BAYESIAN ESTIMATION OF GARCH COEFFICIENTS OF INR/USD EXCHANGE RATE

Ranjit Kumar
Department of Physics, Dyal Singh College, University of Delhi, Delhi 110003, India
du.ranjit@gmail.com

Varun Malik
Department of Physics, Dyal Singh College, University of Delhi, Delhi 110003, India

Abid Hussain Rather
Department of Physics, Dyal Singh College, University of Delhi, Delhi 110003, India

ABSTRACT

In this work the Bayesian estimation have been used to find the coefficients of GARCH(1,1) model with normal and Student’s t innovation. We have also calculated the GARCH coefficient using maximum likelihood estimation and a comparison is made between obtained result. These models are applied in the context of India/USA exchange rate. The results obtained in this paper is based on the work of [17].

Keywords: Volatility, GARCH model, Maximum likelihood estimation, Bayesian statistics, Markov chain monte carlo method.

Introduction

Volatility is measure of risk in finance and its estimation is one of the challenging problem. The ability to estimate and forecast volatilities for different assets leads to a better understanding of current and future financial risk. Accurate knowledge of volatility is used in estimating the value at risk of a portfolio, optimal allocation of asset in a portfolio, hedging the risk of a portfolio, and in other related areas [1].
Volatility ($\sigma$) is defined as the conditional standard deviation of daily returns and it cannot be observed directly. The volatility of an asset can be estimated using the asset price or derivatives or both. In option pricing, volatility is determined using Black-Scholes equation. In Black-Scholes equation, the estimated volatility is also known as implied volatility [2]. Volatility can also be estimated using historical asset price without the knowledge of Black-Scholes equation. The details of this method is given in [2].

Since volatility is directly not observable, one must look at the statistical model. Here we shall discuss the univariate volatility modelling, i.e., single return series.

**Organisation Of The Paper**

In next section we will discuss the characteristics of the volatility and its various models. In particular, ARCH and GARCH model has been discussed in detail. In sections 8 and 9, we have discussed the maximum likelihood estimation of the GARCH(1,1) coefficients with normal and student’s $t$ innovation, respectively. Then, in section 11, Bayesian estimation of GARCH(1,1) coefficients with normal innovation and in section 12 Bayesian estimation of GARCH(1,1) coefficients with student’s $t$ innovation have been discussed. In section 13, we have obtained the GARCH(1,1) coefficients for India/USA exchange rate. Finally, we conclude this paper in section 14.

**Characteristics of Volatility**

Let us first discuss the characteristics of volatility, that helps us to develop the volatility model. Some of the important properties of volatility are [3]:

- One of the important property of volatility is volatility clustering (i.e., volatility may
Some of the volatility models are listed below:

- Moving Average Model
- Exponentially Weighted Moving Average (EWMA) also known as RiskMetrics
- ARCH and GARCH Model and its Extension Models
  - Stochastic Volatility Model
  - Implied Volatility
  - Realised Volatility

**Asset Return**

The market risk or volatility is caused by the movement in asset price and so let us first define asset return. Let us define the daily simple rate of return from the closing price of the asset

$$\eta_{i+1} = \frac{S_{i+1} - S_i}{S_i} = \frac{S_{i+1}}{S_i} - 1$$

where $S_i$ is the value of market variable at the end of day $i$. The log return on asset is defined as [6]
\[ u_{i+1} = \ln(S_{i+1}) - \ln(S_i) = \ln \left( \frac{S_{i+1}}{S_i} \right) \]

Here \( u_{i+1} \) is continuously compounded return during day \( i + 1 \) (between the end of day \( i \) and the end of day \( i + 1 \)). Note that the log return of the asset can also be written as

\[ u_{i+1} = \ln \left( \frac{S_{i+1}}{S_i} \right) = \ln \left( 1 + \frac{S_{i+1} - S_i}{S_i} \right) \approx \frac{S_{i+1} - S_i}{S_i} \]

which is nothing but the relative return of the asset. An unbiased estimate of the variance rate per day, \( \sigma_i^2 \), using the most recent \( m \) observations on the \( u_i \) is

\[ \sigma_i^2 = \frac{1}{m-1} \sum_{\tau=1}^{m} (u_{i-\tau} - \bar{u})^2 \]

where \( \bar{u} \) is the mean of the \( u_i \)’s:

\[ \bar{u} = \frac{1}{m} \sum_{\tau=1}^{m} u_{i-\tau} \]

For practical purpose, Eq. ((1)) is usually changed as

1. \( \bar{u} \) is assumed to be zero.
2. \( m-1 \) is replaced by \( m \).

\textbf{Remark:} (i) One of the property of financial time series is that the autocorrelation function (ACF) of the log return is serially uncorrelated but the ACF of the absolute log return is serially correlated. Thus the financial time series is serially uncorrelated, but dependent [3]. We shall discuss this topic in later section.

(ii) Empirical studies have shown that the distribution of returns is symmetric, but has fatter tails and is more peaked than the Normal, i.e., the distribution is most likely leptokurtic.

\textbf{Volatility Models}

Let us discuss the volatility models listed in Sec. I. We shall start with the simplest model of volatility.

\textbf{Moving Average Model}

Moving average (MA) is the simplest model of volatility forecast. The volatility \( \sigma^2_{i+1} \) on day \( i + 1 \) is given by the equation

\[ \text{(1)} \]
\[ \sigma_{i+1}^2 = \frac{1}{m} \sum_{\tau=1}^{m} u_{i+1-\tau}^2 \]

Here \( u_i \) is observed return on day \( i \), \( \sigma_{i+1} \) is volatility forecast for day \( i + 1 \) and \( m \) is the length of the estimation window, i.e., number of observations used in the calculation. One can see from Eq. ((2)) that in

- MA model all the observation are equally weighted and hence volatility clustering is not explained by the MA model [4].
- MA model is sensitive to the choice of window length. The choice of \( m \) is crucial in deciding the patterns of \( \sigma_{i+1} \).
- A high \( m \) will lead to an excessively smoothly evolving \( \sigma_{i+1} \), and a low \( m \) will lead to an excessively jagged pattern of \( \sigma_{i+1} \) over time.

**Exponentially Weighted Moving Average Model**

Here it is assumed that the past squared rate of return decline exponentially as we move backward in time. This model is also known as JP Morgan’s RiskMetrics. Thus variance in this model can be written as

\[ \sigma_{i+1}^2 = (1 - \lambda) \sum_{\tau=1}^{\infty} \lambda^{\tau-1} u_{i+1-\tau}^2, \quad 0 < \lambda < 1 \]

Thus the square of the past return decline exponentially as we move backward in time. This equation can also be written as

\[ \sigma_{i+1}^2 = (1 - \lambda) \sum_{\tau=1}^{\infty} \lambda^{\tau-1} u_{i+1-\tau}^2 + (1 - \lambda)u_i^2 \]

(3)

where we have separated the \( \tau = 1 \) term in the summation from the rest of the term. Now

\[ \sigma_i^2 = (1 - \lambda) \sum_{\tau=1}^{\infty} \lambda^{\tau-1} u_{i-\tau}^2 \]

\[ = \frac{(1 - \lambda)}{\lambda} \sum_{\tau=2}^{\infty} \lambda^{\tau-2} u_{i+1-\tau}^2 \]

Thus Eq. ((3)) can be written as [7]

\[ \sigma_{i+1}^2 = \lambda \sigma_i^2 + (1 - \lambda)u_i^2 \]

(4)

Thus tomorrow’s volatility is weighted average of today’s volatility and today’s square return. The following remark about JP Morgan RiskMetric are in order:
• In this model, recent return is more important than the distant return, as \( \lambda \) is less than one.

• The model contains only one unknown parameter, namely, \( \lambda \).

• For most cases we set \( \lambda = 0.94 \) for every asset for daily variance forecasting.

• Only few past squared return is needed to calculate the tomorrow’s variance.

**Role of \( \lambda \) in EWMA**

In EWMA model, the value of \( \lambda \) governs the response of daily volatility to the recent daily percentage change in \( u_i \). Suppose there is large move in the market variable on day \( i \), so that \( u_i^2 \) is large. Now according to Eq. ((4)), \( \sigma_{i+1}^2 \) will be large if \( \lambda \) is small (close to zero) and hence the estimates produced for the volatility on successive days are themselves highly volatile (A low value of \( \lambda \) leads to a great deal of weight being given to the \( u_i \) when \( \sigma_{i+1} \) is calculated.). On the other hand, if value of \( \lambda \) is large (close to one), then volatility on successive days will be small [6].

**Remark:** In finance, we make forecast based on past and present information. We want to study the statistical properties of returns given information available at time \( i - 1 \) and create a model of how statistical properties or returns evolve over time.

**The ARCH Model**

ARCH is an acronym meaning AutoRegressive Conditional Heteroscedasticity. This is one of the first model that provides a systematic study of volatility modelling [8]. ARCH model was the first volatility model which attempt to capture the volatility clustering. This model can be specified in terms of the first two conditional moments. The conditional
mean is given by

\[ y_{i+1} = \beta_0 + \sum_{\tau=1}^{n} \beta_{\tau} y_{i+1-\tau} + u_{i+1} \]

where

\[ u_{i} | F_{i-1} : N(0, \sigma_i^2) \]

Here \( F_{i-1} \) represents the past history of the dependent variable. The ARCH conditional variance for this model takes the form:

\[ \sigma_{i+1}^2 = \omega + \sum_{\tau=1}^{q} \alpha_{\tau} u_{i+1-\tau}^2 \]

This is ARCH(\(q\)) model. Here

\[ u_{i} = \sigma_i \varepsilon_{i}, \quad \varepsilon_{i} : i.i.d. N(0,1) \]

Here we assume that mean value of \( u_{i} \) is zero and \( \{\varepsilon_{i}\} \) is a sequence of independent and identically distributed (iid) random variables with mean zero and variance one.

Here \( \varepsilon_{i} \) and \( \sigma_i^2 \) are assumed to be statistically independent. Also, \( \sigma_{i+1}^2 \) is conditional variance that depends on past information, i.e.,

\[ \sigma_{i+1}^2 = \text{var}[u_{i+1} | F_{i}] \]

where \( F_{i} \) is information available up to time \( i \). Thus conditional heteroscedastic models study the evolution of \( \sigma_i^2 \). The following remarks about the ARCH model are as follows:

- **ARCH(\(q\)) model** is given by

\[ \sigma_{i+1}^2 = \omega + \sum_{\tau=1}^{q} \alpha_{\tau} u_{i+1-\tau}^2 \]  

(5)

Thus the variance of the shock \( u_{i} \) is time dependent on the past \( q \) shocks, i.e., \( u_{i}, u_{i-1}, u_{i-2}, \ldots, u_{i-q} \) through their squares.

Note that all the terms on right side of Eq. ((5)) that determines \( \sigma_{i+1}^2 \) are known at time \( i \) and hence \( \sigma_{i+1}^2 \) is the time \( i \) conditional variance and it is in the time \( i \) information set \( F_{i} \) [9].

- If mean of \( u_{i} \) is not zero, then

\[ u_{i} = \mu_i + \sigma_i \varepsilon_{i} \]

where \( \mu_i = E[u_{i}] \).

- Note that

\[ u_{i+1} = \ln \left( \frac{S_{i+1}}{S_{i}} \right) = \ln(S_{i+1}) - \ln(S_{i}) = \sigma_{i+1} \varepsilon_{i+1} \]

and hence

\[ \ln(S_{i+1}) = \ln(S_{i}) + \varepsilon_{i+1} \]
where $\varepsilon_{t+1} = \sigma_{t+1}\varepsilon_{t+1}$. From here we can see that the log of prices tend to follow a random walk.

- We know that in the autoregressive process (AR($p$)-model), we have

$$u_{t+1} = c + \sum_{r=1}^{p} \varphi_{r}u_{t+1-r} + \varepsilon_{t+1}$$

where $p$ is the order of model, $c$ is some constant, $\varphi_{1}, \varphi_{2}, \ldots, \varphi_{p}$ are the parameters and $\varepsilon_{t}$ is a white noise process. Also ARCH(1) process can be shown to be AR(1) process [9]. This can be proved by using the definition of $u_{t} = \varepsilon_{t}\sigma_{t}$. Thus

$$u_{t} = \varepsilon_{t}\sigma_{t}^{2}$$

$$\omega + cau_{t-1}^{2} = \sigma_{t}^{2}$$

Then after subtraction, we get

$$u_{t}^{2} - (\omega + cau_{t-1}^{2}) = \varepsilon_{t}^{2}\sigma_{t}^{2} - \sigma_{t}^{2}$$

which is equivalent to

$$u_{t}^{2} = \omega + cau_{t-1}^{2} + v_{t}$$

where $v_{t} = \sigma_{t}^{2}(\varepsilon_{t}^{2} - 1)$.

- On the other hand, a moving average(MA) process, assumes that the value in a series observed at time $i+1$, is correlated to the mean of the data plus some factors of a white noise. The order of a MA process is given by the number of previous noise terms, $q$, assumed to be correlated to the current result. As such, a MA process, MA($q$), is expressed as follows:

$$u_{i+1} = \mu + \sum_{r=1}^{q} \theta_{r}\varepsilon_{i+1-r} + \varepsilon_{i+1}$$

where $q$ is the order of the model, $\mu$ is mean of the data, $\theta_{1}, \theta_{2}, \ldots, \theta_{q}$ are the parameters and $\varepsilon_{i+1}, \varepsilon_{i}, \ldots$ are white noise.

- In Autoregressive moving average model, (ARMA($p,q$)-model), we have

$$u_{i+1} = c + \sum_{r=1}^{p} \varphi_{r}u_{i+1-r} + \sum_{r=1}^{q} \theta_{r}\varepsilon_{i+1-r} + \varepsilon_{i+1}$$

Note that the ARMA ($p,q$) model, contains both AR($p$) and MA($q$) models. Also, ARMA model is a method that is used to model the data linearly and it does not reflect recent changes. The variance in ARMA model is unconditional variance and it remains constant. On the other hand,
ARCH method is used to model the noise term of ARMA model and the variance is conditional variance and it can be used to forecast the future value which also includes the recent changes. In next section we will show that the ARCH model considers the variance of the current white noise term as a quadratic function of the previous noise terms.

- In particular ARCH(1) model is given by

\[ \sigma_i^2 = \omega + \alpha u_i^2 \]

and hence

\[ u_{i+1} = \sigma_{i+1} \epsilon_{i+1} = \sqrt{\omega + \alpha u_i^2} \epsilon_{i+1} \]

Here \( \omega > 0 \) and \( \alpha > 0 \). ARCH model explains the volatility clustering in the sense that large movements in asset prices tend to follow large movements, of either sign (as the square function only produces positive contributions). The impact of past large movements in prices will be large if \( \alpha \) is large. Thus under ARCH framework, large shocks tend to be followed by another large shock.

- The RiskMetrics model can be taken as special case of ARCH(1) model in which \( \omega = 0 \). Also Riskmetrics attach a weight \( \lambda \) on current variance and a weight \( 1 - \lambda \) to current squared return (see Eq. ((4))). One of the drawbacks of Riskmetrics is that in Riskmetrics, we take only one lag of past squared return and hence it does not produce accurate forecast of variance. Hence to get accurate forecast of variance one needs to use a large number \( q > 1 \) of lags on right side of ARCH representation:

\[ \sigma_{i+1}^2 = \omega + \sum_{r=1}^{q} \alpha_r u_{i+r-1}^2 \]

\[ = \omega + \alpha_1 u_1^2 + \alpha_2 u_2^2 + \cdots + \alpha_q u_{i+q-1}^2 \]

This equation represents ARCH \( (q) \) model.

**Properties of ARCH Variance Model**

Some of the important properties of ARCH variance model are:
(i) **Stationarity Property:** Let us find the unconditional, long-run variance under (6). Because

\[ u_{i+1} = \sigma_{i+1} \epsilon_{i+1} \]

Then unconditional long-run variance is given by

\[ E[u_{i+1}^2] = E[\sigma_{i+1}^2 \epsilon_{i+1}^2] = E[\sigma_{i+1}^2] E[\epsilon_{i+1}^2] = E[\sigma_{i+1}^2] \]

Setting

\[ \sigma^2 = E[\sigma_{i+1}^2] = E[u_{i+1}^2] \quad \forall \tau \]

Thus

\[ \sigma^2 = E[\sigma_{i+1}^2] = \omega + \sum_{\tau=1}^{q} \alpha_{\tau} E[u_{i+1-\tau}^2] \]

\[ = \omega + \sum_{\tau=1}^{q} \alpha_{\tau} \sigma^2 = \omega + \sigma^2 \sum_{\tau=1}^{q} \alpha_{\tau} \]

Thus

\[ \sigma^2 = \frac{\omega}{1 - \sum_{\tau=1}^{q} \alpha_{\tau}} \]

Because unconditional variance makes sense only when \( \sigma^2 > 0 \), this equation implies that when \( \omega > 0 \), the condition

\[ 1 - \sum_{\tau=1}^{q} \alpha_{\tau} > 0 \Rightarrow \sum_{\tau=1}^{q} \alpha_{\tau} < 1 \]

must hold. The condition

\[ \sum_{\tau=1}^{q} \alpha_{\tau} < 1 \]

is the stationary condition. This condition is required for the convergence of the conditional variance.

**Remark:** A process is called **weakly stationary** if all of its second moments are constant. In particular this means that the mean and variance are constants \( \mu_i = \mu \) and \( \sigma_i^2 = \sigma^2 \) that do not depend on the time \( i \). A process is called **strictly stationary** if none of its finite distributions depends on time. A strictly stationary process is not necessarily weakly stationary as its finite distributions, though time-independent, might have infinite moments.

(ii) **Kurtosis:** One of the important property of ARCH model is that the kurtosis of shock \( (u_i) \) is greater than the kurtosis of normal distribution. An intuitive proof of kurtosis is given by [4], [9]
\[ K = \frac{E[u_i^4]}{E[u_i^2]} = \frac{E[E_{i-1}[u_i^4]]}{E[E_{i-1}[\varepsilon_i^2 \sigma_i^2]]^2} = \frac{E[E_{i-1}[\varepsilon_i^4 \sigma_i^4]]}{E[E_{i-1}[\varepsilon_i^2 \sigma_i^2]]^2} = \frac{E[3\varepsilon_i^4]}{E[\sigma_i^4]} = \frac{3E[\sigma_i^4]}{E[\sigma_i^4]} > 3 \]

where

\[ E[u_i^4] = E[\sigma_i^4 \varepsilon_i^4] = 3E[\sigma_i^4] \]

and

\[ E[\varepsilon_i^4] = 3 \]

The kurtosis \( K \) of ARCH model can be shown to be [9]

\[ K = \frac{3(1-\alpha^2)}{(1-3\alpha^2)} > 3, \quad \text{if} \quad 3\alpha^2 < 1 \]

Thus kurtosis of ARCH model is greater than three.

(iii) **Nonlinearity:** We are now going to show that \( u_i \) is nonlinear function of \( (\varepsilon_i, \varepsilon_{i-1}, \ldots) \). To prove this [10], we use the relation \( \sigma_i^2 = \omega + \alpha u_{i-1}^2 \) recursively in the expression \( u_i^2 = \sigma_i^2 \varepsilon_i^2 \), i.e.,

\[ u_i^2 = \sigma_i^2 \varepsilon_i^2 = \varepsilon_i^2 (\omega + \alpha u_{i-1}^2) = \omega \varepsilon_i^2 + \alpha \varepsilon_i^2 u_{i-1}^2 = \omega \varepsilon_i^2 + \alpha \varepsilon_i^2 (\sigma_{i-1}^2 \varepsilon_{i-1}^2) \]

Since \( \alpha < 1 \), the last term of the above equation tends to zero as \( n \to \infty \). Thus in the limit \( n \to \infty \), we obtain

\[ u_i^2 = \omega \sum_{j=0}^{n} \alpha^j \varepsilon_i^2 \varepsilon_{i-j}^2 \varepsilon_{i-j}^2 \]

From here we can see that \( u_i \) is nonlinear function of \( (\varepsilon_i, \varepsilon_{i-1}, \ldots) \).

**Forecasting Volatility Using ARCH(1) Model**

We can use ARCH(1) process to predict the conditional volatility. To do this we write

\[ \sigma_{i+1}^2 = \omega + \alpha u_i^2 \]

\[ = \omega + \alpha \varepsilon_i^2 \]

where \( u_i = \sigma_i \varepsilon_i = \varepsilon_i \). Taking the conditional expectation, we get

\[ E_i[\sigma_{i+1}^2] = E_i[\omega + \alpha \varepsilon_i^2] \]

\[ = \omega + \alpha \varepsilon_i^2 \]
where $E_i$ is the conditional expectation given the information set at time $i$. This is a property common to all ARCH-family models. The two-step ahead forecast is given by conditional expectation of

$$E_i[\sigma_{i+2}^2] = E_i[\omega + \alpha \varepsilon_{i+1}^2] = \omega + \alpha E_i[\varepsilon_{i+1}^2]$$

$$= \omega + \alpha E_i[\varepsilon_{i+1}^2, \sigma_{i+1}^2] = \omega + \alpha E_i[\varepsilon_{i+1}^2] E_i[\sigma_{i+1}^2]$$

$$= \omega + \alpha \cdot 1 \cdot E_i[\sigma_{i+1}^2] = \omega + \alpha (\omega + \alpha \varepsilon_i^2)$$

$$= \omega + \alpha \omega + \alpha^2 \varepsilon_i^2$$

Similarly, the three-step ahead forecast of conditional volatility is given by

$$E_i[\sigma_{i+3}^2] = E_i[\omega + \alpha \varepsilon_{i+2}^2] = \omega + \alpha E_i[\varepsilon_{i+2}^2]$$

$$= \omega + \alpha E_i[\varepsilon_{i+2}^2, \sigma_{i+2}^2] = \omega + \alpha E_i[\varepsilon_{i+2}^2] E_i[\sigma_{i+2}^2]$$

$$= \omega + \alpha \cdot 1 \cdot E_i[\sigma_{i+2}^2]$$

$$= \omega + \alpha (\omega + \alpha \omega + \alpha^2 \varepsilon_i^2)$$

$$= \omega + \alpha \omega + \alpha^2 \omega + \alpha^3 \varepsilon_i^2$$

In a similar fashion, the $h$-step ahead forecast is given by

$$E_i[\sigma_{i+h}^2] = \omega \sum_{j=0}^{h-1} \alpha^j + \alpha^h \varepsilon_i^2$$

The GARCH Variance Model

In ARCH model we need large number of parameters to adequately describe the volatility process of asset return. In some cases twenty to hundred parameters are needed for the volatility process. Hence to overcome this problem, we need some alternative method to describe the volatility process. Bollerslev proposed a useful extension known as the generalized ARCH (GARCH) model [11]. In GARCH variance (1,1) model, the dynamic variance can be written as

$$\sigma_{i+1}^2 = \omega + \alpha \varepsilon_i^2 + \beta \sigma_i^2, \quad \alpha + \beta < 1$$

The equation states that conditional variance of tomorrow’s return is equal to a constant, plus today’s residual squared, plus today’s known variance. Again

$$u_t = \sigma \varepsilon_t$$

and the common choices for $\varepsilon_t$ are normal and student’s $t$ disturbances.
Notice that the Risk Metrics model can be viewed as a special case of the simple GARCH model if we force \( \alpha = 1 - \lambda \), so that \( \alpha + \beta = 1 \), and further \( \omega = 0 \). Thus, the two models appear to be quite similar. However, there is an important difference: We can define the unconditional, or long-run average, variance, \( \sigma^2 \) to be

\[
\sigma^2 = E[\sigma^2_{i+1}] = \omega + \alpha E[u_i^2] + \beta E[\sigma_i^2]
\]

Thus

\[
\sigma^2 = \frac{\omega}{1 - \alpha - \beta}
\]

If \( \alpha + \beta = 1 \) as is the case in the Risk Metrics model, then the long run variance is not well defined in that model. Thus the RiskMetrics model ignores the fact that the long-run average variance tends to be relatively stable over time. The GARCH model, in turn, implicitly relies on \( \sigma^2 \).

This can be seen by solving for \( \omega \) in the long-run variance equation and substituting it into the dynamic variance equation. We get

\[
\sigma^2_{i+1} = E_i[\sigma^2_{i+1}] = (1 - \alpha - \beta)\sigma^2 + \alpha u_i^2 + \beta \sigma_i^2
\]

where \( E_i[\sigma^2_{i+1}] \) indicates volatility on day \( i+1 \) given information on day \( i \). Here

\[
\sigma^2_{i+1} = \text{var}(u_i | F_i) = E[u_i^2 | F_i]
\]

The following remark about the GARCH (1,1) model are in order:

- Under a GARCH(1,1), the forecast of tomorrow’s variance is the long run average variance, adjusted by:

  \( \gamma + \alpha + \beta = 1 \)

Thus

\[
\sigma^2 = \frac{\omega}{1 - \alpha - \beta}
\]

If \( \alpha + \beta = 1 \) as is the case in the Risk Metrics model, then the long run variance is not well defined in that model. Thus the RiskMetrics model ignores the fact that the long-run average variance tends to be relatively stable over time. The GARCH model, in turn, implicitly relies on \( \sigma^2 \).

This can be seen by solving for \( \omega \) in the long-run variance equation and substituting it into the dynamic variance equation. We get

\[
\sigma^2_{i+1} = E_i[\sigma^2_{i+1}] = (1 - \alpha - \beta)\sigma^2 + \alpha u_i^2 + \beta \sigma_i^2
\]

where \( E_i[\sigma^2_{i+1}] \) indicates volatility on day \( i+1 \) given information on day \( i \). Here

\[
\sigma^2_{i+1} = \text{var}(u_i | F_i) = E[u_i^2 | F_i]
\]

The following remark about the GARCH (1,1) model are in order:

- Under a GARCH(1,1), the forecast of tomorrow’s variance is the long run average variance, adjusted by:

  \( \gamma + \alpha + \beta = 1 \)

Thus

\[
\sigma^2 = \frac{\omega}{1 - \alpha - \beta}
\]

If \( \alpha + \beta = 1 \) as is the case in the Risk Metrics model, then the long run variance is not well defined in that model. Thus the RiskMetrics model ignores the fact that the long-run average variance tends to be relatively stable over time. The GARCH model, in turn, implicitly relies on \( \sigma^2 \).

This can be seen by solving for \( \omega \) in the long-run variance equation and substituting it into the dynamic variance equation. We get

\[
\sigma^2_{i+1} = E_i[\sigma^2_{i+1}] = (1 - \alpha - \beta)\sigma^2 + \alpha u_i^2 + \beta \sigma_i^2
\]

where \( E_i[\sigma^2_{i+1}] \) indicates volatility on day \( i+1 \) given information on day \( i \). Here

\[
\sigma^2_{i+1} = \text{var}(u_i | F_i) = E[u_i^2 | F_i]
\]

The following remark about the GARCH (1,1) model are in order:

- Under a GARCH(1,1), the forecast of tomorrow’s variance is the long run average variance, adjusted by:

  \( \gamma + \alpha + \beta = 1 \)

Thus

\[
\sigma^2 = \frac{\omega}{1 - \alpha - \beta}
\]

If \( \alpha + \beta = 1 \) as is the case in the Risk Metrics model, then the long run variance is not well defined in that model. Thus the RiskMetrics model ignores the fact that the long-run average variance tends to be relatively stable over time. The GARCH model, in turn, implicitly relies on \( \sigma^2 \).

This can be seen by solving for \( \omega \) in the long-run variance equation and substituting it into the dynamic variance equation. We get

\[
\sigma^2_{i+1} = E_i[\sigma^2_{i+1}] = (1 - \alpha - \beta)\sigma^2 + \alpha u_i^2 + \beta \sigma_i^2
\]

where \( E_i[\sigma^2_{i+1}] \) indicates volatility on day \( i+1 \) given information on day \( i \). Here

\[
\sigma^2_{i+1} = \text{var}(u_i | F_i) = E[u_i^2 | F_i]
\]
forecasting than for forecasting simply one day ahead.

- GARCH(1,1) forecast can be thought of as a weighted average of unconditional variance, the deviation of last period’s forecast from unconditional variance and the deviation of last period’s squared returns from unconditional variance.

- The Garch variance model is mean reverting process, i.e., over long time horizon, the conditional variance get pull back to a long-run average level of $\sigma^2$ [6].

- GARCH model can be regarded as ARMA model for the squared return series $u_i^2$ [3]. To prove this, we define

$$\eta_i = u_i^2 - \sigma_i^2$$

so that $\sigma_i^2 = u_i^2 - \eta_i$ and hence

$$\sigma_{i-1}^2 = u_{i-1}^2 - \eta_{i-1}$$

Now

$$\sigma_i^2 = \omega + \alpha u_{i-1}^2 + \beta \sigma_{i-1}^2$$

$$= \omega + \alpha u_{i-1}^2 + \beta (u_{i-1}^2 - \eta_{i-1})$$

$$= \omega + (\alpha + \beta) u_{i-1}^2 - \beta \eta_{i-1}$$

This equation is ARMA model for the squared return series $u_i^2$.

### 7.1 Multi-Step Ahead Volatility

Let us now derive two-step-ahead volatility:

$$\sigma_{i+2}^2 = E_i[\sigma_{i+2}^2]$$

$$= E_i[\sigma^2 + \alpha (u_{i+1}^2 - \sigma^2) + \beta (\sigma_{i+1}^2 - \sigma^2)]$$

$$= \sigma^2 + \alpha (E_i[u_{i+1}^2] - \sigma^2) + \beta (\sigma_{i+1}^2 - \sigma^2)$$

$$= \sigma^2 + (\alpha + \beta) (\sigma_{i+1}^2 - \sigma^2)$$

Following similar arguments, the general formula for $n$ periods is given as:

$$\sigma_{i+n}^2 = \sigma^2 + (\alpha + \beta)^{n-1} (\sigma_{i+1}^2 - \sigma^2), \quad n \geq 1$$

From here we can see that the forecast of one-period volatility $n$ periods from now is a weighted average of unconditional variance and the deviation of the one-step forecast from unconditional variance. If $\alpha + \beta < 1$, then in the limit $n \to \infty$ the second term above goes to zero. This implies that longer the forecast horizon, the
closer the forecast will get to unconditional variance. Thus the size of \((\alpha + \beta)\) determines how quickly the predictability of the process subsides: if \((\alpha + \beta)\) is close to zero, predictability will die out very quickly. If \((\alpha + \beta)\) is close to one, predictability will die out slowly.

Thus we can conclude that the coefficient sum \((\alpha + \beta)\) plays a crucial role concerning the forecasting with GARCH models and is commonly called the persistence level/index of the model: a high persistence, \((\alpha + \beta)\) close to 1, implies that shocks which push variance away from its long-run average will persist for a long time, even though eventually the long-horizon forecast will be the long-run average variance, \(\sigma^2\).

The unconditional volatility of the GARCH model was given by (7), but oftentimes a very small or very large shock to the return process can cause conditional volatility to differ greatly from that number. It is often of interest to identify how long it takes for the impact of the shock to subside. The memory of a GARCH model measures how long a shock to the process takes to subside. A measure of memory can be developed by looking at multistep-ahead conditional variance, where \(\sigma^2_{i+2\tau}\) is the volatility on day \(i + 2\) given information on day \(i\).

GARCH \((p,q)\) Model

The natural extension of GARCH(1,1) model is GARCH \((p,q)\) model and is given by

\[
\sigma^2_{i+1} = \omega + \sum_{r=1}^{q} \alpha_r u^2_{i+1-r} + \sum_{r=1}^{p} \beta_r \sigma^2_{i+1-r}
\]

The unconditional variance in this case is given by

\[
\sigma^2 = \frac{\omega}{1 - \sum_{r=1}^{q} \alpha_r - \sum_{r=1}^{p} \beta_r}
\]

Now unconditional variance exist only if \(\sigma^2 > 0\), and hence when \(\omega > 0\), the condition
This condition is also known as stationarity condition.

**GARCH(1,1) Model is Equivalent to ARCH\(^\infty\) Model**

In practice we use only GARCH(1,1) model for volatility forecast. Note that one of the problem of ARCH model is that we need large values of \( q \) for volatility estimation. The reason for the success of simple GARCH(1,1) models is that these can be shown to be equivalent to an ARCH\(^\infty\) model [7]. To prove this we use recursive substitution, i.e.,

\[
\sigma_{i+1}^2 = \omega + \alpha u_i^2 + \beta \sigma_i^2 = \omega + \alpha u_i^2 + \beta \left[ \omega + \alpha u_{i-1}^2 + \beta \sigma_{i-1}^2 \right] \sum_{j=0}^{\infty} \beta^j \\
= \omega(1+\beta) + \alpha u_i^2 + \alpha \beta u_{i-1}^2 + \beta^2 \sigma_{i-1}^2
given by

\[
\sigma_{i+1}^2 = \omega + \alpha u_i^2 + \alpha \beta u_{i-1}^2 + \beta^2 \sigma_{i-1}^2
\]

\[
= \omega(1+\beta + \beta^2) + \alpha u_i^2 + \alpha \beta u_{i-1}^2 + \alpha \beta^2 u_{i-2}^2 + \beta^3 \sigma_{i-2}^2
\]

**Remark:** Note that in RiskMetrics, we had

\[
\sigma_{i+1}^2 = \omega + \alpha u_i^2 + \beta \sigma_{i-1}^2 = \omega + \alpha u_i^2 + \beta \left( \omega + \alpha u_{i-2}^2 + \beta \sigma_{i-2}^2 \right)
\]

and in GARCH(1,1) model, we have

\[
\sigma_{i+1}^2 = \omega + \alpha u_i^2 + \alpha \beta u_{i-1}^2 + \alpha \beta^2 u_{i-2}^2 + \beta^3 \sigma_{i-2}^2
\]
\[ \sigma_{i+1}^2 = \omega + \alpha u_i^2 + \beta \sigma_i^2 \]

From here one can see that RiskMetrics is just a special case of GARCH(1,1) in which \( \omega = 0 \) and \( \alpha = 1 - \beta \) so that, \( (\alpha + \beta) = 1 \).

Thus under RiskMetrics the long-run variance does not exist as

\[ \sigma_{RMetric}^2 = \frac{\omega}{1 - \alpha - \beta} = \frac{0}{1 - 1} = 0 \]

which is indeterminate form.

**Volatility Forecasting Using GARCH(1,1) Model**

Forecast of GARCH(1,1) variance model is given by

\[ \sigma_{i+1}^2 = \omega + \alpha \sigma_i^2 + \beta \epsilon_i^2 \]

\[ = \omega + \alpha \epsilon_i^2 + \beta \sigma_i^2 \]

where \( u_i = \sigma, \epsilon_i = \epsilon_i \). Taking the conditional variance on both sides of this equation, we get

\[ E_i[\sigma_{i+1}^2] = E_i[\omega + \alpha \epsilon_i^2 + \beta \sigma_i^2] \]

\[ = \omega + \alpha \epsilon_i^2 + \beta \sigma_i^2 \]

Similarly,

\[ E_i[\sigma_{i+2}^2] = E_i[\omega + \alpha \epsilon_{i+1}^2 + \beta \sigma_{i+1}^2] \]

\[ = \omega + \alpha \epsilon_{i+1}^2 + \beta \sigma_{i+1}^2 \]

Using Eq. (9), we obtain

\[ E_i[\sigma_{i+2}^2] = \omega + \alpha \cdot E_i[\sigma_{i+1}^2] + \beta E_i[\epsilon_i^2] \]

For three-step ahead forecast, we have

\[ E_i[\sigma_{i+3}^2] = \omega + (\alpha + \beta) E_i[\sigma_{i+2}^2] \]

and hence

\[ E_i[\sigma_{i+3}^2] = \omega + (\alpha + \beta)[\omega + (\alpha + \beta) \omega + (\alpha + \beta) \alpha \epsilon_i^2] \]

\[ = \omega + (\alpha + \beta) \omega + (\alpha + \beta)^2 \omega + (\alpha + \beta)^2 \alpha \epsilon_i^2 + (\alpha + \beta)^2 \alpha \epsilon_i^2 \]

In general, the \( h \)-step ahead forecast is given by

\[ E_i[\sigma_{i+h}^2] = \omega \sum_{j=0}^{h-1} (\alpha + \beta)^j + (\alpha + \beta)^{h-1} (\alpha \epsilon_i^2 + \beta \sigma_i^2) \]  

(9)

**GARCH Parameter Estimation**
Using Gaussian Errors

To find the parameters of GARCH model, we use maximum likelihood estimation (MLE). According to MLE, we are given the data and a model and we have to find the parameters of the probability densities (that the model prescribes) that is most likely to have produced the data [12]. Thus MLE is a method to seek the parameters of the probability distribution that makes the observed data most likely. MLE for GARCH(1,1) model can be estimated as follows:

In GARCH(1,1) model, we have

\[ u_i = \mu_i + \sigma_i \varepsilon_i = \mu_i + \varepsilon_i \]
\[ \sigma_i^2 = \omega + \alpha u_{i-1}^2 + \beta \sigma_{i-1}^2 \]

where \( \varepsilon_i \sim N(0,1) \) are conditionally i.i.d. normal and \( \mu_i = 0 \). Since \( \varepsilon_i \) are normal, we take the likelihood function as

\[
L(u, \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \sigma_i^2}} \exp\left(-\frac{(u_i - \mu_i)^2}{2\sigma_i^2}\right)
\]

and hence the normal log-likelihood function is

\[
l(u, \theta) = \sum_{i=1}^{n} \log[p(u_i | \sigma_i; \theta)]
\]

where \( p \) is the probability density and \( \theta \) is the parameters of the GARCH(1,1) model. Note that in above equation

\[
p(u_i | \sigma_i; \theta) = \frac{1}{\sqrt{2\pi \sigma_i^2}} \exp\left(-\frac{u_i^2}{2\sigma_i^2}\right)
\]

Thus we are given \( u_i \) and we have to determine \( \theta = \{\omega, \alpha, \beta\} \). Using above equation, we obtain

\[
l(u, \theta) = \log L(u, \theta) = \sum_{i=1}^{n} \left[-\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_i^2) - \frac{u_i^2}{2\sigma_i^2}\right]
\]

Now to maximize the log-likelihood function, we set

\[
\frac{\partial l(u, \theta)}{\partial \sigma_i^2} = \sum_{i=1}^{n} \left[-\frac{1}{2 \sigma_i^2} + \frac{u_i^2}{2 \sigma_i^4}\right] = 0
\]

This equation can be written as

\[
\frac{\partial l(u, \theta)}{\partial \sigma_i^2} = \frac{1}{2} \sum_{i=1}^{n} \left[\frac{u_i^2}{\sigma_i^2} - 1\right]
\]

From here we can see that parameters of the
volatility model must be chosen to make
\[ \left( \frac{u_i^2}{\sigma_i^2} - 1 \right) \text{ as close to zero as possible. Now} \]
the parameters of the GARCH(1,1) model are \( \omega, \alpha \) and \( \beta \) and hence we write
\[ \frac{\partial l(u, \theta)}{\partial \theta_i} = \frac{\partial l(u, \theta)}{\partial \sigma_i^2} \frac{\partial \sigma_i^2}{\partial \theta_i} \]
These derivatives can be determined recursively using the relation
\[ \sigma_i^2 = \omega + \alpha u_{i-1}^2 + \beta \sigma_{i-1}^2 \]
\[ = \omega + \alpha u_{i-1}^2 + \beta (\omega + \alpha u_{i-2}^2 + \beta \sigma_{i-2}^2) \]
Thus
\[ \frac{\partial \sigma_i^2}{\partial \omega} = 1 + \beta \frac{\partial \sigma_{i-1}^2}{\partial \omega} \]
\[ \frac{\partial \sigma_i^2}{\partial \alpha} = u_{i-1}^2 + \beta \frac{\partial \sigma_{i-1}^2}{\partial \alpha} \]
\[ \frac{\partial \sigma_i^2}{\partial \beta} = \sigma_{i-1}^2 + \beta \frac{\partial \sigma_{i-1}^2}{\partial \beta} \]
Now
\[ \frac{\partial \sigma_i^2}{\partial \omega} = 1 + \beta \frac{\partial \sigma_{i-1}^2}{\partial \omega} \approx \frac{1}{1 - \beta} \]
Thus using the above equations, we can determine the GARCH(1,1) parameter.

An alternative way to determine GARCH(1,1) model is given in ref. [6].

According to it we maximize
\[ L(u, \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \sigma_i^2}} \exp \left( -\frac{u_i^2}{2\sigma_i^2} \right) \]
which is equivalent to maximizing
\[ l(u, \theta) = \sum_{i=1}^{n} \left[ -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_i^2) - \frac{u_i^2}{2\sigma_i^2} \right] \]

(10)

also
\[ \sigma_i^2 = \omega + \alpha u_{i-1}^2 + \beta \sigma_{i-1}^2 \]

Here we search iteratively to find the parameters that maximizes the expression in Eq. (10). Note that we need \( \sigma_i^2 \) for complete definition of \( l(u, \theta) \). Thus a reasonable guess of \( \sigma_i^2 \) improves accuracy in finite samples. The exact value of \( \sigma_i^2 \) does not matter in large samples, since \( \sigma_i^2 \) converges to its stationary distribution for large \( i \). A reasonable guess of \( \sigma_i^2 \) is sample unconditional variance.

For more details see ref. [6].

GARCH MLE With Student’s \( t \) shocks
In standard GARCH formulation we assume that errors $\varepsilon_i$ are Gaussian. In this section we assume that errors $\varepsilon_i$ are non-Gaussian. One prominent example of GARCH with non-Gaussian errors is the GARCH model with Student’s $t$ error distribution [13]. Note that the Student $t$ disturbance explains the excess kurtosis that is present in financial time series. We assume that $u_i$ follow a student’s $t$ distribution, but still

$$u_i = \sigma_i \varepsilon_i, \quad E_i[\varepsilon_i] = 0, \quad E_i[\varepsilon_i^2] = 1$$

Here the distribution of the error term according to Bollerslev [14] takes the form:

$$p(\varepsilon_i; \nu) = \frac{1}{\sqrt{\pi(\nu-2)}} \frac{\Gamma \left( \frac{\nu+1}{2} \right)}{\sqrt{\nu/2} \Gamma \left( \frac{\nu}{2} \right)} \left( 1 + \frac{\varepsilon_i^2}{\nu-2} \right)$$

where $\Gamma$ is the Gamma function and is given by

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx$$

Thus likelihood function for GARCH(1,1) is given by

$$l(u, \theta) = \sum_{i=1}^n \log \left[ \frac{\Gamma \left( \frac{\nu+1}{2} \right)}{\sqrt{\nu/2} \Gamma \left( \frac{\nu}{2} \right)} \left( 1 + \frac{\varepsilon_i^2}{\nu-2} \right) \right] - \frac{\nu+1}{2} \log \left( 1 + \frac{\varepsilon_i^2}{\nu-2} \right)$$

**Remark:** Note that in both ARCH and GARCH variance model, only magnitude of returns is used and hence they ignore the information on sign of returns. Nelson [15] introduced the Exponential GARCH (EGARCH) model, which was the first in the family of asymmetric GARCH models. There are many variations of GARCH model and have been used successfully in many problems. The GJR model by Glosten, Jaganathan and Runkle [16] account for both the asymmetric relation between stock returns and changes in variance. Also, one of the important extension of GARCH model is multivariate model.

**Bayesian Estimation of GARCH(1,1) Model**

There are certain practical
difficulties in ML estimation of GARCH model. The maximization of likelihood function is achieved via a constrained optimization. Also, the model parameters must be positive to ensure a positive conditional variance [17] and also it must satisfy the covariance stationarity condition, i.e.,

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

for GARCH \((p,q)\) model. For detailed discussion, see ref. [17]. To overcome these difficulties of ML estimation, Bayesian method is used [17]. The Bayesian model uses the rules of probability in order to make prediction. Here we make predictions about the unobserved data on the basis of observed data [19], [18]. In Bayesian estimation, we assume that the observable data is given by the set

$$y = \{y_1, y_2, \ldots, y_n\}^T$$

where \(y\) is \(n \times 1\) vector. Now we know that each model has a number of free parameters. Let \(\theta\) represents the parameter of the model. Here \(\theta\) can be a scalar, a vector or a matrix. In Bayesian estimation, \(\theta\) is considered as a random variable and is characterised by a prior density \(p(\theta)\) (in frequentist approach, \(\theta\) is not a random variable). Here \(p(\theta)\) is our belief in \(\theta\) without the data \(y\). Now using Bayes rule, we can write

$$p(\theta | y) = \frac{p(y | \theta)p(\theta)}{\int p(y | \theta)p(\theta)d\theta}$$

(11)

where \(p(\theta | y)\) is posterior density (it is our belief in \(\theta\) when the data \(y\) is given) and \(p(y | \theta)\) is likelihood function. The likelihood function \(p(y | \theta)\), is the probability that the data could be generated by the model with parameter values \(\theta\). Note that both the prior and likelihood is necessary requirement for Bayesian estimation. To evaluate the integration ((11)), Monte-Carlo method is used.

**Remark:** Note that parameter \(\theta\) is defined in the context of particular model (without a model, \(\theta\) is meaningless). We
assume that the model is well defined. Let us assume that the model is $m$. Then, Bayes rule is given by

$$p(\theta | y, m) = \frac{p(y | \theta, m) p(\theta | m)}{\int p(y | \theta, m) p(\theta | m) d\theta}$$

It is necessary to specify the model, when we have more than one model in our mind. Also, the model is not fully defined until we specify a range for the parameter values, $\theta$. We can also use Bayes rule to compare different models. Let us suppose that we have model set $M$, then posterior probability of particular model $m$ for some observed data $y$ is given by

$$p(m | y) = \frac{p(y | m) p(m)}{p(y)} = \sum_{m \in M} \frac{p(y | m) p(m)}{p(y)}$$

**Bayesian Estimation of GARCH(1,1) Model With Normal Innovation**

Let us first discuss the Bayesian estimation of GARCH(1,1) model with normal innovation [20]. According to the GARCH(1,1) variance model, we have

$$u_i = \sigma_i \varepsilon_i \quad \text{for} \quad i=1,2,\ldots,\tau$$

$$\varepsilon_i \sim N(0,1)$$

$$\sigma_i^2 = \omega + \alpha u_{i-1}^2 + \beta \sigma_{i-1}^2$$

**The Likelihood Function:** Let us define the vector

$$u = (u_1, u_2, \ldots, u_\tau)^T$$

$$\zeta = (\omega, \alpha)^T$$

where $T$ stands for transpose. Now regrouping the parameters, to write

$$\Theta = (\zeta, \beta)$$

Let us also define a $\tau \times \tau$ diagonal matrix

$$\Sigma = \Sigma(\Theta) = \text{diag} \left[ \sigma_i^2(\Theta) \right]_{i=1}^\tau$$

Then the likelihood function of $\Theta$ can be written as

$$l(\Theta | u) \propto \frac{1}{\sqrt{\det \Sigma}} \exp \left[ -\frac{1}{2} u^T \Sigma^{-1} u \right]$$

Here it assumed that initial variance is fixed to $\omega$ and first observation is used as initial condition [20]. The prior on parameters
\[ \zeta, \beta \] is given by
\[ p(\zeta) \propto N_2(\zeta \mid \mu_\zeta, \Sigma_\zeta) 1_{[\zeta > 0]} \]
\[ p(\beta) \propto N(\beta \mid \mu_\beta, \Sigma_\beta) 1_{[\beta > 0]} \]
(12)
Here \( \mu_\zeta, \mu_\beta, \Sigma_\zeta, \Sigma_\beta \) are hyperparameters and \( I \) is an indicator function which is equal to unity if the constraint holds, and zero otherwise. Also, \( N_d \) is the \( d \)-dimensional Normal distribution (here \( d > 1 \)). We also assume that
\[ p(\Theta) = p(\zeta)p(\beta) \]
i.e., \( \zeta \) and \( \beta \) are independent. Using Bayes rule, we can find the joint posterior distribution as
\[ p(\Theta \mid u) \propto l(\Theta \mid u)p(\Theta) \]

**Simulation of Joint Posterior**

Simulation of joint posterior distribution is based on Markov chain Monte-Carlo method [21]. Detailed algorithm of this method is given in ref. [20], [22]. The first step is to draw an initial value
\[ \Theta^{[0]} = (\zeta^{[0]}, \beta^{[0]}) \]
from the joint prior and iteratively generate the \( J \) passes for \( \Theta \). Here, the single pass is decomposed as
\[ \zeta^{[j]} : p(\zeta \mid \beta^{[j-1]}, u) \]
\[ \beta^{[j]} : p(\beta \mid \alpha^{[j]}, u) \]
The proposal densities for \( \zeta \) and \( \beta \) are obtained by writing GARCH(1,1) model as ARMA(1,1) model for \{\( u_i^2 \}\}. Let us first define
\[ w_i = u_i^2 - \sigma_i^2 \]
and hence expression of the conditional variance can be transformed as
\[ \sigma_i^2 = \omega + \alpha u_{i-1}^2 + \beta \sigma_{i-1}^2 \]
Or,
\[ u_i^2 = \omega + (\alpha + \beta) u_{i-1}^2 - \beta w_{i-1} + w_i \]
(13)
Note that \( w_i \) can be written as
\[ w_i = u_i^2 - \sigma_i^2 = \left( \frac{u_i^2}{\sigma_i^2} - 1 \right) \sigma_i^2 \]
\[ = (\chi_i^2 - 1) \sigma_i^2 \]
The likelihood for parameters \( \zeta \) and \( \beta \) is constructed using the expression ((13)).
To do so, the variable \( w_i \) is approximated by the variable \( z_i \). Here \( z_i \) is normally distributed with a mean of zero and a variance of \((\sqrt{2} \sigma_i^2)^2\). Under this assumption, the expression ((13)) can be written as

\[
\omega_i^2 = \omega + (\alpha + \beta)u_{i-1}^2 - \beta z_{i-1} + z_i
\]

Note that \( z_i \) is function of \( \Theta \), and hence we can write

\[
z_i(\Theta) = \omega_i^2 - \omega - (\alpha + \beta)u_{i-1}^2 + \beta z_{i-1}(\Theta)
\]

Let us define a \( \tau \times 1 \) vector

\[
z = (z_1, z_2, \ldots, z_{\tau})^T
\]

and a \( \tau \times \tau \) diagonal matrix

\[
\Lambda(\Theta) = \text{diag}\left[\begin{bmatrix} 2\sigma_1^4(\Theta) & \cdots & 2\sigma_{\tau}^4(\Theta) \end{bmatrix}^T \right]
\]

Thus the likelihood function of \( \Theta \) can be approximated from this auxiliary model as

\[
l(\Theta | u) \propto \frac{1}{\sqrt{\det \Lambda}} \exp\left[-\frac{1}{2}z^T\Lambda^{-1}z\right]
\]

This likelihood function is used for the proposal densities for the parameter \( \zeta \) and \( \beta \). The approximate likelihood function for the parameter \( \zeta \) is given by

\[
l(\zeta | \beta, u) = \frac{1}{\sqrt{\det(\Lambda)}} \exp\left[(v - C\zeta)^T\Lambda^{-1}(v - C\zeta)\right]
\]

where \( v_i = u_i^2 \) and \( v = (v_1, v_2, \ldots, v_{\tau})^T \). For detailed discussion see ref. [20]. The likelihood function for the parameter \( \beta \) is given by

\[
l(\beta | \zeta, u) = \frac{1}{\sqrt{\det(\Lambda)}} \exp\left[(r - \beta \nabla)^T\Lambda^{-1}(r - \beta \nabla)\right]
\]

where

\[
\nabla = (\nabla_1, \ldots, \nabla_{\tau})^T
\]

and

\[
\tau_i = z_i(\bar{\beta}) + \bar{\beta} \nabla_i, \quad \nabla_i = -\left. \frac{dz_i}{d\beta} \right|_{\beta=\bar{\beta}}
\]

See ref. [20] for derivation of this equation.

**Bayesian Estimation of GARCH(1,1) Model With Student’s \( t \) Innovation**

The GARCH(1,1) model with student’s \( t \) innovation is given by

\[
u_i = \varepsilon_i \left(\frac{v - 2}{v} \sigma_i^2\right)^{\frac{1}{2}}
\]

\( \varepsilon_i \): iid \( N(0,1) \)
\[ \sigma_i : i.i.d. IG \left( \frac{\nu}{2}, \frac{\nu}{2} \right) \]

\[ \sigma_i^2 = \omega + \alpha \sigma_{i-1}^2 + \beta \sigma_{i-1}^2 \]

where \( IG \) denote the inverted gamma function and \( \nu > 2 \) [17], [23]. Again, to write the likelihood function, let us define the vectors

\[ u = (u_1, u_2, \ldots, u_r)^T \]

\[ \sigma = (\sigma_1, \sigma_2, \ldots, \sigma_i)^T \]

\[ \zeta = (\omega, \alpha)^T, \quad \Theta = (\zeta, \beta, \nu) \]

then defining the \( r \times r \) diagonal matrix

\[ \Sigma(\Theta, \sigma) = \text{diag} \left[ \sigma_i \left( \frac{\nu - 2}{\nu} \right) \sigma_i^2 (\zeta, \beta) \right] \]

where

\[ \sigma^2(\zeta, \beta) = \omega + \alpha \sigma_{i-1}^2 + \beta \sigma_{i-1}^2 \]

Then the likelihood function of \((\Theta, \sigma)\) can be expressed as [23]

\[ l(\Theta, \sigma | u) \propto \frac{1}{\sqrt{\det \Sigma}} \exp \left\{ -\frac{1}{2} u^T \Sigma^{-1} u \right\} \]

Finally, using the Bayes rule, we can write the posterior density as

\[ p(\Theta, \sigma | u) = \frac{l(\Theta, \sigma | u) p(\Theta, \sigma)}{\int l(\Theta, \sigma | u) p(\Theta, \sigma) d\Theta d\sigma} \]

where \( p(\Theta, \sigma) \) is prior density and \((\Theta, \sigma)\) is considered as random variable.

The prior on parameters \( \zeta, \beta \) is given by Eq. ((12)). Also, the equation

\[ p(\sigma | \nu) = \left( \frac{\nu}{2} \right)^{\frac{\nu}{2}} \left[ \Gamma \left( \frac{\nu}{2} \right) \right]^{-1} \left( \prod_{i=1}^{r} \sigma_i \right)^{\frac{\nu-2}{2}} \exp \left[ -\frac{1}{2} \sum_{i=1}^{r} \frac{\sigma_i^2}{\nu} \right] \]

is the prior distribution of \( \sigma \) conditional on \( \nu \), and

\[ p(\nu) = \lambda \exp \left[ -\lambda (\nu - \delta) \right] I \{ \nu > \delta \} \]

is the prior distribution on degree of freedom \( \nu \) and \( \delta \geq 2 \). When \( \lambda = 100 \) and \( \delta = 500 \), we get the special case of Bayesian estimation with normal innovations. In this special case, the generated value of \( \nu \) are centered around \( \delta = 500 \). The joint prior distribution is given by

\[ p(\Theta | \sigma) = p(\zeta) p(\beta) p(\sigma | \nu) p(\nu) \]

For detailed discussion, see ref. [23]. To simulate Eq. ((15)), Markov chain Monte Carlo method is used. The Markov chain

\[ (\Theta^{[0]}, \sigma^{[0]}), (\Theta^{[1]}, \sigma^{[1]}), \ldots \]

is constructed in the parameter space such (15)
that in the limit \( j \to \infty \), \( (\Theta^{(j)}, \sigma^{(j)}) \) tends
to a distribution whose density is Eq. ((15)).

**INR/USD Exchange Rate**

In this work, we apply the Bayesian estimation to daily observation of Indian rupee vs US dollar exchange rate (see ref. [24], [25]). The time series plot of the daily observation of Indian rupee vs US dollar exchange rate is plotted in first row and first column of Fig. 1. The sample period is taken from the January 2, 2007 to October 31, 2013 (1717 trading days). Note that this sample period also includes the recession period of year 2008.

(Refer Figure 1 Here)

**Differencing of Time Series**

Differencing is done to convert the non-stationary time series to stationary time series [26]. Here the time series is first transformed by logging and then the differencing is performed. The log-differenced time series is also plotted in second row and first column of the Fig. 1. The period of high and low volatility, i.e., volatility clustering can be seen here.

**Autocorrelation Function**

For a weakly stationary stochastic process \( u_t \), the function

\[
\rho_u(k) = \frac{\text{Cov}(u_t, u_{t+k})}{\sigma^2}
\]

is defined as the autocorrelation function with time lag \( k \). Autocorrelation function (ACF) of the INR/USD time series is again plotted in first row and second column of the Fig. 1. Note that the ACF of the INR/USD exchange rate decreases slowly and hence it needs differencing. Second row and second column of Fig. 1 shows the ACF of difference of log of INR/USD exchange rate. From Fig. 1 of the ACF, we can see that there is no strong serial correlation in the log return of the series.

The R-code for Fig. 1 is given below:

```R
ind <- read.csv('fxdata.csv')
```
ML Estimation of GARCH(1,1) Model (with Normal Innovations) for the INR/USD Exchange Rate

Figure 1 confirms the possibility of an GARCH structure in INR/USD exchange rate. Let us find the GARCH(1,1) coefficients of INR/USD exchange return using ML estimation. For this purpose we have used the fGarch package in R [27]. The GARCH(1,1) coefficients of the INR/USD exchange rate are shown in Table 1. These GARCH coefficients are obtained with normal distributed error. The coefficients obtained here will be used in the Bayesian estimation of GARCH(1,1) parameters. Figure 2 shows the plot of conditional standard distribution, standardized residuals, ACF of the standardized residuals and QQ plot of standardized residuals with normal distributed error. Note that the ACF plot in...
the Fig. 2 indicates that the standardized residuals have no serial correlation and hence GARCH model is appropriate for INR/USD exchange rate. The standardized residual is defined as

\[ \hat{\varepsilon}_i = \frac{u_i}{\hat{\sigma}_i} \]

where \( \hat{\sigma}_i \) is the estimated conditional standard deviation. The QQ plot (see Fig. 2) of the standardized residuals indicates that the residuals are not exactly normally distributed (here QQ plots are used to check the normality of the residuals). The Shapiro-Wilk test and the Jarque-Bera test are also helpful for testing the normality of the innovations [28]. The unconditional variance of the log series is given by

\[ \sigma = \sqrt{\frac{\omega}{1 - \alpha - \beta}} = \sqrt{\frac{5.46 \times 10^{-7}}{1 - 0.0934 - 0.8967}} = 0.00743 \]

which is slightly larger than sample variance 0.006 of the data. Finally, the GARCH(1,1) model with normal innovations for the INR/USD exchange rate can be written as

\[ \sigma_{i+1}^2 = \omega + \alpha u_i^2 + \beta \sigma_i^2 \]

\[ = 5.46 \times 10^{-7} + 0.0934u_i^2 + 0.8967\sigma_i^2 \]

The R code for Fig. 2 is given below:

```r
require("fGarch")
fit <- garchFit( : garch(1,1), data=diflog.ind)
summary(fit)
par(mfrow=c(2,2))
plot(fit,which=c(9,10,13,2))
```

**ML Estimation of GARCH(1,1) Model (with Student t Innovations) for the INR/USD Exchange Rate**

Note that the skewness of differenced-log return is -0.189254 and so the log return of INR/USD exchange rate are negatively skewed. Hence to model this skewness, we employ the Student t distribution for the error. Table 2 shows the
GARCH(1,1) coefficients with Student’s $t$ distributed innovations. Figure 3 shows the plot of conditional standard distribution, standardized residuals, ACF of the standardized residuals and QQ plot of standardized residuals with Student’s $t$ distributed innovations.

The QQ plot of standardized residuals in the Fig. 3 indicates that the residuals are not normally distributed and hence it is skewed. Also, the conditional volatility plot of Figs. 2 and 3 are almost same. Thus the correlation between the two volatility models (normal and Student’s $t$ innovation) are close to one. From the plot of conditional SD of Figs. 2 and 3, one can see that the volatility was high around the time index 500 and 1600, that was the end of year 2008 and mid of year 2013 respectively. The R code for for Fig. 3 is given below:

```R
require("fGarch")
fit <- garchFit( : garch(1,1),
data=diflog.ind, cond.dist="std")
summary(fit)
par(mfrow=c(2,2))
plot(fit,which=c(9,10,13,2))
library(fBasics)
basicStats(diflog.ind)
```

(Refer Figure 4 Here)

(Refer Table 2 Here)

Figure 4 shows the plot of the log returns with 95% pointwise predictive intervals. The intervals are calculated by $\hat{\mu} \pm \hat{\sigma}$. Here $\hat{\mu} = -3.070038e-05$ is mean of the log return and it is a constant term. From Fig. 4, one can see that all returns are within the 95% predictive intervals. The R-code [3] for Fig. 4 is given below:

```R
par(mfcol=c(1,1))
plot(diflog.ind,,xlab='Index',ylab='series',type='l',ylim=c(-0.05,0.05))
v1=volatility(fit)
upp = -3.070038e-05+2*v1
low = -3.070038e-05-2*v1
```

par(mfcol=c(1,1))
plot(diflog.ind,,xlab='Index',ylab='series',type='l',ylim=c(-0.05,0.05))

$v1=\text{volatility}(\text{fit})$

$\text{upp} = -3.070038e-05+2*v1$

$\text{low} = -3.070038e-05-2*v1$
Bayesian Estimation

Let us apply the Bayesian Estimation (GARCH(1,1) model with Student’s $t$ innovation) to INR/USD daily exchange log returns [29]. The bayesGARCH package [29] is fully automatic. Here we have scaled the data so that the MCMC sampler does not get stuck at a given value. The R-code for Bayesian estimation is given below:

```r
require(bayesGARCH)
addPriorConditions <- function(psi)
z <- 100*diflog.ind
set.seed(1234)
MCMC <- bayesGARCH(z,
  control = list(l.chain = 10000, n.chain = 2,
  addPriorConditions = addPriorConditions))
gelman.diag(MCMC)
1 - rejectionRate(MCMC)
autocorr.diag(MCMC)
smpl <- formSmpl(MCMC, l.bi = 2500, batch.size = 2)
summary(smpl)
```

In the above R-code, l.chain is the length of each MCMC chain and n.chain is the number of MCMC chains. Here we run two MCMC chains for 10000 passes each. The function addPriorConditions has been used to put the constraint on the parameter. Here we put constraint on parameters $\alpha$ and $\beta$, i.e., $\alpha + \beta < 1$. The function gelman.diag(MCMC) is used to compute the convergence of the sampler, $1 - rejectionRate(MCMC)$ is used for the acceptance rates and autocorr.diag(MCMC) is used to compute the autocorrelations in the chain. The function formSmpl is used to discard the first 2500 draws from the MCMC output (also known as burn in period). In the above code, l.bi is length of the burn in phase. The
posterior statistics is obtained by using the function `summary`. Table 3 gives the values of parameter obtained by using the Bayesian estimation with Student $t$ innovations.

(Refer Table 3 Here)

The unconditional variance $\sigma$ of the scaled data using the Bayesian estimation with Student $t$ innovations is given by

$$\sigma = \sqrt{\frac{\omega}{1-\alpha - \beta}} = \sqrt{\frac{0.007798}{1-0.110215-0.877340}} = 0.791578$$

which is slightly larger than the actual value 0.6001465. For the scaled data, the ML estimate of the GARCH parameter with Student $t$ innovations are given in Table 4.

(Refer Table 4 Here)

The GARCH(1,1) coefficients obtained by ML and Bayesian estimate are close to each other. (see Tables 3, 4). The conditional kurtosis is given by (see Eq. (14))

$$3 \left( \frac{\nu-2}{\nu-4} \right)$$

provided $\nu > 4$. Finally, the Bayesian estimation with normal innovations is obtained by setting $\lambda = 100$ and $\delta = 500$. The length of each MCMC chain is taken as 10000. For detailed discussion, see ref [20]. The Bayesian estimation of GARCH(1,1) coefficients with normal innovations is shown in Table 5.

(Refer Table 5 Here)

The unconditional variance $\sigma$ of the scaled data using the Bayesian estimation with normal innovations is given by

$$\sigma = \sqrt{\frac{\omega}{1-\alpha - \beta}} = \sqrt{\frac{0.006517}{1-0.08999-0.8960}} = 0.682032$$

which is close to the actual value 0.6001465. The R-code for Table 5 is given below:

```R
require(bayesGARCH)
addPriorConditions <- function(psi)
```

z <- 100*diflog.ind

set.seed(1234)

MCMC <- bayesGARCH(z, lambda = 100, delta = 500, control
=list(n.chain = 2,
1.chain = 10000,
addPriorConditions = addPriorConditions))

smpl <- formSmpl(MCMC, l.bi = 2500, batch.size = 2)

summary(smpl)

Concluding Remarks

Estimation of volatility is one of the challenging problem in finance. In this work, we have applied the GARCH(1,1) model to the log return of INR/USD exchange rate. We have obtained the GARCH(1,1) coefficients of INR/USD exchange rate using the normal and Student t innovations. These two models fit the data well. Since the INR/USD log data is negatively skewed, one can see that the Student t innovations give a better fit for the INR/USD exchange rate. We have also obtained the GARCH(1,1) coefficients using the Bayesian estimation with normal and Student t innovations. Also, a comparison is made between the ML and Bayesian estimation with normal and Student t innovations.

The Garch(1,1) coefficients obtained using these two methods are close to each other (see Tables 3, 4). The value of $\alpha + \beta$ is close to one and hence the past variance and past shocks have longer impact on the future conditional variance. Thus the volatility clustering for the INR/USD exchange rate is properly explained by the GARCH(1,1) model. The conditional variance obtained here can be used for the calculation of value at risk (VaR) and for the asset pricing through the Black-Scholes formula.

Acknowledgements

We would like to thank Dyal Singh...
College, University of Delhi for providing the computational facility during the course of this work. We would also like to thank Sukanto Deb for helpful discussions.

References


[26] LyPham, https://talksonmarkets.files.wordpress.com/2012/09/time-series-analysis-with...
-arima-e  


[29] D. Ardia, ps://cran.r-project.org/web/packages/bayesGARCH/bayesGARCH.pdf)

http://CRAN.R-project.org/package=fGarch

List of Figures:
Figure 1: Time series plot of INR/USD exchange rate, ACF of the series, time series plot of log-differenced series, ACF of the log-differenced series, square of the differenced-log return and ACF of the square of the differenced-log return of INR/USD exchange rate for the period 2/1/2007 to 31/12/2013.
Figure 2: Plot of conditional standard deviation, standardized residuals, ACF of the standardized residuals and QQ plot of INR/USD exchange rate with normal error.
Figure 3: Plot of conditional standard deviation, standardized residuals, ACF of the standardized residuals and QQ plot of INR/USD exchange rate with Student’s $t$ distributed error. The QQ-plot shows a straight line pattern.
Figure 4: Time series plot of INR/USD exchange rate for the period 2/1/2007 to 31/12/2013. The dashed lines are pointwise 95% predictive interval with normal innovations.

List of Tables:

Table 1: GARCH(1,1) coefficients with normal distributed error.
### Table 2: GARCH(1,1) coefficients with Student’s $t$ distributed error.

| Coefficients | Estimate | Std. Error | $t$ Value | $Pr(>|t|)$ |
|--------------|----------|------------|-----------|------------|
| $\omega$     | 5.460e-07| 1.495e-07  | 3.653     | 0.000259 *** |
| $\alpha$     | 9.341e-02| 1.368e-02  | 6.826     | 8.75e-12 *** |
| $\beta$      | 8.967e-01| 1.359e-02  | 65.983    | $< 2e-16$ *** |

### Table 3: GARCH(1,1) coefficients using Bayesian estimation with Student’s $t$ distributed error.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Mean</th>
<th>SD</th>
<th>Naive SE</th>
<th>Time-Series SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega$</td>
<td>0.007798</td>
<td>0.00259</td>
<td>2.991e-05</td>
<td>0.000183</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.110215</td>
<td>0.02013</td>
<td>2.325e-04</td>
<td>0.001551</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.877340</td>
<td>0.02033</td>
<td>2.348e-04</td>
<td>0.001625</td>
</tr>
<tr>
<td>$\nu$</td>
<td>5.348165</td>
<td>0.71320</td>
<td>8.235e-03</td>
<td>0.057487</td>
</tr>
</tbody>
</table>

### Table 4: ML estimation of GARCH(1,1) coefficients with Student’s $t$ distributed error.

<p>| Coefficients | Estimate | Std. Error | $t$ Value | $Pr(&gt;|t|)$ |
|--------------|----------|------------|-----------|------------|
| $\omega$     | 0.004946 | 0.001999   | 2.474     | 0.0134 *   |</p>
<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Mean</th>
<th>SD</th>
<th>Naive SE</th>
<th>Time-Series SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega$</td>
<td>6.517e-03</td>
<td>0.001453</td>
<td>1.677e-05</td>
<td>5.713e-05</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>8.999e-02</td>
<td>0.013778</td>
<td>1.591e-04</td>
<td>6.882e-04</td>
</tr>
<tr>
<td>$\beta$</td>
<td>8.960e-01</td>
<td>0.013909</td>
<td>1.606e-04</td>
<td>7.744e-04</td>
</tr>
<tr>
<td>$\nu$</td>
<td>5.000e+02</td>
<td>0.010210</td>
<td>1.179e-04</td>
<td>1.179e-04</td>
</tr>
</tbody>
</table>

Table 5: GARCH(1,1) coefficients using Bayesian estimation with normal innovations.