigma_t^2(\theta_0),\ \sigma_t^2(\theta))', \tag{2.14}$$

then we have:

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

where F indicates the function with the mapping rules determined by equations (2.8) - (2.11). This Y_t process is in the form of the NSS model.

Meyn and Tweedie (2009) proposed a systematic approach to study the stability properties of NSS models. Our goal is to utilize their tools to establish a set of conditions under which the process Y_t is strictly stationary and ergodic. To be more specific, we want to consider a particular form of ergodicity known as the geometric ergodicity.

Loosely speaking, the concept of ergodicity describes the behavior of Markov chains "stabilizing" as the time progresses. Geometric ergodicity is a stronger form of ergodicity which does not only require the chain to stabilize as the time progresses, but the chain also needs to converge to its "stabilized stage" geometrically fast. To formally define this concept we introduce the total variation norm: for some measure ν defined on the state space $(S, \mathcal{B}(S))$, the total variation norm is defined as

$$\|\nu\|_{TV} := \sup_{f:|f|\leq 1} |\nu(f)| = \sup_{f:|f|\leq 1} \left| \int_S f(x)\nu(dx) \right|.$$

The definition of geometric ergodicity is given below. Related definitions such as transitional probability kernel and invariant measure of a Markov Chain can be found in A.3 and A.9 of Appendix A.

Definition 2.5 (Geometric Ergodicity)**.** *A Markov Chain X_t is geometrically ergodic if there exists an invariant measure π and a constant $r \geq 1$ such that*

$$\lim_{n\to\infty} r^n \|P^n(x,\cdot) - \pi\|_{TV} = 0,$$

where $P^n(\cdot,\cdot)$ denotes the n-step transitional probability kernel.

We want to establish the geometric ergodicity property because it has important implications such as the Harris recurrence property (cf. Appendix A, Definition A.8), which will enable us to use a specific form of the ergodic theorem and central limit theorem in the next chapter. To establish the geometric ergodicity property for a Markov chain, one needs to verify a few lower-level stability properties for the chain including ψ-irreducibility (cf. Appendix A, Definition A.4), T-chain property (cf. Appendix A, Definition A.5), aperiodicity (cf. Appendix A, Definition A.6). Their connections are explained in the following result.

Proposition 2.6. *Suppose* $\{\Phi_t\}$ *is a* ψ*-irreducible, aperiodic T-chain. The chain is geometrically ergodic if there exists some compact set* C*, a nonnegative function* $V \geq 1$ *bounded on* C*, and positive constants* $c_1 < 1$*,* $c_2 < \infty$ *satisfying:*

$$\int P(x, dy)V(y) \leq c_1 V(x) + c_2 \mathbb{I}_C(x), \qquad x \in S, \tag{2.15}$$

where S *is the appropriate state space for the Markov chain and* $\mathbb{I}_C(x) = \mathbb{I}(x \in C)$ *is the indicator function.*

Proof. Theorem 6.0.1 of Meyn and Tweedie (2009) showed that for a ψ-irreducible T-chain, every compact set is petite (cf. Appendix A, Definition A.5). This result is then obtained by combining this property with Theorem 19.1.3 of Meyn and Tweedie (2009), which applies to a ψ-irreducible, aperiodic chain with C being a petite set. □

This proposition is the main tool for us to establish the geometrically ergodic property. We also want to point out that it is actually not necessary to cast this result directly on the Y_t process defined in (2.14). Considering equations (2.8) - (2.11), we notice that

$$\begin{aligned}\sigma_t^2(\theta_0) &= \omega_0 + \alpha_0 \epsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2(\theta_0) \\ &= \omega_0 + \alpha_0 \sigma_{t-1}^2(\theta_0) \eta_{t-1}^2 + \beta_0 \sigma_{t-1}^2(\theta_0)\end{aligned}$$

$$= \omega_0 + (\alpha_0 \eta_{t-1}^2 + \beta_0)\sigma_{t-1}^2(\theta_0) \tag{2.16}$$

and

$$\begin{aligned}
\sigma_t^2(\theta) &= \omega + \alpha(y_{t-1} - \lambda - \delta\sigma_{t-1}(\theta))^2 + \beta\sigma_{t-1}^2(\theta) \\
&= \omega + \alpha\{\lambda_0 + \delta_0\sigma_{t-1}(\theta_0) + \sigma_{t-1}(\theta_0)\eta_{t-1} - \lambda - \delta\sigma_{t-1}(\theta)\}^2 + \beta\sigma_{t-1}^2(\theta) \\
&= \omega + \alpha\{\lambda_0 - \lambda + (\delta_0 + \eta_{t-1})\sigma_{t-1}(\theta_0) - \delta\sigma_{t-1}(\theta)\}^2 + \beta\sigma_{t-1}^2(\theta). \qquad (2.17)
\end{aligned}$$

Therefore, recognizing that y_t is in fact a function of $\sigma_t(\theta_0)$, we can reduce the original three dimensional process Y_t to a bivariate process X_t, defined by

$$X_t = (\sigma_t^2(\theta_0),\ \sigma_t^2(\theta))' \tag{2.18}$$

with each component defined by recursions (2.16) and (2.17) respectively.

The processes X_t as in (2.18) and Y_t as in (2.14) are closely related. In fact their components follow the same recursive rule except that Y_t incorporates one extra element y_t which could be completely determined by X_t. This type of structure was noticed by a few authors including Carrasco and Chen (2002), Meitz and Saikkonen (2008). The former paper named such processes "generalized hidden Markov" models. Their finding is that for those processes, the ergodic properties of the "hidden part", meaning the non-observable portion such as X_t in our case, will carry over to the full process that includes both the hidden part and the observable part, like the Y_t process. Therefore we have:

Proposition 2.7. *If the Markov chain X_t defined by (2.18) is geometrically ergodic, Y_t defined by (2.14) also has such property.*

Proof. This is a direct application of Proposition 4 of Carrasco and Chen (2002).

□

In this light our approach will focus on the "reduced" process X_t instead of the three-dimensional process Y_t.

2.3.2 Control model

To study the properties of the X_t process we will make use of the "control model" technique introduced by Meyn and Tweedie (2009). For a generic 2-dimensional column vector $Z = (Z_1,\ Z_2)'$ and a constant $u \in \mathbb{R}$, we define two functions

$$f_1(Z, u) = \omega_0 + (\alpha_0 u^2 + \beta_0) Z_1$$

$$f_2(Z, u) = \omega + \alpha\{\lambda_0 - \lambda + (\delta_0 + u)\sqrt{Z_1} - \delta\sqrt{Z_2}\}^2 + \beta Z_2.$$

Comparing with equations (2.16) and (2.17), it is not difficult to see that for process X_t we have

$$X_{t+1} = F(X_t, \eta_t) = \begin{pmatrix} f_1(X_t, \eta_t) \\ f_2(X_t, \eta_t) \end{pmatrix}. \tag{2.19}$$

In the view of Definition 2.4, X_t is a nonlinear state space model. To study its stability property we define its associated "control model". Generally speaking, the control model is a deterministic version of the original stochastic model. Let $x_0 = (x_{0,1},\ x_{0,2})'$ be a bivariate column vector and $u_1, \ldots, u_k$ be a sequence of constants constrained in an open set O in $\mathbb{R}$. Thus we define recursively:

$$x_1 = F(x_0, u_1)$$

$$x_k = F(x_{k-1}, u_k), \quad k = 2, 3, \ldots$$

where the function F is defined by (2.19). The sequence $\{x_t : t \geq 0\}$ is called the associate control model of the stochastic model X_t. Note that this sequence is essentially depending on x_0 and $u_1, \ldots, u_k$. Therefore we may also defined a group of functions $\{F_k,\ k = 1, 2, \ldots\}$ and re-write the above as:

$$x_1 = F_1(x_0, u_1) = F(x_0, u_1) \tag{2.20}$$

$$x_k = F_k(x_0, u_1, \ldots, u_k) = F(F_{k-1}(x_0, u_1, \ldots, u_{k-1}), u_k). \tag{2.21}$$

The original stochastic model and its associate control model are closely related: Meyn and Tweedie (2009) have shown that properties of the stochastic model could be studied via its deterministic counterpart. Given $x_0 \in S$ with S being the state space and a control sequence $\{u_k : u_k \in O, k \in \mathbb{Z}^+\}$, we define the following matrices:

$$A_{k+1} = \begin{pmatrix} 2\alpha_0 x_{k,1} u_{k+1} \\ 2\alpha\left\{(\lambda_0 - \lambda)\sqrt{x_{k,1}} + (\delta_0 + u_{k+1})x_{k,1} - \delta\sqrt{x_{k,1}x_{k,2}}\right\} \end{pmatrix} \tag{2.22}$$

and

$$B_{k+1} = \begin{pmatrix} \alpha_0 u_{k+1}^2 + \beta_0 & 0 \\ a_{21}(x_k, u_{k+1}) & a_{22}(x_k, u_{k+1}) \end{pmatrix}, \tag{2.23}$$

where

$$a_{21}(x_k, u_{k+1}) = \alpha(\delta_0 + u_{k+1})^2 - \alpha\delta(\delta_0 + u_{k+1})\sqrt{\frac{x_{k,2}}{x_{k,1}}} + \frac{\alpha}{\sqrt{x_{k,1}}}(\delta_0 + u_{k+1})(\lambda_0 - \lambda)$$

$$a_{22}(x_k, u_{k+1}) = \beta + \alpha\delta^2 - \alpha\delta(\delta_0 + u_{k+1})\sqrt{\frac{x_{k,1}}{x_{k,2}}} - \frac{\alpha\delta}{\sqrt{x_{k,2}}}(\lambda_0 - \lambda)$$

with $x_{k,1}$, $x_{k,2}$ indicate the 1st and 2nd component of the 2-dimensional vector x_k defined by the control model (2.20) - (2.21). We are now in a position to present the main result of this chapter.

2.3.3 Geometric ergodicity

Assumptions:

A1 The marginal distribution of η_t possesses a density γ on $\mathbb{R}$ which is supported on an open set and lower semi-continuous with respect to the Lebesgue measure. η_t^2 has a non-degenerate distribution with $E\eta_t^2 = 1$.

A2 For any initial value x_0 within the state space, there exists $k \in \mathbb{Z}^+$ and a sequence $(u_1, \ldots, u_k) \in O^k$ such that the matrix

$$C_{x_0}^k(u_1, \ldots, u_k) = [B_k \cdots B_2 A_1 | B_k \cdots B_3 A_2 | \cdots | B_k A_{k-1} | A_k]$$

has full rank. Here this matrix is written in blocks of 2 dimensional column vectors separated by vertical lines, with matrices A_i and B_i defined by equations (2.22) - (2.23). O is some open set in $\mathbb{R}$.

A3 There exists some value u^* such that for any x_k within the state space:

$$\rho(B_{k+1}^*) < 1,$$

where B_{k+1}^* is the B_{k+1} matrix defined in (2.23) evaluated at x_k and $u_{k+1} = u^*$. $\rho(\cdot)$ is the spectral radius of a matrix.

A4 $\alpha_0 + \beta_0 < 1$.

The above assumptions lead to the geometric ergodicity of the process X_t defined in (2.18). The first assumption is about distributional properties of the innovation process. Being lower semi-continuous is a prerequisite for using the control model technique. The unit second moment is a quite standard assumption. It in fact also connects to other issues such as parameter identifiability, which we will see in the next chapter.

Assumptions A2 and A3 are necessary as they will constrain the parameter space in a way so that the original Markov chain X_t will have certain stability properties such as irreducibility, aperiodicity, etc. The last assumption is related to equation (2.15). From Proposition 2.2 we know that A4 actually ensures the underlying GARCH(1,1) process is second order stationary, which is necessary since the left-hand side of (2.15) is an expectation and will involve the existence of certain moment of the process.

The geometric ergodicity of Y_t will then follow, which also implies the process being Harris recurrent. The result is formally stated below.

Theorem 2.8 (Geometric Ergodicity)**.** *Under the Assumptions (**A1**) - (**A4**), the process Y_t defined by (2.14) is geometric ergodic and Harris recurrent.*

Proof. See the next section. □

This geometrically ergodic process is critical for us to establish asymptotic results later. In the next chapter we will illustrate how this property can facilitate our study of the QML estimator.

2.4 Proof of Theorem

According to Proposition 2.6, in order to show X_t is geometrically ergodic we need to establish two intermediate results:

- X_t is a ψ-irreducible aperiodic T-chain.

- X_t satisfies condition (2.15).

Those two steps will be illustrated in the first two subsections. Following those we will also establish the Harris recurrence property.

2.4.1 Irreducibility, aperiodicity and T-chain property

Lemma 2.9. *Under the conditions (**A1**) - (**A4**), the bivariate process X_t as defined in (2.18) is a ψ-irreducible aperiodic T-chain.*

Proof. We will study the properties of X_t via its associated control model defined by (2.20) - (2.21). We will establish the following three properties:

- The control model x_t is forward accessible.
- The control model x_t has a globally attracting point.
- The control model x_t is aperiodic.

Forward Accessibility

Forward accessibility is in some sense a counter part of the irreducibility property for a Markov model. For $x_0 \in S$, $k \in \mathbb{Z}^+$, we define $A_+^k(x_0)$ to be the set of all states reachable from x_0 at time k by the control model (2.20) - (2.21), i.e. $A_+^0(x_0) = \{x_0\}$ and

$$A_+^k(x_0) = \{F_k(x_0, u_1, \ldots, u_k) : u_i \in O, 1 \leq i \leq k\}.$$

We also define $A_+(x_0)$ to be the set of all states that are reachable from x_0 at some time in the future, given by

$$A_+(x_0) = \bigcup_{k=0}^{\infty} A_+^k(x_0).$$

The control model is called forward accessible if for each $x_0 \in S$, the set $A_+(x_0) \subset S$ has non-empty interior.

Now we proceed to verify the forward accessibility property of the control model. Given $x_0 \in S$ and a control sequence $\{u_k : u_k \in O, k \in \mathbb{Z}^+\}$, calculate the partial derivatives of function F defined in (2.19) and evaluate at x_k, u_{k+1}:

$$\begin{aligned} A(x_k, u_{k+1}) &= \left[\frac{\partial F}{\partial u}\right]_{(x_k, u_{k+1})} \\ &= \begin{pmatrix} \left[\frac{\partial f_1}{\partial u}\right]_{(x_k, u_{k+1})} \\ \left[\frac{\partial f_2}{\partial u}\right]_{(x_k, u_{k+1})} \end{pmatrix} \\ &= \begin{pmatrix} 2\alpha_0 x_{k,1} u_{k+1} \\ 2\alpha\left\{(\lambda_0 - \lambda)\sqrt{x_{k,1}} + (\delta_0 + u_{k+1})x_{k,1} - \delta\sqrt{x_{k,1}x_{k,2}}\right\} \end{pmatrix}, \end{aligned}$$

where $x_{k,1}$, $x_{k,2}$ are the first and second element of the bivariate vector x_k. Suppose that $\frac{\partial f}{\partial x(1)}$ and $\frac{\partial f}{\partial x(1)}$ are the partial derivatives of the function $f(x, u)$ over the first and second element of x, then we have another object:

$$B(x_k, u_{k+1}) = \left[\frac{\partial F}{\partial x}\right]_{(x_k, u_{k+1})}$$

$$= \begin{pmatrix} \left[\frac{\partial f_1}{\partial x(1)}\right]_{(x_k,u_{k+1})} & \left[\frac{\partial f_1}{\partial x(2)}\right]_{(x_k,u_{k+1})} \\ \left[\frac{\partial f_2}{\partial x(1)}\right]_{(x_k,u_{k+1})} & \left[\frac{\partial f_2}{\partial x(2)}\right]_{(x_k,u_{k+1})} \end{pmatrix}$$

$$= \begin{pmatrix} \alpha_0 u_{k+1}^2 + \beta_0 & 0 \\ a_{21}(x_k, u_{k+1}) & a_{22}(x_k, u_{k+1}) \end{pmatrix},$$

where

$$a_{21}(x_k, u_{k+1}) = \alpha(\delta_0 + u_{k+1})^2 - \alpha\delta(\delta_0 + u_{k+1})\sqrt{\frac{x_{k,2}}{x_{k,1}}} + \frac{\alpha}{\sqrt{x_{k,1}}}(\delta_0 + u_{k+1})(\lambda_0 - \lambda)$$

$$a_{22}(x_k, u_{k+1}) = \beta + \alpha\delta^2 - \alpha\delta(\delta_0 + u_{k+1})\sqrt{\frac{x_{k,1}}{x_{k,2}}} - \frac{\alpha\delta}{\sqrt{x_{k,2}}}(\lambda_0 - \lambda).$$

Denote the block matrix $C_{x_0}^k(u_1, \ldots, u_k)$ by

$$C_{x_0}^k(u_1, \ldots, u_k) = \left[B_k \cdots B_2 A_1 | B_k \cdots B_3 A_2 | \cdots | B_k A_{k-1} | A_k\right],$$

where $A_k = A(x_{k-1}, u_k)$, $B_k = B(x_{k-1}, u_k)$. The vertical lines above separate the individual blocks which are 2 dimensional column vectors. By Proposition 7.1.4 of Meyn and Tweedie (2009), the control model is forward accessible if for any x_0, there exists $k \in \mathbb{Z}^+$ and a sequence $(u_1, \ldots, u_k) \in O^k$ such that

$$rank(C_{x_0}^k(u_1, \ldots, u_k)) = 2.$$

Therefore in view of Assumption **A2**, this control model is forward accessible.

Globally Attracting Point

If for any starting value x_0, we can find a control sequence u_k such that $x_t \to x^*$, then x^* is known as a globally attracting point of the control model.

Now we consider the difference $x_t - x_{t-1}$. Under the control sequence $\{u_t = u^*\}$ as in Assumption A3, we can apply the mean value theorem:

$$\begin{aligned} \|x_{t+1} - x_t\| &= \|F(x_t, u^*) - F(x_{t-1}, u^*)\| \\ &= \left\| (x_t - x_{t-1}) \cdot \left[\frac{\partial F}{\partial x}\right]_{(x_t^*, u^*)} \right\| \\ &= \|B(x_t^*, u^*) \cdot (x_t - x_{t-1})\| \,. \end{aligned} \tag{2.24}$$

where x^* is on the cord of x_t and x_{t-1}. Given some vector norm $\|\cdot\|$ on $\mathbb{R}^n$ we can always induce the following norm for a $n \times n$ matrix A:

$$\|A\|_{op} = \sup \left\{ \frac{\|Ax\|}{\|x\|} : x \in \mathbb{R}^n, \|x\| \neq 0 \right\}. \tag{2.25}$$

In general, the spectral radius $\rho(A)$ is bounded above by the operator norm of A. Therefore, we can select a particular vector norm such that its induced matrix norm is very close to the spectral radius. By assumption A3 we know that $\rho\left(B(x_k, u^*)\right) < 1$ for all x_k in the state space. Thus there exists some small positive value ϵ such that

$$\sup_{x_k \in S} \{\rho\left(B(x_k, u^*)\right)\} + \epsilon < 1,$$

where S represents the state space. Denote $\rho_0(\epsilon) = \sup_{x_k \in S} \{\rho\left(B(x_k, u^*)\right)\} + \epsilon$. We select the vector norm which leads to an induced matrix norm satisfying:

$$\|B(x_k, u^*)\|_{op} \leq \rho_0$$

for all $x_k \in S$. By (2.25) the above inequality implies:

$$\|B(x_k, u^*)x\| \leq \rho_0 \|x\|.$$

Substituting the above result back to equation (2.24) we have:

$$\begin{aligned} \|x_{t+1} - x_t\| &= \|B(x_t^*, u^*) \cdot (x_t - x_{t-1})\| \\ &\leq \rho_0 \|x_t - x_{t-1}\| \\ &\leq \rho_0^2 \|x_{t-1} - x_{t-2}\| \\ &\leq \rho_0^t \|x_1 - x_0\|, \end{aligned}$$

with $\rho_0 < 1$ as we previously discussed. Therefore we know

$$\|x_{t+1} - x_t\| \to 0, \quad as \quad t \to \infty.$$

Therefore, for any integers n, k we have that

$$\begin{aligned}
\|x_{n+k} - x_n\| &= \|x_{n+k} - x_{n+k-1} + x_{n+k-1} \cdots x_{n-1} - x_n\| \\
&\leq \|x_{n+k} - x_{n+k-1}\| + \cdots + \|x_{n-1} - x_n\| \\
&\leq \|x_1 - x_0\| \cdot \sum_{t=n}^{n+k-1} \rho_0^t \to 0, \quad n \to \infty
\end{aligned}$$

Hence x_t is a Cauchy sequence and the globally attracting point exists.

Aperiodicity

According to Proposition 7.2.5 of Meyn and Tweedie (2009), the control model is called M-irreducible if it is forward accessible and has a globally attracting point. M-irreducible chains have a set M known as a minimal set (in a sense of being the smallest closed and invariant reachable set, cf. p154 of Meyn and Tweedie (2009)).

Set M has a partition $M = \bigcup U_i$. Those U_i sets are called the "periodic orbit". They are essentially a deterministic counterpart of the d-cycle of a stochastic model (cf. Appendix A, Definition A.6). Since $x_t \to x^*$ where x^* is the globally attracting point, x^* is reachable at almost any time, which means it belongs to each of the U_i orbits. This indicates that only one such U_i set exists. Hence the model is aperiodic.

By Theorem 7.3.5 of Meyn and Tweedie (2009), the NSS model X_t as defined in (2.18) is a ψ-irreducible aperiodic T-chain. □

2.4.2 Drift condition

Lemma 2.10. *Under the conditions (**A1**) - (**A4**), the bivariate process X_t as defined in (2.18) satisfies condition (2.15).*

Proof. Inequality (2.15) is also known as the Foster-Lyapunov drift criteria. For the Markov chain X_t and some non-negative measurable function V, define the drift operator by:

$$\Delta V(x) := \int P(x, dy)V(y) - V(x) = E\{V(X_{t+1})|X_t = x\} - V(x), \quad x \in S,$$

then (2.15) has the following equivalent form:

$$\Delta V(x) \leq -c_3 V(x) + c_2 \mathbb{I}_C(x),$$

where $c_3 > 0$, $c_2 < \infty$. Note that the drift condition needs to be verified on the stochastic model X_t instead of the control model x_t.

Under Assumptions A1 and A4, we know from Jensen's inequality that

$$E\{\log(\alpha_0 \eta_{t-1}^2 + \beta_0)\} \leq \log E(\alpha_0 \eta_{t-1}^2 + \beta_0) < 0$$

since $E(\alpha_0 \eta_{t-1}^2 + \beta_0) = \alpha_0 + \beta_0 < 1$. Hence there exists some $s \in (0, 1)$ such that (cf. Lemma 2.2 of Francq and Zakoïan (2010)):

$$E\{(\alpha_0\eta_{t-1}^2+\beta_0)^s\}<1.$$

Now for a bivariate random vector $Z=(Z_1,\ Z_2)$ we define the function V as $V(Z)=1+|Z_1|^s$. Suppose within the state space we have $x=(x_1,\ x_2)\in S$, by equations (2.16) - (2.17) we obtain:

$$\begin{aligned} E\{V(X_t)|X_{t-1}=x\} &= E\{1+\sigma_t^{2s}(\theta_0)|X_{t-1}=x\} \\ &= 1+E\left\{\left[\omega_0+(\alpha_0\eta_{t-1}^2+\beta_0)x_1\right]^s\right\} \\ &\leq 1+\omega_0^s+x_1^sE\{(\alpha_0\eta_{t-1}^2+\beta_0)^s\}. \end{aligned} \tag{2.26}$$

The last inequality above is due to the C_r-inequality and the fact that $s\in(0,1)$.

We already know that $E\left\{(\alpha_0\eta_{t-1}^2+\beta_0)^s\right\}<1$. Therefore we can set an arbitrary positive number c_1 such that $E\left\{(\alpha_0\eta_{t-1}^2+\beta_0)^s\right\}<c_1<1$ and define the following compact set:

$$C=\left\{x\in R:0\leq x\leq\frac{1-c_1+\omega_0^s}{c_1-E\left\{(\alpha_0\eta_{t-1}^2+\beta_0)^s\right\}}\right\}.$$

By construction, x_1 is restricted to a state space of non-negative numbers since it represents the conditional variance. Therefore for any $x_1 \notin C$ we know

$$x_1 > \frac{1 - c_1 + \omega_0^s}{c_1 - E\left\{(\alpha_0 \eta_{t-1}^2 + \beta_0)^s\right\}},$$

which indicates:

$$\left(c_1 - E\left\{(\alpha_0 \eta_{t-1}^2 + \beta_0)^s\right\}\right) x_1^s > 1 - c_1 + \omega_0^s.$$

Rearranging the terms we obtain:

$$1 + \omega_0^s + x_1^s E\left\{(\alpha_0 \eta_{t-1}^2 + \beta_0)^s\right\} < c_1(1 + x_1^s).$$

Considering equation (2.26) we have shown that when $x_1 \notin C$,

$$E\{V(X_t)|X_{t-1} = x_1\} \leq c_1 V(x_1)$$

for some $c_1 \in (0,\ 1)$.

Now consider the case $x_1 \in C$. We define

$$c_2 = 1 + \omega_0^s + x_1^s E\{(\alpha_0 \eta_{t-1}^2 + \beta_0)^s\}.$$

It is not difficult to verify that $0 < c_2 < \infty$. By (2.26) we have for all $x_1 \in C$

$$E\{V(X_t)|X_{t-1} = x_1\} \leq c_2 \leq c_1 V(x_1) + c_2.$$

Combining both cases we have shown that the drift condition (2.15) holds. □

2.4.3 Harris recurrence

Lemma 2.11. *Geometrically ergodic Markov chains are also Harris recurrent.*

Proof. Theorem 15.0.1 of Meyn and Tweedie (2009) points out that condition (2.15) is in fact necessary and sufficient for the geometric ergodicity property. Therefore if some chain Φ_t is geometrically ergodic, (2.15) has to be satisfied, i.e. there exists $c_1 < 1$, $c_2 < \infty$, $V \geq 1$ and a compact set C such that:

$$\int P(x, dy)V(y) \leq c_1 V(x) + c_2 \mathbb{I}_C(x).$$

Note that this compact set C is not necessarily the set we defined in the last subsection, as we are dealing with a general Markov Chain now. Let $V^*(x) = V(x) - 1$, then

$$\begin{aligned}\int P(x, dy)V^*(y) &= \int P(x, dy)V(y) - 1 \\ &\leq c_1 V(x) + c_2 \mathbb{I}_C(x) - 1 \\ &\leq c_1 V^*(x) + c_2 \mathbb{I}_C(x) - 1 + c_1\end{aligned}$$

$$\leq c_1 V^*(x) + s(x) - f(x),$$

where $f(x) = 1 - c_1$, $s(x) = c_2 \mathbb{I}_C(x)$. Since $c_1 \in (0, 1)$ we have:

$$\int P(x, dy) V^*(y) \leq V^*(x) - f(x) + s(x).$$

Now denote the hitting time for the chain Φ_t to reach C: $\tau = \inf\{t \geq 1 : X_t \in C\}$. By the Comparison Theorem 14.2.2 of Meyn and Tweedie (2009):

$$\begin{aligned} E_{x_0}\left[\sum_{k=0}^{\tau-1} f(\Phi_k)\right] &\leq V^*(x_0) + E_{x_0}\left[\sum_{k=0}^{\tau-1} s(\Phi_k)\right] \\ &\leq V(x_0) + E_{x_0}\left[c_2 \sum_{k=0}^{\tau-1} \mathbb{I}_C(\Phi_k)\right] \\ &= V(x_0) + E_{x_0}\left[c_2 \mathbb{I}_C(x_0)\right] \\ &< \infty, \end{aligned}$$

where $x_0 \in S$ is some initiating point for the chain. The second last line holds since Φ_k will not enter C again until time τ. E_{x_0} indicate the expectation is taken based on initial distribution $\Phi_0 = x_0$.

On the other hand, since $f(x) = 1 - c_1$ we know

$$E_{x_0}\left[\sum_{k=0}^{\tau-1} f(\Phi_k)\right] = (1 - c_1) E_{x_0}(\tau).$$

Therefore we have $E_{x_0}(\tau) < \infty$ which indicates:

$$P(\tau < \infty | X_0 = x_0) = 1.$$

By Theorem 9.1.7 of Meyn and Tweedie (2009), the chain is Harris recurrent. □

2.4.4 Conclusion

Based on Lemmas 2.9 and 2.10, we can use Proposition 2.6 to conclude that X_t defined by (2.18) is a geometrically ergodic chain. By Proposition 2.7 the process Y_t defined by (2.14) is also geometrically ergodic. The Harris recurrence property follows from Lemma 2.11.

Chapter 3

Asymptotic Theory

In this chapter we will establish two important asymptotic results for the QMLE of the GARCH-in-mean model: strong consistency and asymptotic normality. We will show that under certain conditions, the QMLE will converge almost surely to the true parameter value as the sample size increases. Moreover, the distribution of this estimator suitably scaled around the true parameter will also converge to a Gaussian distribution.

3.1 Quasi-maximum Likelihood Estimator

The quasi-maximum likelihood estimation is very popular amongst various GARCH-type models. This approach has the advantage that it does not rely on the distribution information of the process. The procedure starts by imposing a postulated distribution on the i.i.d. innovation process η_t, whose actual distribution is unknown. In practice, the most common substitute is the Gaussian distribution thus the obtained estimator is known as the Gaussian QMLE.

Suppose we have observations $y_1, \ldots, y_n$ generated from a GARCH-M model defined by equations (2.8) - (2.10), with the true parameter being

$$\theta_0 = (\lambda_0,\ \delta_0,\ \omega_0,\ \alpha_0,\ \beta_0)'.$$

To estimate θ_0 we now construct a Gaussian quasi-likelihood function. Since the distribution of η_t is unknown, we assume that

$$\eta_t \sim N(0,1).$$

In this case equations (2.8) - (2.10) imply that

$$y_t|\mathcal{F}_{t-1} \sim N\left(\lambda_0 + \delta_0\sigma_t(\theta_0), \sigma_t^2(\theta_0)\right),$$

where $\mathcal{F}_{t-1}$ is the information set up to time $t-1$. Denoting f_{y_t} as the density for this conditional distribution, we have

$$\log f_{y_t}(x) = -\frac{(x - \lambda_0 - \delta_0\sigma_t(\theta_0))^2}{2\sigma_t^2(\theta_0)} - \frac{1}{2}\log 2\pi\sigma_t^2(\theta_0).$$

Notice that although we have observations $y_1, \ldots, y_n$, the conditional variance $\sigma_t^2(\theta_0)$ is still unknown since it is never observed. In order to construct a quasi-likelihood, one needs to estimate this conditional variance first. The estimation is based on the parametric form (2.11):

$$\sigma_t^2(\theta) = \omega + \alpha(y_{t-1} - \lambda - \delta\sigma_{t-1}(\theta))^2 + \beta\sigma_{t-1}^2(\theta),$$

where $\theta = (\lambda,\ \delta,\ \omega,\ \alpha,\ \beta)'$ is the vector of dummy variables. This object is defined in a similar fashion to the true process $\sigma_t^2(\theta_0)$. By equations (2.8) - (2.10) we may easily verify that $\sigma_t^2(\theta) = \sigma_t^2(\theta_0)$ when $\theta = \theta_0$.

Now we construct the Gaussian quasi-likelihood based on the density of $y_t|\mathcal{F}_{t-1}$ derived above. Define

$$l_t(\theta) = \frac{(y_t - \lambda - \delta\sigma_t(\theta))^2}{\sigma_t^2(\theta)} + \log\sigma_t^2(\theta), \tag{3.1}$$

where $\sigma_t^2(\theta)$ is defined by (2.11). This object is calculated based on the conditional density apart from some constants [1]. The Gaussian quasi-likelihood function can then be constructed based on the joint density of $y_1, \ldots, y_n$, conditional on $\mathcal{F}_0$:

$$I_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} l_t(\theta). \tag{3.2}$$

In practice we do not work directly with $I_n(\theta)$ because it involves $\sigma_1^2(\theta)$ which needs to be calculated based on y_0 and $\sigma_0^2(\theta)$ according to equation (2.11). We

[1] As mentioned in Chapter 1, we work with the negative of the quasi-likelihood. i.e. $l_t = -2\log f_{y_t}(x) + \log 2\pi$ where f denotes the Gaussian density.

do not have information regarding their values. To start the iteration we need to set two initial values y_0 and $\tilde{\sigma}_0(\theta)$. The choice of these two values are almost arbitrary as long as $\tilde{\sigma}_0(\theta)$ takes some positive values considering it represents the volatility. For instance we may use:

$$y_0 = \tilde{\sigma}_0 = \sqrt{\omega}$$

or

$$y_0 = y_1, \quad \tilde{\sigma}_0 = |y_1|.$$

Based on the initial values we could start the following iteration

$$\tilde{\sigma}_t^2(\theta) = \omega + \alpha(y_{t-1} - \lambda - \delta\tilde{\sigma}_{t-1}(\theta))^2 + \beta\tilde{\sigma}_{t-1}^2(\theta). \tag{3.3}$$

This equation is exactly the same as equation (2.11) except that it is based on arbitrarily assigned initial values y_0 and $\tilde{\sigma}_0(\theta)$, while (2.11) is assumed to have an infinite past. It will be shown that the choice of those two initial values does not have any impact on the asymptotic properties of the estimator. However, we want to point out that a good choice of initial values does have its practical value on other aspects such as computational cost, efficiency, etc.

Based on (3.3) we define the following two objects

$$\tilde{l}_t(\theta) = \frac{(y_t - \lambda - \delta\tilde{\sigma}_t(\theta))^2}{\tilde{\sigma}_t^2(\theta)} + \log\tilde{\sigma}_t^2(\theta) \tag{3.4}$$

$$\tilde{I}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\theta) \tag{3.5}$$

which are essentially counterparts of (3.1) and (3.2) but directly workable as they can be explicitly calculated based on assigned initial values. We call $\tilde{I}_n(\theta)$ our Gaussian quasi-likelihood function [1]. The quasi maximum likelihood estimator is then defined as

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} \tilde{I}_n(\theta), \tag{3.6}$$

where Θ is the parameter space. Notice here the estimator is a minimizer of the quasi-likelihood function instead of a maximizer. This is because our quasi-likelihood function is constructed based on the negative of the postulated density function apart from some other constants.

3.2 Consistency

The first asymptotic property we want to investigate is the strong consistency. We are interested in conditions under which the QMLE $\hat{\theta}_n$ defined by (3.6) will converge to the true parameter value θ_0 almost surely.

We make the following assumptions:

B1 $\theta_0 \in \Theta$ and the parameter space Θ is compact.

B2 $\beta < 1$ for $\forall \beta \in \Theta$.

[1]Sometimes we may also call (3.2) the quasi-likelihood by context.

B3 Denote that $M_t(\theta) = \frac{\partial l_t(\theta)}{\partial \theta}$ and $B(\theta, k)$ is some open ball within the interior of Θ with center θ and radius k. Then $\sup_{\theta^* \in B(\theta,k)} M_t(\theta^*)$ is a geometrically ergodic sequence and

$$E\left(\sup_{\theta^* \in B(\theta,k)} \|M_t(\theta^*)\|\right) \leq M < \infty.$$

To establish the consistency of the estimator we need to make use of both pairs $l_t(\theta)$, $I_n(\theta)$ and $\tilde{l}_t(\theta)$, $\tilde{I}_n(\theta)$. Based on the result obtained in Chapter 2, $l_t(\theta)$ is strictly stationary and ergodic under Assumptions (A1) - (A4). Therefore some type of ergodic theorem could apply, which gives us

$$I_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} l_t(\theta) \stackrel{a.s.}{\to} E(l_1(\theta)).$$

We also want to show that θ_0 is a minimizer of $E(l_1(\theta))$. This will as well involve some parameter identifiability issue and the invertibility of the conditional variance equation, which is related to Assumption (B2). On the other hand, the QMLE $\hat{\theta}_n$ minimizes $I_n(\theta)$ by definition. We then conclude the convergence of $\hat{\theta}_n$ to θ_0 using a compactness argument, where Assumption (B1) is needed.

Also note that the QMLE is defined in terms of $\tilde{I}_n(\theta)$ which starts with an arbitrarily appointed initial measure, instead of $I_n(\theta)$ which starts with the stationary measure. The geometric ergodicity and Harris recurrence property established in Chapter 2 will play an important role in connecting these two objects.

Lastly, Assumption (B3) is needed to ensure the equicontinuity of function $\tilde{I}_n(\theta)$, which is necessary for us to use the compactness argument to conclude the consistency. This conditional might be difficult to verify directly, but one can study it by simulation techniques relatively easily.

Our first main result is stated below.

Theorem 3.1. *Under Assumptions (**A1**) - (**A4**) and (**B1**) - (**B3**), the quasi-maximum likelihood estimator defined by (3.6) is strongly consistent, i.e*

$$\hat{\theta}_n \to \theta_0 \quad a.s, \quad n \to \infty.$$

Proof. See Section 3.4. □

3.3 Asymptotic Normality

In the last section we established a number of conditions under which the QMLE $\hat{\theta}_n$ is strongly consistent. In this section we further investigate this estimator by studying its distribution. We are interested to see whether under certain conditions, the distribution of $\hat{\theta}_n$ around the true parameter approaches a normal distribution as the sample size increases. In other words, suppose J is some matrix representing the asymptotic covariance structure. We would like to find conditions that lead to the following property:

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

where θ_0 is the true parameter and $\hat{\theta}_n$ is the QMLE defined by (3.6). The notation $\xrightarrow{D}$ indicates the convergence in distribution.

The approach to study this problem involves the Taylor's expansion of the quasi-likelihood function $\tilde{I}_n(\hat{\theta}_n)$ around the true parameter value θ_0. Since $\hat{\theta}_n$ is the minimizer of $\tilde{I}_n(\theta)$ by definition, we know that its first order derivative evaluated at this minimizer should be zero under suitable conditions. In other words we have:

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial \tilde{l}_t(\hat{\theta}_n)}{\partial \theta} = 0,$$

where $\frac{\partial \tilde{l}_t(\hat{\theta}_n)}{\partial \theta}$ is the partial derivative $\frac{\partial \tilde{l}_t(\theta)}{\partial \theta}$ evaluated at $\theta = \hat{\theta}_n$.

This first order derivative can be further expanded around θ_0 by applying Taylor's expansion. This gives us the following second-order condition:

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\frac{\partial \tilde{l}_t(\hat{\theta}_n)}{\partial \theta} &= \frac{1}{n}\sum_{t=1}^{n}\{\frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta} + \frac{\partial^2 \tilde{l}_t(\theta^*)}{\partial \theta \partial \theta'}(\hat{\theta}_n - \theta_0)\} \\ &= 0,\end{aligned}$$

where θ^* is on the cord of θ_0 and $\hat{\theta}_n$. Rearranging the terms we have:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \left[-\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 \tilde{l}_t(\theta^*)}{\partial\theta\partial\theta'} \right]^{-1} \cdot \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}. \tag{3.7}$$

Therefore the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ can be studied if we can understand the properties of both objects on the right-hand side.

The ergodic theorem and central limit theorem for martingale difference sequences are the most important tools we need to study processes on the right hand side of equation (3.7). The Harris recurrence and geometrically ergodicity property established in Chapter 2 will help to eliminate the asymptotic impact of arbitrarily assigned initial values. The distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ could then be derived following the Slutsky's theorem.

Introduce the following two matrices:

$$A = E\left(\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}\right), \quad B = E\left(\frac{\partial l_t(\theta_0)}{\partial\theta}\frac{\partial l_t(\theta_0)}{\partial\theta'}\right) \tag{3.8}$$

and we assume that:

C1 $\theta_0 \in \overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ.

C2 $E(\eta_t) = 0$, $E(\eta_t^4) < \infty$.

C3 $\beta_0^2 + \alpha_0^2\delta_0^2 < 1$ and $3\alpha_0^2 + 2\alpha_0\beta_0 + \beta_0^2 < 1$.

C4 The matrix A defined in (3.8) is nonsingular.

Our second main result of this chapter is stated below.

Theorem 3.2. *Under Assumptions (**A1**) - (**A4**), (**B1**) - (**B3**) and (**C1**) - (**C4**), the QMLE $\hat{\theta}_n$ defined by (3.6) has an asymptotic normal distribution around the true parameter value θ_0, and*

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

where $\xrightarrow{D}$ indicates the convergence in distribution and A, B are matrices defined in (3.8).

Proof. See Section 3.5 □

Strong consistency is necessary for us to further conclude the asymptotic normality of the estimator, therefore assumptions (A1) - (A4), (B1) - (B3) are necessary. (C1) is more restrictive than B1 as it precludes θ_0 from being on the boundary so that the first-order condition will hold. (C2) is quite standard in the literature, cf. Francq and Zakoïan (2004), Berkes et al. (2003). It is a necessary moment condition to ensure certain objects in matrices A and B have finite expectations. Assumption (C3) concerns certain moment properties of our GARCH-M model. Studying the second order derivatives will involve solving a random coefficient AR type of structure. For us to conclude the convergence of such a structure we will require the existence of the 4th moment of the process, which is ensured by the second part of Assumption (C3). On the other hand, the first part of the Assumption (C3) is related to the AR parameter and will lead to a solution of the random coefficient process with desired properties. Together with Assumption C2 they ensure the matrices A and B are well-defined. The last assumption is necessary for us to apply the Slutsky's theorem.

We will see later that the matrices A and B are limits related to the first order and second order derivatives of the quasi-likelihood. It is sometimes difficult to obtain closed form solutions of A and B. However, certain numerical analysis can be applied to approximate those derivatives and an estimate of the asymptotic covariance matrix $A^{-1}BA^{-1}$ can be obtained. The algorithm will be discussed in more detail in Chapter 4.

3.4 Proof of Consistency

In this section we prove Theorem 3.1. Two intermediate results will be established to facilitate the proof. First of all we want to show a certain expectation is well-defined, and it is minimized at the true parameter value θ_0. This expectation will be our limit criterion. Secondly, we will show that the quasi maximum likelihood function $\tilde{I}_n(\theta)$ converges to this limit criterion under any appointed initial values.

To ease our presentation we introduce an additional notation:

$$\tilde{a}_t(\theta) = y_t - \lambda_t - \delta_t \tilde{\sigma}_t(\theta), \tag{3.9}$$

where $\tilde{\sigma}_t(\theta) = \sqrt{\tilde{\sigma}_t^2(\theta)}$ with $\tilde{\sigma}_t^2(\theta)$ defined by equation (3.3). Notice that $\tilde{a}_t(\theta)$ is essentially defined in the same way as $a_t(\theta)$ of equation (2.12), but starts with arbitrarily assigned initial values instead of an infinite past.

3.4.1 Limit criterion

Lemma 3.3. *Under Assumptions (**A1**) - (**A4**) and (**B1**) - (**B2**), the expectation $E_{\theta_0} l_1(\theta)$ is well defined in $\mathbb{R} \cup \{+\infty\}$ for $\forall\ \theta \in \Theta$ and in $\mathbb{R}$ at $\theta = \theta_0$. Moreover, $E_{\theta_0} l_1(\theta) > E_{\theta_0} l_1(\theta_0)$ for $\forall\ \theta \neq \theta_0$, $\theta \in \Theta$.*

Proof. We establish the following three steps in sequence:

- $E_{\theta_0} l_1(\theta)$ is well-defined.
- It is minimized at $\sigma_1^2(\theta_0)$ and $a_1(\theta_0)$.
- The parameters are identifiable.

The Expectation is Well Defined in $\mathbb{R} \cup \{+\infty\}$

Define notations $x^- = \max(-x, 0)$ and $x^+ = \max(x, 0)$. Then we have

$$\begin{aligned} E_{\theta_0} l_1^-(\theta) &= E_{\theta_0}\{\max(-\log \sigma_1^2(\theta) - \frac{a_1^2(\theta)}{\sigma_1^2(\theta)}, 0)\} \\ &\leq E_{\theta_0}\{\max(-\log \sigma_1^2(\theta), 0)\} \\ &\leq \max(-\log E_{\theta_0} \sigma_1^2(\theta), 0) \\ &\leq \max(-\log \omega, 0) \end{aligned}$$

$$< \infty,$$

where $a_t(\theta)$ is defined by (2.12). Note that the third step above involves the application of Jensen's inequality on the logarithm function.

Since $E_{\theta_0} l_1^-(\theta) < \infty$, we know that $E_{\theta_0} l_1(\theta) \neq -\infty$. Therefore $E_{\theta_0} l_1(\theta) \in \mathbb{R} \cup \{+\infty\}$ for any parameter value within the parameter space. Now consider the particular case when $\theta = \theta_0$:

$$\begin{aligned} E_{\theta_0} l_1(\theta_0) &= E_{\theta_0} \left\{ \frac{a_1^2(\theta_0)}{\sigma_1^2(\theta_0)} + \log \sigma_1^2(\theta_0) \right\} \\ &= E_{\theta_0} \left\{ \frac{\sigma_1^2(\theta_0)\eta_1^2}{\sigma_1^2(\theta_0)} + \log \sigma_1^2(\theta_0) \right\} \\ &= 1 + E_{\theta_0} \log \sigma_1^2(\theta_0). \end{aligned}$$

Assumption A4 ensures the underlying GARCH process is second order stationary. Hence there exists some positive number $s \in (0, 1)$, such that $E_{\theta_0} \sigma_1^{2s}(\theta_0) < \infty$. By Jensen's inequality we have

$$E_{\theta_0} \log \sigma_1^2(\theta_0) = E_{\theta_0} \frac{1}{s} \log \sigma_1^{2s}(\theta_0) \leq \frac{1}{s} \log E_{\theta_0} \sigma_1^{2s}(\theta_0) < \infty,$$

which implies that $E_{\theta_0} l_1(\theta)$ is finite at $\theta = \theta_0$. In summary, $E_{\theta_0} l_1(\theta)$ is well defined in $\mathbb{R} \cup \{+\infty\}$.

The Expectation is Minimized at $a_1(\theta_0)$ and $\sigma_1^2(\theta_0)$

Now for an arbitrary $\theta \in \Theta$, we have

$$\begin{aligned}
E_{\theta_0} l_1(\theta) - E_{\theta_0} l_1(\theta_0) &= E_{\theta_0}\left\{\frac{a_1^2(\theta)}{\sigma_1^2(\theta)} - \frac{a_1^2(\theta_0)}{\sigma_1^2(\theta_0)}\right\} + E_{\theta_0}\log\frac{\sigma_1^2(\theta)}{\sigma_1^2(\theta_0)} \\
&= E_{\theta_0}\left\{\frac{a_1^2(\theta)}{\sigma_1^2(\theta)} - \frac{\sigma_1^2(\theta_0)\eta_1^2}{\sigma_1^2(\theta_0)}\right\} + E_{\theta_0}\log\frac{\sigma_1^2(\theta)}{\sigma_1^2(\theta_0)} \\
&= E_{\theta_0}\frac{a_1^2(\theta)}{\sigma_1^2(\theta)} - 1 + E_{\theta_0}\left\{\log\frac{\sigma_1^2(\theta)}{\sigma_1^2(\theta_0)} + \frac{\sigma_1^2(\theta_0)}{\sigma_1^2(\theta)} - \frac{\sigma_1^2(\theta_0)}{\sigma_1^2(\theta)}\right\} \\
&= E_{\theta_0}\left\{\frac{a_1^2(\theta)}{\sigma_1^2(\theta)} - \frac{\sigma_1^2(\theta_0)}{\sigma_1^2(\theta)}\right\} + E_{\theta_0}\left\{\log\frac{\sigma_1^2(\theta)}{\sigma_1^2(\theta_0)} + \frac{\sigma_1^2(\theta_0)}{\sigma_1^2(\theta)} - 1\right\}.
\end{aligned}$$

First consider the second term on the righthand side. Notice the fact that

$$\log x \le x - 1, \ \forall x > 0$$

with the equality if and only if $x = 1$. Therefore:

$$E_{\theta_0}\left\{\log\frac{\sigma_1^2(\theta)}{\sigma_1^2(\theta_0)} + \frac{\sigma_1^2(\theta_0)}{\sigma_1^2(\theta)} - 1\right\} \ge E_{\theta_0}\left\{\log\frac{\sigma_1^2(\theta)}{\sigma_1^2(\theta_0)} + \log\frac{\sigma_1^2(\theta_0)}{\sigma_1^2(\theta)}\right\} = 0$$

with equality holds if and only if

$$\frac{\sigma_1^2(\theta_0)}{\sigma_1^2(\theta)} = 1, \quad i.e.\ \sigma_1^2(\theta) = \sigma_1^2(\theta_0).$$

Now consider the first term

$$
\begin{aligned}
E_{\theta_0}\left\{\frac{a_1^2(\theta)}{\sigma_1^2(\theta)}-\frac{\sigma_1^2(\theta_0)}{\sigma_1^2(\theta)}\right\} &= E_{\theta_0}\left\{\frac{a_1^2(\theta)}{\sigma_1^2(\theta)}-\frac{a_1^2(\theta_0)}{\sigma_1^2(\theta)}+\frac{a_1^2(\theta_0)}{\sigma_1^2(\theta)}-\frac{\sigma_1^2(\theta_0)}{\sigma_1^2(\theta)}\right\} \\
&= E_{\theta_0}\left\{\frac{a_1^2(\theta)}{\sigma_1^2(\theta)}-\frac{a_1^2(\theta_0)}{\sigma_1^2(\theta)}\right\}+E_{\theta_0}\frac{\sigma_1^2(\theta_0)\eta_1^2-\sigma_1^2(\theta_0)}{\sigma_1^2(\theta)} \\
&= E_{\theta_0}\left\{\frac{a_1^2(\theta)}{\sigma_1^2(\theta)}-\frac{a_1^2(\theta_0)}{\sigma_1^2(\theta)}\right\}+E_{\theta_0}\frac{\sigma_1^2(\theta_0)}{\sigma_1^2(\theta)}E_{\theta_0}(\eta_1^2-1) \\
&= E_{\theta_0}\frac{a_1^2(\theta)-a_1^2(\theta_0)}{\sigma_1^2(\theta)} \\
&= E_{\theta_0}\frac{(a_1(\theta)-a_1(\theta_0))^2}{\sigma_1^2(\theta)}+E_{\theta_0}\frac{2a_1(\theta)a_1(\theta_0)-2a_1^2(\theta_0)}{\sigma_1^2(\theta)} \\
&= E_{\theta_0}\frac{(a_1(\theta)-a_1(\theta_0))^2}{\sigma_1^2(\theta)}+E_{\theta_0}\frac{2\eta_1\sigma_1(\theta_0)(a_1(\theta)-a_1(\theta_0))}{\sigma_1^2(\theta)}.
\end{aligned}
$$

Note that for the third step above we used the fact η_1 is independent of $\sigma_1^2(\theta_0)$ and $\sigma_1^2(\theta)$. We know for the first term on the righthand side of the equation:

$$
E_{\theta_0}\frac{(a_1(\theta)-a_1(\theta_0))^2}{\sigma_1^2(\theta)} \geq 0
$$

with equality holds if and only if $a_1(\theta) = a_1(\theta_0)$. We also know that $\sigma_1(\theta_0)$ and $\sigma_1(\theta)$ belong to the information set at time $t = 0$, as well as $a_1(\theta) - a_1(\theta_0) = \lambda_0 - \lambda + \delta_0\sigma_1(\theta_0) - \delta\sigma_1(\theta)$. Hence

$$
E_{\theta_0}\left\{\frac{2\eta_1\sigma_1(\theta_0)(a_1(\theta)-a_1(\theta_0))}{\sigma_1^2(\theta)}\right\} = E_{\theta_0}(\eta_1)E_{\theta_0}\left\{\frac{2\sigma_1(\theta_0)(a_1(\theta)-a_1(\theta_0))}{\sigma_1^2(\theta)}\right\} = 0.
$$

Therefore

$$E_{\theta_0}\left\{\frac{a_1^2(\theta)}{\sigma_1^2(\theta)} - \frac{\sigma_1^2(\theta_0)}{\sigma_1^2(\theta)}\right\} \geq 0$$

with equality holds if and only if $a_1(\theta) = a_1(\theta_0)$. Along with the previous results we have shown that

$$E_{\theta_0} l_1(\theta) - E_{\theta_0} l_1(\theta_0) \geq 0, \ \forall \ \theta \in \Theta$$

with equality holds if and only if

$$a_1(\theta) = a_1(\theta_0), \quad \sigma_1^2(\theta) = \sigma_1^2(\theta_0).$$

Identifiability of The Parameters

The last step is to show the above equations imply $\theta = \theta_0$. Denote $\sigma_1^2(\theta) = \sigma_1^2(\theta_0) = \sigma_1^2$. Since $a_1(\theta) = a_1(\theta_0)$ we have

$$y_1 - \lambda - \delta\sigma_1 = y_1 - \lambda_0 - \delta_0\sigma_1,$$

which implies

$$\lambda - \lambda_0 = (\delta_0 - \delta)\sigma_1.$$

If $\delta_0 - \delta \neq 0$ then $\sigma_1 = \frac{\lambda - \lambda_0}{\delta_0 - \delta}$ is a constant. Because $\sigma_t^2(\theta_0)$ is a strictly stationary and ergodic process we know $\sigma_t^2(\theta_0)$ remains to be a constant at all time. On the other hand, based on the GARCH-in-mean specification:

$$\sigma_t^2(\theta_0) = \omega_0 + \alpha_0\epsilon_{t-1}^2 + \beta_0\sigma_{t-1}^2(\theta_0) = \omega_0 + \alpha_0\sigma_{t-1}^2(\theta_0)\eta_{t-1}^2 + \beta_0\sigma_{t-1}^2(\theta_0).$$

If $\sigma_t^2(\theta_0)$ is a constant, from the above relation η_t^2 needs to be a constant as well. This contradicts the assumption that η_t^2 has a non-degenerate distribution. Therefore we have $\delta_0 - \delta = 0$, i.e. $\delta_0 = \delta$ and as a result, $\lambda_0 = \lambda$.

Now we consider the other parameters within the θ. Define polynomials

$$\mathcal{A}_\theta(z) = \alpha z,\ \mathcal{B}_\theta(z) = 1 - \beta z$$
$$\mathcal{A}_{\theta_0}(z) = \alpha_0 z,\ \mathcal{B}_{\theta_0}(z) = 1 - \beta_0 z.$$

Denote L the lag operator, we have the following representations:

$$\mathcal{B}_\theta(L)\sigma_1^2(\theta) = \omega + \mathcal{A}_\theta(L)a_0(\theta)$$
$$\mathcal{B}_{\theta_0}(L)\sigma_1^2(\theta_0) = \omega_0 + \mathcal{A}_{\theta_0}(L)a_0(\theta_0).$$

Notice $a_1(\theta) = a_1(\theta_0) = \epsilon_1$. Along with the invertibility assumption $\beta < 1$ we have the following representations:

$$\sigma_1^2(\theta) = \frac{\omega}{\mathcal{B}_\theta(1)} + \frac{\mathcal{A}_\theta(L)}{\mathcal{B}_\theta(L)}\epsilon_1^2$$
$$\sigma_1^2(\theta_0) = \frac{\omega_0}{\mathcal{B}_{\theta_0}(1)} + \frac{\mathcal{A}_{\theta_0}(L)}{\mathcal{B}_{\theta_0}(L)}\epsilon_1^2.$$

Based on the fact that $\sigma_1^2(\theta) = \sigma_1^2(\theta_0)$ we have:

$$\left\{ \frac{\mathcal{A}_\theta(L)}{\mathcal{B}_\theta(L)} - \frac{\mathcal{A}_{\theta_0}(L)}{\mathcal{B}_{\theta_0}(L)} \right\} \epsilon_1^2 = \frac{\omega_0}{\mathcal{B}_{\theta_0}(1)} - \frac{\omega}{\mathcal{B}_\theta(1)}.$$

If $\frac{\mathcal{A}_\theta(z)}{\mathcal{B}_\theta(z)} - \frac{\mathcal{A}_{\theta_0}(z)}{\mathcal{B}_{\theta_0}(z)} \neq 0$ within the unit circle, there would exist a constant linear combination of the ϵ_{1-j}^2, $j \geq 0$. In this case $\epsilon_1^2|\psi_0$ would be a constant and it would imply

$$\epsilon_1^2 - E_{\theta_0}(\epsilon_1^2|\psi_0) = 0,$$

where ψ_0 denotes the information set at time $t = 0$. However, this contradicts the condition that η_1^2 has a non-degenerate distribution as we recognize under such assumption

$$\epsilon_1^2 - E_{\theta_0}(\epsilon_1^2|\psi_0) = \sigma_1^2(\theta_0)(\eta_1^2 - 1) \neq 0 \ \text{ with positive probability.}$$

Therefore for all $|z| < 1$:

$$\frac{\mathcal{A}_\theta(z)}{\mathcal{B}_\theta(z)} = \frac{\mathcal{A}_{\theta_0}(z)}{\mathcal{B}_{\theta_0}(z)}, \quad \frac{\omega}{\mathcal{B}_\theta(1)} = \frac{\omega_0}{\mathcal{B}_{\theta_0}(1)}.$$

Under Assumption A4 the polynomials $\mathcal{A}_\theta(z)$ and $\mathcal{B}_\theta(z)$ do not have common roots. Therefore it follows that $\mathcal{A}_\theta(z) = \mathcal{A}_{\theta_0}(z)$, $\mathcal{B}_\theta(z) = \mathcal{B}_{\theta_0}(z)$ and $\omega = \omega_0$.

Thus we have shown that $\sigma_t^2(\theta) = \sigma_t^2(\theta_0)$ implies

$$(\omega,\ \alpha,\ \beta)' = (\omega_0,\ \alpha_0,\ \beta_0)'$$

Together with the previous results, we have shown that $a_1(\theta) = a_1(\theta_0)$ and $\sigma_1^2(\theta) = \sigma_1^2(\theta_0)$ imply $\theta = \theta_0$.

To sum up, the expectation $E_{\theta_0} l_1(\theta)$ is uniquely minimized at $\theta = \theta_0$. □

3.4.2 Convergence to the criterion

Lemma 3.4. *Under Assumptions (**A1**) - (**A4**) and (**B1**) - (**B2**),*

$$\tilde{I}_n(\theta) \to E_{\theta_0} l_1(\theta) \quad a.s. \quad ,\theta \in \Theta$$

Proof. Under the above assumptions, we know from Theorem 2.8 that the multivariate process $(y_t,\ \sigma_t^2(\theta_0),\ \sigma_t^2(\theta))'$ is stationary and geometrically ergodic. Since $l_t(\theta)$ is a measurable function of this process, it is also ergodic.

From lemma 3.3, there are two cases that require consideration, depending on whether the expectation $E_{\theta_0} l_1(\theta)$ is finite.

Case 1: when $E_{\theta_0} l_1(\theta)$ is finite

Due to the stationarity and ergodicity of $l_t(\theta)$, we can incur the standard ergodic theorem for a stationary sequence, cf. Doob (1990), which yields:

$$\frac{1}{n}\sum_{t=1}^{n} l_t(\theta) \to E_{\theta_0} l_1(\theta) \quad a.s. \ ,$$

where $l_t(\theta)$ is assumed to start with the invariant measure π.

$\tilde{l}_t(\theta)$ is exactly the same sequence except that it starts with some arbitrarily assigned initial measure. According to Proposition 17.1.6 of Meyn and Tweedie (2009), the above convergence result also holds for any initial distribution provided the chain is Harris recurrent and geometrically ergodic. Therefore we have

$$\frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\theta) \to E_{\theta_0} l_1(\theta) \quad a.s. \ .$$

Case 2: when $E_{\theta_0} l_1(\theta)$ is positive infinite

When $E_{\theta_0} l_1(\theta) = +\infty$ the ergodic theorem does not apply directly. We consider the following truncated sequences

$$l_t(\theta, k) = l_t(\theta) I_{\{l_t(\theta) \leq k\}}, \quad \tilde{l}_t(\theta, k) = \tilde{l}_t(\theta) I_{\{l_t(\theta) \leq k\}},$$

where $k > 0$ and $k \to +\infty$. For all $k > 0$ we have $l_t(\theta, k) \leq k$, and hence the expectation of $l_t(\theta, k)$ is finite. By the standard ergodic theorem we know

$$\frac{1}{n}\sum_{t=1}^{n} l_t(\theta, k) \to E_{\theta_0} l_1(\theta, k) \ \ a.s. \quad (n \to \infty).$$

We apply the same argument in the previous case and Proposition 17.1.6 of Meyn and Tweedie (2009) yields:

$$\frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\theta, k) \to E_{\theta_0} l_1(\theta, k) \ \ a.s., \quad (n \to \infty).$$

When $k \to \infty$, notice that $\tilde{l}_1(\theta, k) \to \tilde{l}_1(\theta)$. Therefore by Beppo Levi's theorem

$$E_{\theta_0} l_1(\theta, k) \to E_{\theta_0} l_1(\theta) = +\infty, \quad (k \to \infty).$$

Hence we conclude that

$$\frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\theta) \geq \frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\theta, k) \to +\infty, \quad (k \to \infty,\ n \to \infty).$$

In summary, when $E_{\theta_0} l_1(\theta) \in \mathbb{R} \cup \{+\infty\}$, then

$$\frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\theta) \to E_{\theta_0} l_1(\theta), \quad (n \to \infty).$$

$\square$

3.4.3 Equicontinuity of quasi-likelihood

Lemma 3.5. *Under Assumptions (**A1**) - (**A4**) and (**B1**) - (**B3**), the function $\tilde{I}_n(\theta)$ defined in Equation (3.2) is equicontinous.*

Proof. Suppose an open ball $B(\theta, k)$ within the interior of Θ, with center θ and radius k. Then for any $\theta^{(1)}$, $\theta^{(2)}$ in $B(\theta, k)$, we apply the mean value theorem:

$$\begin{aligned}\left|\tilde{I}_n(\theta^{(1)}) - \tilde{I}_n(\theta^{(2)})\right| &\leq \frac{1}{n}\sum_{t=1}^{n}\left|\tilde{l}_t(\theta^{(1)}) - \tilde{l}_t(\theta^{(2)})\right| \\ &\leq \left(\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta^*\in B(\theta,k)}\left\|\tilde{M}_t(\theta^*)\right\|\right)\cdot\left\|\theta^{(1)} - \theta^{(2)}\right\|\end{aligned}$$

where $\tilde{M}_t(\theta) = \frac{\partial \tilde{l}_t(\theta)}{\partial \theta}$. By the law of large numbers for geometric ergodic sequence, we know that

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta^*\in B(\theta,k)}\left\|\tilde{M}_t(\theta^*)\right\| \overset{a.s}{\to} E\left(\sup_{\theta^*\in B(\theta,k)}\left\|\frac{\partial l_t(\theta)}{\partial\theta}\right\|\right) = E\left(\sup_{\theta^*\in B(\theta,k)}\|M_t(\theta^*)\|\right)$$

which is bounded above by M according to the assumption. On the other hand $\|\theta^{(1)} - \theta^{(2)}\|$ is also bounded above since they are restricted within the same neighborhood of θ. Therefore we know that

$$|\tilde{I}_n(\theta^{(1)}) - \tilde{I}_n(\theta^{(2)})| < M\left\|\theta^{(1)} - \theta^{(2)}\right\|$$

for any $\theta^{(1)}$ and $\theta^{(2)}$ within $B(\theta, k)$. Thus the equicontinuity property holds. □

3.4.4 Conclusion

For all $\theta \in \Theta$, let $B_k(\theta)$ be an open ball with center θ and radius $1/k$. Notice that for any neighborhood $V(\theta_0)$ of θ_0, we have

$$\limsup_{n\to\infty} \inf_{\theta^*\in V(\theta_0)} \tilde{I}_n(\theta^*) \leq \lim_{n\to\infty} \tilde{I}_n(\theta_0) = E_{\theta_0} l_1(\theta_0). \tag{3.10}$$

For some $\theta \neq \theta_0$, we could apply a similar argument as in the proof of Lemma 3.4. Applying the ergodic theorem on the sequence $\{\inf_{\theta^*\in B_k(\theta)\cap\Theta} l_t(\theta^*)\}$, we have

$$\liminf_{n\to\infty} n^{-1} \sum_{t=1}^{n} \inf_{\theta^*\in B_k(\theta)\cap\Theta} l_t(\theta^*) = E_{\theta_0} \inf_{\theta^*\in B_k(\theta)\cap\Theta} l_1(\theta^*).$$

By Beppo Levi's theorem, when $k \to \infty$ we have

$$E_{\theta_0} \inf_{\theta^*\in B_k(\theta)\cap\Theta} l_1(\theta^*) \to E_{\theta_0} l_1(\theta). \tag{3.11}$$

In the view of Lemma 3.3, we know that for any $\theta \neq \theta_0$, there exists some neighborhood $B(\theta)$ such that

$$\liminf_{n\to\infty} \inf_{\theta^*\in B(\theta)} \tilde{I}_n(\theta^*) > E_{\theta_0} l_1(\theta_0) \quad a.s. \,. \tag{3.12}$$

Note that the validity of the above equation also requires the equicontinuity of $\tilde{I}_n(\theta)$, which is established in Lemma 3.5.

Since the parameter space is compact it can be covered by unions of finite open sets. Suppose Θ is covered by the union of an arbitrary neighborhood $B(\theta_0)$ of θ_0 and a finite sequence of balls $B(\theta_1)$, …, $B(\theta_k)$ satisfying relation (3.12). Those balls form a finite subcover of Θ. It is not difficult to see that

$$\inf_{\theta\in\Theta} \tilde{I}_n(\theta) = \min_{i=0,1,\ldots,k} \inf_{\theta\in\Theta\cap B(\theta_i)} \tilde{I}_n(\theta).$$

By relations (3.10) - (3.11), we know almost surely $\hat{\theta}_n$ belongs to $V(\theta_0)$ when n goes to infinity. Since this is true for an arbitrary neighborhood of θ_0, we know $\hat{\theta}_n$ converges to θ_0 almost surely. The proof is completed.

3.5 Proof of Asymptotic Normality

In this section we prove Theorem 3.2. To establish the asymptotic normality property, one important step is to show the matrices A and B as in (3.8) are well-defined. Then we can apply certain type of limit theorems to the objects on the right-hand side of (3.7). The proof is concluded by applying the Slutsky's theorem.

3.5.1 First order derivatives

Lemma 3.6. *Under Assumptions (**A1**) - (**A4**), (**B1**) - (**B3**) and (**C1**) - **C3**) we have*

$$E\left\|\frac{\partial l_t(\theta_0)}{\partial\theta}\frac{\partial l_t(\theta_0)}{\partial\theta'}\right\| < \infty. \tag{3.13}$$

Proof. Recall the parameter form definitions:

$$l_t(\theta) = \frac{a_t^2}{\sigma_t^2} + \log \sigma_t^2, \tag{3.14}$$

where

$$\sigma_t^2 = \omega + \alpha a_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{3.15}$$

$$a_t = y_t - \lambda - \delta \sigma_t \tag{3.16}$$

and $\theta = (\lambda,\ \delta,\ \omega,\ \alpha,\ \beta)'$ with the true parameter being $\theta_0 = (\lambda_0,\ \delta_0,\ \omega_0,\ \alpha_0,\ \beta_0)'$. Moving forward we will use the following shorthand $a_t = a_t(\theta)$, $\sigma_t = \sigma_t(\theta)$, $l_t = l_t(\theta)$ unless stated otherwise.

Now take derivative with respect to θ on both sides of equation (3.14) and it yields:

$$\frac{\partial l_t}{\partial \theta} = \frac{2a_t}{\sigma_t^2} \cdot \frac{\partial a_t}{\partial \theta} - \frac{a_t^2}{\sigma_t^4} \cdot \frac{\partial \sigma_t^2}{\partial \theta} + \frac{1}{\sigma_t^2} \cdot \frac{\partial \sigma_t^2}{\partial \theta}. \tag{3.17}$$

From equation (3.16) and the relation $\frac{\partial \sigma_t^2}{\partial \theta} = 2\sigma_t \cdot \frac{\partial \sigma_t}{\partial \theta}$ we obtain:

$$\begin{aligned}
\frac{\partial a_t}{\partial \lambda} &= -1 - \frac{\delta}{2\sigma_t} \cdot \frac{\partial \sigma_t^2}{\partial \lambda} \\
\frac{\partial a_t}{\partial \delta} &= -\sigma_t - \frac{\delta}{2\sigma_t} \cdot \frac{\partial \sigma_t^2}{\partial \delta} \\
\frac{\partial a_t}{\partial u_i} &= -\frac{\delta}{2\sigma_t} \cdot \frac{\partial \sigma_t^2}{\partial u_i},
\end{aligned}$$

where u_i denotes an arbitrary pure GARCH parameter (any parameter other than

$\lambda, \delta)$. Let $d(\theta) = (-1,\ -\sigma_t,\ 0,\ 0,\ 0)'$, then we may re-write the above results as:

$$\frac{\partial a_t}{\partial \theta} = d(\theta) - \frac{\delta}{2\sigma_t} \cdot \frac{\partial \sigma_t^2}{\partial \theta}. \tag{3.18}$$

Now substitute (3.18) back into equation (3.17), then evaluate the derivative at the true parameter value $\theta = \theta_0$:

$$\begin{aligned}
\frac{\partial l_t(\theta_0)}{\partial \theta} &= \frac{2a_t(\theta_0)}{\sigma_t^2(\theta_0)} \cdot \left[d(\theta_0) - \frac{\delta_0}{2\sigma_t(\theta_0)} \cdot \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \right] - \frac{a_t^2(\theta_0)}{\sigma_t^4(\theta_0)} \cdot \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} + \frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \\
&= \frac{2\eta_t}{\sigma_t(\theta_0)} \cdot d(\theta_0) - \frac{\eta_t \delta_0}{\sigma_t^2(\theta_0)} \cdot \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} + \frac{1-\eta_t^2}{\sigma_t^2(\theta_0)} \cdot \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \\
&= \frac{2\eta_t}{\sigma_t(\theta_0)} \cdot d(\theta_0) + \frac{1-\eta_t^2 - \eta_t \delta_0}{\sigma_t^2(\theta_0)} \cdot \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta}.
\end{aligned}$$

Note that the above calculation involves the fact that

$$a_t(\theta_0) = \epsilon_t = \sigma_t(\theta_0)\eta_t.$$

Given the above results we have:

$$\begin{aligned}
\frac{\partial l_t(\theta_0)}{\partial \theta} \frac{\partial l_t(\theta_0)}{\partial \theta'} &= 4\eta_t^2 \cdot \frac{d(\theta_0)}{\sigma_t(\theta_0)} \frac{d'(\theta_0)}{\sigma_t(\theta_0)} + \frac{(1-\eta_t^2-\eta_t\delta_0)^2}{\sigma_t^4(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'} \\
&\quad + \frac{2\eta_t(1-\eta_t^2-\eta_t\delta_0)}{\sigma_t^2(\theta_0)} \cdot \left[\frac{d(\theta_0)}{\sigma_t(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'} + \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \frac{d'(\theta_0)}{\sigma_t(\theta_0)} \right].
\end{aligned} \tag{3.19}$$

We consider the above equation term by term. First notice that $\frac{1}{\sigma_t(\theta_0)^k} \leq \frac{1}{\omega_0^k}$

for any $k > 0$. Therefore we know

$$E\left\|\frac{d(\theta_0)}{\sigma_t(\theta_0)}\right\| = E\left\|(-\frac{1}{\sigma_t(\theta_0)} \quad -1 \;\; 0 \; \dots \; 0)\right\| < \infty.$$

It is also not difficult to verify that

$$E\left\|\frac{d'(\theta_0)}{\sigma_t(\theta_0)}\right\| < \infty, \quad E\left\|\frac{d(\theta_0)}{\sigma_t(\theta_0)}\frac{d'(\theta_0)}{\sigma_t(\theta_0)}\right\| < \infty.$$

Consider the right hand side of equation (3.19). Given Assumption C2, it is not difficult to realize that the first term, the first parts of the second and third terms have finite expectations. Assume in addition we have

$$E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\right\| < \infty,$$

which would also imply

$$E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\| < \infty, \quad E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\right\| < \infty.$$

Therefore one could easily verify that every term on the righthand side of (3.19) has finite expectations. By applying a simple triangular inequality type of argument we could show that $\|\frac{\partial l_t(\theta_0)}{\partial\theta}\frac{\partial l_t(\theta_0)}{\partial\theta'}\|$ has a finite expectation thus inequality (3.13) holds.

On the other hand, applying the Cauchy-Schwarz inequality gives us:

$$E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\right\| \leq \sqrt{E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\|^2}\cdot\sqrt{E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\right\|^2},$$

which suggests that $E\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\|$ is finite if we have

$$E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\|^2 < \infty. \tag{3.20}$$

Based on the above discussion, we only need to establish equation (3.20) in order to show that (3.13) holds.

Now, by (3.14) - (3.15) and (3.18):

$$\begin{aligned}
\frac{\partial\sigma_t^2}{\partial\omega} &= 1 + 2\alpha a_{t-1}\frac{\partial a_{t-1}}{\partial\omega} + \beta\frac{\partial\sigma_{t-1}^2}{\partial\omega}\\
&= 1 + 2\alpha a_{t-1}(-\frac{\delta}{2\sigma_t}\cdot\frac{\partial\sigma_{t-1}^2}{\partial\omega}) + \beta\frac{\partial\sigma_{t-1}^2}{\partial\omega}\\
&= 1 - \alpha\delta\frac{a_{t-1}}{\sigma_{t-1}}\frac{\partial\sigma_{t-1}^2}{\partial\omega} + \beta\frac{\partial\sigma_{t-1}^2}{\partial\omega}\\
&= 1 + (\beta - \alpha\delta\frac{a_{t-1}}{\sigma_{t-1}})\frac{\partial\sigma_{t-1}^2}{\partial\omega}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{\partial\sigma_t^2}{\partial\alpha} &= a_{t-1}^2 + \alpha\frac{\partial a_{t-1}^2}{\partial\alpha} + \beta\frac{\partial\sigma_{t-1}^2}{\partial\alpha}\\
&= a_{t-1}^2 + 2\alpha a_{t-1}(-\frac{\delta}{2\sigma_{t-1}}\frac{\partial\sigma_{t-1}^2}{\partial\alpha}) + \beta\frac{\partial\sigma_{t-1}^2}{\partial\alpha}\\
&= a_{t-1}^2 + (\beta - \alpha\delta\frac{a_{t-1}}{\sigma_{t-1}})\frac{\partial\sigma_{t-1}^2}{\partial\alpha},
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial \sigma_t^2}{\partial \beta} &= \alpha \frac{\partial a_{t-1}^2}{\partial \beta} + \beta \frac{\partial \sigma_{t-1}^2}{\partial \beta} + \sigma_{t-1}^2 \\
&= \sigma_{t-1}^2 + 2\alpha a_{t-1}(-\frac{\delta}{2\sigma_{t-1}} \frac{\partial \sigma_{t-1}^2}{\partial \beta}) + \beta \frac{\partial \sigma_{t-1}^2}{\partial \beta} \\
&= \sigma_{t-1}^2 + (\beta - \alpha\delta \frac{a_{t-1}}{\sigma_{t-1}}) \frac{\partial \sigma_{t-1}^2}{\partial \beta}, \\
\frac{\partial \sigma_t^2}{\partial \lambda} &= \alpha \frac{\partial a_{t-1}^2}{\partial \lambda} + \beta_0 \frac{\partial \sigma_{t-1}^2}{\partial \lambda} \\
&= 2\alpha a_{t-1}(-1 - \frac{\delta}{2\sigma_{t-1}} \frac{\partial \sigma_{t-1}^2}{\partial \lambda}) + \beta \frac{\partial \sigma_{t-1}^2}{\partial \lambda} \\
&= -2\alpha a_{t-1} + (\beta - \alpha\delta \frac{a_{t-1}}{\sigma_{t-1}}) \frac{\partial \sigma_{t-1}^2}{\partial \lambda}, \\
\frac{\partial \sigma_t^2}{\partial \delta} &= \alpha \frac{\partial a_{t-1}^2}{\partial \delta} + \beta \frac{\partial \sigma_{t-1}^2}{\partial \delta} \\
&= 2\alpha a_{t-1}(-\sigma_{t-1} - \frac{\delta}{2\sigma_{t-1}} \frac{\partial \sigma_{t-1}^2}{\partial \delta}) + \beta \frac{\partial \sigma_{t-1}^2}{\partial \delta} \\
&= -2\alpha a_{t-1}\sigma_{t-1} + (\beta - \alpha\delta \frac{a_{t-1}}{\sigma_{t-1}}) \frac{\partial \sigma_{t-1}^2}{\partial \delta}.
\end{aligned}
$$

Denote θ_k an arbitrary element of the parameter vector θ. Observing the above results we realize that $\frac{\partial \sigma_t^2}{\partial \theta_k}$ satisfies a general recursive equation:

$$
\frac{\partial \sigma_t^2}{\partial \theta_k} = b_{t-1}(\theta) \cdot \frac{\partial \sigma_{t-1}^2}{\partial \theta_k} + e_{t-1}(\theta), \tag{3.21}
$$

where $b_t(\theta) = \beta - \alpha\delta \cdot \frac{a_t}{\sigma_t}$ and

$$e_t(\theta) = \begin{cases} 1, & \text{if } \theta_k = \omega \\ a_t^2, & \text{if } \theta_k = \alpha \\ \sigma_t^2, & \text{if } \theta_k = \beta \\ -2\alpha a_t, & \text{if } \theta_k = \lambda \\ -2\alpha a_t \sigma_t, & \text{if } \theta_k = \delta \end{cases} \cdot$$

A process defined by the recursion $X_t = a_{t-1} + b_{t-1}X_{t-1}$ with a_t, b_t being random processes is also known as the random coefficient autoregressive process. Properties of such process has been studied by a few authors, cf. Brandt (1986), Bougerol and Picard (1992b), Aue et al. (2006). The solution of such a process can be obtained by repeatedly applying the recursive relation. For our case, we know that the solution of (3.21) has the following form

$$\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta_k} = \sum_{i=1}^{\infty} e_{t-i}(\theta_0) \prod_{j=1}^{i-1} b_{t-j}(\theta_0). \tag{3.22}$$

Moreover, if $\{b_t(\theta_0)\}$, $\{e_t(\theta_0)\}$ are strictly stationary and ergodic processes, and they satisfy

$$E(\log^+ |e_0(\theta_0)|) < \infty, \quad E(\log^+ |b_0(\theta_0)|) < \infty, \quad E(\log |b_0(\theta_0)|) < 0, \tag{3.23}$$

where $\log^+ x = \max(0, \log x)$, then the solution (3.22) is also strictly stationary and ergodic.

From Theorem 2.8 we know that $b_t(\theta_0)$, $e_t(\theta_0)$ are indeed stationary and ergodic under our assumptions. Now we suppose (3.23) does hold thus the stationary solution (3.22) exists. By Minkowski inequality we know

$$\begin{aligned}\left\{E\left|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta_k}\right|^2\right\}^{1/2} &= \left\{E\left|\sum_{i=1}^{\infty} e_{t-i}(\theta_0)\prod_{j=1}^{i-1} b_{t-j}(\theta_0)\right|^2\right\}^{1/2} \\ &\leq \sum_{i=1}^{\infty}\left\{E\left|e_{t-i}(\theta_0)\prod_{j=1}^{i-1} b_{t-j}(\theta_0)\right|^2\right\}^{1/2} \\ &= \{E|e_0(\theta_0)|^2\}^{1/2}\sum_{i=1}^{\infty}\{E|b_0(\theta_0)|^2\}^{(i-1)/2}\end{aligned}$$

Therefore, if

$$E|e_0(\theta_0)|^2 < \infty,\ E|b_0(\theta_0)|^2 < 1, \tag{3.24}$$

then we have

$$E\left|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta_k}\right|^2 < \infty,$$

which shows that (3.20) holds.

Compare conditions (3.24) and (3.23). Notice that when $E|e_0(\theta_0)|^2 < \infty$ we know $E|e_0(\theta_0)| < \infty$, which also implies the first inequality of (3.23) holds considering $E(\log^+|e_0(\theta_0)|) \leq E|e_0(\theta_0)|$. Using the same argument we know $E(\log^+|b_0(\theta_0)|)$ is finite whenever $E|b_0(\theta_0)|$ is finite. Suppose $E|b_0(\theta_0)|^2 < 1$, by

Jensen's inequality we know:

$$E(\log|b_0(\theta_0)|) = \frac{1}{2}E(\log|b_0(\theta_0)|^2) \leq \frac{1}{2}\log E|b_0(\theta_0)|^2 < 0.$$

From the above argument we see that if condition (3.24) holds, conditions (3.23) are satisfied automatically.

Bollerslev (1986) has shown that the GARCH process has finite 4th moments given the second inequality of Assumption C3. Along with Assumption C2 they ensure the validity of the first inequality in (3.24). Also, by the first inequality of Assumption C3 we know

$$\begin{aligned} E|b_0(\theta_0)|^2 &= E|\beta_0 - \alpha_0\delta_0\frac{a_0(\theta_0)}{\sigma_0(\theta_0)}|^2 \\ &= E(\beta_0 - \alpha_0\delta_0\eta_0)^2 \\ &= \beta_0^2 + \alpha_0^2\delta_0^2 E(\eta_0^2) - 2\alpha_0\beta_0\delta_0 E(\eta_0) \\ &= \beta_0^2 + \alpha_0^2\delta_0^2 \\ &< 1. \end{aligned}$$

Thus (3.24) is true under the specified assumptions. As already argued (3.20) will follow thus (3.13) is proved. □

3.5.2 Second order derivatives

Lemma 3.7. *Under Assumptions (**A1**) - (**A4**), (**B1**) - (**B3**) and (**C1**) - **C3**) we have:*

$$E\left\|\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}\right\| < \infty. \tag{3.25}$$

Proof. To prove this lemma we will need to make use of some intermediate results from the last section. The first order derivative is given by equation (3.17). Taking derivative with respective to θ' yields:

$$\begin{aligned}\frac{\partial^2 l_t}{\partial\theta\partial\theta'} &= (\frac{2}{\sigma_t^2}\cdot\frac{\partial a_t}{\partial\theta'} + \frac{2a_t}{\sigma_t^4}\cdot\frac{\partial\sigma_t^2}{\partial\theta'})\cdot\frac{\partial a_t}{\partial\theta} + \frac{2a_t}{\sigma_t^2}\cdot\frac{\partial^2 a_t}{\partial\theta\partial\theta'} \\ &- (\frac{2a_t}{\sigma_t^4}\cdot\frac{\partial a_t}{\partial\theta'} - \frac{2a_t^2}{\sigma_t^6}\cdot\frac{\partial\sigma_t^2}{\partial\theta'})\cdot\frac{\partial\sigma_t^2}{\partial\theta} - \frac{a_t^2}{\sigma_t^4}\cdot\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta'} \\ &- \frac{1}{\sigma_t^4}\cdot\frac{\partial\sigma_t^2}{\partial\theta'}\cdot\frac{\partial\sigma_t^2}{\partial\theta} + \frac{1}{\sigma_t^2}\cdot\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta'}.\end{aligned} \tag{3.26}$$

In order to study the finiteness of the expectation of this object we will investigate each term on the right hand side of the above equation. By equation (3.18) we know that

$$\begin{aligned}\frac{\partial\sigma_t^2}{\partial\theta'}\cdot\frac{\partial a_t}{\partial\theta} &= \frac{\partial\sigma_t^2}{\partial\theta'}\cdot d(\theta) - \frac{\delta}{2\sigma_t}\cdot\frac{\partial\sigma_t^2}{\partial\theta'}\cdot\frac{\partial\sigma_t^2}{\partial\theta} \\ \frac{\partial a_t}{\partial\theta'}\cdot\frac{\partial a_t}{\partial\theta} &= \left[d'(\theta) - \frac{\delta}{2\sigma_t}\cdot\frac{\partial\sigma_t^2}{\partial\theta'}\right]\cdot\left[d(\theta) - \frac{\delta}{2\sigma_t}\cdot\frac{\partial\sigma_t^2}{\partial\theta}\right] \\ &= d'(\theta)d(\theta) + \frac{\delta^2}{4\sigma_t^2}\cdot\frac{\partial\sigma_t^2}{\partial\theta'}\cdot\frac{\partial\sigma_t^2}{\partial\theta} - \frac{\delta}{2}\left[\frac{d'(\theta)}{\sigma_t}\cdot\frac{\partial\sigma_t^2}{\partial\theta} + \frac{\partial\sigma_t^2}{\partial\theta'}\cdot\frac{d(\theta)}{\sigma_t}\right].\end{aligned}$$

Notice that $\frac{1}{\sigma_t^k} \leq \frac{1}{\omega^k}$ for some $k > 1$. Using the same arguments from the last section it is not difficult to show that objects like $\|\frac{d(\theta)}{\sigma_t}\|$, $\|\frac{d'(\theta)d(\theta)}{\sigma_t^2}\|$ are well-bounded above. From the proof of Lemma 3.6, we also know that under our assumptions:

$$E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\right\| < \infty, \quad E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\|^2 < \infty.$$

Based on those facts it is obvious that:

$$\begin{aligned} E\left\|\frac{a_t(\theta_0)}{\sigma_t^4(\theta_0)}\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\cdot\frac{\partial a_t(\theta_0)}{\partial\theta}\right\| &\le \frac{E|\eta_t|}{\omega_0^2}\cdot E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\right\|\cdot E\left\|\frac{d(\theta_0)}{\sigma_t(\theta_0)}\right\| \\ &+ \frac{E|\eta_t|}{\omega_0^2}\cdot\frac{\delta_0}{2\omega_0^2}\cdot E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\| \\ &< \infty \end{aligned}$$

and

$$\begin{aligned} E\left\|\frac{1}{\sigma_t^2(\theta_0)}\cdot\frac{\partial a_t(\theta_0)}{\partial\theta'}\cdot\frac{\partial a_t(\theta_0)}{\partial\theta}\right\| &\le E\left\|\frac{d'(\theta_0)d(\theta_0)}{\sigma_t^2(\theta_0)}\right\| + \frac{\delta_0^2}{4\omega_0^4}\cdot E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\| \\ &+ \frac{\delta_0}{2\omega_0^2}\cdot\left[E\left\|\frac{d'(\theta_0)}{\sigma_t(\theta_0)}\right\|\cdot E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\| + E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\right\|\cdot E\left\|\frac{d(\theta_0)}{\sigma_t(\theta_0)}\right\|\right] \\ &< \infty. \end{aligned}$$

Therefore, for the first term on the right-hand side of equation (3.26):

$$E\left\|\left[\frac{2}{\sigma_t^2(\theta_0)}\cdot\frac{\partial a_t(\theta_0)}{\partial\theta'} + \frac{2a_t(\theta_0)}{\sigma_t^4(\theta_0)}\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\right]\cdot\frac{\partial a_t(\theta_0)}{\partial\theta}\right\| < \infty.$$

We could use similar arguments and conclude that

$$\begin{aligned}
E\left\|\frac{2a_t(\theta_0)}{\sigma_t^4(\theta_0)}\cdot\frac{\partial a_t(\theta_0)}{\partial\theta'}\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\| &= E\left\|\frac{2a_t(\theta_0)}{\sigma_t^3(\theta_0)}\cdot(\frac{d'(\theta_0)}{\sigma_t(\theta_0)}-\frac{\delta_0}{2\sigma_t^2(\theta_0)}\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'})\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\| \\
&\leq \frac{2E|\eta_t|}{\omega_0^2}\cdot E\left\|\frac{d'(\theta_0)}{\sigma_t(\theta_0)}\right\|+\frac{\delta_0}{2\omega_0^2}\cdot E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\| \\
&< \infty
\end{aligned}$$

and

$$\begin{aligned}
E\left\|\frac{2a_t^2(\theta_0)}{\sigma_t^6(\theta_0)}\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\| &\leq \frac{2E(\eta_t^2)}{\omega_0^4}E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\| \\
&< \infty.
\end{aligned}$$

Therefore for the 3rd term on the right-hand side of equation (3.26):

$$E\left\|\left[\frac{2a_t(\theta_0)}{\sigma_t^4(\theta_0)}\cdot\frac{\partial a_t(\theta_0)}{\partial\theta'}-\frac{2a_t^2(\theta_0)}{\sigma_t^6(\theta_0)}\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\right]\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\| < \infty.$$

Now consider the 5th term on the right-hand side of equation (3.26), obviously:

$$E\left\|\frac{1}{\sigma_t^4(\theta_0)}\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\| \leq \frac{1}{\omega_0^4}E\left\|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\|,$$

which is also finite. For the 2nd, 4th and 6th terms on the right-hand side of equation (3.26) we know:

$$E\left\|\frac{2a_t(\theta_0)}{\sigma_t^2(\theta_0)}\cdot\frac{\partial^2 a_t(\theta_0)}{\partial\theta\partial\theta'}\right\| \le \frac{2E|\eta_t|}{\omega_0}\cdot E\left\|\frac{\partial^2 a_t(\theta_0)}{\partial\theta\partial\theta'}\right\|,$$
$$E\left\|\frac{a_t^2(\theta_0)}{\sigma_t^4(\theta_0)}\cdot\frac{\partial^2 \sigma_t^2(\theta_0)}{\partial\theta\partial\theta'}\right\| \le \frac{2E(\eta_t^2)}{\omega_0^2}\cdot E\left\|\frac{\partial^2 \sigma_t^2(\theta_0)}{\partial\theta\partial\theta'}\right\|,$$
$$E\left\|\frac{1}{\sigma_t^2(\theta_0)}\cdot\frac{\partial^2 \sigma_t^2(\theta_0)}{\partial\theta\partial\theta'}\right\| \le \frac{1}{\omega_0^2}\cdot E\left\|\frac{\partial^2 \sigma_t^2(\theta_0)}{\partial\theta\partial\theta'}\right\|.$$

Based on the above discussion, inequality (3.25) can be shown by applying a triangular inequality type of argument on (3.26), if we have the following 2 additional conditions

$$E\left\|\frac{\partial^2 \sigma_t^2(\theta_0)}{\partial\theta\partial\theta'}\right\| < \infty, \quad E\left\|\frac{\partial^2 a_t(\theta_0)}{\partial\theta\partial\theta'}\right\| < \infty.$$

First consider the second inequality above. From equation (3.18) we know:

$$\frac{\partial^2 a_t}{\partial\theta\partial\theta_j} = \frac{\partial d(\theta)}{\partial\theta_j} - \frac{\delta}{2\sigma_t}\cdot\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta_j} + \frac{\delta}{4\sigma_t^3}\cdot\frac{\partial\sigma_t^2}{\partial\theta_j}\cdot\frac{\partial\sigma_t^2}{\partial\theta}, \tag{3.27}$$

when $\theta_j \neq \delta$, and

$$\begin{aligned}\frac{\partial^2 a_t}{\partial\theta\partial\delta} &= \frac{\partial d(\theta)}{\partial\delta} - \frac{\delta}{2\sigma_t}\cdot\frac{\partial^2\sigma_t^2}{\partial\theta\partial\delta} - \frac{\sigma_t - \frac{\delta}{2\sigma_t}\cdot\frac{\partial\sigma_t^2}{\partial\delta}}{2\sigma_t^2}\cdot\frac{\partial\sigma_t^2}{\partial\theta} \\ &= \frac{\partial d(\theta)}{\partial\delta} - \frac{\delta}{2\sigma_t}\cdot\frac{\partial^2\sigma_t^2}{\partial\theta\partial\delta} - \frac{1}{2\sigma_t}\cdot\frac{\partial\sigma_t^2}{\partial\theta} + \frac{\delta}{4\sigma_t^3}\cdot\frac{\partial\sigma_t^2}{\partial\delta}\cdot\frac{\partial\sigma_t^2}{\partial\theta}.\end{aligned} \tag{3.28}$$

Notice that according to the definition of $d(\theta)$ in (3.18):

$$E\left\|\frac{\partial d(\theta)}{\partial \theta'}\right\| = E\left\|(0,\ \frac{\partial \sigma_t(\theta)}{\partial \theta'},\ 0,\ \ldots,\ 0)'\right\|$$

which is finite at θ_0 from the proof of Lemma 3.6. For the other objects in equation (3.27) - (3.28):

$$E\left\|\frac{\delta_0}{2\sigma_t(\theta_0)}\right\| \leq \frac{|\delta_0|}{2\sqrt{\omega_0}} < \infty, \quad E\left\|\frac{\delta_0}{4\sigma_t^3(\theta_0)}\right\| \leq \frac{|\delta_0|}{4\omega_0^{3/2}} < \infty.$$

In the view of (3.27) - (3.28), we know that

$$E\left\|\frac{\partial^2 a_t(\theta_0)}{\partial \theta \partial \theta'}\right\| < \infty,$$

whenever

$$E\left\|\frac{\partial^2 \sigma_t^2(\theta_0)}{\partial \theta \partial \theta'}\right\| < \infty. \tag{3.29}$$

Therefore to show that inequality (3.25) holds we only need to show the validity of inequality (3.29).

Denoting θ_k, θ_l to be arbitrary elements of the parameter vector θ, we want to consider objects like $\frac{\partial^2 \sigma_t^2}{\partial \theta_k \partial \theta_l}$. Based on equality (3.21) we have

$$\frac{\partial^2 \sigma_t^2}{\partial \theta_k \partial \theta_l} = \frac{\partial e_{t-1}(\theta)}{\partial \theta_l} + \frac{\partial b_{t-1}(\theta)}{\partial \theta_l} \cdot \frac{\partial \sigma_{t-1}^2}{\partial \theta_k} + b_{t-1}(\theta) \cdot \frac{\partial^2 \sigma_{t-1}^2}{\partial \theta_k \partial \theta_l}.$$

We want to use the same arguments as in the last section. Notice that the above equation also defines a random coefficient autoregressive model

$$\frac{\partial^2 \sigma_t^2}{\partial\theta_k \partial\theta_l} = b_{t-1}(\theta) \cdot \frac{\partial^2 \sigma_{t-1}^2}{\partial\theta_k \partial\theta_l} + e_{t-1}^*(\theta), \tag{3.30}$$

where

$$e_t^*(\theta) = \frac{\partial e_t(\theta)}{\partial\theta_l} + \frac{\partial b_t(\theta)}{\partial\theta_l} \cdot \frac{\partial\sigma_t^2}{\partial\theta_k}$$

with function $e_t(\theta)$ being defined in (3.21). This representation is in the same fashion of equation (3.21) of the last section. Applying the same arguments used before, we know that

$$E\left|\frac{\partial^2 \sigma_t^2}{\partial\theta_k \partial\theta_l}\right| < \infty$$

if we have the following:

$$E|e_t^*(\theta_0)| < \infty,\ E|b_t(\theta_0)| < 1. \tag{3.31}$$

Those two conditions could be derived by applying the Minkowski's inequality in the same fashion as when we were deriving (3.24). To verify the second inequality above we follow the same argument as in the last subsection given Assumption A4. To show the first inequality, from the proof of Lemma 3.6 we know that

$$E|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta_k}| < \infty$$

under the assumed conditions. Therefore to show (3.31) holds we only need to prove the following:

$$E|\frac{\partial e_t(\theta_0)}{\partial \theta_l}| < \infty,\ E|\frac{\partial b_t(\theta_0)}{\partial \theta_l}| < \infty. \tag{3.32}$$

First we consider the object $\frac{\partial b_t(\theta_0)}{\partial \theta_l}$. Based on the definition of b_t in (3.21) we can calculate all of its first order partial derivatives. The particular form varies as we take θ_l to be different elements of θ. For example, when $\theta_l = \omega$ or $\theta_l = \lambda$, we have

$$\begin{aligned}
\left|\frac{\partial b_t(\theta_0)}{\partial \theta_l}\right| &= \left|\beta_0 - \alpha_0\delta_0 \cdot \frac{\partial\left(a_t(\theta_0)/\sigma_t(\theta_0)\right)}{\partial \alpha}\right| \\
&= \left|\beta_0 - \alpha_0\delta_0 \cdot \frac{\frac{\partial a_t(\theta_0)}{\partial \theta_l} \cdot \sigma_t(\theta_0) - \frac{\partial \sigma_t(\theta_0)}{\partial \theta_l} \cdot a_t(\theta_0)}{\sigma_t^2(\theta_0)}\right| \\
&\leq \beta_0 + \left|\frac{\alpha_0\delta_0}{\sigma_t(\theta_0)} \cdot \frac{\partial a_t(\theta_0)}{\partial \theta_l}\right| + \left|\frac{\alpha_0\delta_0}{\sigma_t^2(\theta_0)} \cdot \frac{a_t(\theta_0)}{2\sigma_t(\theta_0)} \cdot \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta_l}\right| \\
&\leq \beta_0 + \left|\frac{\alpha_0\delta_0}{\sqrt{\omega_0}}\right| \cdot \left|\frac{\partial a_t(\theta_0)}{\partial \theta_l}\right| + \left|\frac{\alpha_0\delta_0}{\omega_0}\right| \cdot \left|\frac{\eta_t}{2} \cdot \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta_l}\right|.
\end{aligned}$$

From the proof of Lemma 3.6, we know that under our assumptions

$$E\left|\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta_l}\right| < \infty$$

for any θ_l being ω or λ. In this case $|\frac{\partial a_t(\theta_0)}{\partial \theta_l}|$ will also have finite expectation due to equation (3.18). Therefore when $\theta_l = \omega$ or $\theta_l = \lambda$ we have

$$E\left|\frac{\partial b_t(\theta_0)}{\partial \theta_l}\right| < \infty.$$

Now consider the other cases. When $\theta_l = \beta$:

$$\frac{\partial b_t(\theta_0)}{\partial \beta} = 1 - \alpha_0 \delta_0 \cdot \frac{\partial\left(a_t(\theta_0)/\sigma_t(\theta_0)\right)}{\partial \beta}.$$

When $\theta_l = \alpha$:

$$\begin{aligned}\frac{\partial b_t(\theta_0)}{\partial \alpha} &= \beta_0 - \alpha_0 \delta_0 \cdot \frac{\partial\left(a_t(\theta_0)/\sigma_t(\theta_0)\right)}{\partial \alpha} - \delta_0 \cdot \frac{a_t(\theta_0)}{\sigma_t(\theta_0)} \\ &= \beta_0 - \alpha_0 \delta_0 \cdot \frac{\partial\left(a_t(\theta_0)/\sigma_t(\theta_0)\right)}{\partial \alpha} - \delta_0 \eta_t.\end{aligned}$$

Similarly when $\theta_l = \delta$:

$$\frac{\partial b_t(\theta_0)}{\partial \delta} = \beta_0 - \alpha_0 \delta_0 \cdot \frac{\partial\left(a_t(\theta_0)/\sigma_t(\theta_0)\right)}{\partial \alpha} - \alpha_0 \eta_t.$$

We see the key component for those three forms is still the derivative $\frac{\partial(a_t/\sigma_t)}{\partial \theta_l}$ evaluated at θ_0. Therefore it is not difficult for us to apply the same argument as for the $\theta_l = \omega$, $\theta_l = \lambda$ cases and conclude

$$E\left|\frac{\partial b_t(\theta_0)}{\partial \theta_l}\right| < \infty$$

for any component θ_l of vector θ.

Now consider the first inequality in (3.32). According to the definition of $e_t(\theta)$ from (3.21) we know it has five different forms. For example, if θ_k in (3.30) is set as $\theta_k = \omega$, then $e_t(\theta) = 1$. In this case $\frac{\partial e_t(\theta_0)}{\partial \theta_l} = 0$ for any θ_l.

When $\theta_k = \beta$, we have $e_t(\theta) = \sigma_t^2(\theta)$. In this case, we have:

$$E\left|\frac{\partial e_t(\theta_0)}{\partial \theta_l}\right| = E\left|\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta_l}\right|$$

for all θ_l. we could conclude from the proof of Lemma 3.6 that this expectation is finite.

When $\theta_k = \alpha$, we have $e_t(\theta) = a_t^2(\theta)$. Therefore by (3.18):

$$\begin{aligned} E\left|\frac{\partial e_t(\theta_0)}{\partial \theta_l}\right| &= E\left|2a_t(\theta_0) \cdot \frac{\partial a_t(\theta_0)}{\partial \theta_l}\right| \\ &= E\left|2\epsilon_t \cdot \left(d_l(\theta_0) - \frac{\delta_0}{2\sigma_t(\theta_0)} \cdot \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta_l}\right)\right|, \end{aligned}$$

where d_l is the respective component of $d(\theta)$ defined in (3.18). One could use the same argument for the $\theta_k = \beta$ case to conclude that this expectation is also finite.

When $\theta_k = \lambda$, $e_t(\theta) = -2\alpha a_t$. In this case $E\left|\frac{\partial e_t(\theta_0)}{\partial \theta_l}\right|$ will essentially be determined by $\frac{\partial a_t(\theta_0)}{\partial \theta_l}$, which is has been shown to be finite in the previous case.

The last case is $\theta_k = \delta$, which indicates $e_t(\theta) = -2\alpha a_t \sigma_t$. Take the derivative with respect to θ_l ($\theta_l \neq \alpha$) and we have

$$E\left|\frac{\partial e_t(\theta_0)}{\partial \theta_l}\right| = E\left|2\alpha_0\sigma_t(\theta_0)\cdot\frac{\partial a_t(\theta_0)}{\partial \theta_l} + 2\alpha_0 a_t(\theta_0)\frac{\partial \sigma_t(\theta_0)}{\partial \theta_l}\right|$$
$$\leq E\left|2\alpha_0\sigma_t(\theta_0)\cdot\frac{\partial a_t(\theta_0)}{\partial \theta_l}\right| + \left|\alpha_0\eta_t\cdot\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta_l}\right|.$$

In previous cases we already discussed terms like $\frac{\partial a_t(\theta_0)}{\partial \theta_l}$ and $\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta_l}$. Notice that $\sigma_t(\theta_0)$ also have finite moments due to the 2nd order stationarity of the GARCH process. Therefore $E|\frac{\partial e_t(\theta_0)}{\partial \theta_l}|$ is finite in this case. Note that when $\theta_l = \alpha$ the we have

$$E\left|\frac{\partial e_t(\theta_0)}{\partial \theta_l}\right| \leq E\left|2\epsilon_t\sigma_t(\theta_0)\right| + \left|2\alpha_0\cdot\frac{\partial a_t(\theta_0)\sigma_t(\theta_0)}{\theta_l}\right|.$$

The second term on the right hand side is the same object we just considered, which we know is finite. The first term is also finite due to the 2nd order stationarity property of the GARCH process. Based on all those scenarios we have shown that for any θ_l:

$$E\left|\frac{\partial e_t(\theta_0)}{\partial \theta_l}\right| < \infty.$$

Therefore (3.32) is satisfied, which will lead to (3.31). In the view of (3.29), we have proved (3.25). □

3.5.3 Conclusion

Now we proceed to prove the theorem. Denoting the information set up to time t as $\mathcal{F}_t$, by equation (3.17) it is not difficult to verify that $\frac{\partial l_t(\theta_0)}{\partial \theta}$ is a martingale difference sequence:

$$
\begin{aligned}
E\left(\frac{\partial l_t(\theta_0)}{\partial\theta}|\mathcal{F}_{t-1}\right) &= E\left[\frac{1}{\sigma_t^2(\theta_0)}\left(2\epsilon_t\cdot\frac{\partial a_t(\theta_0)}{\partial\theta} - \frac{\epsilon_t^2}{\sigma_t^2(\theta_0)}\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta} + \frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right)|\mathcal{F}_{t-1}\right] \\
&= \frac{1}{\sigma_t^2(\theta_0)}\left\{E(2\epsilon_t|\mathcal{F}_{t-1})\cdot\frac{\partial a_t(\theta_0)}{\partial\theta} - E(1-\eta_t^2)\cdot\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\} \\
&= 0.
\end{aligned}
$$

Note that this sequence is well-defined in its first and second-order structures guaranteed by Lemma 3.6. It is also ergodic and stationary. By the central limit theorem for stationary and ergodic martingale difference sequences from Billingsley (1961), we can conclude that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\theta}\xrightarrow{D} N\left(0,B\right),$$

where $B = E(\frac{\partial l_t(\theta_0)}{\partial\theta}\frac{\partial l_t(\theta_0)}{\partial\theta'})$ as given in (3.8). By Lemma 3.6 we know this matrix is well-defined. We also know that under the specified conditions, our process is geometrically ergodic by Theorem 2.8. Therefore Proposition 17.1.6 of Meyn and Tweedie (2009) applies so that the above convergence result also holds for $l_t(\theta)$ started from any arbitrary initial values. Applying a similar argument as in the proof of Lemma 3.4 we have:

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}\xrightarrow{D} N\left(0,B\right), \tag{3.33}$$

where B is the matrix defined above.

Next we consider the other part on the righthand side of equation (3.7). Perform a Taylor's expansion around θ_0:

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t(\theta^*)}{\partial\theta_i\partial\theta_j} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta_i\partial\theta_j} + o(\theta^* - \theta_0)$$

where θ^* is on the cord of θ_0 and $\hat{\theta}_n$. Since $\hat{\theta}_n$ is strongly consistent, it converges to θ_0 when n goes large. Therefore θ^* also converges to θ_0 in the meantime. The second term $o(\theta^* - \theta_0)$ on the right hand side converges to zero faster than $\theta^* - \theta_0$, which happens as n becoming larger due to the consistency. For the first term, the ergodic theorem applies to the sequence $\frac{\partial^2 l_t(\theta_0)}{\partial\theta_i\partial\theta_j}$. As we have argued previously, this result does not rely on the choice of initial values due to the geometric ergodicity property. Therefore we could extend this result and conclude that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta_i\partial\theta_j} \stackrel{a.s.}{\to} A_{ij}, \tag{3.34}$$

where A_{ij} is the respective element of the matrix $A = E(\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'})$ as defined in (3.8). It is a well-defined object as shown by Lemma 3.7.

To complete the proof we apply the Slutsky's theorem. In the view of equations (3.7), (3.33) and (3.34) the theorem is proved.

Chapter 4

Simulation Study

This chapter includes a few examples to examine stylized facts of financial series, fit data to GARCH-in-mean models and numerically investigate the asymptotic behaviors of the quasi-maximum likelihood estimator.

4.1 Stylized facts of financial series

We consider the daily log-returns of S&P 500 indices for years 1990 - 2012 [1]. Suppose p_t is the adjusted close price of the index of day t, the log-return series could be easily calculated from $\log(p_t/p_{t-1})$. Throughout this chapter we work with a re-scaled series obtained by $y_t = 100\log(p_t/p_{t-1})$.

Figure 4.1 shows a time series plot of this re-scaled log-returns of S&P 500 indices.

[1] Online data obtained from yahoo finance: http://ca.finance.yahoo.com/q/hp?s=%5EGSPC

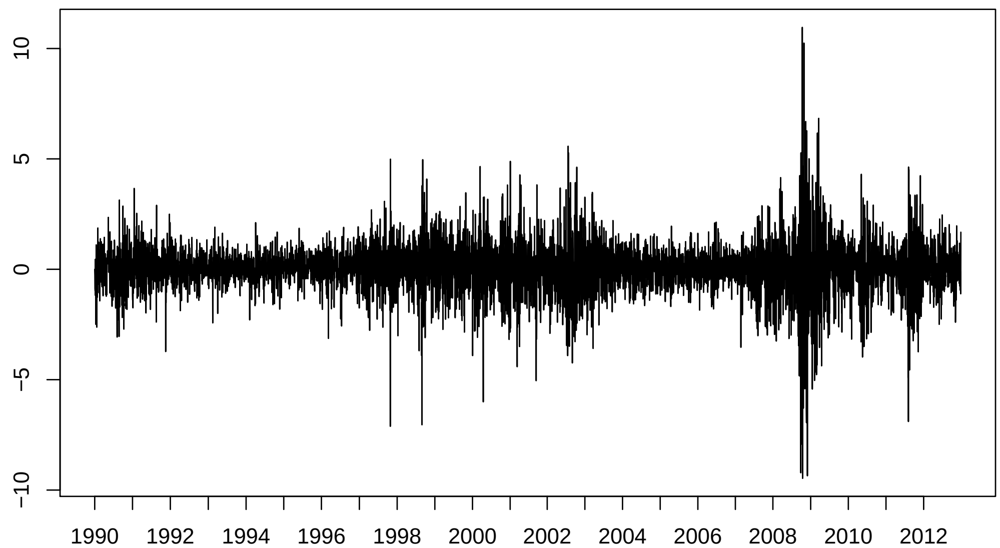

Figure 4.1: Re-scaled daily S&P 500 log-returns: Jan 1990 - Dec 2012

One preliminary analysis we want to perform is to examine this series and identify a few properties commonly known for financial series, as we discussed in the first chapter. From Figure 4.1 one may already notice a few stylized facts here. First of all, the series is apparently nonstationary. The series has more turbulent subperiods such as years 2008 - 2010 as well as relatively quiet subperiods such as 2004 - 2006, which is a fair reflection of the economic states at those times. Moreover, those subperiods representing high or low financial volatility tend to appear in clusters, which is consistent with the volatility clustering phenomenon.

Financial series also frequently exhibit interesting properties in their autocorrelations. For example, a return series generally has very small autocorrelations that resembles a white noise process, while strong autocorrelations are frequently witnessed for higher order structures. The graph below shows the sample acf for the original series y_t and the squared series y_t^2.

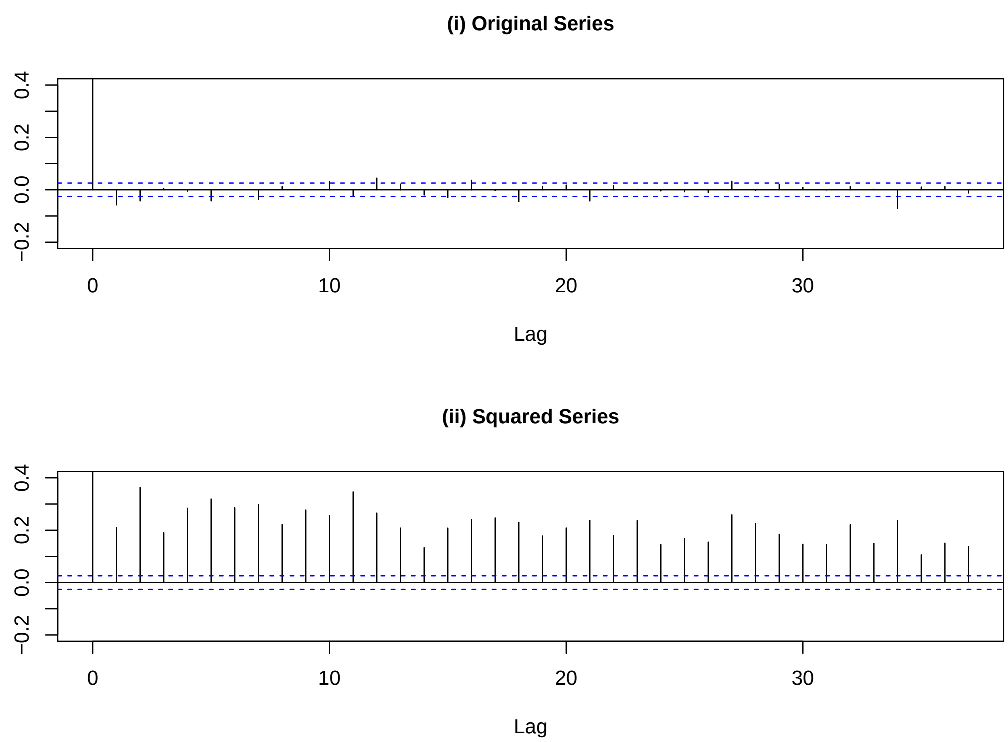

Figure 4.2: Sample ACFs: re-scaled S&P 500 log-returns

The graph above clearly shows strong autocorrelations within the 2nd order structure while such dependence is not as apparent within the original series itself.

We may also want to examine some distributional properties. As already mentioned in the first chapter, the marginal distributions of financial series are often found to be leptokurtic, which indicates sharper peaks and fatter tails compared to a normal distribution. The graph below shows the estimated density of the y_t series (solid line) along with a reference Gaussian distribution (dashed line) with mean and variance equal to the sample mean and variance of the series.

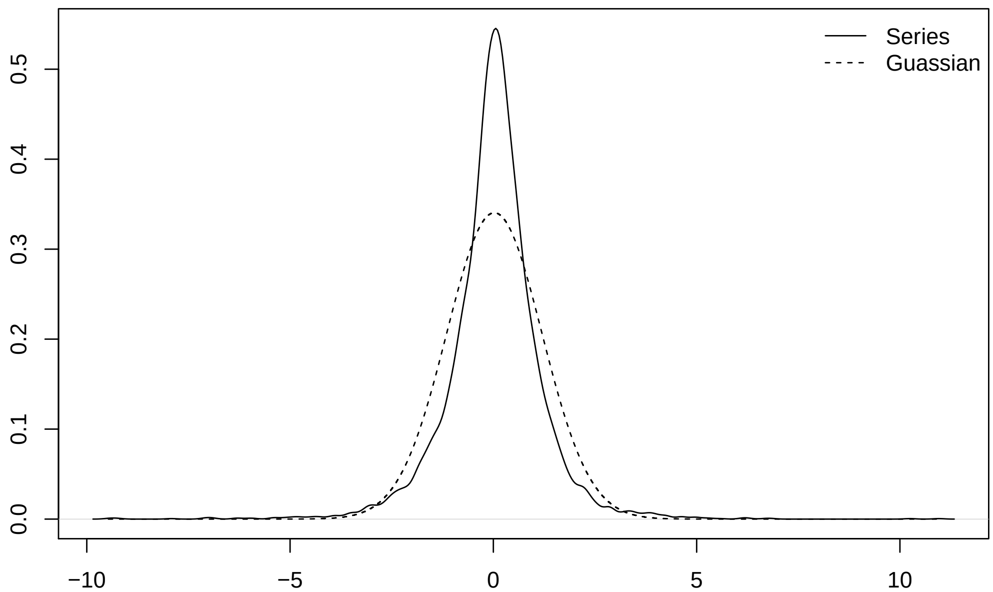

Figure 4.3: Estimated density of re-scaled S&P 500 returns v.s. fitted Gaussian density

It is quite apparent that the marginal distribution of the series does not look like Gaussian. It shows a much sharper peak and slightly fatter tails than Gaussian. The sample kurtosis of the series is found to be 8.464948, which is excessive comparing to a normal distribution.

GARCH-type processes are usually found helpful for capturing frequently observed stylized facts of financial series. We now fit the GARCH-M(1,1) model specified by (2.8) - (2.10) by quasi-maximum likelihood estimation discussed in Chapter 3. The minimizing algorithm often requires a starting value for the parameter θ. To find a suitable value the following 3-step algorithm is implemented:

1. Fit the data using a standard GARCH model with a nonzero mean, i.e.

$$y_t = \mu + \epsilon_t,$$

   where $\epsilon_t \sim$ GARCH(1,1) and μ is some constant. Obtain the parameter estimate $\hat{\omega}$, $\hat{\alpha}$ and $\hat{\beta}$.

2. Estimate the volatility process $\hat{\sigma}_t$ based on the model obtained in the last step. Then fit the linear regression model

$$E(y_t) = c_0 + c_1 \hat{\sigma}_t.$$

   Obtain parameter estimate $\hat{c}_0$ and $\hat{c}_1$.

3. Fit the GARCH-M(1,1) model with the starting value $\lambda = \hat{c}_0$, $\delta = \hat{c}_1$ and $\hat{\omega}$, $\hat{\alpha}$, $\hat{\beta}$.

Applying the above algorithm to the y_t series we can obtain GARCH-M(1,1) parameter estimates. We also fit the data without imposing the "in-mean" structure by simply performing another fit using a pure GARCH(1,1). The results are shown below.

Table 4.1: QMLEs of GARCH-M(1,1) fit

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
λ	0.042393	0.054499	0.777873	0.436644
δ	0.009726	0.060179	0.161618	0.871607
ω	0.010131	0.003519	2.878911	0.003991
α	0.074515	0.008714	8.550776	0.000000
β	0.917898	0.009408	97.562232	0.000000

Table 4.2: QMLEs of GARCH(1,1) fit

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
ω	0.009729	0.003467	2.805787	0.005019
α	0.072567	0.008602	8.435850	0.000000
β	0.920160	0.009266	99.301892	0.000000

The estimated value of $\hat{\delta}$ is relatively small suggesting that the GARCH-in-mean effect is not very strong in the series. This fact is confirmed by relatively large p-values for first two parameters which are unique to GARCH-in-mean. The parameter estimates for ω_0, α_0 and β_0 given by both models are found to be quite close. The GARCH-in-mean model has slightly larger values of $\hat{\omega}$ and $\hat{\alpha}$ while getting a slightly smaller estimate $\hat{\beta}$. The standard errors of the estimates from GARCH(1,1) fit are smaller than GARCH-M(1,1) case, but the difference is quite minor. The results suggest that their performances are very close.

The standard errors shown in Tables 4.1 and 4.2 are obtained from a parametric bootstrap of 5000 repetitions. Common computer packages in R such as `tseries` would output standard errors as well, but based on the diagonal elements of the estimated Hessian matrix. As we are dealing with the QMLE here, we are no longer able to directly use such numbers, because the variance matrix now is specified as $A^{-1}BA^{-1}$ as stated in Theorem 3.2.

Given the parameter estimates above, one important application is to reconstruct the volatility process which is never observed at any time. The fitted volatility series can be obtained by iterating the following equation:

$$\hat{\sigma}_t^2 = \hat{\omega} + \hat{\alpha}(y_{t-1} - \hat{\lambda} - \hat{\delta}\hat{\sigma}_{t-1})^2 + \hat{\beta}\hat{\sigma}_{t-1}^2.$$

The graph below shows the estimated volatility process $\hat{\sigma}_t$ based on our GARCH-in-mean model obtained above.

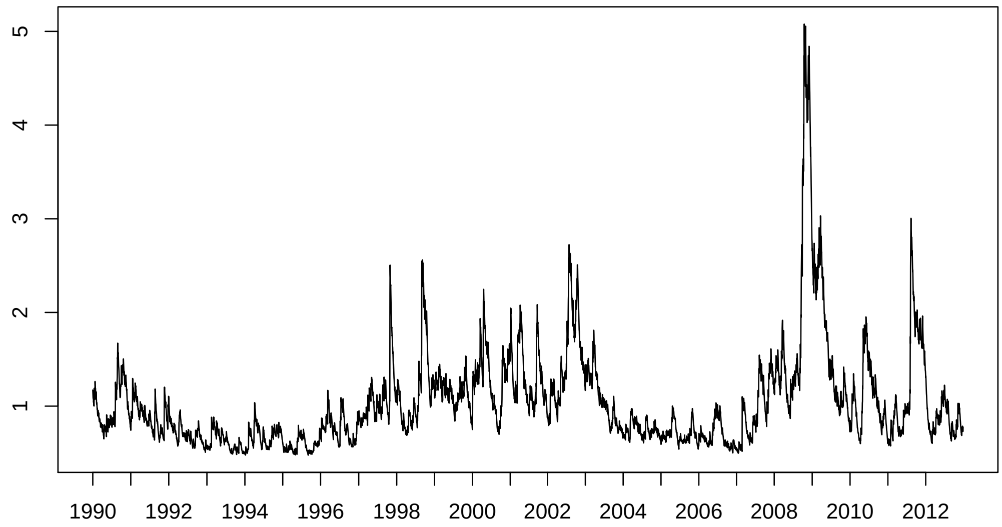

Figure 4.4: Estimated volatilities of re-scaled S&P 500 returns by GARCH-M(1,1)

We see that this graph is consistent with our S&P 500 data series shown in Figure 4.1. Periods with high volatilities such as years 2008 - 2011 seems to be more unstable in Figure 4.1. Quiet subperiod such as 2004 - 2006 shows relatively low volatility values.

The fitted volatility series from the pure GARCH(1,1) model obtained in Table 4.1 is very close to the series we showed above. It is not distinguishable if we overlay both series on a single plot. The average difference between the two volatility series is found to be 0.012372 and it is the pure GARCH model that has slightly larger estimated volatilities on average. It may due to the fact that GARCH-M model has additional structure in its mean structure thus some variability has already been explained in the mean equation.

The fitted volatilities have important implications in finance since the volatility represents the financial risk, but its true dynamics are not directly observable. Estimating the underlying volatilities will help practitioners to better understand the current state of market and make better decisions accordingly.

4.2 Asymptotic evaluation

The goal of this section is to investigate the asymptotic properties of the QMLE of GARCH-M models using simulation techniques. Unlike in the last section, we no longer use the real S&P 500 data because we need to have precise knowledge of the true parameter value to assess the convergence of QMLEs. Therefore we simulate the observable data y_t according to GARCH-M equations (2.8) - (2.10), with Gaussian innovations and the true parameter value θ_0 given by

$$\lambda_0 = 0.1, \quad \delta_0 = 0.1, \quad \omega_0 = 0.05, \quad \alpha_0 = 0.12, \quad \beta_0 = 0.8.$$

Those values are chosen to be in vicinities of the parameter estimates we obtained in Table 4.1. We also considerably enlarged the values of λ_0 and δ_0 in the hope of magnifying the "in-mean" effects of the generated data.

The experiment starts with repeatedly simulating sample paths of length n according to the above specification. We denote each sample path by $y_1^{(i)}, \ldots, y_n^{(i)}$ where $i = 1, \ldots, K$ with K being the total number of repetitions. For every simulated sample path we can obtain a QMLE denoted by $\hat{\theta}_n^{(i)}$. As n increases, we may investigate the convergence of the QMLE as well as its distributional properties.

The choices of initial values are unimportant asymptotically. However, we need to point out that properly choused initial values do enjoy certain advantages in terms of computational efficiency, speed of convergence etc. All the experiments below adopt the following initial values: $y_0^{(i)} = y_1^{(i)}$, and $\tilde{\sigma}_0^{(i)}$ equals to the sample standard deviation of the corresponding sample path.

4.2.1 Convergence of the estimates

We start with a sample size $n = 250$ and gradually increase the size up to 5000. For each n we replicate the simulation-estimation process for 10000 times, i.e. $K = 10000$, and calculate the average of the estimates by:

$$\bar{\theta}_n = \frac{1}{K}\sum_{i=1}^{K}\hat{\theta}_n^{(i)}.$$

As n increases, we observe the change in $|\bar{\theta}_n - \theta_0|$ as well as in the root mean square error (RMSE). For example, the RMSE with respect to estimates of λ_0 can be calculated by

$$RMSE(\hat{\lambda}_n) = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(\hat{\lambda}_n^{(i)} - \bar{\lambda}_n)^2}.$$

The graph below shows 8 estimates of θ_0 at different sample sizes: $\bar{\theta}_{250}$, $\bar{\theta}_{500}$, $\bar{\theta}_{750}$, $\bar{\theta}_{1000}$, $\bar{\theta}_{1500}$, $\bar{\theta}_{2000}$, $\bar{\theta}_{3000}$ and $\bar{\theta}_{5000}$. The respective true value is also marked in the graph as a dashed reference line.

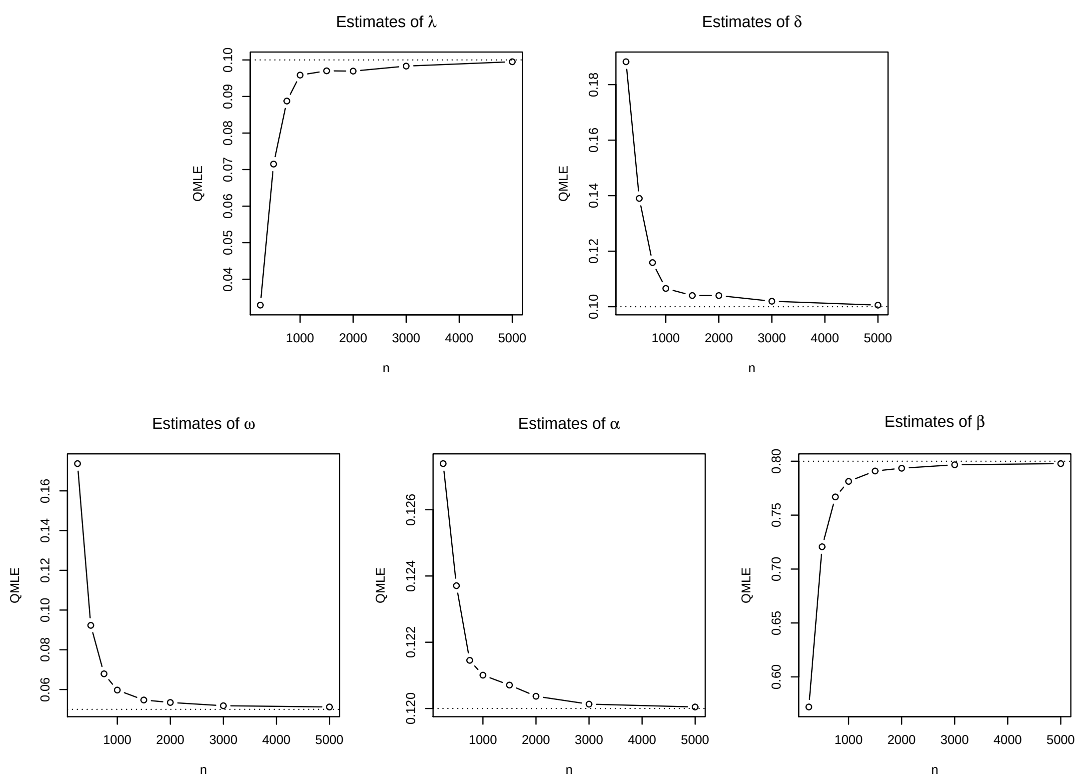

Figure 4.5: QMLEs under different sample sizes

From the graph we can see that each parameter converges to the true value as the size of the sample paths increases. The rate of convergence is also relatively fast: most of the estimates (except $\bar{\beta}$) are within 0.01 of their respective true values when the sample size n reaches 1000. The table below shows the absolute difference $|\bar{\theta}_n - \theta_0|$ under different sample sizes.

Table 4.3: Absolute difference $|\bar{\theta}_n - \theta_0|$ under different sample sizes

n	$\lvert\bar{\lambda}_n - \lambda_0\rvert$	$\lvert\bar{\delta}_n - \delta_0\rvert$	$\lvert\bar{\omega}_n - \omega_0\rvert$	$\lvert\bar{\alpha}_n - \alpha_0\rvert$	$\lvert\bar{\beta}_n - \beta_0\rvert$
250	0.067083	0.088231	0.123725	0.007393	0.227890
500	0.028497	0.039018	0.042291	0.003708	0.079322
750	0.011245	0.015896	0.017902	0.001454	0.033091
1000	0.004148	0.006584	0.009721	0.001006	0.018704
1500	0.002981	0.004030	0.004718	0.000707	0.009063
2000	0.003058	0.004023	0.003468	0.000370	0.006505
3000	0.001687	0.001978	0.001829	0.000131	0.003305
5000	0.000501	0.000574	0.001207	0.000046	0.002210

The overall convergence trend is quite apparent from the table above. When the sample size increases to 5000, the estimates of λ_0, δ_0 and α_0 are within 0.001 of their true values while estimates of ω_0 and β_0 are within 0.01. To evaluate the QMLEs one could also investigate the RMSE introduced before, which are shown as below.

Table 4.4: RMSEs for QMLEs

n	RMSE($\hat{\lambda}_n$)	RMSE($\hat{\delta}_n$)	RMSE($\hat{\omega}_n$)	RMSE($\hat{\alpha}_n$)	RMSE($\hat{\beta}_n$)
250	1.483099	2.041421	0.184566	0.074593	0.315838
500	0.493331	0.674302	0.110266	0.046238	0.200055
750	0.250876	0.339313	0.064250	0.034958	0.121762
1000	0.161955	0.220632	0.038580	0.029880	0.079090
1500	0.120937	0.163953	0.019695	0.024035	0.046832
2000	0.102243	0.138894	0.016318	0.020510	0.039153
3000	0.080994	0.109585	0.011317	0.016524	0.029149
5000	0.060648	0.082361	0.008712	0.012795	0.022430

Table 4.4 clearly shows that the RMSE decreases steadily as the sample size increases. This observation is consistent with our findings from Figure 4.5. It suggests that a larger sample size enables us to obtain more accurate estimates, and the estimates will eventually converge to the true value as sample size increases.

4.2.2 Limiting distribution

We are also interested in finite-sample distributions of the QMLEs and want to investigate whether they converge to normal distributions as our theorem stated. According to Theorem 3.2, the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is a Gaussian distribution with zero mean and variance determined by $A^{-1}BA^{-1}$, where A and B are defined in (3.8).

The first and second order derivatives of the quasi-likelihood $l_t(\theta)$ have complex forms and are difficult to evaluate algebraically. Nevertheless, we may conduct a simulation study and provide estimates of these two matrices. According to (3.8) we have the definition of matrix A as:

$$A = E\left(\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}\right).$$

From Chapter 3 we know that this object is also the limit of $\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}$ when n goes to infinity. Therefore to numerically estimate this matrix, we can simulate a number of GARCH-M observations and take the average of the individual Hessian matrices. Similarly, to approximate matrix B we need to numerically evaluate the outer products of the gradients and then take the average. Following this algorithm we can obtain an estimate of the matrix $A^{-1}BA^{-1}$, and its diagonal elements can be regarded as estimates of the asymptotic variances. Below shows the estimated asymptotic standard deviations for elements of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ under different sample sizes.

Table 4.5: Estimated Asymptotic SDs for each element of $\sqrt{n}(\hat{\theta}_n - \theta_0)$

n	$\sqrt{n}(\hat{\lambda}_n - \lambda_0)$	$\sqrt{n}(\hat{\delta}_n - \delta_0)$	$\sqrt{n}(\hat{\omega}_n - \omega_0)$	$\sqrt{n}(\hat{\alpha}_n - \alpha_0)$	$\sqrt{n}(\hat{\beta}_n - \beta_0)$
1000	5.209461	7.012085	1.034789	1.773574	2.991136
2000	4.525709	6.168533	0.726148	1.107946	1.946490
3000	4.368645	5.968818	0.648966	1.001053	1.727769
5000	4.31964	5.894882	0.621891	0.956338	1.644865
7000	4.297810	5.832972	0.608512	0.911009	1.569210
10000	4.283571	5.820551	0.600515	0.902701	1.560754

We see that the estimated standard deviations decreases as the sample size increases. If we keep increasing the sample size we may still expect some slight drops in the estimates. However, the differences between $n = 10000$ case and $n = 7000$ case are already relatively small. Therefore we treat the last row of Table 4.5 as our final estimates of asymptotic standard deviations.

Now we investigate the distributions of QMLEs we obtained. By the design of our simulation study, we can obtain 10000 estimates of each parameter vector given a fixed sample size n, which are denoted $\hat{\theta}_n^{(1)}, \ldots, \hat{\theta}_n^{(10000)}$ as we mentioned in the previous subsection. We estimate the sample standard deviation of each element, for instance $\sqrt{n}(\hat{\lambda}_n - \lambda_0)$ based on 10000 data points, and compare them to our estimated asymptotic standard deviations. Define s_{1n} being the sample standard deviation of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ based on QMLEs $\hat{\theta}_n^{(1)}, \ldots, \hat{\theta}_n^{(10000)}$ and s_{2n} being the standard deviation estimates from the last row of Table 4.5 [1]. We calculate the difference $s_{1n} - s_{2n}$ for each parameter and the result is shown below.

Table 4.6: Differences between Asymptotic and Sample Standard Deviations

n	$\sqrt{n}(\hat{\lambda}_n - \lambda_0)$	$\sqrt{n}(\hat{\delta}_n - \delta_0)$	$\sqrt{n}(\hat{\omega}_n - \omega_0)$	$\sqrt{n}(\hat{\alpha}_n - \alpha_0)$	$\sqrt{n}(\hat{\beta}_n - \beta_0)$
250	19.167459	26.455760	2.317871	0.276782	3.433337
500	6.748202	9.255058	1.865242	0.131268	2.912826
750	2.587308	3.469395	1.159121	0.054716	1.773999
1000	0.838151	1.153783	0.619567	0.042243	0.940418
1500	0.400518	0.526637	0.162301	0.028224	0.253129
2000	0.289122	0.388306	0.129300	0.014565	0.190322
3000	0.152898	0.178963	0.019378	0.002399	0.035883
5000	0.058364	0.044462	0.015539	0.002081	0.025378
8000	0.005101	0.000515	0.008559	0.001070	0.016597

The above result clearly shows that the sample standard deviation derived from QMLEs converges to the asymptotic standard deviations as the sample size increase. We want to point out that the asymptotic standard deviations we used here are actually estimates of the true “theoretical” ones hence are not necessarily quite precise. However, they still provide us valuable insights around the convergence of sample standard deviations.

Lastly, we want to compare the distributions of QMLEs against respective

[1]For convenience, we refer those estimates as the asymptotic standard deviations moving forward

asymptotic normal distributions. To do this, the kernel density estimation of each element of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is applied based on $\hat{\theta}_n^{(1)}, \ldots, \hat{\theta}_n^{(10000)}$ for different n. The comparable normal distributions are of zero means and variances determined by the respective elements from the last row of Table 4.5. We plot the distributions of QMLEs as solid lines, along with a dashed reference line representing the comparable Gaussian distributions. The results are shown below.

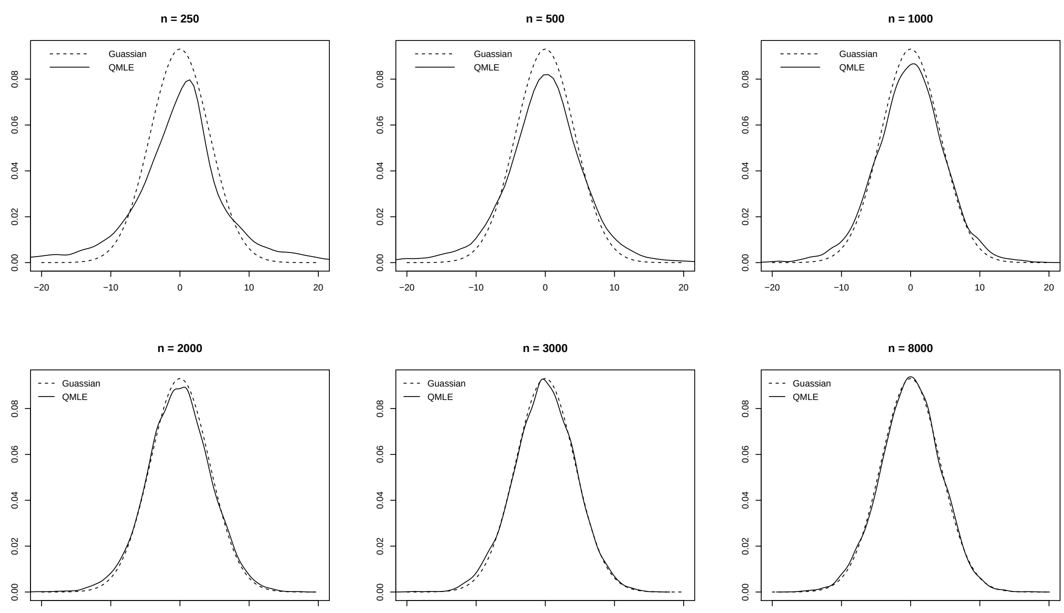

Figure 4.6: Distribution of $\sqrt{n}(\hat{\lambda}_n - \lambda_0)$

We can see from the graph that the distribution of the estimates converges to a normal distribution as the sample size increases. At $n = 3000$ the distribution of QMLEs is already quite close to a normal distribution. When $n = 8000$ those two lines almost coincide.

Similar behavior could be observed from the estimates of $\sqrt{n}(\hat{\delta}_n - \delta_0)$ as well. The figure below shows the situation as the sample size increases. We could notice that when $n = 3000$ the distribution of QMLEs quite resembles a Gaussian distribution. $n = 8000$ provides a slightly improved approximation.

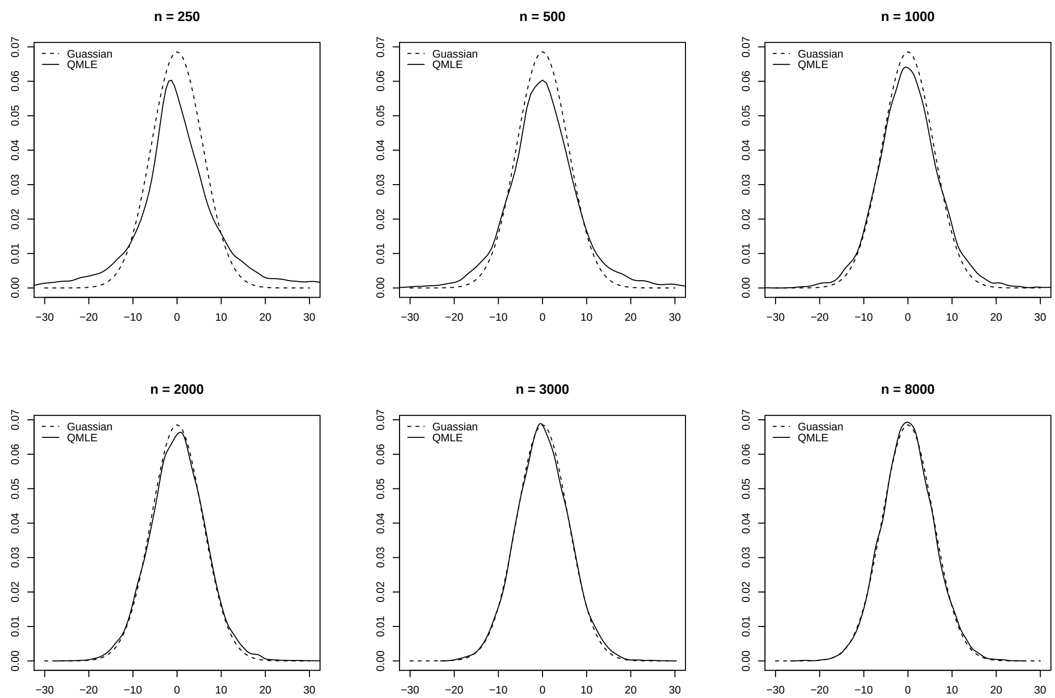

Figure 4.7: Distribution of $\sqrt{n}(\hat{\delta}_n - \delta_0)$

Following the same procedure we can investigate estimates for the other parameters. The graph below shows the distributions of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ for the rest of the parameters ω, α and β.

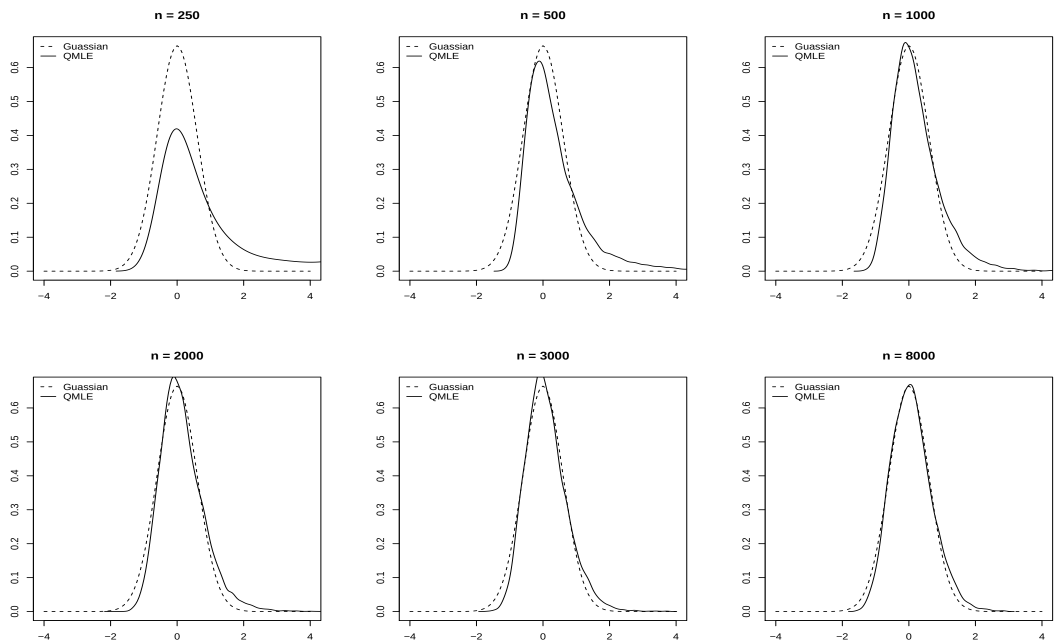

Figure 4.8: Distribution of $\sqrt{n}(\hat{\omega}_n - \omega_0)$

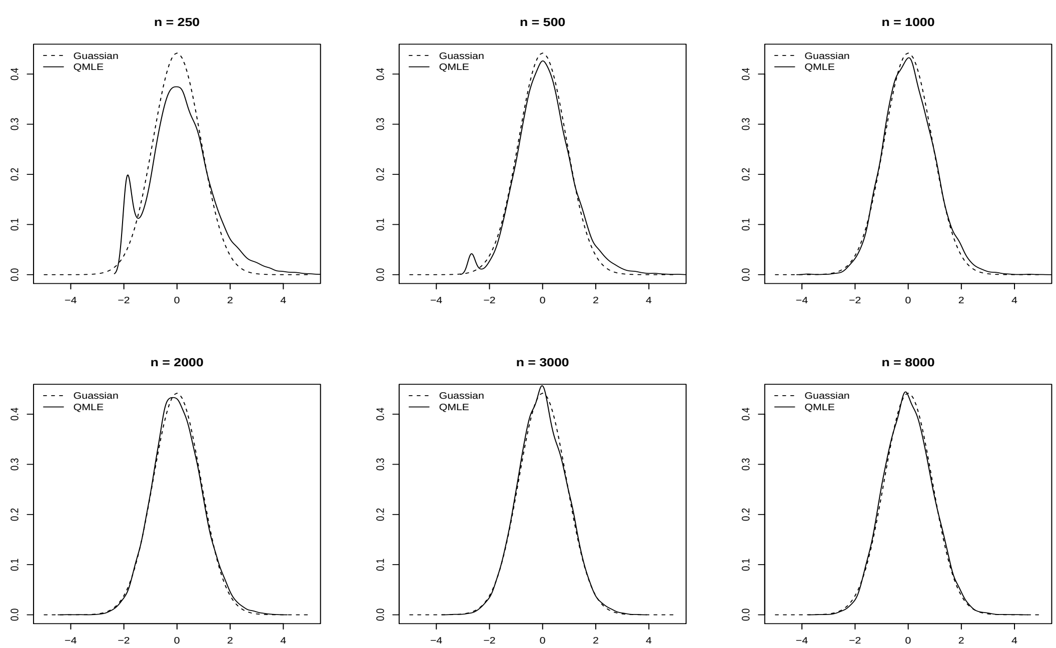

Figure 4.9: Distribution of $\sqrt{n}(\hat{\alpha}_n - \alpha_0)$

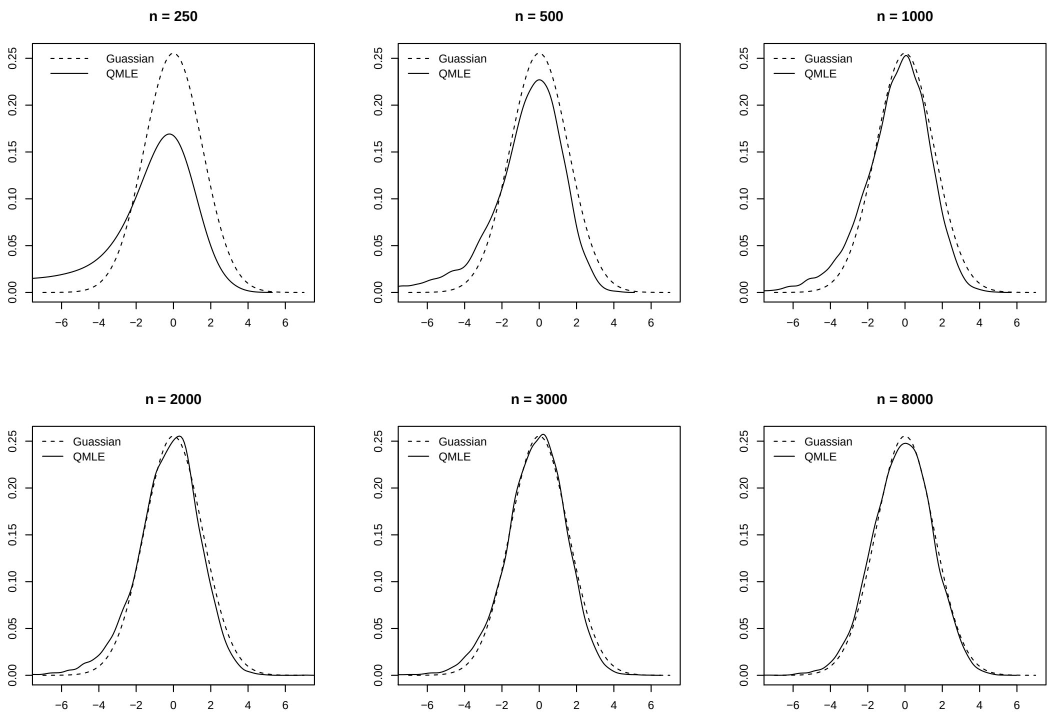

Figure 4.10: Distribution of $\sqrt{n}(\hat{\beta}_n - \beta_0)$

The results above clearly illustrated the overall convergence trend of QMLEs: as the sample sizes increases, the solid line becomes more and more smooth and close to the dashed reference line representing comparable Gaussian distributions. Notice that when $n = 8000$ the smoothness of the distribution is not equally good for each parameter. This indicates the difference in the their speeds of convergence. We may still notice certain levels of deviation even at $n = 8000$ but the overall trend of convergence is quite evident. Given all the results above, it is very convincing that the distribution of QMLE will eventually converge to a Gaussian distribution.

4.2.3 Impact of the true distribution

The observable series y_t simulated in the last subsection are based on Gaussian innovations. As we already discussed earlier, marginal distributions of financial series are usually deemed non-Gaussian by empirical studies. In this section, we will re-perform the previous experiment but based on innovations following a t-distribution.

We adopt the same true parameter values and simulate the GARCH-M innovations based on a $t(8)$ distribution to ensure the process has a finite 4th moment. We also need to satisfy the condition $E(\eta_t^2) = 1$. Thus the simulated innovations are re-scaled by dividing by $\sqrt{4/3}$, i.e. $\eta_t \overset{i.i.d}{\sim} X/\sqrt{4/3}$ where $X \sim t(8)$. This density is shown in the graph below against a standard normal density.

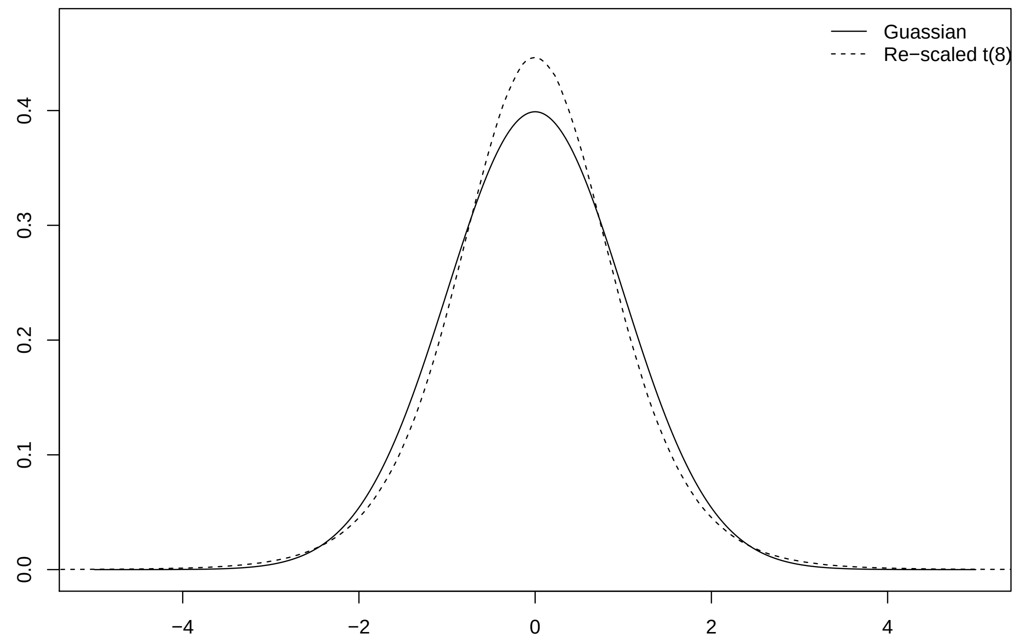

Figure 4.11: Gaussian density v.s. re-scaled t(8) density

From Figure 4.11 we see the re-scaled t-distribution clearly has a higher peak and slightly fatter tails than the normal distribution. When fitting QMLEs we will still stay with the Gaussian kernel so there is a certain level of disparity between the true distribution of the process and the postulated distribution underlying the quasi-likelihood. We want to understand whether this disparity will have any impact on the asymptotic properties of the estimator.

We first investigate the consistency property. Like in Figure 4.5, we plot the averages of QMLEs $\bar{\theta}_n$ against the sample size n. The graph below shows the estimates given the sample size 250, 500, 1000, 2000, 3000 and 5000, with the true values indicated by the dashed reference lines.

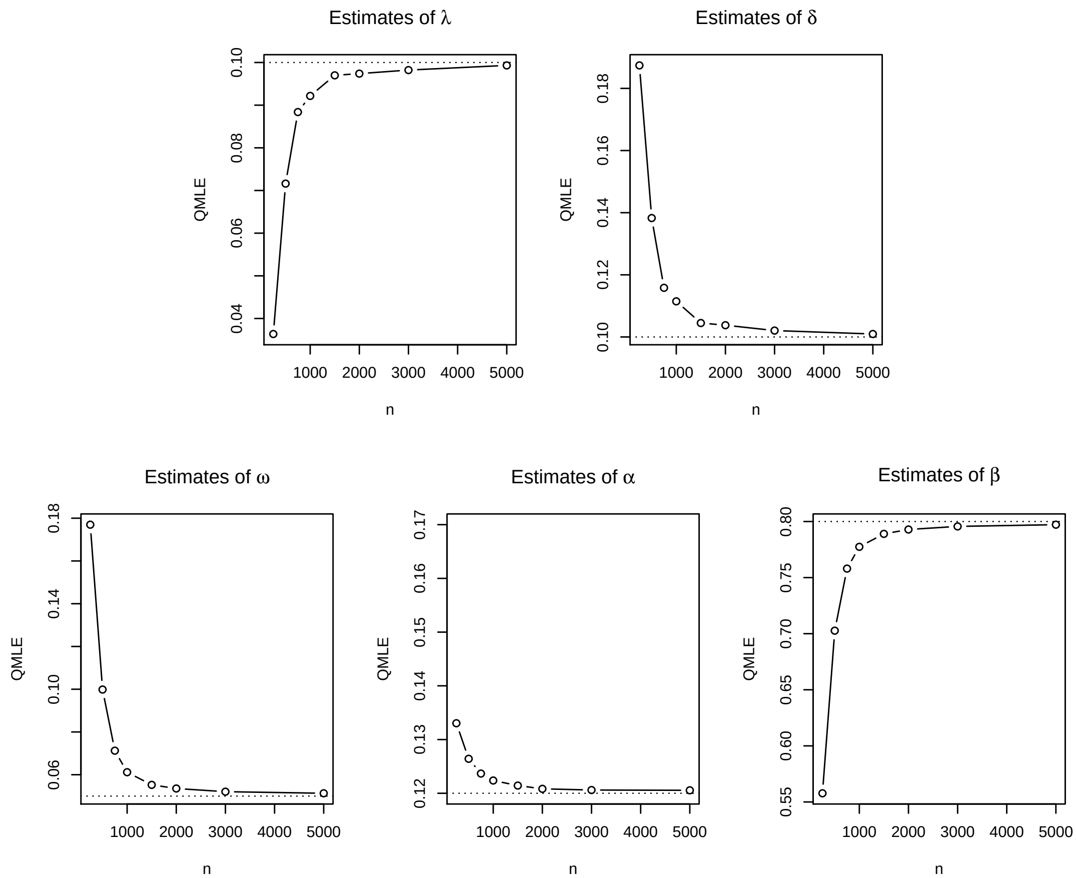

Figure 4.12: Convergence of the QMLEs: t-distributed GARCH-M

Comparing Figures 4.5 and 4.12 we see they are very similar. The overall convergence trend to the true values is quite apparent. This experiment confirms the fact that QMLEs are consistent even if the true distribution of the process does not agree with the postulated distribution used to construct the quasi-likelihood.

We may also calculate the biases of our estimates analogous to Table 4.3 and the results are shown below.

Table 4.7: Absolute difference $|\bar{\theta}_n - \theta_0|$: t-distributed GARCH-M

n	$\lvert\bar{\lambda}_n - \lambda_0\rvert$	$\lvert\bar{\delta}_n - \delta_0\rvert$	$\lvert\bar{\omega}_n - \omega_0\rvert$	$\lvert\bar{\alpha}_n - \alpha_0\rvert$	$\lvert\bar{\beta}_n - \beta_0\rvert$
250	0.063637	0.087397	0.126976	0.013036	0.242258
500	0.028387	0.038281	0.049856	0.006420	0.097303
750	0.011619	0.015829	0.021258	0.003670	0.041965
1000	0.007834	0.011472	0.011184	0.002376	0.022566
1500	0.003011	0.004496	0.005267	0.001436	0.011016
2000	0.002612	0.003791	0.003557	0.000829	0.007119
3000	0.001778	0.002068	0.002068	0.000603	0.004368
5000	0.000687	0.000955	0.001326	0.000553	0.002812

We have two observations from the above chart. First of all, the difference between the true values and QMLEs shrinks as we increase the sample size, with different rate of convergence for each parameter. At $n = 5000$, each estimate is within 0.01 away from the true values. The consistency of the estimates is quite evident. Secondly, comparing with Table 4.3, we see that under the t-distributed innovations, the rate of convergence is slower in general: for instance at $n = 5000$, the differences shown from Table 4.3 are consistently lower than the comparable numbers above. This fact indicates that although the true distribution of the innovation does not impact our final asymptotic results, it may have an influence on other aspects such as efficiency of estimates, rate of convergence, etc.

We can also examine the RMSEs of the estimates which are shown below.

Table 4.8: RMSEs for QMLEs: t-distributed GARCH-M

n	RMSE($\hat{\lambda}_n$)	RMSE($\hat{\delta}_n$)	RMSE($\hat{\omega}_n$)	RMSE($\hat{\alpha}_n$)	RMSE($\hat{\beta}_n$)
250	1.444368	2.042924	0.185752	0.087721	0.321014
500	0.535705	0.746663	0.119723	0.054628	0.221354
750	0.288251	0.400470	0.071293	0.041222	0.138842
1000	0.200866	0.282390	0.045529	0.034558	0.094269
1500	0.135323	0.186070	0.023441	0.027554	0.055074
2000	0.095475	0.131431	0.016349	0.023598	0.041576
3000	0.075754	0.104347	0.012549	0.019314	0.032999
5000	0.058399	0.080379	0.009476	0.015016	0.025093

Comparing with Table 4.4, the RMSEs shown above are close to what we seen when studying the Gaussian GARCH-M. Closer examination of the two charts found that in general, the RMSEs from Table 4.8 are still slightly larger than those from Table 4.4 especially when the sample size is relatively small. This finding is consistent with our discussion above. This suggests that although the choice of the estimation kernel does not matter in terms of the asymptotic results, it may affect other aspects of the estimator.

Now we study the distribution of the estimators. Following the same approach of calculating $A^{-1}BA^{-1}$ as in Table 4.5, we found the estimated asymptotic standard deviation of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ to be

$$(3.920569, 5.449024, 0.641962, 1.020419, 1.711424)' \tag{4.1}$$

when $n = 15000$. We compare this result to the sample standard deviations of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ derived from QMLEs $\hat{\theta}_n^{(1)}$, ..., $\hat{\theta}_n^{(10000)}$. The following result is obtained.

Table 4.9: Difference: Sample SDs − Asymptotic SDs: t-distributed GARCH-M

n	$\sqrt{n}(\hat{\lambda}_n - \lambda_0)$	$\sqrt{n}(\hat{\delta}_n - \delta_0)$	$\sqrt{n}(\hat{\omega}_n - \omega_0)$	$\sqrt{n}(\hat{\alpha}_n - \alpha_0)$	$\sqrt{n}(\hat{\beta}_n - \beta_0)$
250	18.918034	26.854064	2.280488	0.366646	3.364509
500	8.058753	11.247711	2.020555	0.201170	3.238448
750	3.973891	5.518859	1.295874	0.108556	2.091099
1000	2.431685	3.481370	0.783173	0.072470	1.269764
1500	1.320713	1.757796	0.251263	0.046790	0.421687
2000	0.349394	0.429045	0.074539	0.034981	0.147997
3000	0.228830	0.266585	0.030723	0.037508	0.096116
5000	0.209064	0.234896	0.013456	0.041422	0.063004
8000	0.077187	0.053979	0.008682	0.028557	0.033283

From the table we see that the differences are obviously decreasing as the sample size increases. It is convincing that the sample standard deviations will eventually converge to the theoretical ones. Comparing with Table 4.6, one may notice that although they show the same trend of convergence, the differences shown in Table 4.9 are in general larger than the ones from Table 4.6, which indicates a difference in the speeds of convergence. This is consistent with our observations before.

Lastly we investigate the distribution of each element of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. We construct reference Gaussian distributions of mean zero and standard deviations specified by (4.1). Then comparisons are performed between the estimated densities from QMLEs and those reference distributions in the same fashion as Figures 4.6 - 4.10. The results are shown below, with solid lines representing estimated densities derived from QMLEs and dashed lines being the reference Gaussian densities.

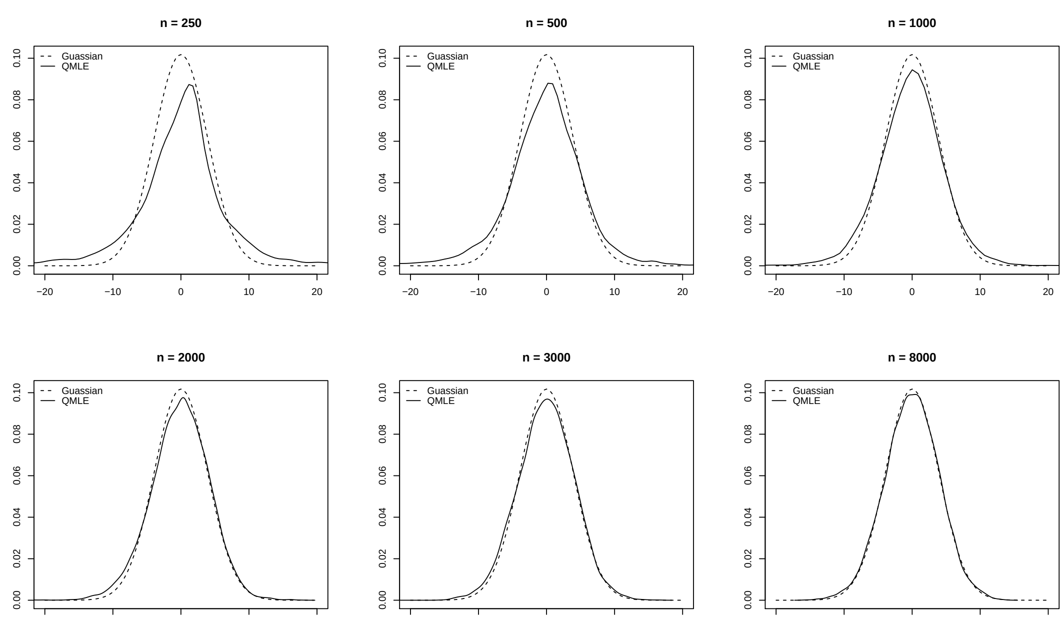

Figure 4.13: Distribution of $\sqrt{n}(\hat{\lambda}_n - \lambda_0)$: t-distributed GARCH-M

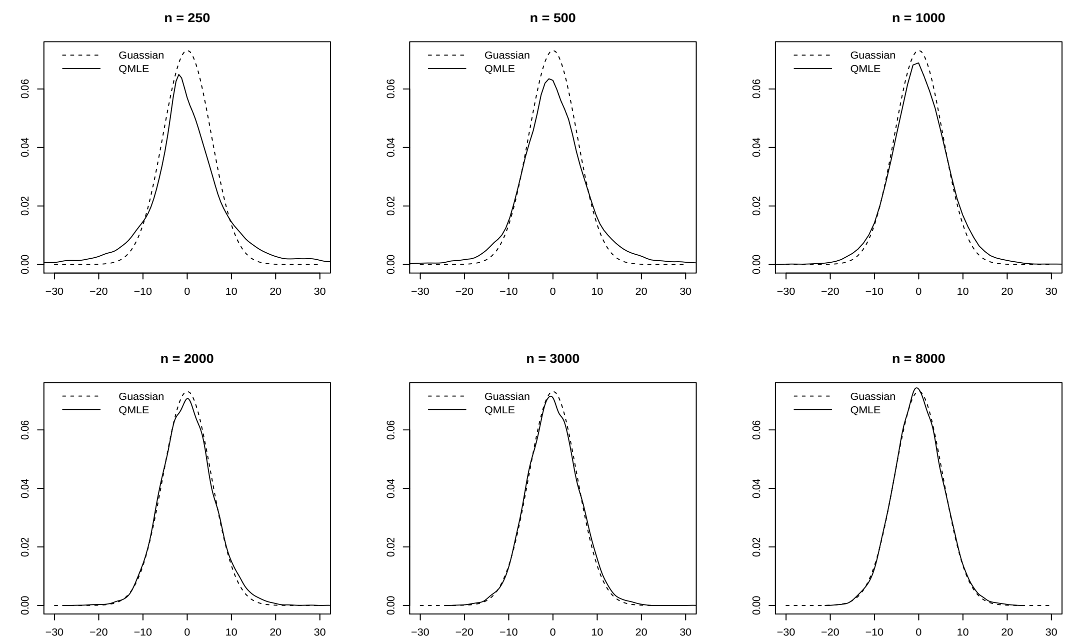

Figure 4.14: Distribution of $\sqrt{n}(\hat{\delta}_n - \delta_0)$: t-distributed GARCH-M

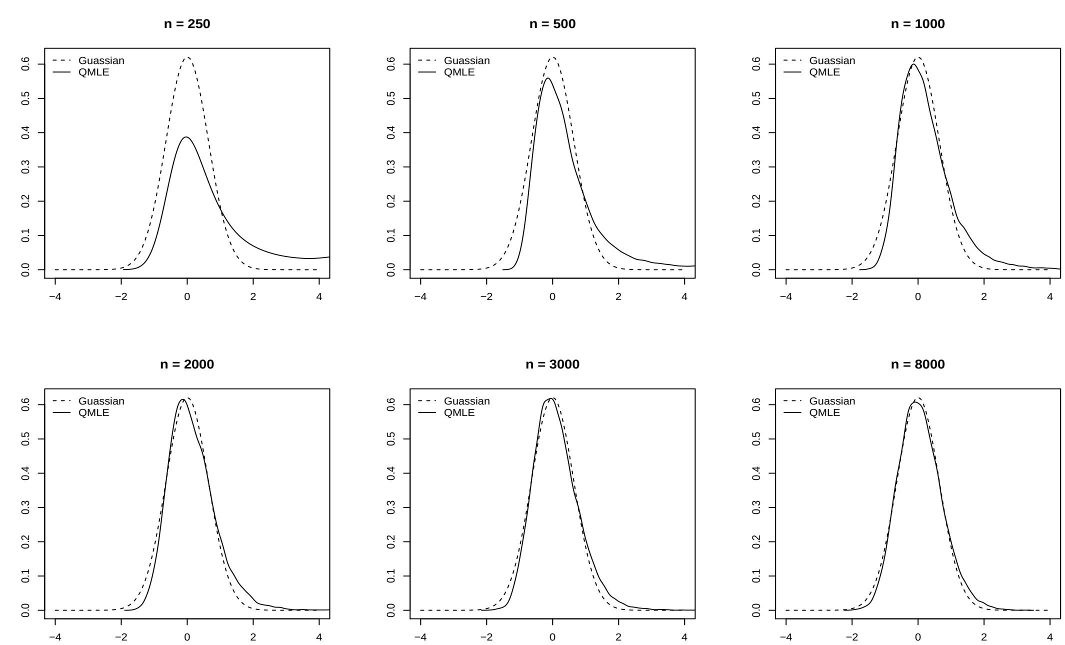

Figure 4.15: Distribution of $\sqrt{n}(\hat{\omega}_n - \omega_0)$: t-distributed GARCH-M

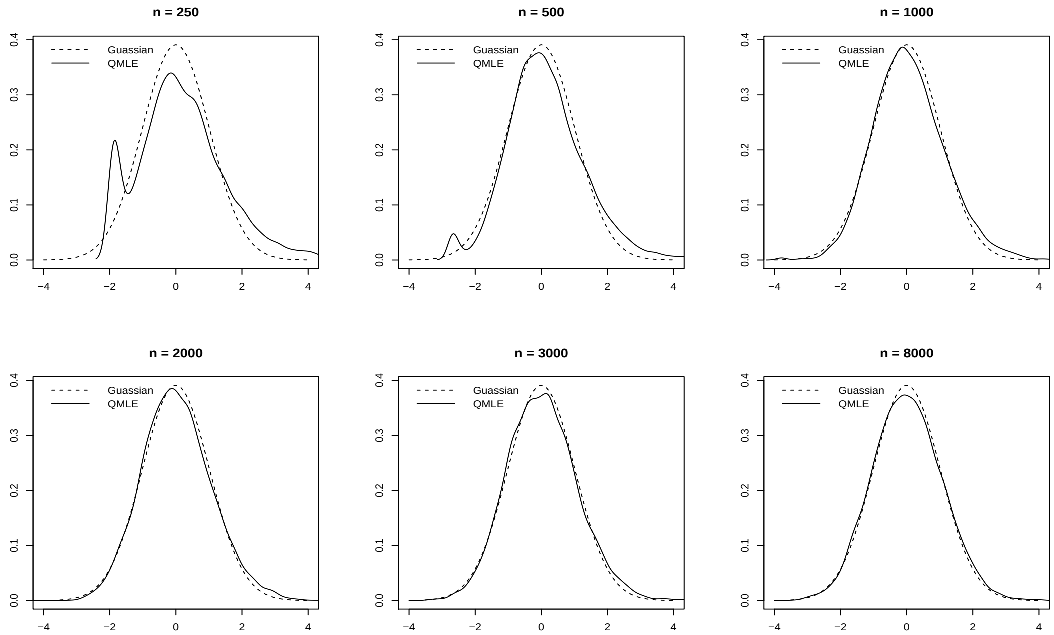

Figure 4.16: Distribution of $\sqrt{n}(\hat{\alpha}_n - \alpha_0)$: t-distributed GARCH-M

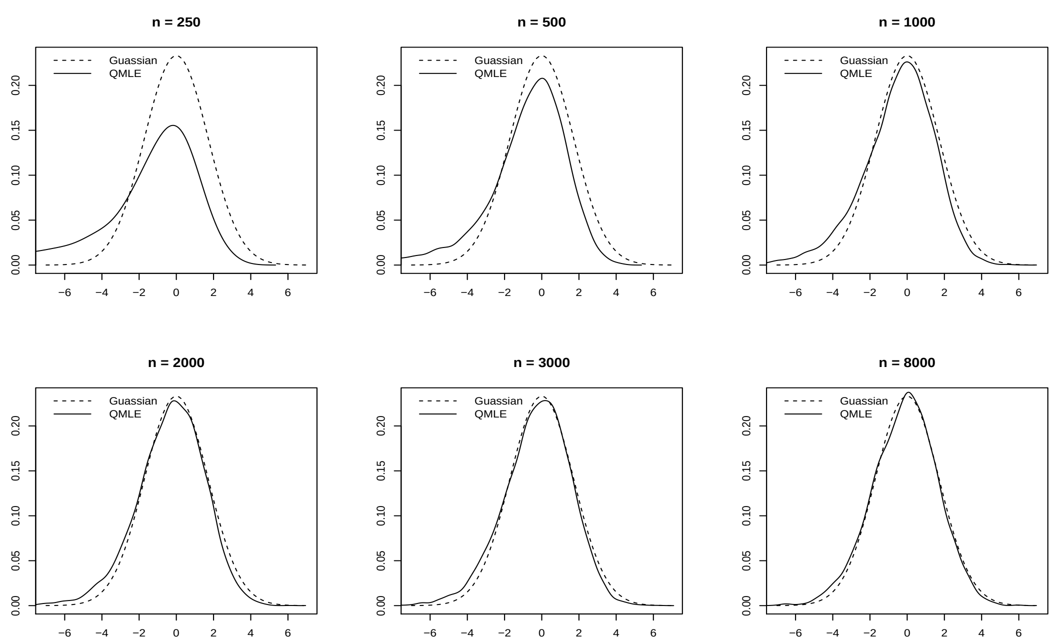

Figure 4.17: Distribution of $\sqrt{n}(\hat{\beta}_n - \beta_0)$: t-distributed GARCH-M

The above graphes exhibit similar trend as Figures 4.6 - 4.10. The solid lines are getting more smooth as the sample size increases, and their discrepancies from dashed reference lines become less noticeable. Although certain discrepancies can still be spotted even at the level $n = 8000$, the overall trend of converging to the normal distribution is convincing. To assess the normality, one may choose to use normal Q-Q plots. Below shows the plot at the level $n = 12000$.

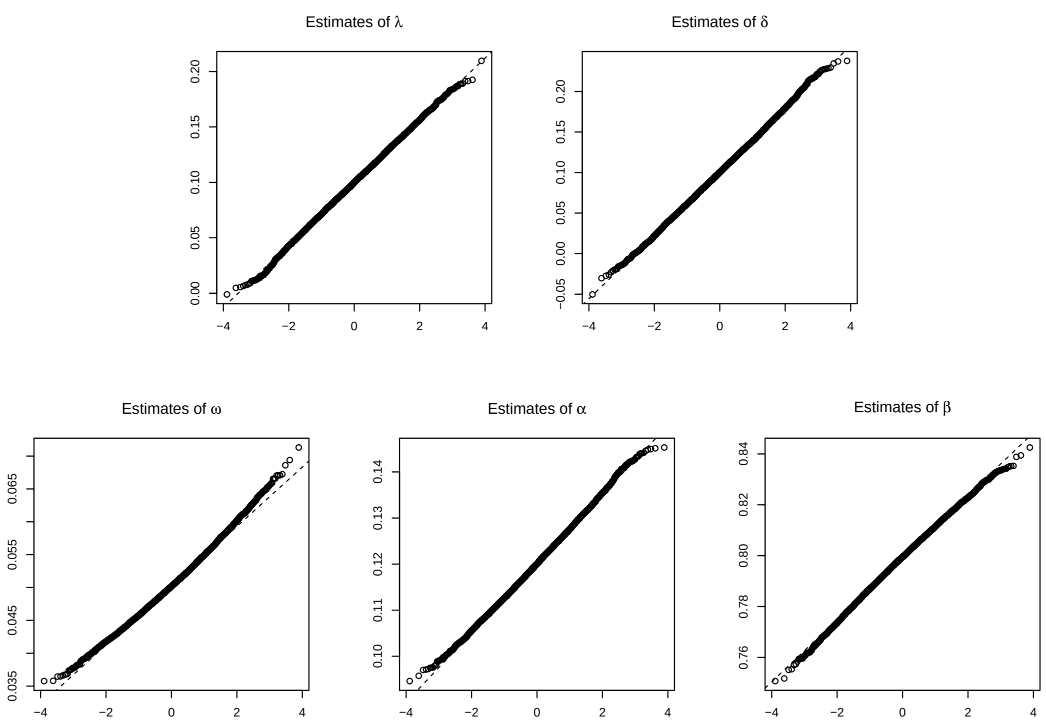

Figure 4.18: Normal Q-Q plot for QMLEs: t-distributed GARCH-M

We could notice a low level of deviation from the graph for ω estimates. However, given the overall trend of convergence shown in previous figures, it is convincing that this parameter will eventually converges to a Gaussian distribution like the others when we keep increasing the sample size.

4.3 The rank condition

Geometric ergodicity is the key property we exploited leading to asymptotic properties of the QMLE. We have imposed a few conditions in Theorem 2.8 to ensure the process is geometrically ergodic. Amongst those conditions the full-rank assumption A2 is relatively abstract and difficult to examine directly.

The rank condition A2 involves both the true parameter θ_0 and the dummy variable θ. It provides necessary constraints for θ_0 which generates the true process as well as sheds light on what the parameter space Θ should look like. This condition is not easy to verify directly because of its algebraic complexity. It involves verifying the existence of some k so the matrix $C_{x_0}^k$ has full rank. In other words, in the event of a failure at one level, one can keep increasing the integer k which decides the dimension of $C_{x_0}^k$, which leads to additional mathematical complexity.

Although it is not easy to study the rank of those matrices algebraically, we could still check this condition to a certain extent by numeric method. It is difficult for us to see what does the parameter space Θ exactly look like, but given a specific value of θ we can check if this value is compliant with Assumption A2. For example, given the true parameter θ_0 specified in Section 4.2, suppose one wants to verify if the following θ is within the parameter space

$$\lambda = 0.5, \quad \delta = 0.5, \quad \omega = 0.2, \quad \alpha = 0.3, \quad \beta = 0.6.$$

We will start with the simplest case $k = 2$. Given the values of θ_0 and θ we could evaluate the matrices A_1, A_2 and B_2 defined in (2.22) - (2.23). Therefore, the matrix $C^2_{x_0}$ becomes a 2×2 matrix depending on the starting value $x_0 = (x_{01},\ x_{02})'$ and control values u_1, u_2.

We could compute the determinant of $C^2_{x_0}$ using packages that are capable of symbolic calculations such as Mathematica. The matrix is full-ranked if the determinant does not equal to zero. Therefore we want to search for roots u_1, u_2 of the equation

$$\left|C^2_{x_0}\right| = 0$$

given any x_0 within the state space. We could also use a 3-D plot to assist us identifying roots. For example, we may set $u_1 = u_2 = 5$. the following graph is based on x_{01} and x_{02} in the range of (0.01, 10).

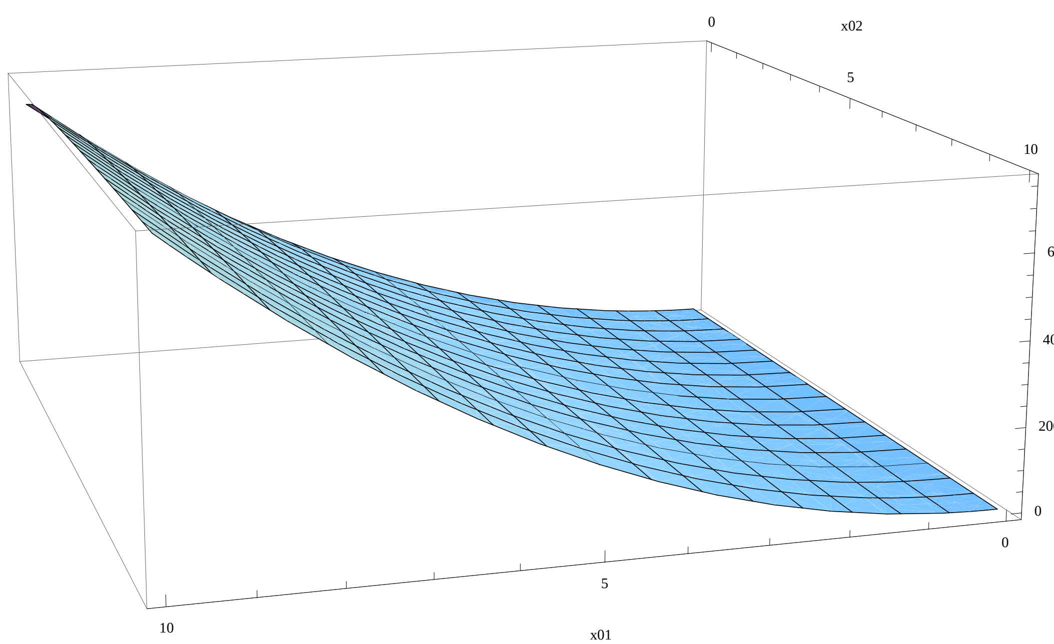

Figure 4.19: 3D visualization of $|C^2_{x_0}|$ fixing u_1 and u_2

In the graph above, values of x_{01} and x_{02} are represented by x and y-axis, while the values of the determinant of $C^2_{x_0}$ on the z-axis. From the graph we could tell when we fix the control values $u_1 = u_2 = 5$, the determinant seems increasing as the values of x_{01} and x_{02} gets large. We could also tell from the graph that all values are strictly positive when x_{01} and x_{02} is relatively far from zero.

To investigate what happens for small positive values of x_{01} and x_{02}, we can fix x_{01} and x_{02} and see how the values of u_1 and u_2 impact the determinant. The graph below shows the values of determinant given u_1 and u_2 ranging from -10 to 10, while fixing $x_{01} = x_{02} = 0.01$.

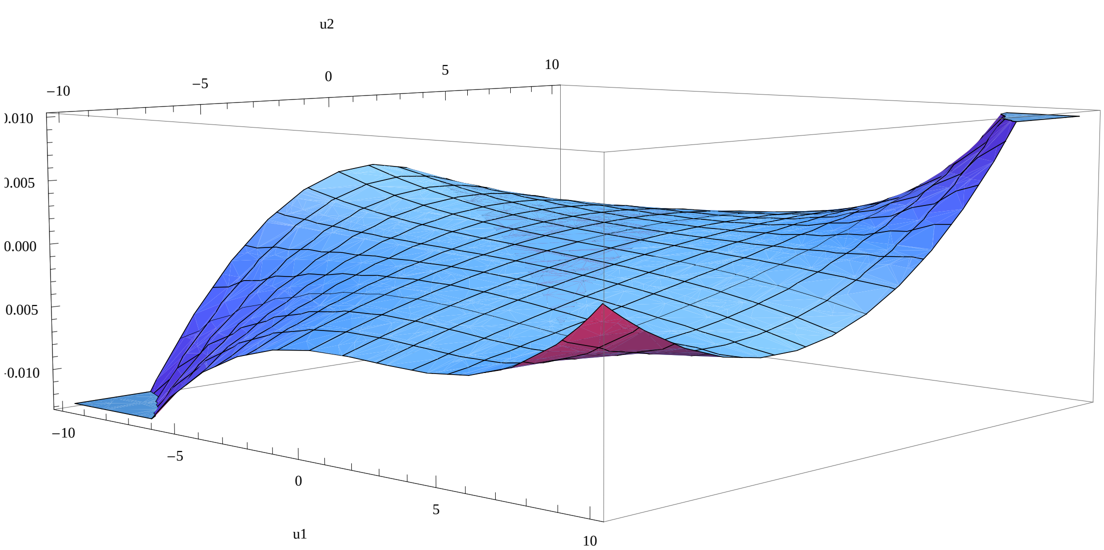

Figure 4.20: 3D visualization of $|C^2_{x_0}|$ fixing x_{01} and x_{02}

From the graph we see that there exists multiple control values within (-10, 10) under which the determinant is nonzero for our fixed starting value. Combined with the previous figure we are confident that for any starting values $x_{01} > 0$, $x_{02} > 0$ we could find certain control values u_1 and u_2 to make the determinant of $C^2_{x_0}$ nonzero. Therefore this particular θ satisfies the full-rank condition A2.

Lastly we want to emphasis that if no nonzero solution could be found for $C^2_{x_0}$, it does not imply the condition is violated. We may continue with $k = 3$ following the same procedure. One may apply Gaussian elimination to check the rank of this 2×3 matrix, or separately check the determinants of the matrices that consist 2 of the 3 column vectors.

Chapter 5

Concluding Remarks

In this thesis we studied the asymptotic properties of the quasi-maximum likelihood estimator of the GARCH-in-mean process. We have found conditions under which this QMLE will be strongly consistent and the distribution around the true parameter will be asymptotically normal.

One difficulty we encountered is the nonlinear structure of the process $\sigma_t^2(\theta)$ that we need to construct the quasi-likelihood. Under the GARCH-in-mean specification this object is defined by a recursion which does not yield an obvious infinite-past representation. This reason made us part ways with traditional approaches that are applicable to GARCH-type models. Instead we constructed a three dimensional Markov model including the observable process y_t, the true conditional variance process $\sigma_t^2(\theta_0)$ and its parametric form $\sigma_t^2(\theta)$. We tackled this Markov model following a systematic approach introduced by Meyn and Tweedie (2009) and concluded its stability properties. The consistency and asymptotic normality can then be concluded by applying appropriate limit theorems.

To conclude the geometric ergodicity property we proposed a few conditions in Chapter 2. The condition A2 is relatively abstract and difficult to verify directly. However, given specific parameter values we could still verify this condition by numeric experiment, which is demonstrated in Chapter 4. Note that this rank condition is in fact related to the ψ-irreducibility of the chain. There's some other work available that could help verifying this property, eg. Cline and Pu (1998). However, their theorem imposed certain assumptions on the innovation process that seems too restrictive for financial series. Doukhan (1994) also includes a number of results around the issue of geometric ergodicity which are more applicable to processes with relatively simpler structures.

The results obtained in this thesis is important for both theoretical research and practical applications. It helps researchers to further study statistical inference and other problems for GARCH-in-mean models, as well as assists practitioners to better understand their estimates hence improve their practices.

Bibliography

Arvanitis, S. and A. Demos (2004). Time dependence and moments of a family of time-varying parameter GARCH in mean models. *Journal of Time Series Analysis 25*, 1–25.

Aue, A., L. Horváth, and J. Steinebach (2006). Estimation in random coefficient autoregressive models. *Journal of Time Series Analysis 27*(1), 61–76.

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

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

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

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

Billingsley, P. (1995). *Probability and Measure.* New York, NY: John Wiley & Sons, Inc.

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

Bollerslev, T. (2010). Glossary to ARCH(GARCH). In T. Bollerslev, J. Russell, and M. Watson (Eds.), *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*. Oxford, U.K.: Oxford University Press.

Bollerslev, T., R. Y. Chou, and K. F. Kroner (1992). ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics 52*, 5–59.

Bougerol, P. and N. Picard (1992a). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics 52*, 115–127.

Bougerol, P. and N. Picard (1992b). Strict stationarity of generalized autoregressive processes. *The Annals of Probability 20*, 1714–1730.

Brandt, A. (1986). The stochastic equation Yn+1 = AnYn+Bn with stationary coefficients. *Advances in Applied Probability 18*(1), 211–220.

Brewer, E., J. M. Carson, E. Elyasiani, I. Mansur, and W. L. Scott (2007). Interest rate risk and equity values of life insurance companies: A GARCH-M model. *The Journal of Risk and Insurance 74*(2), 401–423.

Brockwell, P. J. and R. A. Davis (1987). *Time Series: Theory and Methods.* New York, NY: Springer-Verlag.

Carrasco, M. and X. Chen (2002). Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory 18*, 17–39.

Christensen, B. J., C. M. Dahl, and E. M. Iglesias (2012). Semiparametric inference in a GARCH-in-mean model. *Journal of Econometrics 167*, 458–472.

Cline, D. B. and H. H. Pu (1998). Verifying irreducibility and continuity of a nonlinear time series. *Statistics & Probability Letters 40*(2), 139–148.

Devaney, M. (2001). Time varying risk premia for real estate investment trusts: A GARCH-M model. *The Journal of Risk and Insurance 41*(3), 335–346.

Doob, J. L. (1990). *Stochastic processes.* New York, NY: John Wiley & Sons, Inc.

Doukhan, P. (1994). *Mixing : properties and examples.* New York, NY: Springer-Verlag.

Duan, J. (1995). The GARCH option pricing model. *Mathematical Finance 5*(1), 13–32.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*(4), 987–1007.

Engle, R. F. (1983). Estimates of the variance of U.S. inflation based on ARCH model. *Journal of Money Credit and Banking 15*(3), 286–301.

Engle, R. F. (2001). GARCH 101: the use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives 15*(4), 157–168.

Engle, R. F. (2004). Risk and volatility: econometric models and financial practice. *The American Economic Review 94*(3), 405–420.

Engle, R. F. and D. Kraft (1983). Multiperiod forecast error variance of inflation estimated from ARCH models. In A. Zellner (Ed.), *Applied Time Series Analysis of Economic Data*, Bureau of the Census (Washiongton DC), pp. 293–302.

Engle, R. F., D. M. Lilien, and R. P. Robins (1987). Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica 55*(2), 391–407.

Escanciano, J. C. (2009). Quasi-maximum likelihood estimation of semi-strong GARCH models. *Econometric Theory 25*(2), 561–570.

Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business 38*(1), 34–105.

Feigin, P. D. and R. L. Tweedie (1985). Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments. *Journal of Time Series Analysis 6*, 1–14.

Francq, C. and J.-M. Zakoïan (2004). Maximum Likelihood Estimation of Pure Garch and ARMA-GARCH Processes. *Bernoulli 10*(4), 605–637.

Francq, C. and J.-M. Zakoïan (2010). *GARCH models: structure, statistical inference and financial applications*. West Sussex, U.K.: John Wiley & Sons Ltd.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance 48*(5), 1779–1801.

Gouriéroux, C. (1997). *ARCH models and financial applications*. New York, NY: Springer-Verlag.

Grier, K. B. and M. J. Perry (2000). The effects of real and nominal uncertainty on inflation and output growth: Some GARCH-M evidence. *Journal of Applied Econometrics 15*(1), 45–58.

Hamilton, J. D. and R. Susmel (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics 64*(1-2), 307–333.

Hong, E. P. (1991). The autocorrelation structure for the GARCH-M process. *Economics Letters 37*, 129–132.

Iglesias, E. M. and G. D. A. Phillips (2012). Estimation, testing, and finite sample properties of quasi-maximum likelihood estimators in GARCH-M models. *Econometric Reviews 31*, 532–557.

Jensen, S. T. and A. Rahbek (2004a). Asymptotic inference for nonstationary GARCH. *Econometrica 20*, 1203–1226.

Jensen, S. T. and A. Rahbek (2004b). Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. *Econometrica 72*, 641–646.

Kontonikas, A. (2004). Inflation and inflation uncertainty in the united kingdom: evidence from garch modelling. *Economic Modelling 21*, 525–543.

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

Ling, S. (2007). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics 140*, 849–873.

Ling, S. and W. K. Li (1998). Limiting distributions of maximum likelihood estimators for unstable ARMA model with GARCH errors. *Annals of Statistics 26*, 84–125.

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

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business 36*(4), 394–419.

Meitz, M. and P. Saikkonen (2008). Ergodicity, mixing and existence of moments of a class of markov models with applications to GARCH and ACD models. *Econometric Theory 24*, 1291–1320.

Meyn, S. and R. L. Tweedie (2009). *Markov chains and stochastic stability*. Cambridge,UK: Cambridge University Press.

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

Newey, W. K. and D. McFadden (1994). Chapter 35: Large sample estimation and hypothesis testing. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4. Elsevier Science.

Straumann, D. (2005). *Estimation in Conditionally Heteroscedastic Time Series Models*. New York, NY: Springer-Verlag.

Taylor, S. J. (2005). *Asset price dynamics, volatility and prediction*, Chapter 4. Princeton, NJ: Princeton University Press.

Tjøstheim, D. (1990). Non-linear time series and markov chains. *Advances in Applied Probability 22*, 587–611.

Tsay, R. (2010). *Analysis of financial time series* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

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

Appendix A

Markov Chain Definitions

First of all we need to define two important stochastic stability properties.

Definition A.1 (Strict Stationarity). *The stochastic process $\{X_t\}$ is strictly stationary if the joint distributions of $(X_{t_1}, \dots, X_{t_k})'$ and $(X_{t_1+h}, \dots, X_{t_k+h})'$ are the same for all positive integers k and for all $t_1, \dots, t_k, h \in \mathbb{Z}$.*

Definition A.2 (Ergodicity). *A strictly stationary process $\{X_t\}$ is said to be ergodic if and only if, for any Borel set B and any integer k,*

$$n^{-1} \sum_{t=1}^{n} \mathbf{1}_B(X_t, X_{t+1}, \dots, X_{t+k}) \to P\{(X_1, \dots, X_{1+k}) \in B\}$$

with probability one. Here $\mathbf{1}_B$ is the indicator function.

The definition of ergodicity is in fact much more general and could be extended to nonstationary process, for example cf. Billingsley (1995). However, throughout this thesis we only deal with strictly stationary and ergodic process.

Now proceed with definitions around various terminologies used in the Markov model theory. The main reference here is Meyn and Tweedie (2009).

Definition A.3 (Markov Chain)**.** *The time-homogenous Markov chain is defined as a stochastic process* $\{X_t,\ t \in \mathbb{Z}\}$ *evolving on a state space* X *with a* σ*-algebra* $\mathcal{B}(X)$*, satisfying*

$$P(X_{t+1} \in A \mid X_r, r < t; X_t = x) = P(x, A), \quad \forall t \in \mathbb{Z},\ x \in X, A \in \mathcal{B}(X)$$

where $P = \{P(x, A), x \in X, A \in \mathcal{B}(X)\}$ *is known as the transition probability kernel satisfying*

(i) for each $A \in \mathcal{B}(X)$*,* $P(\cdot, A)$ *is a non-negative measurable function on* X

(ii) for each $x \in X$*,* $P(x, \cdot)$ *is a probability measure on* $\mathcal{B}(X)$

For our purpose of study, throughout the thesis we only consider Markov Chains defined on a general state space X, equipped with a countably generated σ-field $\mathcal{B}(X)$.

The first level of the stability of a Markov Chain is related to whether the chain has the ability to visit any sizable set in the σ-field. Formally it is known as the irreducibility property.

Definition A.4 (Irreducibility)**.** *We call a Markov Chain* X_t φ*-irreducible if* *there exists a measure ϕ on $\mathcal{B}(X)$ such that, whenever $\varphi(A) > 0$, we have*

$$L(x, A) > 0, \quad \forall x \in X$$

where $L(x, A)$ denotes the probability that the chain starts from $x \in X$ and ever enters $A \in \mathcal{B}(X)$. Note that whenever a Markov Chain is φ-irreducible, there exists a maximum irreducibility measure (in the sense that it dominates any other irreducible measures) ψ such that X_t is also ψ-irreducible.

We also want to introduce the the definition of petite sets and the T-chain concept.

Definition A.5 (Petite Set)**.** *A set $C \in \mathcal{B}(X)$ is ν_a-petite if the sampled chain satisfy the bound*

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

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

The T-chain concept is connected with the so-called sampling chain: the Markov chain with transitional probability $K_d := \sum_{n=0}^{\infty} P^n(x, B)d(n)$. A Markov chain is called a T-chain if there exists a sampling distribution d such that K_d possesses a continuous component. For a ψ-irreducible T-chain, every compact set is petite.

Periodicity is another important property for a Markov chain. For a ψ-

irreducible chain on $(X, \mathcal{B}(X))$, there exists some positive integer d and disjoint sets $D_1, \ldots, D_d \in \mathcal{B}(X)$ (d-cycle), such that

i for $x \in D_i$, $P(x, D_{i+1}) = 1$

ii the set $N = [\cup_{i=1}^{d} D_i]^c$ is ψ-null.

We have the following definition for periodicity.

Definition A.6 (Aperiodic Chain)**.** *The largest possible integer d among all d-cycles of a chain X_t is called the period of X_t. If $d = 1$ the X_t is said to be an aperiodic chain.*

Next we introduce the concept of recurrence and Harris recurrence.

Definition A.7 (Recurrence)**.** *A ψ-irreducible chain is called recurrent if for every $x \in X$ and $A \in \mathcal{B}^+(X)$,*

$$E_x \left[\sum_{t=1}^{\infty} I_{(X_t \in A)} \right] = \infty$$

where $\mathcal{B}^+(X)$ includes all sets in $\mathcal{B}(X)$ that are ψ-positive. E_x indicates the chain is initiated by $X_0 = x$. I is the indicator function.

Note that the quantity $\sum\limits_{t=1}^{\infty} I_{(X_t \in A)}$ is also known as the occupation time, representing the number of visits by X_t to A after time zero. Now we define a stronger form of recurrence.

Definition A.8 (Harris Recurrence). *A ψ-irreducible chain is called Harris recurrent if for every $x \in X$ and $A \in \mathcal{B}^+(X)$,*

$$P(\left[\sum_{t=1}^{\infty} I_{(X_t \in A)}\right] = \infty | X_0 = x) = 1$$

where $\mathcal{B}^+(X)$ includes all sets in $\mathcal{B}(X)$ that are ψ-positive. I is the indicator function.

Lastly we define the invariant measure for a chain. Sometimes it is also called the stationary measure.

Definition A.9 (Invariant Measure). *A σ-finite measure π on $\mathcal{B}(X)$ with the property*

$$\pi(A) = \int_X \pi(dx) P(x, A), \quad A \in \mathcal{B}(X)$$

is called invariant.

Curriculum Vitae

Name:	Weiwei Liu
Post-Secondary Education and Degrees:	University of Western Ontario London, ON, Canada 2009 - 2013, Ph.D Candidate
	University of Western Ontario London, ON, Canada 2008 - 2009, M.Sc
	Huazhong University of Science and Technology Wuhan, China 2003 - 2007, B.Sc
Honours and Awards:	Western Graduate Research Scholarship 2008 - 2013
Related Work Experience:	Teaching Assistant and Research Assistant University of Western Ontario 2008 - 2013