

Univariate volatility models

by Kevin Kotzé

There are many reasons for wanting to analyse the risk that is associated with an asset, where the use of volatility is perhaps the most commonly used measure of risk. The objective of this section is to consider some of the characteristics of models that may be used to describe the evolution of volatility. In addition, we will also discuss some of the applications of such models.

As many of you would be probably know, measures of volatility are used in many important financial and economic models. It is a critical metric in the *value-at-risk* calculation and is one of the key variables in options pricing and asset allocation models. For example, the well known Black-Scholes method for valuing the price of a call option is largely dependent on the measure of volatility. Therefore, if we are able to derive a more accurate measure of volatility then we could possibly identify miss-priced derivatives. In addition, portfolio allocation measures such as the Markowitz mean-variance framework also depend on volatility, so a better measure of this statistic would lead to a more efficient allocation of capital. More recently, some volatility indices have also been introduced on many of the financial markets. For example, the VIX index of the Chicago Board Options Exchange (CBOE) started to trade in futures on 26 March 2004. Similarly the South African Volatility Index (SAVI) was launched in 2007, as a measure of the implied volatility in the FTSE/JSE Top40 index.

Although we are able to define various aspects or types of volatility, this characteristic is not directly observable in practice. What we observe are the respective prices of assets. To estimate the volatility that is associated with this variable we would either need to transform this data, or derive a suitable model. As we will see, this feature of the data has several important implications for this particular research area. In addition, the choice of model or transformation technique is also of importance when trying to describe the degree of volatility in various types of time series data.

There are many different types of univariate volatility models that are described in what is an extremely broad literature. These include the autoregressive conditional heteroscedastic (ARCH) model of Engle (1982), the generalised autoregressive conditional heteroscedastic (GARCH) model of Bollerslev (1986), the exponential generalised autoregressive conditional heteroscedastic (EGARCH) model of Nelson (1991), the threshold generalised autoregressive conditional heteroscedastic (TGARCH) model of Glosten, Jagannathan, and Runkle (1993) and Zakoian (1994), the non-symmetric generalised autoregressive conditional heteroscedastic (NGARCH) model of Engle and Ng (1993) and Duan (1995), and the stochastic volatility (SV) models of Melino and Turnbull (1990), Taylor (1994), Harvey, Ruiz, and Shephard (1994), and Jacquier, Polson, and Rossi (1994).

In what follows, we discuss the main advantages and weaknesses of these volatility models. In addition, we will also consider their applicability to a number of potentially interesting investigations.

1 Characteristics of volatility

Although volatility is not directly observable, it has a number of characteristics that are present in many time series variables. Firstly, volatility is usually high for certain periods of time and low for other periods. This gives rise to volatility clusters. Secondly, volatility evolves over time in a continuous manner, where volatility jumps are rare. Thirdly, volatility does not diverge to infinity, as

we note that it varies within some fixed range. This may provide an indication that volatility may be stationary. Fourthly, volatility seems to react differently to a big price increase and a big price drop with the latter having a greater impact. This phenomenon may be referred to as the asymmetric effect that is due to leverage. These properties play an important role in the development of volatility models. For example, the EGARCH and TGARCH models were developed to capture the asymmetric effects in volatility that is induced by the differences in positive and negative price movements.

In financial applications, we typically estimate the volatility of an asset using the price of the security, its derivatives, or a combination of these two measures. To measure the daily volatility of a particular share that is quoted on a financial exchange we observe (i) the daily return for each trading day, (ii) tick-by-tick data for intra-day transactions and quotes, and (iii) the prices of options contingent on this particular share price. These three data sources give rise to three types of volatility measures for this share. They are as follows:

- **Traditional volatility:** Derived from the conditional standard deviation of daily returns. This is the usual definition of volatility and is the focus of volatility models that we will discuss in what follows.
- **Implied volatility:** Using prices from options markets, one can use a pricing formula, for example, the Black-Scholes pricing formula, to deduce the volatility of the share price. This volatility measure is called the *implied volatility*. Since implied volatility is derived under certain assumptions that relate the price of an option to that of the underlying share price, it is often criticized as being model dependent. Several studies have shown that implied volatility of an asset return tends to be larger than that obtained by using daily returns and a volatility model. This may be due to the risk-premium for volatility on options markets or it may be caused by the way that daily returns are calculated. The VIX on the CBOE and the SAVI on the FTSE/JSE are examples of implied volatility measures.
- **Realised volatility:** With the availability of high frequency financial data, one can use intra-day returns, for example, 5-min returns, to estimate the daily volatility. This volatility measure is called *realised volatility*, which has been a subject of intensive research over the last few years.

In most practical applications, the time interval used to measure volatility is a year. Thus, volatility is often annualised. If daily returns are used to estimate volatility, one can obtain the annualised volatility by multiplying daily volatility by $\sqrt{252}$, where it is assumed that there are typically 252 trading days on most financial markets.

2 Structure of a model

Let y_t be the observed value of a variable, that is assumed to be stationary. This could represent the natural logarithm of the return of a security, at a time period, t . The basic idea behind a study of volatility is that while the first moment of the series is either serially uncorrelated or has a minor lower order serial correlation, the second order moment of the variable displays some form of serial dependence.

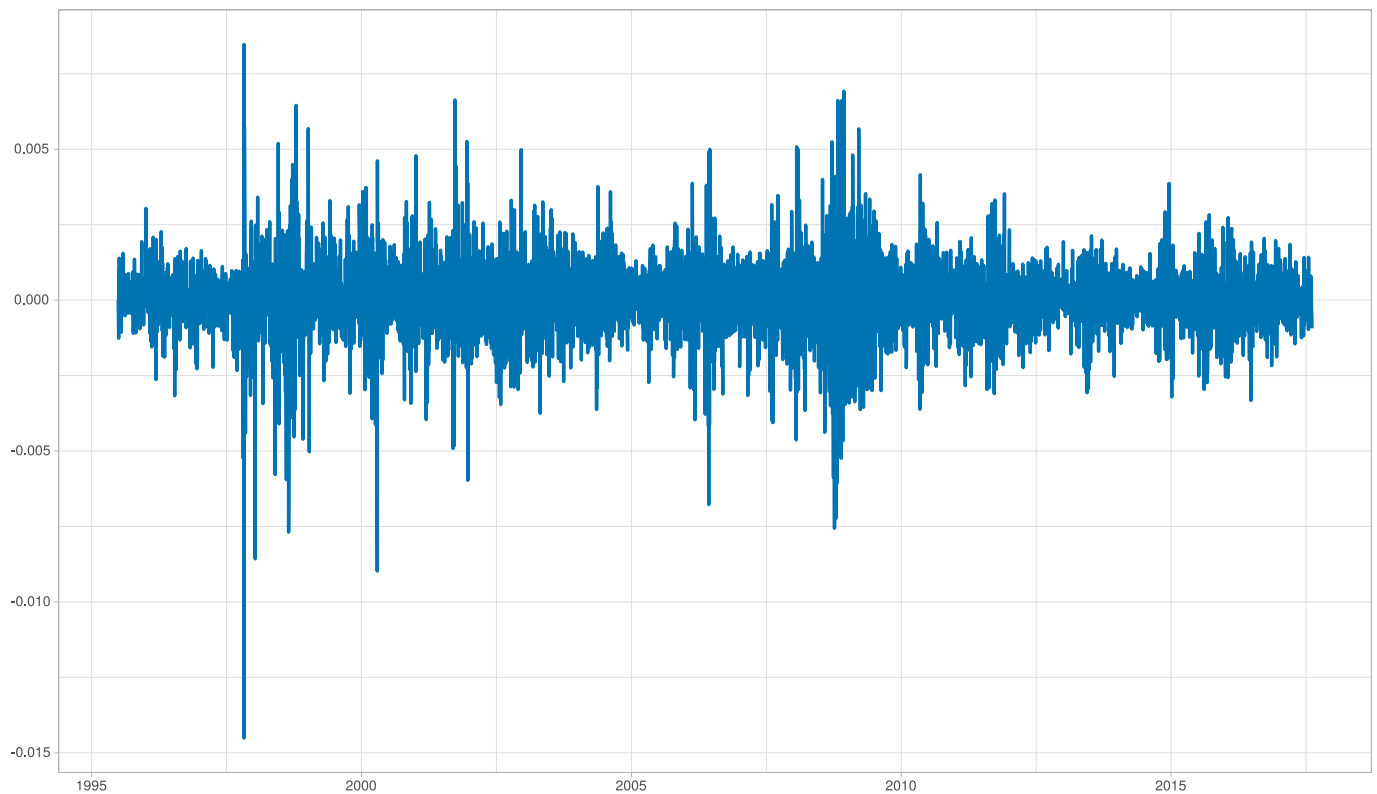


Figure 1: **FTSE/JSE All Share - daily returns**

For illustrative purposes, consider the daily logarithmic returns for the FTSE/JSE All Share, which is displayed in Figure 1. This time series appears to be stationary and random, and one could use a sample ACF to confirm this feature of the data. Indeed, the Ljung-Box statistic suggests that $Q(12) = 18.68$ with a p value of 0.10 for y_t . However, the sample ACF of the absolute value of the natural logarithm of returns (i.e. $|y_t|$) suggests that this transformation of the variable is serially correlated. The Ljung-Box statistics of $|y_t|$ suggest that $Q(12) = 124.91$ with a p value close to zero. Consequently, the logarithmic returns of the FTSE/JSE All Share are serially uncorrelated, but dependent through the second moment. This feature of the data may usually be described by a volatility model.

To put these volatility models in proper perspective, it is useful to consider the conditional mean and variance of y_t given the information set, I_{t-1} . We could then summarise the first two moments as,

$$\begin{aligned}\mu_t &= E[y_t | I_{t-1}] \\ \sigma_t^2 &= \text{var}[y_t | I_{t-1}] = E[(y_t - \mu_t)^2 | I_{t-1}]\end{aligned}\tag{2.1}$$

where μ_t is the expected mean and we use the information set available at time $t - 1$. Typically, I_{t-1} consists of all linear functions of the past observed values of this variable. When seeking to model the volatility of a financial security, we note that in most instances, the serial dependence in the first moment of the data is usually relatively weak, if there is any evidence of such serial correlation at all. Therefore, the equation for μ_t in equation (2.1) would usually make use of a relatively parsimonious form. For example, consider the logarithmic returns of the FTSE/JSE All Share. In this case the Ljung-Box statistics suggest that the variable is not serially correlated, however, the mean of y_t is significantly different from zero. More specifically, the t-ratio of testing $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$ is 2.38 with a p value of 0.018. Therefore, when seeking to model the first-moment of the FTSE/JSE All Share returns, we have $y_t = \mu_t + a_t$, where $\mu_t = \mu$, since we are going to assume that the process has a constant mean.

In general, when seeking to model other data, we may find that it would be more appropriate to describe the first moment of y_t with the aid of an ARMA(p, q) model so that $y_t = \mu_t + a_t$, where μ_t is given by

$$\mu_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} - \sum_{j=1}^q \theta_j a_{t-j} \quad (2.2)$$

When the use of explanatory variables may provide a more accurate description of the first-moment of y_t , then μ_t would take the following functional form:

$$\mu_t = \phi_0 + \sum_{i=1}^k \beta_i x_{i,t-1} + \sum_{i=1}^p \phi_i z_{t-i} - \sum_{j=1}^q \theta_j a_{t-j} \quad (2.3)$$

where $z_{t-i} = y_{t-i} - \phi_0 - \sum_{i=1}^k \beta_i x_{i,t-i-1}$ denotes the adjusted return series after removing the effect of explanatory variables, and $x_{i,t-j}$ are explanatory variables available at time $t - j$.

The values for p and q in the ARMA(p, q) model may depend on the frequency of the time series data. For example, daily returns of a market index may display a minor amount of serial correlation, but monthly returns of the index may not contain any significant serial correlation. Note also that the functional form of the explanatory variables, $x_{i,t-j}$, in equation (2.3) could be relatively flexible. For example, a dummy variable could be used for Mondays when using daily data to study the effect of the previous weekend on stock returns. In the capital asset pricing model (CAPM), the mean equation of y_t can be written as $y_t = \phi_0 + \beta r_{m,t} + a_t$, where $r_{m,t}$ denotes the market return.

Note that since $\sigma_t^2 = \text{var}[y_t | I_{t-1}] = E[(y_t - \mu_t)^2 | I_{t-1}]$, and since in the case of the FTSE/JSE All Share returns, we have $y_t = \mu_t + a_t$, then by combining these equations, we have

$$\sigma_t^2 = \text{var}(y_t | I_{t-1}) = \text{var}(a_t | I_{t-1}) \quad (2.4)$$

This functional form will be used for all of the conditional heteroscedastic models that are described below, as they are all concerned with the evolution of σ_t^2 . The manner under which σ_t^2 evolves over time is what distinguishes one volatility model from another.

Conditional heteroscedastic models can be classified into two general categories. Those in the first category use an exact function to govern the evolution of σ_t^2 , whereas those in the second category use a stochastic equation to describe σ_t^2 . The ARCH/GARCH models belongs to the first category, while the SV model belongs in the second category.

In what follows, a_t is referred to as the shock or innovation to variable y_t at time t . The model for μ_t in equation (2.3) is referred to as the mean equation for y_t and the model for σ_t^2 is the volatility equation for y_t . Therefore, modelling conditional heteroscedasticity amounts to augmenting a traditional time series model with a dynamic equation that governs the time evolution of the conditional variance of the variable.

2.1 Constructing a volatility model

The construction of a traditional volatility model consists of the following four steps:

1. Specify a mean equation after testing for serial dependence in the data. If necessary, make use of an econometric model (e.g. an ARMA model) for the time series variable to remove any linear dependence.

2. Make use of the residuals from the mean equation to test for ARCH effects.
3. Specify a volatility equation. If ARCH effects are statistically significant then perform a joint estimation of the mean and volatility equations.
4. Check the fitted model carefully and refine it where necessary.

In what follows, we are going to focus our attention on the first two steps of the modelling procedure. In addition, we will also introduce various extensions for a number of popular volatility models.

2.1.1 Specifying the Mean Equation

As we have already noted, for most variables that measure the daily returns for a particular asset, the serial correlation in the mean of the process is usually weak and may not be significantly different from zero. Therefore, the construction of a mean equation usually amounts to removing the sample mean from the data, if the sample mean is found to be significantly different from zero. For other time series variables, a simple autoregressive model might be required to remove the serial correlation in the first moment of the process. In some cases, the mean equation may also incorporate an additional explanatory variable, such as an dummy variable for weekend or January effects.

2.1.2 Testing for ARCH effects

For ease of notation, we assume that a_t , which is equal to $y_t - \mu_t$, represents the residuals from the mean equation. The squared series a_t^2 is then used to check for conditional heteroscedasticity. This procedure provides an indication of possible ARCH effects. If we suspect that there are possible ARCH effects, we could then make use of two formal tests. The first test, which is discussed in McLeod and Li (1983), applies the usual Ljung-Box statistics $Q(m)$ to the a_t^2 series. The null hypothesis of the test statistic is that the first m lags of the ACF for the a_t^2 time series are zero. The second test for conditional heteroscedasticity makes use of a Lagrange multiplier and is discussed in Engle (1982). This test is equivalent to the usual F-statistic for testing $\alpha_i = 0$, ($i = 1, \dots, m$) in the linear regression,

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_m a_{t-m}^2 + v_t, \quad t = m+1, \dots, T$$

where v_t denotes the error term, m is a pre-specified positive integer, and T is the sample size. Specifically, the null hypothesis is $H_0 : \alpha_1 = \dots = \alpha_m = 0$ and the alternative hypothesis is $H_1 : \alpha_i \neq 0$ for some i between 1 and m .

Now if we calculate the sum of square residuals under the null and alternative hypotheses, such that $SSR_0 = \sum_{t=m+1}^T (a_t^2 - \bar{\omega})^2$, where $\bar{\omega} = (1/T) \sum_{t=1}^T a_t^2$ is the sample mean of a_t^2 , and $SSR_1 = \sum_{t=m+1}^T \hat{v}_t^2$, where \hat{v}_t is the least squares residual of the prior linear regression. Then we can calculate an appropriate value for the F-statistic, as follows,

$$F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T - 2m - 1)}$$

which follows an F-distribution with degrees of freedom that are equal to m and $T - 2m - 1$ under H_0 . When T is sufficiently large, one can use mF as the test statistic, which has an asymptotic chi-squared distribution with m degrees of freedom under the null hypothesis. The decision rule is to reject the null hypothesis if $mF > \chi_m^2(\alpha)$, where $\chi_m^2(\alpha)$ is the upper $100(1 - \alpha)$ th percentile of χ_m^2 , or the p value of mF is less than α , for a Type I error.

3 Autoregressive Conditional Heteroscedastic Model

The first model that we will consider was introduced in Engle (1982). It provides an intuitive and systematic framework for the modelling of volatility and is commonly termed the ARCH model. The basic idea of ARCH models is that (i) the shock to the time series is serially uncorrelated, but dependent, and (ii) the dependence of a_t can be described by a simple quadratic function of its lagged values. Specifically, an ARCH(m) model assumes that,

$$\begin{aligned} a_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_m a_{t-m}^2 \end{aligned} \quad (3.1)$$

where ε_t is an independent and identically distributed random variable with an assumed mean of 0 and variance of 1. All of the coefficients in the volatility equation would then take on positive values, such that $\alpha_0 > 0$, and $\alpha_i \geq 0$ for $i > 0$. The coefficients α_i must satisfy some regularity conditions to ensure that the unconditional variance of a_t is finite. In practice, ε_t is often assumed to follow either of the standard normal distribution, standardised Student t-distribution, or a generalised error distribution (GED). In some applications, a skew-distribution is also used for ε_t .

Given this structure for the model, it is evident that large values for past squared shocks, $\{a_{t-i}^2\}_{i=1}^m$ would provide a large conditional variance σ_t^2 for the innovation a_t . This means that, under the ARCH framework, large shocks tend to be followed by other large shocks. Hence, the probability of obtaining a large variance is greater than obtaining a smaller variance, when the immediate past includes relatively large measures for the variance. This feature may induce the volatility clusterings that is observed for many financial asset returns.

Remark. Some authors follow the original notation of Engle (1982) and use h_t to denote the conditional variance in equation (3.1). In this case, the shock becomes $a_t = \sqrt{h_t} \varepsilon_t$.

3.1 Properties of ARCH Models

To consider the properties of ARCH models, it is useful to study the ARCH(1) model in some detail. This model may be written as

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2,$$

where $\alpha_0 > 0$ and $\alpha_1 \geq 0$. Note that the unconditional mean of a_t remains zero, since

$$E[a_t] = E\{E[a_t | I_{t-1}]\} = E\{\sigma_t E[\varepsilon_t]\} = 0.$$

Secondly, the unconditional variance of a_t can then be derived as

$$\begin{aligned} \text{var}[a_t] &= E[a_t^2] = E\{E[a_t^2 | I_{t-1}]\} \\ &= E[\alpha_0 + \alpha_1 a_{t-1}^2] = \alpha_0 + \alpha_1 E[a_{t-1}^2] \end{aligned}$$

Note that a_t is a stationary process, with $E[a_t] = 0$, $\text{var}[a_t] = \text{var}[a_{t-1}] = E[a_{t-1}^2]$. Therefore, we have $\text{var}[a_t] = \alpha_0 + \alpha_1 \text{var}[a_t]$ and $\text{var}[a_t] = \alpha_0 / (1 - \alpha_1)$. Now since the variance of a_t must be positive, we would need to satisfy the condition that $0 \leq \alpha_1 < 1$.

Thirdly, in some applications we may need to make use of higher order moments of a_t . Therefore, we can show that α_1 must also satisfy some additional constraints. For instance, to study the tails of the distribution for volatility, we require that the fourth moment of a_t must be finite. Under the normality assumption of ε_t in equation (3.1), we have

$$E[a_t^4 | I_{t-1}] = 3 \{E[a_t^2 | I_{t-1}]\}^2 = 3(\alpha_0 + \alpha_1 a_{t-1}^2)^2$$

Therefore,

$$\begin{aligned} E[a_t^4] &= E\{E[a_t^4 | I_{t-1}]\} = 3E(\alpha_0 + \alpha_1 a_{t-1}^2)^2 \\ &= 3E(\alpha_0^2 + 2\alpha_0\alpha_1 a_{t-1}^2 + \alpha_1^2 a_{t-1}^4) \end{aligned}$$

If a_t is fourth-order stationary with $m_4 = E[a_t^4]$, then we have

$$\begin{aligned} m_4 &= 3(\alpha_0^2 + 2\alpha_0\alpha_1 \text{var}[a_t] + \alpha_1^2 m_4) \\ &= 3\alpha_0^2 \left(1 + 2\frac{\alpha_1}{1 - \alpha_1}\right) + 3\alpha_1^2 m_4 \end{aligned}$$

Consequently,

$$m_4 = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}$$

This result has two important implications: (i) as the fourth moment of a_t is positive, we see that α_1 must also satisfy the condition $1 - 3\alpha_1^2 > 0$, where $0 \leq \alpha_1^2 < \frac{1}{3}$; and (ii) the unconditional kurtosis of a_t is

$$\frac{E[a_t^4]}{(\text{var}[a_t])^2} = 3 \frac{\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)} \times \frac{(1 - \alpha_1)^2}{\alpha_0^2} = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3$$

Therefore, the excess kurtosis of a_t is positive so the tail of the distribution of a_t is heavier than that of a normal distribution. In other words, it is more likely that the shock a_t , in a conditional Gaussian ARCH(1) model, will produce *outliers*, when compared to the case of a traditional Gaussian white noise process. This would motivate for the use of ARCH models for asset returns, since *outliers* appear more often in such data, than in an i. i. d. sequence that is produced by a random normal variable.

These properties continue to hold for general ARCH models, but the formulas become more complicated as the order for the ARCH model increases. For example, when working with higher-order models, the condition $\alpha_i \geq 0$ in equation (3.1) can be relaxed. However, we would still need to ensure that the conditional variance σ_t^2 is positive for all t . As an alternative, a simple way to achieve positive conditional variance is to rewrite an ARCH(m) model as

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + A'_{m,t-1} \Omega A_{m,t-1} \quad (3.2)$$

where $A_{m,t-1} = (a_{t-1}, \dots, a_{t-m})'$ and Ω is an $m \times m$ non-negative definite matrix. The ARCH(m) model in equation (3.1) would then require that Ω is diagonal. Therefore, we could show that the representation of Engle (1982) makes use of a parsimonious approach to approximate a quadratic function.¹

3.2 Advantages and Weaknesses of ARCH Models

The key advantages of making use of an ARCH model to describe the volatility in a particular variable include the following:

1. The model can produce volatility clusters
2. The shocks in the model that are summarised by a_t have heavy tails

Both of these characteristics are consistent with the stylised features of the data that is provided by the returns for most financial assets. However, these models also have a number of weaknesses, which include:

1. The model assumes that positive and negative shocks have the same effects on volatility, since it depends on the square of the previous shocks. In most cases we can show that the price of a financial asset responds differently to positive and negative shocks.
2. The ARCH model is rather restrictive. For instance, the α_1^2 parameter in an ARCH(1) model must be within the interval $[0, \frac{1}{3}]$, if the measure of volatility has a finite fourth moment. This constraint becomes complicated for higher order ARCH models and would limit their ability to capture excess kurtosis (particularly where we assume Gaussian innovations).
3. The ARCH model does not provide any new insight into our understanding of the source of variations in the time series variable. It merely provides a mechanical way of describing the behaviour of the conditional variance.
4. ARCH models are likely to over-predict the volatility because they respond slowly to large isolated shocks in the time series.

3.3 Making use of an ARCH Model

As may be gathered from the above discussion, it is relatively straightforward to specify an ARCH model. When doing so, it is worth taking note of the following details:

Order Determination. If an ARCH effect is found to be significant, one can use the PACF of a_t^2 to determine the ARCH order. Using PACF of a_t^2 to select the ARCH order can be justified as follows. From the model in equation (3.1), we have

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_m a_{t-m}^2$$

For a given sample, a_t^2 is an unbiased estimate of σ_t^2 . Therefore, we expect that a_t^2 is linearly related to $a_{t-1}^2, \dots, a_{t-m}^2$ in a manner similar to that of an autoregressive model of order m . Note that a single a_t^2 is generally not an efficient estimate of σ_t^2 , but it can serve as an approximation that could be informative in specifying the order m .

Alternatively, define $\eta_t = a_t^2 - \sigma_t^2$. It can be shown that η_t is an uncorrelated series with mean 0. The ARCH model then becomes

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_m a_{t-m}^2 + \eta_t$$

which is in the form of an AR(m) model for a_t^2 , except that η_t is not an i.i.d. series. As noted previously, the PACF of a_t^2 is a useful tool to determine the order m . Since η_t are not identically distributed, the least squares estimates of the prior model are consistent, but not efficient. Hence, the PACF of a_t^2 may not be effective when the sample size is small.

Estimation. Several likelihood functions are commonly used in ARCH estimation, depending on the distributional assumption of ε_t . When dealing with more complicated models it is worth noting that: starting values, choice of optimisation algorithm, use of analytic or numerical derivatives, and convergence criteria all influence the resulting parameter estimates.²

Model Checking. For a properly specified ARCH model, the standardised residuals are given by

$$\tilde{a}_t = \frac{a_t}{\sigma_t}$$

which approximate a sequence of i. i. d. random variables. Therefore, one can check the adequacy of a fitted ARCH model by examining the time series \tilde{a}_t . In particular, the Ljung-Box statistics of \tilde{a}_t can be used to check the adequacy of the mean equation, while the test on \tilde{a}_t^2 can be used to test the validity of the volatility equation. The skewness, kurtosis, and quantile-to-quantile plot (i.e. QQ plot) of \tilde{a}_t could then be used to check the validity of the assumption relating to the distribution of the residual.

Forecasting. Forecasts from the ARCH model in equation (3.1) can be obtained recursively, as was the case with an AR model. Consider an ARCH(m) model at point t , where the 1-step ahead forecast of $\sigma_{h=1}^2$ is

$$\sigma_{h=1}^2 = \alpha_0 + \alpha_1 a_t^2 + \dots + \alpha_m a_{t+1-m}^2$$

The 2-step ahead forecast is then

$$\sigma_{h=2}^2 = \alpha_0 + \alpha_1 \sigma_{h=1}^2 + \alpha_2 a_t^2 + \dots + \alpha_m a_{t+2-m}^2$$

and the H-step ahead forecast for $\sigma_{h=H}^2$ is

$$\sigma_{h=H}^2 = \alpha_0 + \sum_{i=1}^m \alpha_i \sigma_{h=H-i}^2 \quad (3.3)$$

where $\sigma_{h=H-i}^2 = a_{h+H-i}^2$ if $H-i \leq 0$.

4 The GARCH model

Although the ARCH model has a simple functional form, many parameters estimates may be required to adequately describe the volatility process of a financial asset, as m would need to take on a reasonably large value. To simplify the model, Bollerslev (1986) employs a useful extension that lead to the development of the generalised ARCH (GARCH) model. To describe the features of this

model, we assume that the mean equation for the y_t time series variable, may be described by $a_t = y_t - \mu_t$, once again. This would allow for the basic formulation of the GARCH(m, s) model, that may take the form,

$$\begin{aligned} a_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^m \alpha_i a_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 \end{aligned} \quad (4.1)$$

where ε_t is a sequence of i. i. d. random variables with an assume mean of 0 and a variance of 1, while $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and $\sum_{i,j=1}^{m,s} (\alpha_i + \beta_j) < 1$. Here, we assume that $\alpha_i = 0$ for $i > m$ and $\beta_j = 0$ for $j > s$. The latter constraint on $\alpha_i + \beta_j$ implies that the unconditional variance of a_t is finite, whereas its conditional variance σ_t^2 evolves over time. As before, we would usually assume that ε_t may follow either a standard normal distribution, standardised Student t-distribution, GED, etc. Equation (4.1) would then reduce to a pure ARCH(m) model if $s = 0$. The α_i and β_j are referred to as ARCH and GARCH parameters, respectively.

To understand the properties of GARCH models, we could make use of the following representation. Let $\eta_t = a_t^2 - \sigma_t^2$ so that $\sigma_t^2 = a_t^2 - \eta_t$ and substituting $\sigma_{t-i}^2 = a_{t-i}^2 - \eta_{t-i}$ ($i = 0, \dots, s$) into equation (4.1), we can rewrite the GARCH model as

$$a_t^2 = \alpha_0 + \sum_{i,j=1}^{m,s} (\alpha_i + \beta_j) a_{t-i}^2 + \eta_t - \sum_{j=1}^s \beta_j \eta_{t-j} \quad (4.2)$$

It is easy to verify that η_t is a martingale difference series (i.e. $E[\eta_t] = 0$ and $\text{cov}(\eta_t, \eta_{t-j}) = 0$ for $j \geq 1$). However, η_t in general is not an i. i. d. sequence. Equation (4.2) would effectively take on an ARMA form for the squared series a_t^2 . Therefore, a GARCH model could be regarded as an application of the ARMA idea to the squared series a_t^2 . Using the unconditional mean of an ARMA model, we have

$$E[a_t^2] = \frac{\alpha_0}{1 - \sum_{i,j=1}^{m,s} (\alpha_i + \beta_j)}$$

provided that the denominator of the prior fraction is positive.

The strengths and weaknesses of GARCH models may be considered by focusing on the simplest GARCH(1, 1) framework for the model, where

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad 0 \leq \alpha_1, \quad \beta_1 \leq 1, \quad (\alpha_1 + \beta_1) < 1. \quad (4.3)$$

Note firstly that large values for a_{t-1}^2 or σ_{t-1}^2 would give rise to a large values for subsequent σ_t^2 . This implies that a large a_{t-1}^2 tends to be followed by another large a_t^2 , which would generate the well-known behaviour of volatility clustering. Secondly, it can be shown that if $1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$, then

$$\frac{E(a_t^4)}{(E[a_t^2])^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3$$

Consequently, and similar to the case of ARCH models, the tail of the distribution of a GARCH(1, 1) process is heavier than that of a normal distribution. Thirdly, this model provides a simple parametric function that is more parsimonious than an equivalent ARCH(m) model that may be used

to describe the evolution of volatility in a variable.

Forecasts from a GARCH model can be obtained using methods similar to those of an ARMA or ARCH(m) model. Consider the GARCH(1, 1) model in equation (4.3) and assume that the forecast origin is t . For a 1-step ahead forecast, we have

$$\sigma_{h=1}^2 = \alpha_0 + \alpha_1 a_t^2 + \beta_1 \sigma_t^2$$

where a_t and σ_t^2 are known at the time index t .

As $E(\varepsilon_{h+1}^2 - 1 | I_h) = 0$, the 2-step ahead volatility forecast at the forecast origin t could be derived from the equation

$$\sigma_{h=2}^2 = \alpha_0 + (\alpha_1 + \beta_1) \sigma_{h=1}^2$$

In general, we have

$$\sigma_{h=H}^2 = \alpha_0 + (\alpha_1 + \beta_1) \sigma_{h=H-1}^2, \quad H > 1 \quad (4.4)$$

This result is analogous to that of an ARMA(1, 1) model. By repeated substitutions in equation (4.4), we would obtain the H -step ahead forecast, which can be written as

$$\sigma_{h=H}^2 = \frac{\alpha_0 [1 - (\alpha_1 + \beta_1)^{H-1}]}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{H-1} \sigma_{h=1}^2.$$

Therefore,

$$\sigma_{h=H}^2 \rightarrow \frac{\alpha_0}{1 - \alpha_1 - \beta_1}, \quad \text{as } H \rightarrow \infty$$

provided that $\alpha_1 + \beta_1 < 1$. Consequently, the multi-step ahead volatility forecasts of a GARCH(1, 1) model would converge on the unconditional variance of a_t , as the forecast horizon increases to infinity, provided that $\text{var}[a_t]$ exists.

The literature on GARCH models is extremely large. For example, see Bollerslev, Chou, and Kroner (1992), Bollerslev, Engle, and Nelson (1994) and the references therein. In most cases, the basic model encounters the same weaknesses as the ARCH model. For instance, it responds equally to positive and negative shocks. In addition, recent empirical studies that make use of high frequency financial time series data suggest that the behaviour of the tail in most GARCH models are nevertheless too short, even with standardised Student t innovations.³

4.1 Forecasting evaluation

As the volatility of a time series variable is not directly observable, comparing the forecasting performance of different volatility models can be problematic. Some researchers use out-of-sample forecasts and compare the volatility forecasts, $\sigma_{h=H}^2$, with the shock, $a_{h=H}^2$, to assess the forecasting performance of a volatility model. This approach often finds a low correlation coefficient between $a_{h=H}^2$ and $\sigma_{h=H}^2$, which would equate to a low R^2 when regressing these estimates on one another. However, such a finding is not surprising because $a_{h=H}^2$ alone is not an adequate measure of the volatility at time $t + H$.

For example, consider the 1-step ahead forecasts. From a statistical point of view, $E[a_{t+1}^2 | I_h] = \sigma_{h+1}^2$ so that a_{t+1}^2 is a consistent estimate of σ_{h+1}^2 . However, this measure is not an accurate estimate of σ_{h+1}^2 , since a single observation of a random variable with a known mean value cannot provide an accurate estimate of its variance. Consequently, such an approach to evaluate the forecasting performance of volatility models is strictly speaking not proper. For more information concerning forecasting evaluation of GARCH models, the interested reader is referred to Andersen and Bollerslev (1998).

4.2 Two-pass estimation method

On the basis of equation (4.2), a two-pass estimation method can be used to estimate the parameters in GARCH models. Firstly, if we ignore any volatility effects, we could estimate the mean equation of a variable with the aid of an ARMA representation and/or additional explanatory variables. This would provide a vector of residuals, which we could denote a_t . Secondly, by treating a_t^2 as an observed time series, we could estimate the parameters in equation (4.2). In this case we could denote the autoregressive and moving average coefficient estimates by $\hat{\phi}_i$ and $\hat{\theta}_i$. The GARCH estimates are then obtained as $\hat{\beta}_i = \hat{\theta}_i$ and $\hat{\alpha}_i = \hat{\phi}_i - \hat{\theta}_i$. Obviously, such estimates are approximations of the true parameter values and their statistical properties would need to be rigorously investigated. However, limited experience shows that this simple approach often provides good approximations, especially when the sample size is moderate or large.

5 Extensions to the GARCH framework

5.1 The integrated GARCH model

When the AR polynomial of the GARCH representation in equation (4.2) has a unit root, then we have an integrated-GARCH (IGARCH) model. Thus, IGARCH models take the form of unit-root GARCH models. Similar to ARIMA models, a key feature of IGARCH models is that the impact of past squared shocks $\eta_{t-i} = a_{t-i}^2 - \sigma_{t-i}^2$ for $i > 0$ on a_t^2 does not dissipate.

An IGARCH(1, 1) model could then be written as

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + (1 - \beta_1) a_{t-1}^2$$

where ε_t is defined as before and $1 > \beta_1 > 0$.

Hence, the IGARCH(1, 1) model is equivalent to a GARCH(1, 1) when $\alpha_1 + \beta_1 = 1$. Repeated substitutions in equation (4.4) give

$$\sigma_{h=H}^2 = \sigma_{h=1}^2 + (H - 1)\alpha_0, \quad H \geq 1, \quad (5.1)$$

where t is the forecast origin. Consequently, the effect of $\sigma_{h=1}^2$ on future volatilities is also persistent, and the volatility forecasts form a straight line with slope α_0 . Nelson (1990) studied the probabilistic properties of the volatility process σ_t^2 when described by an IGARCH model. He found that the process σ_t^2 is a martingale sequence for which some nice results are available in the literature. Under certain conditions, the volatility process is strictly stationary, but not weakly stationary because it does not have the first two moments.

The case of $\alpha_0 = 0$ is of particular interest, when studying the IGARCH(1, 1) model. In this instance, the volatility forecasts are simply $\sigma_{h=1}^2$ for all forecast horizons; see equation (5.1). This special IGARCH(1, 1) model is the volatility model that is used in the popular RiskMetrics model, which is

an approach that is used to calculate value-at-risk.⁴ The model also takes on the characteristics of an exponential smoothing model for the a_t^2 series. To see this, we can rewrite the model as follows:

$$\begin{aligned}\sigma_t^2 &= (1 - \beta_1)a_{t-1}^2 + \beta_1\sigma_{t-1}^2 \\ &= (1 - \beta_1)a_{t-1}^2 + \beta_1[(1 - \beta_1)a_{t-2}^2 + \beta_1\sigma_{t-2}^2] \\ &= (1 - \beta_1)a_{t-1}^2 + (1 - \beta_1)\beta_1a_{t-2}^2 + \beta_1^2\sigma_{t-2}^2\end{aligned}$$

By repeated substitutions, we have

$$\sigma_t^2 = (1 - \beta_1)[a_{t-1}^2 + \beta_1a_{t-2}^2 + \beta_1^2a_{t-3}^2 + \dots]$$

which takes the well-known exponential smoothing form, where β_1 may describe the discount factor. Therefore, exponential smoothing methods may be used to estimate these IGARCH(1, 1) models.

5.2 The GARCH-M model

In several financial applications, the return of an asset may depend on its volatility. To model such a phenomenon, we could make use of the popular GARCH-M model, where the “M” stands for GARCH-in-mean. A simple example of a GARCH-M(1, 1) model could be expressed with the aid of the following three equations:

$$\begin{aligned}y_t &= \mu + c\sigma_t^2 + a_t \\ a_t &= \sigma_t\varepsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1a_{t-1}^2 + \beta_1\sigma_{t-1}^2\end{aligned}\tag{5.2}$$

where μ and c are constants. The parameter c is called the risk-premium parameter. A positive c indicates that the return is positively related to its past volatility. Other specifications of risk-premium have also been used in the literature, including $y_t = \mu + c\sigma_t + a_t$ and $y_t = \mu + c\log(\sigma_t^2) + a_t$.

The formulation of the GARCH-M model in equation (5.2) implies that there are serial correlations in the time series y_t . These serial correlations are introduced by those in the volatility process, σ_t^2 . Therefore, the existence of risk-premium is another reason that the returns for some financial assets have serial correlations.

5.3 The exponential GARCH model

To allow for asymmetric effects between positive and negative asset returns, Nelson (1991) extended the GARCH framework to derive the exponential-GARCH (EGARCH) model. He considered the weighted innovation

$$g(\varepsilon_t) = \theta\varepsilon_t + \gamma[|\varepsilon_t| - E[|\varepsilon_t|]]\tag{5.3}$$

where θ and γ are real constants. Both ε_t and $|\varepsilon_t| - E[|\varepsilon_t|]$ are zero-mean i.i.d. sequences with continuous distributions. Therefore, $E[g(\varepsilon_t)] = 0$, and the asymmetry of $g(\varepsilon_t)$ can be easily seen by rewriting it as

$$g(\varepsilon_t) = \begin{cases} (\theta + \gamma)\varepsilon_t - \gamma E(|\varepsilon_t|) & \text{if } \varepsilon_t \geq 0, \\ (\theta - \gamma)\varepsilon_t - \gamma E(|\varepsilon_t|) & \text{if } \varepsilon_t < 0 \end{cases}$$

Remark. For the standard Gaussian random variable ε_t , $E[|\varepsilon_t|] = \sqrt{2/\pi}$. For the standardised Student t-distribution in equation (8.1), we have

$$E[|\varepsilon_t|] = \frac{2\sqrt{v-2}((v+1)/2)}{(v-1)(v/2)\sqrt{\pi}}$$

An EGARCH(m, s) model could then be written as

$$\begin{aligned} a_t &= \sigma_t \varepsilon_t, \\ \log(\sigma_t^2) &= \alpha_0 + \frac{1 + \beta_1 L + \dots + \beta_{s-1} L^{s-1}}{1 - \alpha_1 L - \dots - \alpha_m L^m} g(\varepsilon_{t-1}) \end{aligned} \quad (5.4)$$

where α_0 is a constant, L is the back-shift (or lag) operator such that $Lg(\varepsilon_t) = g(\varepsilon_{t-1})$, and $1 + \beta_1 L + \dots + \beta_{s-1} L^{s-1}$ and $1 - \alpha_1 L - \dots - \alpha_m L^m$ are stationary polynomials and have no common factors. Again, equation (5.4) uses the usual ARMA parameterisation to describe the evolution of the conditional variance of a_t . On the basis of this representation, some properties of the EGARCH model can be obtained in a similar manner as those of the GARCH model. For instance, the unconditional mean of $\log(\sigma_t^2)$ is α_0 . However, the model differs from the GARCH model in several ways.

Firstly, it uses the natural logarithm of the conditional variance to relax the positiveness constraint of model coefficients. Secondly, the use of $g(\varepsilon_t)$ enables the model to respond asymmetrically to positive and negative lagged values of ε_t . Some additional properties of the EGARCH model can be found in Nelson (1991).

To better understand the EGARCH model, we consider a simple model that has lag-order (1, 1):

$$a_t = \sigma_t \varepsilon_t, \quad (1 - \alpha L) \log(\sigma_t^2) = (1 - \alpha) \alpha_0 + g(\varepsilon_{t-1}) \quad (5.5)$$

where the ε_t are i. i. d. standard normal and the subscript of α_1 is omitted. In this case, $E[|\varepsilon_t|] = \sqrt{2/\pi}$ and the model for $\log(\sigma_t^2)$ becomes

$$(1 - \alpha L) \log(\sigma_t^2) = \begin{cases} \alpha_* + (\gamma + \theta) \varepsilon_{t-1} & \text{if } \varepsilon_{t-1} \geq 0, \\ \alpha_* + (\gamma - \theta) (-\varepsilon_{t-1}) & \text{if } \varepsilon_{t-1} < 0 \end{cases} \quad (5.6)$$

where $\alpha_* = (1 - \alpha) \alpha_0 - \sqrt{2/\pi} \gamma$. This is a nonlinear function similar to that of the threshold autoregressive (TAR) model of Tong (1978), Tong (1990). It suffices to say that for this simple EGARCH model the conditional variance evolves in a nonlinear manner depending on the sign of a_{t-1} . Specifically, we have

$$\sigma_t^2 = \sigma_{t-1}^{2\alpha} \exp(\alpha_*) \begin{cases} \exp[(\gamma + \theta) \frac{a_{t-1}}{\sigma_{t-1}}] & \text{if } a_{t-1} \geq 0, \\ \exp[(\gamma - \theta) \frac{|a_{t-1}|}{\sigma_{t-1}}] & \text{if } a_{t-1} < 0 \end{cases}$$

The coefficients $(\gamma + \theta)$ and $(\gamma - \theta)$ describe the asymmetry in response to positive and negative a_{t-1} . Therefore, the model is nonlinear if $\theta \neq 0$. Since negative shocks tend to have larger impacts, we expect that θ will be negative. For higher order EGARCH models, the nonlinearity becomes much more complicated.⁵

5.3.1 An alternative model form

For ease in estimation, one can use an alternative form of the EGARCH(m, s) model:

$$\log(\sigma_t^2) = \alpha_0 + \sum_{i=1}^m \alpha_i \frac{|a_{t-i}| + \gamma_i a_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^s \beta_j \log(\sigma_{t-j}^2) \quad (5.7)$$

In this case, a positive a_{t-i} provides a contribution of $\alpha_i(1 + \gamma_i)|\varepsilon_{t-i}|$ to the natural logarithm of volatility, whereas a negative a_{t-i} provides a contribution of $\alpha_i(1 - \gamma_i)|\varepsilon_{t-i}|$, where $\varepsilon_{t-i} = a_{t-i}/\sigma_{t-i}$. Therefore, the γ_i parameter signifies the leverage effect of a_{t-i} , where we expect that γ_i will be negative in most applications that involve financial data.

5.4 The threshold GARCH model

Another volatility model that is commonly used to handle leverage effects is the threshold generalised autoregressive conditional heteroscedastic (or TGARCH) model; see Glosten, Jagannathan, and Runkle (1993) and Zakoian (1994). Such a TGARCH(m, s) model may assume the form

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m (\alpha_i + \gamma_i N_{t-i}) a_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 \quad (5.8)$$

where N_{t-i} is an indicator for negative a_{t-i} , that is,

$$N_{t-i} = \begin{cases} 1 & \text{if } a_{t-i} < 0, \\ 0 & \text{if } a_{t-i} \geq 0 \end{cases}$$

and α_i , γ_i , and β_j are non-negative parameters that satisfying conditions that are similar to those of traditional GARCH models. From the above representation, it may be noted that a positive a_{t-i} contributes $\alpha_i a_{t-i}^2$ to σ_t^2 , whereas a negative a_{t-i} would be responsible for a larger impact on the volatility, since such a shock would make use of the representation $(\alpha_i + \gamma_i) a_{t-i}^2$, where $\gamma_i > 0$. The model would usually make use of zero as the threshold to separate the impact of past shocks. Other threshold values can also be used.⁶ This model is also called the GJR model, in recognition of the contribution of Glosten, Jagannathan, and Runkle (1993).

5.5 Asymmetric power ARCH model

The TGARCH model belongs to the class of asymmetric power autoregressive conditional heteroscedastic (APARCH) models of Ding, Granger, and Engle (1993). A general APARCH(m, s) model could be written as

$$y_t = \mu_t + a_t, \quad a_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim D(0, 1) \\ \sigma_t^\delta = \omega + \sum_{i=1}^m \alpha_i (|a_{t-i}| + \gamma_i a_{t-i})^\delta + \sum_{j=1}^s \beta_j \sigma_{t-j}^\delta \quad (5.9)$$

where μ_t is the conditional mean, $D(0, 1)$ denotes a distribution with mean zero and variance 1, δ is a positive real number, and the coefficients ω , α_i , γ_i , and β_j satisfy some regularity conditions so that the volatility is positive. Similar to GARCH models, the APARCH(1, 1) model is often used in practice. Three special cases of the APARCH models are of interest. When $\delta = 2$, the APARCH model

reduces to a TGARCH model. When $\delta = 1$, the model uses volatility directly in the volatility equation. The case of $\delta = 0$ is at the limit of $\delta \rightarrow 0$ and in this case the model becomes the EGARCH model of Nelson (1991).

The power function in equation (5.9) is a transformation used to improve the goodness of fit of the model. This appears to be a sensible approach if one is interested in prediction. On the other hand, except for some special values, it seems hard to find a good interpretation for the power parameter δ .

5.6 Non-symmetric GARCH model

Another GARCH family model that is able capture asymmetric volatility responses to past positive and negative shocks is described in Engle and Ng (1993) and Duan (1995). It takes the form

$$\begin{aligned} y_t &= \mu_t + a_t, \quad a_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim D(0, 1) \\ \sigma_t^2 &= \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 (a_{t-1} - \theta \sigma_{t-1})^2 \end{aligned} \quad (5.10)$$

where μ_t is the conditional mean, $D(0, 1)$ denotes a distribution with mean 0 and variance 1, β_i contains non-negative parameters with $\beta_0 > 0$, while θ is a leverage parameter. The model in equation (5.10) is referred to as a non-symmetric GARCH(1, 1), or NGARCH(1, 1) model. It reduces to a GARCH(1, 1) model if $\theta = 0$.

To study the properties of the NGARCH(1, 1) model, we rewrite equation (5.10) as

$$\sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-1}^2 (\varepsilon_{t-1} - \theta)^2 \quad (5.11)$$

Taking expectation and where we assume that ε_{t-1} and σ_{t-1} are independent, we have

$$\begin{aligned} E[\sigma_t^2] &= \beta_0 + \beta_1 E[\sigma_{t-1}^2] + \beta_2 E[\sigma_{t-1}^2] E(\varepsilon_{t-1} - \theta)^2 \\ &= \beta_0 + \beta_1 E[\sigma_{t-1}^2] + \beta_2 E[\sigma_{t-1}^2] (1 + \theta^2) \end{aligned}$$

If y_t is weakly stationary, then $E[\sigma_t^2] = E[\sigma_{t-1}^2]$ and we have

$$E[\sigma_t^2] = \frac{\beta_0}{1 - \beta_1 - \beta_2(1 + \theta^2)}$$

which is the unconditional variance of y_t . Consequently, we would require that $1 - \beta_1 - \beta_2(1 + \theta^2) > 0$ when making use of the NGARCH(1, 1) model. Multiplying equation (5.11) by ε_{t-1} and taking expectation, we obtain

$$\begin{aligned} E(\varepsilon_{t-1} \sigma_t^2) &= -2\theta \beta_2 E[\sigma_{t-1}^2] \\ &= \frac{-2\theta \beta_0 \beta^2}{1 - \beta_1 - \beta_2(1 + \theta^2)} \end{aligned}$$

This result suggests that if $\theta > 0$ and $\beta_2 > 0$, then ε_{t-1} is negatively related to σ_t^2 . Therefore, θ is a leverage parameter and should be positive. Finally, it can be shown that, under certain conditions, the shock a_t of a NGARCH(1, 1) model has heavy tails even if ε_t is Gaussian.⁷

6 The stochastic volatility model

An alternative approach to describe the evolution of volatility in a time series is to introduce a stochastic innovation to the conditional variance equation of a_t . This additional stochastic term could be used to explain the unexpected shocks to the volatility process and is considered in Melino and Turnbull (1990), Taylor (1994), Harvey, Ruiz, and Shephard (1994), and Jacquier, Polson, and Rossi (1994). The resulting model is called a stochastic volatility (SV) model and it makes use of a relatively straightforward functional form. Similar to EGARCH models, to ensure positiveness of the conditional variance, SV models use $\log(\sigma_t^2)$ instead of σ_t^2 , where an SV model may be defined as

$$\begin{aligned} a_t &= \sigma_t \varepsilon_t \\ (1 - \alpha_1 L - \dots - \alpha_m L^m) \log(\sigma_t^2) &= \alpha_0 + v_t \end{aligned} \quad (6.1)$$

where the distribution for ε_t may take the form i. i. d.N(0, 1), while the distribution of v_t may be described by i. i. d.N(0, σ_v^2). Note that ε_t and v_t are independent, α_0 is a constant, and all the polynomials, $1 - \sum_{i=1}^m \alpha_i L^i$, ensure stationarity. Adding the innovation v_t would increase the flexibility of the model and its ability to describe the evolution of σ_t^2 . However, it also increases the difficulty in parameter estimation. To estimate an SV model, we could evaluate the likelihood function with the aid of a Kalman filtering technique, or in most cases, the use of a non-linear filter may be preferable. Jacquier, Polson, and Rossi (1994) provide some comparison of estimation results between quasi-likelihood and Markov chain Monte Carlo (MCMC) methods that may be used for parameter estimation. The difficulty with estimating the parameters in a SV model is understandable since for each shock a_t the model uses two innovations ε_t and v_t .⁸

Limited experience suggests that SV models often provided improvements in their in-sample ability, however, their contributions to out-of-sample volatility forecasts are somewhat mixed.

6.1 Long-memory Stochastic Volatility Model

More recently, the SV model has been extended to allow for long memory in volatility. A time series is a long-memory process if its autocorrelation function decays at a hyperbolic, instead of an exponential, rate as the lag increases. The extension to long-memory models in volatility studies is motivated by the fact that the autocorrelation function of the squared or absolute-valued series of an asset return often decays slowly, even though the return series has no serial correlation; see Ding, Granger, and Engle (1993).

A simple long-memory stochastic volatility (LMSV) model can be written as

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t = \sigma \exp(u_t/2), \quad (1 - L)^d u_t = \eta_t \quad (6.2)$$

where $\sigma > 0$, the ε_t are i. i. d.N(0, 1), the η_t are i. i. d.N(0, σ_η^2) and independent of ε_t , and $0 < d < 0.5$. The feature of long memory stems from the fractional difference $(1 - L)^d$, which implies that the ACF of u_t decays slowly at a hyperbolic, instead of an exponential, rate as the lag increases. This would imply that to model a LMSV, we would have

$$\begin{aligned} \log(a_t^2) &= \log(\sigma^2) + u_t + \log(\varepsilon_t^2) \\ &= [\log(\sigma^2) + E[\log \varepsilon_t^2]] + u_t + [\log(\varepsilon_t^2) - E[\log \varepsilon_t^2]] \equiv \mu + u_t + e_t \end{aligned}$$

Thus, the $\log(a_t^2)$ would take the form of a Gaussian long-memory signal plus a non-Gaussian white noise; see Breidt, Crato, and de Lima (1998). Using the natural logarithm of a time series for the squared daily returns for companies in the S&P 500 index, Bollerslev and Jubinski (1999) and Ray and Tsay (2000) found that the median estimate of d is about 0.38. For applications, Ray and Tsay

(2000) studied common long-memory components in daily stock volatilities of groups of companies classified by various characteristics. They found that companies in the same industrial or business sectors tend to have more common long-memory components (e.g. big US national banks and financial institutions).

7 Conclusion

Statistical models that seek to describe volatility have been applied to a wide range of analyses that involve time series variables, where a number of applications in finance have been particularly successful. Since many financial decisions are partially based upon the amount of risk that may be incurred, it is not surprising to note that an analysis into the evolution of volatility constitutes an integral part of most asset pricing and portfolio analysis theories. During the course of this lecture, we considered the use of various ARCH, GARCH and stochastic volatility models, where particular attention was directed towards the relative strength and weakness of each of the characterisations.

8 Appendix: The likelihood function of a ARCH(m) model

Under the normality assumption, the likelihood function of an ARCH(m) model is

$$\begin{aligned} f(a_1, \dots, a_T | \alpha) &= f(a_T | I_{T-1}) f(a_{T-1} | I_{T-2}) \dots f(a_{m+1} | I_m) f(a_1, \dots, a_m | \alpha) \\ &= \prod_{t=m+1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{a_t^2}{2\sigma_t^2}\right) \times f(a_1, \dots, a_m | \alpha), \end{aligned}$$

where $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_m)'$ and $f(a_1, \dots, a_m | \alpha)$ is the joint probability density function of a_1, \dots, a_m . As the exact form of $f(a_1, \dots, a_m | \alpha)$ is complicated, it is commonly dropped from the prior likelihood function, especially when the sample size is sufficiently large. This results in using the conditional likelihood function

$$f(a_{m+1}, \dots, a_T | \alpha, a_1, \dots, a_m) = \prod_{t=m+1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{a_t^2}{2\sigma_t^2}\right)$$

where σ_t^2 can be evaluated recursively. We refer to estimates obtained by maximizing the prior likelihood function as the conditional maximum likelihood estimates (MLEs) under normality.

Maximizing the conditional likelihood function is equivalent to maximizing its logarithm, which is easier to handle. The conditional log likelihood function is

$$\ell(a_{m+1}, \dots, a_T | \alpha, a_1, \dots, a_m) = \sum_{t=m+1}^T \left[-\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_t^2) - \frac{1}{2} \frac{a_t^2}{\sigma_t^2} \right]$$

As the first term $\log(2\pi)$ does not involve any parameters, the log likelihood function becomes

$$\ell(a_{m+1}, \dots, a_T | \alpha, a_1, \dots, a_m) = - \sum_{t=m+1}^T \left[\frac{1}{2} \log(\sigma_t^2) + \frac{1}{2} \frac{a_t^2}{\sigma_t^2} \right]$$

where $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_m a_{t-m}^2$ can be evaluated recursively.

In some applications, it is more appropriate to assume that ε_t follows a heavy-tailed distribution such as a standardised Student t-distribution. Let x_v be a Student t-distribution with v degrees of freedom. Then $\text{var}[x_v] = v/(v-2)$ for $v > 2$, and we use $\varepsilon_t = x_v/\sqrt{v/(v-2)}$. The probability density function of ε_t is

$$f(\varepsilon_t|v) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{(v-2)\pi}} \left(1 + \frac{\varepsilon_t^2}{v-2}\right)^{-(v+1)/2}, \quad v > 2, \quad (8.1)$$

where $\Gamma(x)$ is the usual gamma function (i.e. $\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy$). Using $a_t = \sigma_t \varepsilon_t$, we obtain the conditional likelihood function of a_t as

$$f(a_{m+1}, \dots, a_T | \alpha, A_m) = \prod_{t=m+1}^T \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{(v-2)\pi}} \frac{1}{\sigma_t} \left(1 + \frac{a_t^2}{(v-2)\sigma_t^2}\right)^{-(v+1)/2}$$

where $v > 2$ and $A_m = (a_1, a_2, \dots, a_m)$. We refer to the estimates that maximise the prior likelihood function as the conditional MLEs under t-distribution. The degrees of freedom of the t-distribution can be specified a priori or estimated jointly with other parameters. A value between 3 and 6 is often used if it is prespecified.

If the degrees of freedom v of the Student t-distribution is prespecified, then the conditional log likelihood function is

$$\ell(a_{m+1}, \dots, a_T | \alpha, A_m) = - \sum_{t=m+1}^T \left[\frac{v+1}{2} \log\left(1 + \frac{a_t^2}{(v-2)\sigma_t^2}\right) + \frac{1}{2} \log(\sigma_t^2) \right] \quad (8.2)$$

If one wishes to estimate v jointly with other parameters, then the log likelihood function becomes

$$\begin{aligned} \ell(a_{m+1}, \dots, a_T | \alpha, v, A_m) \\ = (T - m) [\log(\Gamma(v+1)/2)) - \log(\Gamma(v/2)) - 0.5 \log((v-2)\pi)] \\ + \ell(a_{m+1}, \dots, a_T | \alpha, A_m) \end{aligned}$$

where the second term is given in equation (8.2).

Besides fat tails, empirical distributions of asset returns may also be skew. To handle this additional characteristic of asset returns, the Student t-distribution has been modified to become a skew Student t-distribution. There are multiple versions of skew Student t-distribution, but we shall adopt the approach of Fernandez and Steel (1998), which can introduce skewness into any continuous unimodal and symmetric (with respect to 0) univariate distribution. Specifically, for the innovation ε_t of an ARCH process, Lambert and Laurent (2001) apply the method of Fernandez and Steel (1998) to the standardised Student t-distribution in equation (8.1) to obtain a standardised skew Student t-distribution. The resulting probability density function is

$$g(|\varepsilon_t|, \xi, v) = \begin{cases} \frac{2}{\xi} + \frac{1}{\xi} \varrho f[\xi(\varrho \varepsilon_t + \varpi)|v] & \text{if } \varepsilon_t < -\varpi/\varrho \\ \frac{2}{\xi} + \frac{1}{\xi} \varrho f[(\varrho \varepsilon_t + \varpi)/\xi|v] & \text{if } \varepsilon_t \geq -\varpi/\varrho \end{cases} \quad (8.3)$$

where $f(\cdot)$ is the pdf of the standardised Student t-distribution in equation (8.1), ξ is the skewness parameter, $v > 2$ is the degrees of freedom, and the parameters ϱ and ϖ are given below

$$\varpi = \frac{\Gamma((v-1)/2)\sqrt{v-2}}{\sqrt{\pi}\Gamma(v/2)}\left(\xi - \frac{1}{\xi}\right)$$

$$q^2 = \left(\xi^2 + \frac{1}{\xi^2} - 1\right) - \varpi^2$$

In equation (8.3), ξ^2 is equal to the ratio of probability masses above and below the mode of the distribution and, hence, it is a measure of the skewness.

9 References

- Andersen, T. G., and T. Bollerslev. 1998. "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts." *International Economic Review* 39: 885–905.
- Bollerslev, T. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31: 307–27.
- Bollerslev, T., R. Y. Chou, and K. F. Kroner. 1992. "ARCH Modeling in Finance." *Journal of Econometrics* 52: 5–59.
- Bollerslev, T., R. F. Engle, and D. B. Nelson. 1994. "Handbook of Econometrics Iv." In, edited by R. F. Engle and D. C. McFadden, 2959–3038. Amsterdam: Elsevier Science.
- Bollerslev, T., and D. Jubinski. 1999. "Equality Trading Volume and Volatility: Latent Information Arrivals and Common Long-Run Dependencies." *Journal of Business and Economic Statistics* 19: 9–21.
- Breidt, F. J., N. Crato, and P. de Lima. 1998. "On the Detection and Estimation of Long Memory in Stochastic Volatility." *Journal of Econometrics* 83: 325–48.
- Cao, C., and Tsay R. S. 1992. "Nonlinear Time Series Analysis of Stock Volatilities." *Journal of Applied Econometrics* 7: 165–85.
- Ding, Z., C. W. J. Granger, and R. F. Engle. 1993. "A Long Memory Property of Stock Returns and a New Model." *Journal of Empirical Finance* 1: 83–106.
- Duan, J. 1995. "The Garch Option Pricing Model." *Mathematical Finance* 5: 13–32.
- Engle, R. F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50: 987–1007.
- Engle, R. F., and V. Ng. 1993. "Measuring and Testing the Impact of News on Volatility." *Journal of Finance* 48: 1749–78.
- Fernandez, C., and M. F. J. Steel. 1998. "On Bayesian Modelling of Fat Tails and Skewness." *Journal of American Statistical Association* 93: 1779–1801.
- Glosten, L. R., R. Jagannathan, and D. E. Runkle. 1993. "On the Relation Between the Expected Value and the Volatility of Nominal Excess Return on Stocks." *Journal of Finance* 48: 1779–1801.
- Harvey, A. C., E. Ruiz, and N. Shephard. 1994. "Multivariate Stochastic Variance Models." *Review of Econometric Studies* 61: 247–64.
- Jacquier, E., N. G. Polson, and P. Rossi. 1994. "Bayesian Analysis of Stochastic Volatility Models (with Discussion)." *Journal of Business and Economic Statistics* 12: 371–417.

- Lambert, P., and S. Laurent. 2001. “Modelling Financial Time Series Using Garch-Type Models and a Skewed Student Density.” Working paper. Universite de Liege.
- McLeod, A. I., and W. K. Li. 1983. “Diagnostic Checking Arma Time Series Models Using Squared-Residual Autocorrelations.” *Journal of Time Series Analysis* 4: 269–73.
- Melino, A., and S. M. Turnbull. 1990. “Pricing Foreign Currency Options with Stochastic Volatility.” *Journal of Econometrics* 45: 318–34.
- Nelson, D. B. 1990. “Stationarity and Persistence in the Garch(1,1) Model.” *Econometric Theory* 6: 318–34.
- . 1991. “Conditional Heteroskedasticity in Asset Returns: A New Approach.” *Econometrica* 59: 347–70.
- Ray, B. K., and R. S. Tsay. 2000. “Long-Range Dependence in Daily Stock Volatilities.” *Journal of Business and Economic Statistics* 18: 254–62.
- Taylor, S. J. 1994. “Modeling Stochastic Volatility: A Review and Comparative Study.” *Mathematical Finance* 4: 183–204.
- Tong, H. 1978. “Pattern Recognition and Signal Processing.” In, edited by C. H. Chen. Amsterdam: Sijhoff & Noordhoff.
- . 1990. *Non-Linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.
- Tsay, R. S. 2010. *Analysis of Financial Time Series*. Hoboken (NJ): John Wiley & Sons.
- Zakoian, J. M. 1994. “Threshold Heteroscedastic Models.” *Journal of Economic Dynamics and Control* 18: 931–55.

-
1. Tsay (2010) contains additional details on this representation.↵
 2. Further discussions of the likelihood function for the ARCH(m) model is contained in the appendix.↵
 3. For further information about kurtosis of GARCH models, see Tsay (2010).↵
 4. See, Tsay (2010) for further details of this measure.↵
 5. See, Cao and S. (1992) for how to obtain multi-step ahead volatility forecasts for nonlinear models, (including EGARCH models).↵
 6. See, Tsay (2010) for the general concept of threshold models.↵
 7. See, Duan (1995) for further details.↵
 8. Tsay (2010) contains details relating to the use of MCMC methods to estimate SV models. For more discussions on SV models, see Taylor (1994).↵