8-12-2016

Different Estimations of Time Series Models and Application for Foreign Exchange in Emerging Markets

Jingjing Wang

Different estimations of time series models and application for foreign exchange in emerging markets

By

Jingjing Wang

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Statistics
in the Department of Mathematics and Statistics

Mississippi State, Mississippi

August 2016

Different estimations of time series models and application for foreign exchange in emerging markets

By

Jingjing Wang

Approved:

Tung-Lung Wu
(Major Professor)

T. Len Miller
(Minor Professor)

Matthew S. McBride
(Committee Member)

Mohammad Sepehrifar
(Committee Member)

Corlis P. Johnson
(Graduate Coordinator)

Rick Travis
Interim Dean
College of Arts & Sciences

Name: Jingjing Wang

Date of Degree: August 12, 2016

Institution: Mississippi State University

Major Field: Statistics

Major Professor: Tung-Lung Wu

Title of Study: Different estimations of time series models and application for foreign exchange in emerging markets

Pages in Study: 36

Candidate for Degree of Master of Science

Time series models have been widely used in simulating financial data sets. Finding a nice way to estimate the parameters is really important. One of the traditional ways is to use maximum likelihood estimation to make an approach. However, when the error terms don't have normality, MLE would be less efficient. Quasi maximum likelihood estimation, also regarded as Gaussian MLE, would be more efficient. Considering the heavy-tailed financial data sets, we can use non-Gaussian quasi maximum likelihood, which needs less assumptions and conditions. We use real financial data sets to compare these estimators.

DEDICATION

To my dear parents and professors who have taught me. Without their help, through these years, I couldn't finish this.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

With the rise and development of emerging market economics, emerging markets play a more and more crucial role in the global market. Analyzing the financial data set from emerging markets become an interesting but difficult task. We always need to build models under different conditions and assumptions. One good approach is to build time series models.

The Autoregressive Moving Average (ARMA) model was first described by Whittle (1951) and became popularized by E. P. Box and Jenkins (1971). The ARMA model is combined with the AR model and the MA model, and one of the most important properties of the ARMA model is stationary. When we analyze a sequence which is not stationary but we still want to use the ARMA model to model it, we first need to make this sequence "stationary". For example, we can take the logarithm of this sequence and then take a difference to remove the non-stationary property. There are a lot of applications of the ARMA model. One may be familiar with the technical index in stock and option markets—MACD (Moving Average Convergence Divergence), which is a direct use of the ARMA model to summarize the past and make predictions.

The Autoregressive Conditional Heteroscedasticity (ARCH) model was first proposed by Engle (1982) and their generation was introduced by Bollerslev (1986). Bollerslev (1987) became the first person to model the daily exchange rates conditional variance by using the GARCH model. ARCH-type models sometimes are considered in the family of stochastic volatility models. In particular the error term of the ARCH model is related to the squares of previous innovations.

Maximum likelihood estimation (MLE) is good if we know the right distribution of the data sets. For time series models, least squares is a better fit, and the application of the MLE in some developed markets attains good results. However it is not so efficient for emerging markets. The quasi-MLE comes without the normality assumption of the error term, which leads to more powerful techniques to estimate parameters. The extension of the QMLE--Non-Gaussian quasi MLE, which is widely used in GARCH models, is QMLE with non-Gaussian likelihood family, so it doesn't depend on consistent estimates. Numerous papers studied asymptotic distributions and QMLE estimators. Weiss (1984), Lee and Hansen (1994) focused on deriving the asymptotic distributions of the estimators, Berkes et al. (2003) and Francq and Zacoian (2005) obtained the minimal assumptions required for the consistency and asymmetric normality of the QMLE. In addition, Hall and Yao (2003) derived the asymptotic distribution of the QMLE with infinite variance under the restriction of the parameter space. Fan, Qi and Xiu (2014) gives details in using non-Gaussian quasi MLE.

The rest of thesis will be organized as follows: Chapter 2 will introduce the time series model, that is, the related definitions and properties of AR(p), MA(q), ARMA(p,q), ARCH(p) and GARCH(p,q); in Chapter 3, I will give details about different estimations of different time series models, these estimations are based on maximum likelihood and generated by maximum likelihood, including MLE, quasi-MLE and non-Gaussian quasi MLE; in Chapter 4, I will use the data sets of foreign exchange rates in emerging markets, BRL/USD and ARS/USD, and make comparisons; in the last Chapter, I will make some conclusions on the different estimations.

CHAPTER II

RELATED DEFINITIONS AND PORPERTIS OF TIME SERIES MODELS

2.1 AR/MA/ARMA Model

2.1.1 AR Models

The Autoregressive (AR) model is a kind of linear prediction, that is, when an observation from a time series can be regressed as on previous observations from the same time series. So the essence of this model is similar to linear regression, in which the response variable in the previous time period becomes the predictor and the error terms are under some assumptions. In statistics, the AR model represents a type of random process, which predicts time-varying processes in nature, economics and so on.

The notation AR(p) means an AR Model of order p. The form of AR(p) model is:

$$X_t = C + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t .$$

The notation AR(p) means an AR model, where C is the constant, φ_i, $i = 1,2,\ldots,p$ is the parameter and the error term ε_t is i.i.d. $N(0,1)$.

We calculate the expectation of X_t as

$$\mu_p = E(X_t) = \frac{C}{1 - \sum_{i=1}^{p} \varphi_i},$$

and the covariance matrix:

$$V_p = \begin{bmatrix} \alpha_0 & \alpha_1 & . & . & . & \alpha_{p-1} \\ \alpha_1 & \alpha_2 & . & . & . & \alpha_{p-2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \alpha_{P-1} & \alpha_{p-2} & . & . & . & \alpha_0 \end{bmatrix}.$$

The joint distribution of the first p observations is given by:

$$f_{X_p, X_{p-1}, \ldots, X_1}(X_p, X_{p-1}, \ldots, X_1; \Theta)$$

$$= (2\pi)^{-1/2} \,|\, \sigma^2 V_p \,|^{-1/2} \exp\left[-\frac{1}{2\sigma^2} (X_p - \mu_p)' V_p^{-1} (X_p - \mu_p) \right] \quad (2.1)$$

for $t > p$,

$$f_{X_t | X_{t-1}, \ldots, X_1}(X_t \mid X_{t-1}, \ldots, X_1; \Theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{(X_t - C - \sum_{i=1}^{p} \varphi_i X_{t-i})^2}{2\sigma^2} \right). \quad (2.2)$$

We can derive that the joint distribution of $X_t, X_{t-1}, \ldots, X_1$ is

$$f_{X_t, X_{t-1}, \ldots, X_1}(X_t, X_{t-1}, \ldots, X_1; \Theta) = f_{X_p, X_{p-1}, \ldots, X_1}(X_p, X_{p-1}, \ldots, X_1; \Theta) \times$$

$$\prod_{t=p+1}^{T} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(X_t - C - \sum_{i=1}^{p} \varphi_i X_{t-i})}{2\sigma^2} \right]. \quad (2.3)$$

2.1.2 MA Models

The Moving-Average (MA) model is used as a common approach for modeling univariate time series. In this model, we use the linear combination of the random disturbance or prediction error of the past to derive the current predictive value.

The notation MA(q) means an MA model with order q, and the structure of MA(q) is

$$X_t = u + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} ,$$

where u is the mean of the series, $\theta_j, j = 1,2,...,q$ is the parameters of the model, $\varepsilon_t, \varepsilon_{t-j}, j = 1,2,...,q$ is the error terms with mean 0, variance σ^2 and $\mathrm{cov}(\varepsilon_t, \varepsilon_{t-j}) = 0$.

By the above assumption, we can conduct that:

$$E(X_t) = u ,$$

$$Var(X_t) = (1 + \sum_{j=1}^{q} \theta_j^2)\sigma^2 ,$$

and

$$Cov(X_t, X_{t-j}) = 0 (j > q) .$$

2.1.3 ARMA Models

The Autoregressive Moving Average (ARMA) model was first introduced by Peter Whittle (1951) and then becomes popular by George Box and Gwilym Jenkins. It is combined with the AR model and the MA model. The notation ARMA(p,q) represents the ARMA model with order p of AR Model and order q of MA model.

So the form of the ARMA(p,q) model is

$$X_t = u + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} .$$

An ARMA(p,q) needs to satisfy the stationary property, i.e., all the roots of

$$1 - \varphi_1 L - ... - \varphi_p L^p = 0$$

and

$$1+\theta_1 L+\ldots+\theta_q L^q=0$$

lie outside the unit circle, where ε_t is i.i.d $N(0,\sigma^2)$.

Also, we can calculate the mean μ_p of X_t

$$\mu_p=E(X_t)=\frac{u}{1-\sum_{i=1}^{p}\varphi_i}$$

2.2 ARCH/GARCH Model

2.2.1 ARCH Model

The Autoregressive Conditional Heteroscedasticity (ARCH) model was first proposed by Engle (1982), which developed rapidly in econometrics later on. Roughly speaking, this model takes all available current information as the condition, and then uses an autoregressive form to describe the variation of the variance. For a time series at different times, the available information is different, so the relevant conditional variance is also different. By using the ARCH model, we can characterize the time dependent conditional variance.

The structure of ARCH(q) model, q means the order, is

$$\varepsilon_t=\sigma_t v_t$$

$$\sigma_t^2=\alpha_0+\sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2,$$

where $t=1,2,\ldots,T$, $\alpha_0\geq 0$, $\alpha_i\geq 0, i=1,2,\ldots,q$ is constant, v_t is i.i.d. with mean 0 and variance 1, v_t and ε_t are independent.

Let F_{t-1} be the set of available information at $t-1$. Then the conditional expectation of ε_t is

$$E(\varepsilon_t \mid F_{t-1}) = \sigma_t \cdot E(\nu_t \mid F_{t-1}) = 0 \tag{2.4}$$

which implies the conditional variance of ε_t is

$$Var(\varepsilon_t \mid F_{t-1}) = E\{[\varepsilon_t - E(\varepsilon_t \mid F_{t-1})]^2 \mid F_{t-1}\}$$

$$= E(\varepsilon_t^2 \mid F_{t-1})$$

$$= E(\sigma_t^2 \nu_t^2 \mid F_{t-1})$$

$$= E(\nu_t^2) E(\alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \mid F_{t-1})$$

$$= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 = \sigma_t^2 \tag{2.5}$$

Next, we conduct the unconditional mean and variance of ε_t

$$E(\varepsilon_t) = E[E(\varepsilon_t \mid F_{t-1})] = 0$$

$$Var(\varepsilon_t) = Var[E(\varepsilon_t \mid F_{t-1})] + E[(Var(\varepsilon_t \mid F_{t-1})]$$

$$= \alpha_0 + \sum_{i=1}^{q} \alpha_i Var(\varepsilon_{t-i}) .$$

Moreover, if ε_t^2 is covariance stationary, we have

$$1 - \alpha_1 x - \alpha_2 x^2 - ... - \alpha_q x^q = 0$$

and all of the roots of this equation lie outside the unit circle. Moreover, the unconditional variance is stable, so

$$Var(\varepsilon_t) = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i} \text{ where } \sum_{i=1}^{q} \alpha_i < 1 .$$

However, if $\sum_{i=1}^{q} \alpha_i \geq 1$, the variance does not exist and the process is no longer covariance-stationary.

2.2.2 GARCH Model

The General Autoregressive Conditional Heteroscedasticity (GARCH) model is generated by the ARCH model, which was set up by Bollerslev (1986). The GARCH(0,q) is equal to ARCH(q). Since Engle (1982) proposed a heteroscedasticity property of the ARCH model in analyzing time series, T. Bollerslev (1986) proposed the GARCH Model, which is a specialized regression model for financial data. Besides the similarities of ordinary regression models, the GARCH model goes a step further in modeling the variance of error. This makes it especially suitable for the analysis and prediction of volatility. This kind of analysis can play a very significant role in the decisions of investors. Its meaning is more than the value of their own analysis and forecasting.

The structure of GARCH(p,q) Model is

$$\begin{aligned} \varepsilon_t &= \sigma_t v_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 , \end{aligned} \tag{2.6}$$

where integer $p \geq 0, q \geq 0$, $\alpha_0 \geq 0$, $\alpha_i \geq 0, i = 1,2,...,q$; $\beta_j \geq 0, j = 1,2,...,p$.

It can be seen that GARCH(0,q) is ARCH(q), so the GARCH model has same characteristics of ARCH model, that is, they can simulate the price volatility clustering phenomenon; the difference is that the conditional variance of the GARCH model is not only the linear function of the lagged squared perturbation, but also the linear function of the lagged conditional variance.

Similarly for ARCH(q), we first define F_{t-1} as the set of available information at time $t-1$, then we conduct the conditional mean and the variance of ε_t

$$E(\varepsilon_t \mid F_{t-1}) = \sigma_t E(\nu_t \mid F_{t-1}) = 0$$

$$\begin{aligned} Var(\varepsilon_t \mid F_{t-1}) &= E\{[\varepsilon_t - E(\varepsilon_t \mid F_{t-1})]^2 \mid F_{t-1}\} \\ &= E(\varepsilon_t^2 \mid F_{t-1}) = E(\sigma_t^2 \nu_t^2 \mid F_{t-1}) \\ &= E(\nu_t^2) E\left(\alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \mid F_{t-1}\right) \\ &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 = \sigma_t^2 . \end{aligned}$$

Next, we compute the unconditional mean and variance of ε_t

$$E(\varepsilon_t) = E[E(\varepsilon_t \mid F_{t-1})] = 0$$

$$\begin{aligned} E(\varepsilon_t^2) = E(\sigma_t^2 \nu_t^2) &= E[\nu_t^2 (\alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2)] \\ &= \alpha_0 + \sum_{i=1}^{q} \alpha_i E(\varepsilon_{t-i}^2) + \sum_{j=1}^{p} \beta_j E(\varepsilon_{t-j}^2) . \end{aligned}$$

So the unconditional variance of ε_t is

$$Var(\varepsilon_t) = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j} .$$

This implies

$$\sum_{i=1}^{q}\alpha_i+\sum_{j=1}^{p}\beta_j<1, \tag{2.7}$$

moreover,

$$\sum_{i=1}^{q}\alpha_i<1 \text{ and } \sum_{j=1}^{p}\beta_j<1.$$

Under condition (2.6), we can conclude that:

$$\sigma_t^2=\frac{\alpha_0}{1-\sum_{j=1}^{p}\beta_j}+\sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2+\sum_{i=1}^{q}\alpha_i\sum_{k=1}^{\infty}\sum_{j_1=1}^{p}\cdots\sum_{j_k=1}^{p}\beta_{j_1}\cdots\beta_{j_k}\varepsilon^2{}_{t-i-j_1-\ldots-j_k}.$$

Notice σ_t is a function of ε_s, where $-\infty<s\le t-1$

Then, with given α_0, p, q, the GARCH model is

$$y_t=C+\sum_{i=1}^{k}\alpha_i y_{t-i}+\varepsilon_t$$

where the conditional probability of ε_t is

$$P(\varepsilon_t\mid\varepsilon_{t-1},\ldots,\varepsilon_0)=\frac{1}{\sqrt{2\pi}\sigma_t}e^{-\frac{\varepsilon_t^2}{2\sigma_t^2}}.$$

CHAPTER III

ESTIMATION METHODS OF TIME SERIES MODELS

3.1 Maximum Likelihood Estimation

There are two important properties of Maximum Likelihood Estimation (MLE): consistency and asymptotic normality.

1. Consistency

We say that an estimate $\hat{\theta}$ is consistent if $\hat{\theta} \to \theta_0$ in probability a.s. when $n \to \infty$, where θ_0 is the 'true' unknown parameter of the distribution of the sample.

2. Asymptotic Normality

We say that an estimate $\hat{\theta}$ is asymptotically if

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \sigma_{\theta_0}^2),$$

where $\sigma_{\theta_0}^2$ is called the asymptotic variance of the estimator $\hat{\theta}$. Asymptotic normality says that the estimator not only converges to the unknown parameter, but it converges fast enough, at rate $1/\sqrt{n}$. In this paper, we have T observations, so we use T instead n.

3.1.1 MLE for AR(p) Model

By using equation (2.1), we derive the MLE of the AR(p) model

$$l(\theta) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 + \frac{1}{2}|V_p^{-1}| - \frac{1}{2\sigma^2}(X_p - \mu_p)' V_p^{-1}(X_p - \mu_p)$$

$$-\sum_{t=p+1}^{T}\frac{\left(X_t-C-\sum_{i=1}^{p}\varphi_i X_{t-i}\right)^2}{2\sigma^2}. \tag{3.1}$$

However, directly using equation (3.1), it is hard to find the MLE of σ^2, so we use the conditional MLE to find the estimation of θ in the AR(p) model.

Using equation (2.3), we compute

$$l^*(\theta)=-\frac{T-p}{2}\ln(2\pi)-\frac{T-p}{2}\ln(2\sigma^2)-\sum_{t=p+1}^{T}\frac{\left(X_t-\sum_{i=1}^{p}\varphi_i X_{t-i}-C\right)^2}{2\sigma^2}.$$

Then the estimate of σ^2 comes from

$$\frac{\partial l^*}{\partial\sigma^2}=-\frac{T-p}{2\cdot 2\sigma^2}+\frac{1}{2\sigma^4}\sum_{t=p+1}^{T}\left(X_t-\sum_{i=1}^{p}\varphi_i X_{t-i}-C\right)^2=0$$

thus, the estimation of σ^2 is

$$\hat{\sigma}^2=\frac{1}{T-p}\sum_{t=p+1}^{T}\left(X_t-\sum_{i=1}^{p}\varphi_i X_{t-i}-C\right)^2=\frac{1}{T-p}\sum_{t=p+1}^{T}\varepsilon_t^2.$$

The estimate of φ_i is

$$\frac{\partial l^*}{\partial\varphi_i}=\frac{1}{\sigma^2}\sum_{t=p+1}^{T}\varepsilon_t X_{t-i}=0$$

which equals to

$$\sum_{t=p+1}^{T}\left(X_t-\sum_{i=1}^{p}\varphi_i X_{t-i}-C\right)X_{t-i}=0.$$

In order to get the information matrix, we first find the second derivative of $l^*(\theta)$:

$$\frac{\partial^2 l^*}{\partial\varphi_i\partial\varphi_j}=-\frac{1}{\sigma^2}\sum_{t=p+1}^{T}X_{t-i}X_{t-j}-\frac{1}{\sigma^2}\sum_{t=p+1}^{T}\varepsilon_t X_{t-i-j}$$

$$\frac{\partial^2 l^*}{\partial \varphi_i \partial \sigma^2} = -\frac{1}{\sigma^4} \sum_{t=p+1}^{T} \varepsilon_t X_{t-i}$$

$$\frac{\partial^2 l^*}{\partial \sigma^4} = -\frac{T}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{t=p+1}^{T} \varepsilon_t^2 .$$

Second, we find the negative expectation of the above three equations to obtain the information matrix, denoted as $I(\theta)$:

$$-E\left(\frac{\partial^2 l^*}{\partial \varphi_i \partial \varphi_j} \right) = \frac{1}{\sigma^2} \sum_{i=1}^{p} E\left(X_{t-i} X_{t-j} \right) = V_P ,$$

$$-E\left(\frac{\partial^2 l^*}{\partial \varphi_i \partial \sigma^2} \right) = \frac{1}{\sigma^4} E\left(\sum_{i=1}^{p} \varepsilon_t X_{t-i} \right) = 0 ,$$

$$-E\left(\frac{\partial^2 l^*}{\partial \sigma^4} \right) = \left(\frac{T}{2\sigma^2} + \frac{1}{\sigma^6} T\sigma^2 \right) / T = \frac{1}{2\sigma^4} .$$

Then the information matrix is

$$\sqrt{T}(\hat{\Theta} - \Theta) \to N(0, V),$$

where

$$V = \begin{bmatrix} V_p^{-1} & 0 \\ 0 & 2\sigma^4 \end{bmatrix} .$$

3.1.2 MLE for MA(q) Model

First, we define $\Theta = (u, \theta_1, \theta_2, ..., \theta_q, \sigma^2)'$.

Consider

$$X_1 = u + \varepsilon_1 + \sum_{j=1}^{q} \theta_j \varepsilon_{1-j} ,$$

where we assume that the first q values of ε are all zero in order to make a simple approach:

$$\varepsilon_0 = \varepsilon_{-1} = ... = \varepsilon_{-q+1} = 0$$

then

$$(X_1 \mid \varepsilon_0 = \varepsilon_{-1} = .. = \varepsilon_{-q+1} = 0) \sim N(u, \sigma^2)$$

i.e.

$$f_{X_1|\varepsilon_0=\varepsilon_{-1}=...=\varepsilon_{-q+1}}(x_1 \mid \varepsilon_0 = \varepsilon_{-1} = ... = \varepsilon_{-q+1} = 0; \Theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_1 - u)^2}{2\sigma^2}\right]$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\varepsilon_1^2}{2\sigma^2}\right].$$

Consider the following: the second observation X_2 conditional on the first observation

$$X_1 = x_1$$

$$X_2 = u + \varepsilon_2 + \sum_{j=1}^{q} \theta_j \varepsilon_{2-j}$$

also, given $X_1 = x_1$, we can obtain the value of ε_1

$$\varepsilon_1 = x_1 - u$$

and

$$\varepsilon_0 = \varepsilon_{-1} = ... = \varepsilon_{-q+2} = 0$$

therefore

$$(X_2 \mid X_1 = x_1, \varepsilon_0 = \varepsilon_{-1} = ... = \varepsilon_{-q+2}) \sim N((u + \theta_1 \varepsilon_1), \sigma^2)$$

i.e.

$$f_{X_2|X_1,\varepsilon_0=\varepsilon_{-1}=\ldots=\varepsilon_{-q+2}=0}(x_2 \mid x_1, \varepsilon_0=\varepsilon_{-1}=\ldots=\varepsilon_{-q+2}=0;\Theta)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x_2-u-\theta_1\varepsilon_1)^2}{2\sigma^2}\right]$$

$$=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{\varepsilon_2^2}{2\sigma^2}\right]$$

ε_2 can be calculated as:

$$\varepsilon_2 = x_2 - u - \theta_1\varepsilon_1 .$$

In this way, the full sequence of $\{\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_T\}$ can be obtained by $\{x_1,x_2,..,x_T\}$, i.e.

$$\varepsilon_t = x_t - u - \sum_{j=1}^{q}\theta_j\varepsilon_{t-q} .$$

The likelihood of the complete sample size T can be calculated as

$$f_{X_T,X_{T-1},\ldots,X_1|\varepsilon_0=\varepsilon_{-1}=\ldots=\varepsilon_{-q+1}}(x_T,x_{T-1},\ldots,x_1 \mid \varepsilon_0=\varepsilon_{-1}=\ldots=\varepsilon_{-q+1}=0;\Theta)$$

$$=f_{X_1|\varepsilon_0=\varepsilon_{-1}=\ldots=\varepsilon_{-q+1}=0}(x_1 \mid \varepsilon_0=\varepsilon_{-1}=\ldots=\varepsilon_{-q+1}=0;\Theta)\cdot$$

$$\prod_{t=2}^{T} f_{X_t|X_{t-1},..,X_1,\varepsilon_0=\varepsilon_{-1}=\ldots=\varepsilon_{-q+1}=0}(x_t \mid x_{t-1},\ldots,x_1,\varepsilon_0=\varepsilon_{-1}=\ldots=\varepsilon_{-q+1}=0;\Theta) .$$

We can derive the conditional log-likelihood function:

$$l^*(\Theta)=\ln f_{X_T,X_{T-1},\ldots,X_1,\varepsilon_0=\varepsilon_{-1}=\ldots=\varepsilon_{-q+1}=0}(x_T,x_{T-1},\ldots,x_1 \mid \varepsilon_0=\varepsilon_{-1}=\ldots=\varepsilon_{-q+1}=0;\Theta)$$

$$=-\frac{T}{2}\ln(2\pi)-\frac{T}{2}\ln(\sigma^2)-\sum_{t=1}^{T}\frac{\varepsilon_t^2}{2\sigma^2} .$$

The first derivative of $l^*(\Theta)$ is

$$\frac{\partial l^*}{\partial\theta_i}=\frac{1}{\sigma^2}\sum_{t=1}^{T}\varepsilon_t\varepsilon_{t-i} ,$$

$$\frac{\partial l^*}{\partial\sigma^2}=\frac{T}{2\sigma^2}+\left(\frac{1}{2\sigma^4}\right)\sum_{t=1}^{T}\varepsilon_t^2 .$$

Then we calculate the second derivative of $l^*(\Theta)$ and find the expectation:

$$\frac{\partial^2 l^*}{\partial\theta_i\partial\theta_j} = \left(\frac{1}{\sigma^2}\right)\sum_{t=p+1}^{T}\left(\frac{\partial\varepsilon_t}{\partial\theta_j}\right)\varepsilon_{t-i} + \left(\frac{1}{\sigma^2}\right)\sum_{t=p+1}^{T}\varepsilon_t\left(\frac{\partial\varepsilon_{t-i}}{\partial\theta_j}\right)$$

$$= \left(\frac{1}{\sigma^2}\right)\sum_{t=p+1}^{T}\varepsilon_{t-i}\varepsilon_{t-j} + \left(\frac{1}{\sigma^2}\right)\sum_{t=p+1}^{T}\varepsilon_t\varepsilon_{t-i-j} .$$

Note that

$$E(\varepsilon_t\varepsilon_{t-i-j}) = 0,$$

which implies that

$$\lim_{T\to\infty}\left(-E\frac{\partial^2 l^*}{\partial\theta_i\partial\theta_j}\right) = \lim_{T\to\infty}\left(-\left(\frac{1}{\sigma^2}\right)\sum_{t=p+1}^{T}\varepsilon_{t-i}\varepsilon_{t-j}\right) = \lambda_{ij} .$$

We calculate:

$$\frac{\partial^2 l^*}{\partial\theta_i\partial\sigma^2} = -\frac{1}{\sigma^4}\sum_{t=1}^{T}\varepsilon_t\varepsilon_{t-i},$$

$$E\left(-\frac{\partial^2 l^*}{\partial\theta_i\partial\sigma^2}\right) = -\frac{1}{\sigma^4}\sum_{t=1}^{T}E(\varepsilon_t\varepsilon_{t-i}) = 0 ;$$

and

$$\frac{\partial^2 l^*}{\partial\sigma^4} = -\frac{1}{\sigma^6}\sum_{t=1}^{T}\varepsilon_t^2 + \frac{T}{2\sigma^4} ,$$

$$E\left(-\frac{\partial^2 l^*}{\partial\sigma^4}\right) = -E\left(-\frac{1}{\sigma^6}\sum_{t=1}^{T}\varepsilon_t^2 + \frac{T}{2\sigma^4}\right) = \frac{T}{2\sigma^4} .$$

Then the information matrix of MA(q) is

$$V = \begin{bmatrix} (\lambda_{ij}) & \vec{0} \\ \vec{0} & 2\sigma^4 \end{bmatrix}.$$

3.1.3 MLE for ARMA(p,q) Model

First, we define $\Theta = (\mu, \varphi_1, \varphi_{2,} ..., \varphi_p, \theta_1, ..., \theta_q, \sigma^2)$ as the vector of observations parameters. By conditioning on X 's and ε's, we can obtain the approximation to the likelihood function of the ARMA(p,q) model.

The $(p+1)$th observation is

$$X_{p+1} = u + \varepsilon_{p+1} + \sum_{i=1}^{p} \varphi_i X_{p+1-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{q+1-j},$$

conditional on $X_1 = x_1, X_2 = x_2, ..., X_p = x_p$, we have

$$X_{p+1} \sim N\left(\left(u + \sum_{i=1}^{p} \varphi_i X_{p+1-i}\right), \sigma^2\right).$$

Then for $t = (p+1, p+2, ..., T)$, we calculate the conditional maximum likelihood

$$l^*(\Theta) = \ln f(x_T, ..., x_{p+1} \mid x_p, ..., x_1, \varepsilon_p = \varepsilon_{p-1} = ... = \varepsilon_{p+1-q} = 0; \Theta)$$

$$= -\frac{T-p}{2}\ln(2\pi) - \frac{T-p}{2}\ln(\sigma^2) - \sum_{t=p+1}^{T} \frac{\left(X_t - u - \sum_{i=1}^{p} \varphi_i X_{t-i} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}\right)^2}{2\sigma^2}$$

where $\{\varepsilon_{p+1}, ..., \varepsilon_T\}$ can be calculated from $\{x_1, x_2, ..., x_T\}$:

$$\varepsilon_t = X_t - u - \sum_{i=1}^{p} \varphi_i X_{t-i} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, t = p+1, ..., T .$$

So

$$l^*(\Theta) = \ln f(x_T, ..., x_{p+1} \mid x_p, ..., x_1, \varepsilon_p = \varepsilon_{p-1} = ... = \varepsilon_{p+1-q} = 0; \Theta)$$

$$= -\frac{T-p}{2}\ln(2\pi) - \frac{T-p}{2}\ln(\sigma^2) - \sum_{t=p+1}^{T} \frac{\varepsilon_t^2}{2\sigma^2}.$$

We note that by using conditional MLE, we only use $T-p$ observations.

Next

$$\frac{\partial l^*}{\partial \varphi_i} = \frac{1}{\sigma^2} \sum_{t=p+1}^{T} \varepsilon_t X_{t-i} ,$$

$$\frac{\partial l^*}{\partial \theta_j} = \frac{1}{\sigma^2} \sum_{t=p+1}^{T} \varepsilon_t \varepsilon_{t-j}$$

and

$$\frac{\partial l^*}{\partial \sigma^2} = -\left(\frac{T-p}{2\sigma^2}\right) + \left(\frac{1}{2\sigma^4}\right) \sum_{t=p+1}^{T} \varepsilon_t^2 .$$

We can obtain the information matrix from the second derivatives, i.e.

$$\frac{\partial^2 l^*}{\partial \varphi_i \partial \varphi_j} = \left(\frac{1}{\sigma^2}\right) \sum_{t=p+1}^{T} X_{t-i} \left(\frac{\partial \varepsilon_t}{\partial \varphi_j}\right) + \left(\frac{1}{\sigma^2}\right) \sum_{t=p+1}^{T} \varepsilon_t \left(\frac{\partial \varepsilon_{t-i}}{\partial \varphi_j}\right)$$

$$= \left(\frac{1}{\sigma^2}\right) \sum_{t=p+1}^{T} X_{t-i} X_{t-j} + \left(\frac{1}{\sigma^2}\right) \sum_{t=p+1}^{T} \varepsilon_t X_{t-i-j} .$$

Note that

$$E(\varepsilon_t X_{t-i-j}) = 0$$

which implies that

$$\lim_{T\to\infty} \left(-E \frac{\partial^2 l^*}{\partial \varphi_i \partial \varphi_j} \right) = \lim_{T\to\infty} \left(-\left(\frac{1}{\sigma^2}\right) \sum_{t=p+1}^{T} X_{t-i} X_{t-j} \right) = \gamma_{ij} \text{ and } (\gamma_{ij}) = V_p$$

which is the autocovariance of the process X

$$\frac{\partial^2 l^*}{\partial \theta_i \partial \theta_j} = \left(\frac{1}{\sigma^2}\right) \sum_{t=p+1}^{T} \left(\frac{\partial \varepsilon_t}{\partial \theta_j}\right) \varepsilon_{t-i} + \left(\frac{1}{\sigma^2}\right) \sum_{t=p+1}^{T} \varepsilon_t \left(\frac{\partial \varepsilon_{t-i}}{\partial \theta_j}\right)$$

$$= \left(\frac{1}{\sigma^2}\right) \sum_{t=p+1}^{T} \varepsilon_{t-i} \varepsilon_{t-j} + \left(\frac{1}{\sigma^2}\right) \sum_{t=p+1}^{T} \varepsilon_t \varepsilon_{t-i-j} .$$

Note that

$$E(\varepsilon_t \varepsilon_{t-i-j}) = 0,$$

which implies that

$$\lim_{T\to\infty}\left(-E\frac{\partial^2 l^*}{\partial\theta_i\partial\theta_j}\right) = \lim_{T\to\infty}\left(-\left(\frac{1}{\sigma^2}\right)\sum_{t=p+1}^{T}\varepsilon_{t-i}\varepsilon_{t-j}\right) = \lambda_{ij}$$

also

$$\frac{\partial^2 l^*}{\partial\varphi_i\partial\theta_j} = \left(\frac{1}{\sigma^2}\right)\sum_{t=p+1}^{T}\left(\frac{\partial\varepsilon_t}{\partial\theta_j}\right)X_{t-i} + \left(\frac{1}{\sigma^2}\right)\sum_{t=p+1}^{T}\varepsilon_t\left(\frac{\partial X_{t-i}}{\partial\theta_j}\right)$$

$$= \left(\frac{1}{\sigma^2}\right)\sum_{t=p+1}^{T}\varepsilon_{t-j}X_{t-i} + \left(\frac{1}{\sigma^2}\right)\sum_{t=p+1}^{T}\varepsilon_t\left(\frac{\partial X_{t-i}}{\partial\theta_j}\right),$$

and

$$\lim_{T\to\infty}\left(-E\frac{\partial^2 l^*}{\partial\varphi_i\partial\theta_j}\right) = \lim_{T\to\infty}\left(\frac{1}{\sigma^2}\right)\sum_{t=p+1}^{T}\varepsilon_{t-j}X_{t-i} = \eta_{ij}\,.$$

Moreover, we have

$$\frac{\partial^2 l^*}{\partial\sigma^4} = -\frac{1}{\sigma^6}\sum_{t=p+1}^{T}\varepsilon_t^2 + \frac{T-p}{2\sigma^4}$$

$$-E\left(\frac{\partial^2 l^*}{\partial\sigma^4}\right) - E\left(-\frac{1}{\sigma^6}\sum_{t=p+1}^{T}\varepsilon_t^2 + \frac{T-p}{2\sigma^4}\right) = \frac{T-p}{2\sigma^4}\,.$$

Finally we calculate

$$\frac{\partial^2 l^*}{\partial\varphi_i\partial\sigma^2} = -\frac{1}{\sigma^4}\sum_{t=p+1}^{T}\varepsilon_t X_{t-i}\,,$$

$$\frac{\partial^2 l^*}{\partial\theta_j\partial\sigma^2} = -\frac{1}{\sigma^4}\sum_{t=p+1}^{T}\varepsilon_t\varepsilon_{t-j}\,.$$

We obtain

$$E\left(-\frac{\partial^2 l^*}{\partial\varphi_i\partial\sigma^2}\right)=-\frac{1}{\sigma^4}\sum_{t=p+1}^{T}E(\varepsilon_t X_{t-i})=0$$

$$E\left(-\frac{\partial^2 l^*}{\partial\theta_j\partial\sigma^2}\right)=-\frac{1}{\sigma^4}\sum_{t=p+1}^{T}E(\varepsilon_t\varepsilon_{t-j})=0\,.$$

Then the information matrix is

$$\sqrt{T}(\hat{\Theta}-\Theta)\to N(0,V)$$

where

$$V=\begin{bmatrix}\begin{bmatrix}(\gamma_{ij}) & (\lambda_{ij})\\ (\lambda_{ij}) & (\eta_{ij})\end{bmatrix}^{-1} & \vec{0}\\ \vec{0} & 2\sigma^4\end{bmatrix}.$$

Obviously AR(p) is a special case of ARMA(p,q), when $(\lambda_{ij})=\vec{0}$ and $(\eta_{ij})=\vec{0}$.

3.1.4 MLE for ARCH(p) Model

First, we set the log-likelihood function of parameter vector

$$\Theta=(\alpha_0,\alpha_1,...,\alpha_q)'\ .$$

Under the assumption that v_t is Gaussian, we obtain the conditional density function of ε_t:

$$f(\varepsilon_t\,|\,F_{t-1})=\sigma_t f(v_t\,|\,F_{t-1})\ \text{and}\ f(v_t\,|\,F_{t-1})\sim N(0,1)$$

so we have

$$f(\varepsilon_t\,|\,F_{t-1})\sim N(0,\sigma_t^2)\,.$$

Given sample size T, using the conditional mean (2.5) and variance (2.6), we conduct the likelihood function as

$$f(\varepsilon_T,\varepsilon_{T-1},...,\varepsilon_1) = f(\varepsilon_T \mid F_{T-1}) f(\varepsilon_{T-1} \mid F_{T-2}) .... f(\varepsilon_{q+1} \mid F_q) f(\varepsilon_q,\varepsilon_{q-1},...,\varepsilon_1)$$

$$= \prod_{t=q+1}^{T} \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right) f(\varepsilon_q,\varepsilon_{q-1},...,\varepsilon_1)$$

Since the exact form of $f(\varepsilon_q,\varepsilon_{q-1},...,\varepsilon_1)$ is hard to obtain, and the sample size is large enough, we use the conditional likelihood function instead,

$$f(\varepsilon_T,\varepsilon_{T-1},...,\varepsilon_{q+1} \mid \varepsilon_q,\varepsilon_{q-1},...,\varepsilon_1) = \prod_{q+1}^{T} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right).$$

Then the conditional log-likelihood function is

$$l^*(\Theta) = \sum_{t=q+1}^{T} \left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_t^2) - \frac{\varepsilon_t^2}{2\sigma_t^2} \right].$$

The first derivative of $l^*(\Theta)$ is

$$\frac{\partial l^*(\Theta)}{\partial \Theta} = \sum_{t=q+1}^{T} \left(\frac{\varepsilon_t^2}{2\sigma_t^4} - \frac{1}{2\sigma_t^2} \right) \frac{\partial \sigma_t^2}{\partial \Theta}.$$

The second derivative of $l^*(\Theta)$ is

$$\frac{\partial^2 l^*(\Theta)}{\partial \Theta \partial \Theta'} = \sum_{t=q+1}^{T} \left[\left(-\frac{\varepsilon_t^2}{2\sigma_t^2} \right) \frac{\partial \sigma_t^2}{\partial \Theta} \frac{\partial \sigma_t^2}{\partial \Theta} + \left(\frac{\varepsilon_t^2}{2\sigma_t^4} - \frac{1}{2\sigma_t^2} \right) \frac{\partial^2 \sigma_t^2}{\partial \Theta \partial \Theta'} \right].$$

From (2.5) and (2.6), we obtain the conditional expectation of ε_t^2,

$$E(\varepsilon_t^2 \mid F_{t-1}) = \sigma_t^2$$

Then we calculate the information matrix

$$I = \sum_{t=q+1}^{T} E\left[\left(-\frac{\varepsilon_t^2}{2\sigma_t^6} \right) \frac{\partial \sigma_t^2}{\partial \Theta} \frac{\partial \sigma_t^2}{\partial \Theta} + \left(\frac{\varepsilon_t^2}{2\sigma_t^4} - \frac{1}{2\sigma_t^2} \right) \frac{\partial^2 \sigma_t^2}{\partial \Theta \partial \Theta'} \right]$$

$$= \sum_{t=q+1}^{T} E\left(-\frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \Theta} \frac{\partial \sigma_t^2}{\partial \Theta'} \right)$$

3.1.5 MLE for GARCH(p,q) Model

First, we set the log-likelihood function of parameter vector as

$$\Theta = (\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)' \ .$$

The log-likelihood of (2.8) is:

$$l(\Theta) = -\frac{1}{2}\ln 2\pi - \ln\sigma_t - \frac{\varepsilon_t^2}{2\sigma_t^2} \ ,$$

so the MLE for Θ is:

$$l^*(\Theta) = \sum_{t=q+1}^{T}\left(-\frac{1}{2}\ln 2\pi - \ln\sigma_t - \frac{\varepsilon_t^2}{2\sigma_t^2}\right).$$

The first and second derivation of $l^*(\Theta)$ is

$$\frac{\partial l^*(\Theta)}{\partial\Theta} = \sum_{t=q+1}^{T}\left(\frac{\varepsilon_t^2}{2\sigma_t^4} - \frac{1}{2\sigma_t^2}\right)\frac{\partial\sigma_t^2}{\partial\Theta}$$

$$\frac{\partial^2 l^*(\Theta)}{\partial\Theta\partial\Theta'} = \sum_{t=q+1}^{T}\left[\left(\frac{\varepsilon_t^2}{2\sigma_t^4} - \frac{1}{2\sigma_t^2}\right)\frac{\partial^2\sigma_t^2}{\partial\Theta\partial\Theta'} + \left(\frac{1}{2\sigma_t^4} - \frac{\varepsilon_t^2}{\sigma_t^6}\right)\frac{\partial\sigma_t^2}{\partial\Theta}\frac{\partial\sigma_t^2}{\partial\Theta'}\right].$$

And $\frac{\partial\sigma_t^2}{\partial\Theta}$ has the expression as following

$$\frac{\partial\sigma_t^2}{\partial\Theta} = \sum_{j=1}^{p}\beta_j\frac{\partial\sigma_{t-j}^2}{\partial\Theta} + (1, \varepsilon_{t-1}^2, ..., \varepsilon_{t-q}^2, \sigma_{t-1}^2, ..., \sigma_{t-p}^2)' ,$$

then we can conduct the information matrix

$$I = \sum_{t=q+1}^{T}E\left[\left(\frac{\varepsilon_t^2}{2\sigma_t^4} - \frac{1}{2\sigma_t^2}\right)\frac{\partial^2\sigma_t^2}{\partial\Theta\partial\Theta'} + \left(\frac{1}{2\sigma_t^4} - \frac{\varepsilon_t^2}{\sigma_t^6}\right)\frac{\partial\sigma_t^2}{\partial\Theta}\frac{\partial\sigma_t^2}{\partial\Theta'}\right]$$

$$= \frac{1}{2}\sum_{t=q+1}^{T}E\left(\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\Theta}\frac{\partial\sigma_t^2}{\partial\Theta'}\right)$$

3.2 Quasi Maximum Likelihood Estimation (QMLE)

In section 3.1 we discussed several time series models using maximum likelihood under normality assumption. Here we will use quasi likelihood estimation, regarded as Gaussian MLE, which allows for over dispersion. For example, in GARCH(p,q), when we use MLE under the assumption that ν_t is i.i.d random variables such that $\nu_t \sim N(0,1)$. However, if we use QMLE, the ν_t's don't need to be subjected to normal distribution.

Then quasi maximum likelihood (QML) estimation can be conducted on a QML estimator $\hat{\Theta}_T$ of Θ is a solution $\hat{\Theta}_T$ to

$$\arg\max_{\Theta} \frac{1}{T}\sum_{t=1}^{T} l_t(\Theta).$$

The QML estimator is asymptotically normal distributed as

$$\sqrt{T}(\hat{\Theta}_T - \Theta_0) \xrightarrow{d} N(0, D^{-1}VD^{-1}),$$

where

$$D = -\frac{1}{T} E_{\Theta_0}\left[\frac{\partial^2 l(\Theta)}{\partial\Theta\partial\Theta'}\right]$$

and

$$V = \frac{1}{T} E_{\Theta_0}\left(\frac{\partial l(\Theta)}{\partial\Theta}\frac{\partial l(\Theta)}{\partial\Theta'}\right).$$

D and V are matrices.

3.2.1 QMLE for ARMA(p,q) Model

We obtain the QMLE of ARCH(q), i.e.

$$\arg\max_{\Theta} \frac{1}{T}\sum_{t=1}^{T} l_t(\Theta)$$

$$= \arg\max_{\Theta} \frac{1}{T}\sum_{t=1}^{T}\left[-\frac{1}{2}\ln(\sigma^2(\Theta)) - \frac{\varepsilon_t^2(\Theta)}{2\sigma^2(\Theta)}\right].$$

Without loss of generality, we rewrite the ARMA(p,q) as

$$X_t - \mu = \sum_{i=1}^{p} \varphi_i (X_{t-i} - \mu) + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

where μ is the mean of X_t.

Then we use matrix form to rewrite ARMA(p,q) as

$$\varphi(X - \mu 1) = \theta\varepsilon$$

where $\varphi = (1, -\varphi_1, -\varphi_2, ..., -\varphi_p)$, $\theta = (\theta_1, ..., \theta_q)$, $X = (X_t, X_{t-1}, ..., X_{t-p})'$ and

$\varepsilon = (\varepsilon_t, \varepsilon_{t-1}, ..., \varepsilon_{t-q})'$.

By Bao (2015) we obtain the asymmetric covariance matrix V of QMLE $\hat{\Theta}$ of

$\Theta = (\mu, \varphi_1, ..., \varphi_p, \theta_1, ...\theta_q, \sigma^2)$

$$V = \begin{pmatrix} \frac{T\sigma^2}{1'\varphi'(\theta^{-1})'\theta^{-1}\varphi 1} & 0'_{p+q} & \frac{E(\mu_t^3)1'\theta^{-1}\varphi 1}{1'\varphi'(\theta^{-1})'\theta^{-1}\varphi 1} \\ 0_{p+q} & B^{-1} & 0_{p+q} \\ \frac{E(\mu_t^3)1'\theta^{-1}\varphi 1}{1'\varphi'(\theta^{-1})'\theta^{-1}\varphi 1} & 0'_{p+q} & E(\mu_t^4) - \sigma^4 \end{pmatrix},$$

where

$$B^{-1} = \begin{bmatrix} (\gamma_{ij}) & (\lambda_{ij}) \\ (\lambda_{ij}) & (\eta_{ij}) \end{bmatrix}^{-1} \text{ and } E(\mu_t^3) < \infty, E(\mu_t^4) < \infty.$$

3.2.2 QMLE for GARCH(p,q) Model

We obtain the QMLE of GARCH(p,q), i.e.

$$\arg\max_{\Theta} \frac{1}{T}\sum_{t=1}^{T} l_t(\Theta)$$

$$= \arg\max_{\Theta} \frac{1}{T}\sum_{t=1}^{T}\left[-\frac{1}{2}\ln(\sigma_t^2(\Theta)) - \frac{\varepsilon_t^2(\Theta)}{2\sigma_t^2(\Theta)}\right].$$

Then we calculate the first derivative of the likelihood function for each t

$$\frac{\partial l_t}{\partial \Theta} = \frac{\partial\left(-\ln(\sigma_t^2(\Theta) - \frac{\varepsilon_t^2(\Theta)}{2\sigma_t^2(\Theta)}\right)}{\partial \Theta} = -\frac{1}{2\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \Theta} - \frac{\varepsilon_t}{\sigma_t^2}\frac{\partial \varepsilon_t}{\partial \Theta} + \frac{\varepsilon_t^2}{2\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \Theta}.$$

Here we rewrite $\varepsilon_t(\Theta) = y_t - \mu_t(\Theta)$, y_t is the observation at time t, $\mu_t(\Theta)$ is the mean of

y_t,

so

$$\frac{\partial l_t}{\partial \Theta} = -\frac{1}{2\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \Theta} + \frac{\varepsilon_t}{\sigma_t^2}\frac{\partial \mu_t}{\partial \Theta} + \frac{\varepsilon_t^2}{2\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \Theta}.$$

The second derivative of the likelihood function for each t is

$$\frac{\partial^2 l_t}{\partial \Theta \partial \Theta'} = +\frac{1}{2\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \Theta}\frac{\partial \sigma_t^2}{\partial \Theta'} - \frac{1}{2\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \Theta \partial \Theta'}$$

$$+\frac{\varepsilon_t^2}{2\sigma_t^4}\frac{\partial^2 \sigma_t^2}{\partial \Theta \partial \Theta'} - \frac{\varepsilon_t^2}{\sigma_t^6}\frac{\partial \sigma_t^2}{\partial \Theta}\frac{\partial \sigma_t^2}{\partial \Theta'} - \frac{\varepsilon_t}{\sigma_t^4}\frac{\partial \mu_t}{\partial \Theta}\frac{\partial \sigma_t^2}{\partial \Theta'}$$

Given $E\left(\frac{\varepsilon_t}{\sigma_t} \mid F_{t-1}\right) = 0$ and $E\left(\frac{\varepsilon_t^2}{\sigma_t^2} \mid F_{t-1}\right) = 1$.

So we can compute

$$D_t = E_{\Theta_0}\left(-\frac{\partial^2 l_t}{\partial\Theta\partial\Theta'}\right) = E_{\Theta_0}\left(-\frac{1}{2\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\Theta}\frac{\partial\sigma_t^2}{\partial\Theta'} + \frac{1}{2\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\Theta\partial\Theta'}\right)$$

$$+E_{\Theta_0}\left[-\frac{1}{2\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\Theta\partial\Theta'}E\left(\frac{\varepsilon_t^2}{\sigma_t^2}\mid F_{t-1}\right)\right] + E_{\Theta_0}\left[\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\Theta}\frac{\partial\sigma_t^2}{\partial\Theta'}E\left(\frac{\varepsilon_t^2}{\sigma_t^2}\mid F_{t-1}\right)\right]$$

$$+E_{\Theta_0}\left[-\frac{1}{\sigma_t}\frac{\partial^2\mu_t}{\partial\Theta\partial\Theta'}E\left(\frac{\varepsilon_t}{\sigma_t}\mid F_{t-1}\right)\right] + E_{\Theta_0}\left[\frac{1}{\sigma_t^2}\frac{\partial\mu_t}{\partial\Theta}\frac{\partial\mu_t}{\partial\Theta'}\right]$$

$$+E_{\Theta_0}\left[-\frac{1}{\sigma_t^3}\frac{\partial\mu_t}{\partial\Theta}\frac{\partial\sigma_t^2}{\partial\Theta'}E\left(\frac{\varepsilon_t}{\sigma_t}\mid F_{t-1}\right)\right] + E_{\Theta_0}\left[\frac{1}{\sigma_t^5}\frac{\partial\sigma_t^2}{\partial\Theta}\frac{\partial\mu_t}{\partial\Theta'}E\left(\frac{\varepsilon_t}{\sigma_t^2}\mid F_{t-1}\right)\right]$$

$$= E_{\Theta_0}\left[\frac{1}{2\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\Theta}\frac{\partial\sigma_t^2}{\partial\Theta'}\right] + E_{\Theta_0}\left[\frac{1}{\sigma_t^2}\frac{\partial\mu_t}{\partial\Theta}\frac{\partial\mu_t}{\partial\Theta'}\right]$$

and $D = \sum_{t=1}^{T} D_t$.

Next we calculate

$$V_t = E_{\Theta_0}\left(\frac{\partial l(\Theta)}{\partial\Theta}\frac{\partial l(\Theta)}{\partial\Theta'}\right) = E_{\Theta_0}\left[\left(-\frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\Theta} + \frac{\varepsilon_t}{\sigma_t^2}\frac{\partial\mu_t}{\partial\Theta} + \frac{\varepsilon_t^2}{2\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\Theta}\right)\times\right.$$

$$\left.\left(-\frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\Theta} + \frac{\varepsilon_t}{\sigma_t^2}\frac{\partial\mu_t}{\partial\Theta} + \frac{\varepsilon_t^2}{2\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\Theta}\right)'\right]$$

$$= E_{\Theta_0}\left[\frac{1}{4\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\Theta}\frac{\partial\sigma_t^2}{\partial\Theta'} + \frac{1}{4\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\Theta}\frac{\partial\sigma_t^2}{\partial\Theta'}E\left(\frac{\varepsilon_t^4}{\sigma_t^4}\mid F_{t-1}\right)\right] + E_{\Theta_0}\left[-\frac{1}{2\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\Theta}\frac{\partial\sigma_t^2}{\partial\Theta'}\right]$$

$$+E_{\Theta_0}\left[\frac{1}{\sigma_t^2}\frac{\partial\mu_t}{\partial\Theta}\frac{\partial\mu_t}{\partial\Theta'}E\left(\frac{\varepsilon_t^2}{\sigma_t^2}\mid F_{t-1}\right)\right] + E_{\Theta_0}\left[\left(\frac{1}{2\sigma_t^3}\frac{\partial\sigma_t^2}{\partial\Theta}\frac{\partial\mu_t}{\partial\Theta'} + \frac{1}{2\sigma_t^3}\frac{\partial\mu_t}{\partial\Theta}\frac{\partial\sigma_t^2}{\partial\Theta'}\right)E\left(\frac{\varepsilon_t^3}{\sigma_t^3}\mid F_{t-1}\right)\right]$$

$$= E_{\Theta_0}\left\{\frac{1}{4\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\Theta}\frac{\partial\sigma_t^2}{\partial\Theta'}\left[E\left(\frac{\varepsilon_t^4}{\sigma_t^2}\mid F_{t-1}\right)-1\right]\right\} + E_{\Theta_0}\left(\frac{1}{\sigma_t^2}\frac{\partial\mu_t}{\partial\Theta}\frac{\partial\mu_t}{\partial\Theta'}\right)$$

$$+E_{\Theta_0}\left[\frac{1}{2\sigma_t^6}\frac{\partial\mu_t}{\partial\Theta}\frac{\partial\sigma_t^2}{\partial\Theta'}E\left(\frac{\varepsilon_t^3}{\sigma_t^3}\mid F_{t-1}\right)\right]+E_{\Theta_0}\left[\frac{1}{2\sigma_t^6}\frac{\partial\sigma_t^2}{\partial\Theta}\frac{\partial\mu_t}{\partial\Theta'}E\left(\frac{\varepsilon_t^3}{\sigma_t^3}\mid F_{t-1}\right)\right],$$

where

$$E\left(\frac{\varepsilon_t^3}{\sigma_t^3}\mid F_{t-1}\right)<\infty \text{ and } E\left(\frac{\varepsilon_t^4}{\sigma_t^4}\mid F_{t-1}\right)<\infty,$$

and $V=\sum_{t=1}^{T}V_t$.

Through matrices D and V, we can obtain the information matrix of GARCH(p,q).

3.3 Non-Gaussian Quasi Maximum Likelihood Estimation (NGQMLE)

The non-Gaussian maximum likelihood estimation is widely used in GARCH models in order to capturing the heavy-tailed data set. Numerous financial data sets are a kind of "heavy-tailed". In general heteroscedastic models, Newey and Steigerwald (1997) first considered the condition for consistency of the heteroscedastic parameter with an NGQMLE. Sometimes we prefer an NGQMLE method since it is robust against density misspecification and more efficient than GQMLE.

3.3.1 NGQMLE for GARCH(p,q) Model

In the paper of Fan, Qi, and Xiu (2014), they discussed a three-step estimation procedure of NGQMLE, i.e. reparametrize the GARCH(p,q) model (2.7) as:

$$\begin{aligned}\varepsilon_t &= \sigma\cdot\sigma_t v_t\\ \sigma_t^2 &= 1+\sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2+\sum_{j=1}^{q}\beta_j\sigma_{t-j}^2\end{aligned}.$$

First step: conduct Gaussian quasi-likelihood estimation:

$$\widetilde{\Theta}_T=\arg\max_{\Theta}\frac{1}{T}\sum_{t=1}^{T}l_1(\bar{\varepsilon}_t,\Theta)$$

$$= \arg\max_{\Theta} \frac{1}{T}\sum_{t=1}^{T}\left[-\ln(\sigma\sigma_t) - \frac{\varepsilon_t^2}{2\sigma^2\sigma_t^2}\right].$$

Second step: obtain $\hat{\eta}_f$ by maximizing the equation

$$\eta_f = \arg\max_{\eta>0} E\left[-\ln(\eta) + \ln f\left(\frac{\varepsilon}{\eta}\right)\right],$$

where the expectation is taken under the true density, with estimated residuals from the first step:

$$\hat{\eta}_f = \arg\max_{\eta} \frac{1}{T}\sum_{t=1}^{T} l_2\left(\bar{\varepsilon}_t, \tilde{\Theta}_T, \eta\right)$$

$$= \arg\max_{\eta} \frac{1}{T}\sum_{t=1}^{T}\left[-\ln(\eta) + \ln f\left(\frac{\tilde{\varepsilon}_t}{\eta}\right)\right],$$

where $\tilde{\varepsilon}_t$ is the GQMLE residuals in the first step.

Third step: maximize non-Gaussian quasi-likelihood by replacing η_f by $\hat{\eta}_f$ and obtain $\hat{\Theta}_T$

$$\hat{\Theta}_T = \arg\max_{\Theta} \frac{1}{T}\sum_{t=1}^{T} l_3\left(\bar{\varepsilon}_t, \hat{\eta}_f, \Theta\right)$$

$$= \arg\max_{\Theta} \frac{1}{T}\sum_{t=1}^{T}\left[\ln\left(\hat{\eta}_f\sigma\sigma_t\right) + \ln f\left(\frac{\varepsilon_t}{\hat{\eta}_f\sigma\sigma_t}\right)\right].$$

We then obtain the NGQMLE estimator $\hat{\Theta}_T$.

CHAPTER IV

REAL FINANCIAL DATA APPLICATIONS

In order to use these different estimations, I first chose the one-year daily foreign exchange rate BRL/USD and ARS/USD and then 3-year daily foreign exchange rate BRL/USD and ARS/USD.

By finding the lowest AIC and SBC, I found that GARCH model fits better. So I used GARCH model to do the estimation.

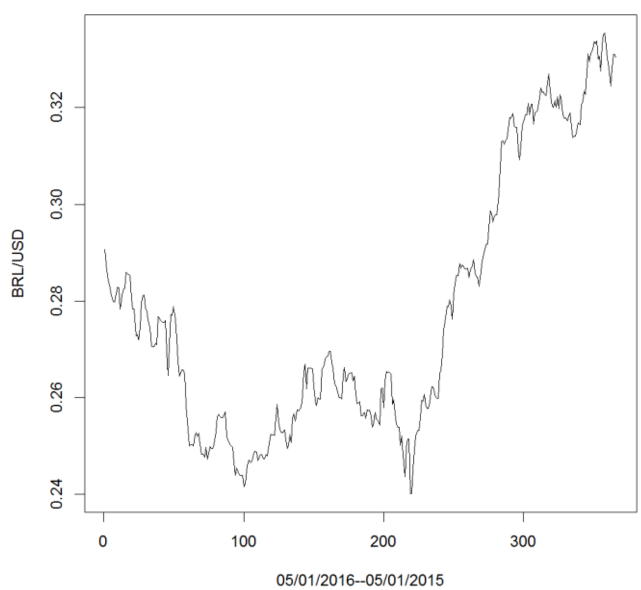

Figure 4.1 The foreign exchange rate of BRL/USD from 05/01/2015—05/01/2016

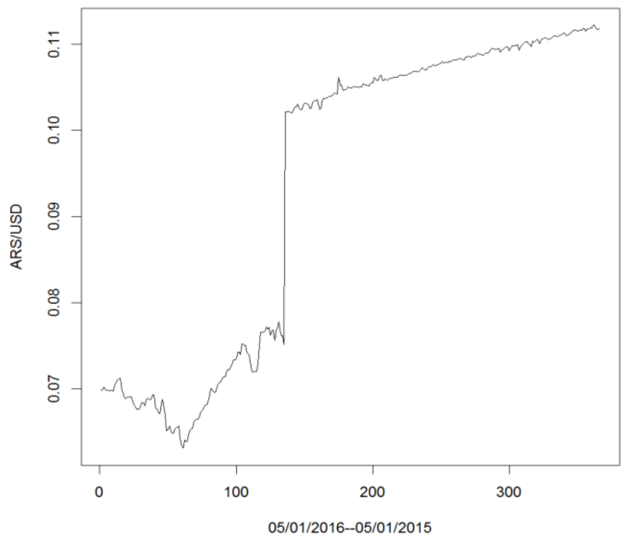

Figure 4.2 The foreign exchange rate of ARS/USD from 05/01/2015—05/01/2016

Table 4.1 Comparison of one-year data sets

Foreign exchange rate	True Value (Var)	MLE Average	QMLE Average	NGQMLE Average
BRL/USD	0.0007417	0.0007833	0.0007423	0.0007425
ARS/USD	0.0003366	0.0003389	0.0003342	0.0003337

Here we can see that the QMLE and NGQMLE have very little deference in estimating the variance of BRL/USD and ARS/USD. But compared with MLE, QMLE and NGQ-MLE is more efficient.

The following are the graphs of 3-year daily data of BRL/USD and ARS/USD

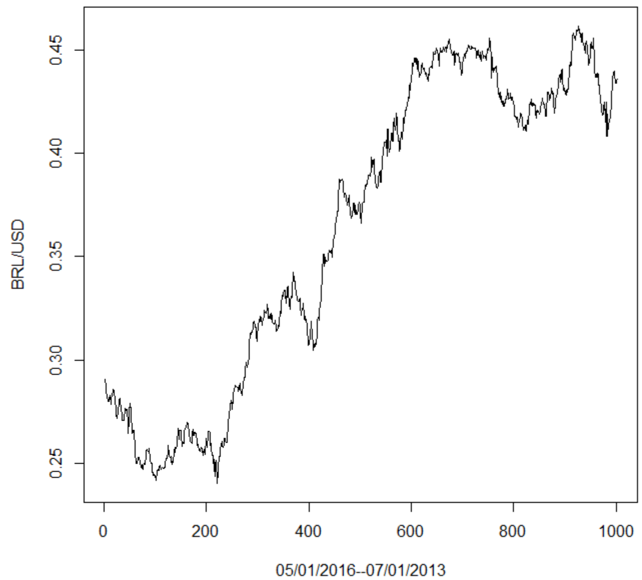

Figure 4.3 The foreign exchange rate of BRL/USD from 07/01/2013—05/01/2016

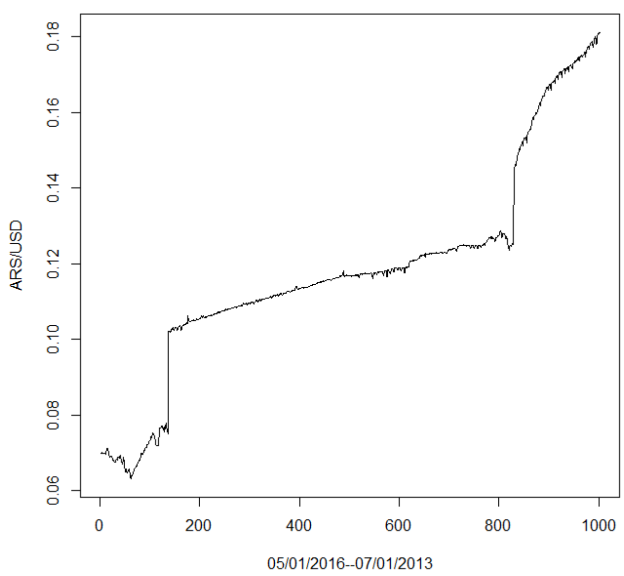

Figure 4.4 The foreign exchange rate of ARS/USD from 07/01/2013—05/01/2016

Table 4.2 Comparison of three-year data sets

Foreign Exchange	True Value (Var)	MLE Average	QMLE Average	NGQMLE Average
BRL/USD	0.00552	0.006613	0.005917	0.005392
ARS/USD	0.000775	0.0010746	0.0008049	0.0007853

In this table, it is clear that NGQMLE is the most powerful estimating tool for the variance of BRL/USD and ARS/USD. However, MLE is less effective in estimating variance, which implies the normality assumption is not proper.

CHAPTER V

CONCLUSIONS

From Chapter 4, we can see that the GARCH model is better in modeling the financial time series since financial time series have been found to be heteroscedastic with autocorrelated volatility. Compared with MLE, both QMLE and NGQMLE are more powerful in estimating σ_t^2. When the sample size of the data set is not large, MLE, QMLE and NGQMLE would differ very little; when the sample size becomes larger, the MLE is not a nice estimation, but the NGQMLE becomes the most powerful. The exchange rates in emerging markets are more changeable than in developed markets, which is due to the incomplete financial policy, lack of oversight, and compacted with speculators. So using the normality assumption is not proper in estimating parameters, which would lead to less effectiveness of the MLE.

REFERENCES

[1] Andrews, B. (2012), "Rank-based Estimation for ARCH Process", *Econometric Theory*, 28, 1037-1064.

[2] Bao, Y. (2016) "The Asymptotic Covariance Matrix of the QMLE in ARMA Models", *Econometric Reviews*

[3] Berkes et al. (2003), "GARCH Process: Structure and Estimation", *Bernoulli*, 9, pp. 201-207.

[4] Bollerslev, T., and Wooldbridge, J. M. (1992), "Quasi-maximum Likelihood Estimation and Inference in Dynamic Models With Time-varying Covariances," *Econometric Reviews*, 11, 143–172.

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

[6] Brockwell, P. J. (1993), "Threshold ARMA processes in continuous time", *Dimension Estimation and Models*, 170–90.

[7] Bollerslev, T., and Wooldbridge, J. M. (1992), "Quasi-maximum Likelihood Estimation and Inference in Dynamic Models With Time-varying Covariances," *Econometric Reviews*, 11, 143–172.

[8] Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

[9] Fan, J., Qi, L., Xiu, D. (2014), "Quasi Maximum Likelihood Estimation of GARCH Models with Heavy-tailed Likelihoods", *Journal of Business and Economic Statistics*, Vol. 32, No. 2.

[10] Francq, C., and Lepage, G., and Zakoian, J. M. (2011), "Two-stage Non Gaussian QML Estimation of GARCH Models and Testing the Efficiency of the Gaussian QMLE," *Journal of Econometrics*, 165, 246–257.

[11] Francq, C. (2004), "Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes," *Bernoulli*, 10, 605–637.

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