Chapter 9

Modelling volatility and correlation

Models for Volatility

- Modelling and forecasting stock market volatility has been the subject of vast empirical and theoretical investigation
- There are a number of motivations for this line of inquiry:
  - Volatility is one of the most important concepts in finance
  - Volatility, as measured by the standard deviation or variance of returns, is often used as a crude measure of the total risk of financial assets
  - Many value-at-risk models for measuring market risk require the estimation or forecast of a volatility parameter
  - The volatility of stock market prices also enters directly into the Black–Scholes formula for deriving the prices of traded options
- We will now examine several volatility models.

Historical Volatility

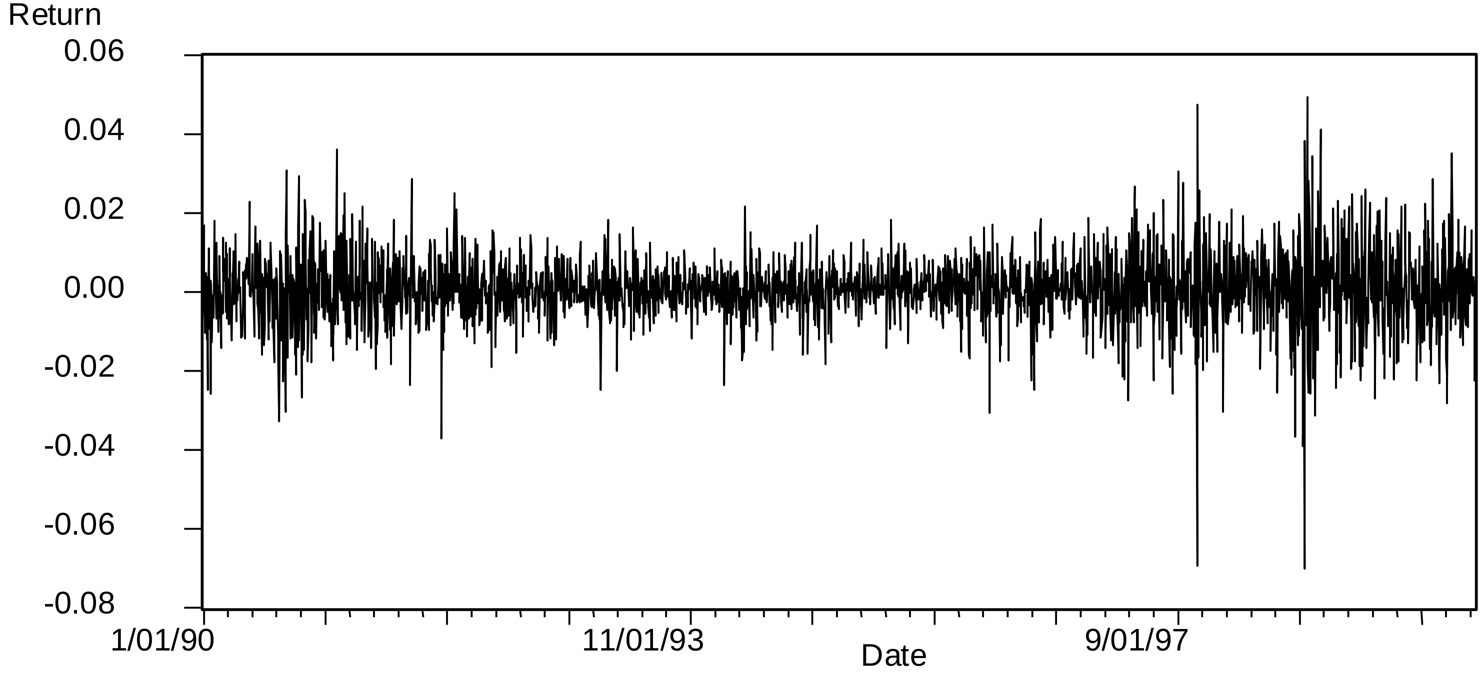
- The simplest model for volatility is the historical estimate
- Historical volatility simply involves calculating the **variance (or standard deviation) of returns** in the usual way over some historical period
- This then becomes the volatility forecast for all future periods
- Evidence suggests that the use of volatility predicted from more sophisticated time series models will lead to more accurate forecasts and option valuations
- Historical volatility is still useful as a **benchmark for comparing the forecasting ability of more complex time models**

Implied Volatility

- All pricing models for financial options require a volatility estimate or forecast as an input.
- Given the price of a traded option obtained from transactions data, it is possible to determine the **volatility** forecast over the lifetime of the option **implied** by the option's valuation.
- For example, if the standard **Black–Scholes model** is used, the option price, the time to maturity, a risk-free rate of interest, the strike price and the current value of the underlying asset, are all either specified in the details of the options contracts or are available from market data
- Therefore, given all of these quantities, it is possible to use a numerical procedure to derive the **volatility implied by the option**

A Sample Financial Asset Returns Time Series

Daily S&P 500 Returns for January 1990 – December 1999



1. An Excursion into Non-linearity Land

- Our "traditional" structural model could be something like:

  $y_t = \beta_1 + \beta_2 x_{2t} + ... + \beta_k x_{kt} + u_t$, or $y = X\beta + u$.

  We also assume $u_t \sim N(0,\sigma^2)$.

- Motivation: the linear structural (and time series) models cannot explain a number of important features common to much financial data:

  - leptokurtosis: returns have dist. with higher peak and fatter tail

  - volatility clustering or volatility pooling (bunches of volatility, autocorrelation) where volatility is variance of returns

  - leverage effects (asymmetry): following a price fall, the volatility is higher than following a price rise of the same magnitude

Non-linear Models: A Definition

- Campbell, Lo and MacKinlay (1997) define a non-linear data generating process as one that can be written

  $$y_t = f(u_t, u_{t-1}, u_{t-2}, \ldots)$$

  where u_t is an iid error term and f is a non-linear function.

- They also give a slightly more specific definition as

  $$y_t = g(u_{t-1}, u_{t-2}, \ldots) + u_t\sigma^2(u_{t-1}, u_{t-2}, \ldots)$$

  where g is a function of past error terms only and σ^2 is a variance term.

- Models with nonlinear g(•) are "non-linear in mean", while those with nonlinear σ^2(•) are "non-linear in variance".

Types of non-linear models

- The linear paradigm is a useful one. Many apparently non-linear relationships can be made linear by a suitable transformation. On the other hand, it is likely that many relationships in finance are intrinsically non-linear.

- There are many types of non-linear models, e.g.

  - ARCH / GARCH: useful for modelling and forecasting volatility

  - switching models: series following different processes at different points in time

Testing for Non-linearity

- The "traditional" tools of time series analysis (acf's, spectral analysis) may find no evidence that we could use a linear model, but the data may still not be independent.
- Portmanteau tests for non-linear dependence have been developed. The simplest is Ramsey's RESET test, which took the form:

$$y_t = \alpha_1 + \alpha_2 \hat{y}_t^2 + \alpha_3 \hat{y}_t^3 + \ldots + \alpha_p \hat{y}_t^p + \sum \beta_i x_{it} + v_t$$

- Many other non-linearity tests are available, e.g. the "BDS test" and the bispectrum test.
- One particular non-linear model that has proved very useful in finance is the ARCH model due to Engle (1982).

Heteroscedasticity Revisited

- An example of a structural model is

  $$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$$

  with $u_t \sim N(0, \sigma_u^2)$.

- The assumption that the variance of the errors is constant is known as homoscedasticity, i.e. Var $(u_t) = \sigma_u^2$.

- What if the variance of the errors is not constant?

  - heteroscedasticity

  - would imply that standard error estimates could be wrong.

- Is the variance of the errors likely to be constant over time? Not for financial data.

7. Autoregressive Conditionally Heteroscedastic (ARCH) Models

- So use a model which does not assume that the variance is constant.
- Definition of the conditional variance of u_t:

  $$\sigma_t^2 = \mathrm{Var}(u_t \mid u_{t-1}, u_{t-2},\ldots) = \mathrm{E}[(u_t-\mathrm{E}(u_t))^2 \mid u_{t-1}, u_{t-2},\ldots]$$

  We usually assume that $\mathrm{E}(u_t) = 0$

  so $\sigma_t^2 = \mathrm{Var}(u_t \mid u_{t-1}, u_{t-2},\ldots) = \mathrm{E}[u_t^2 \mid u_{t-1}, u_{t-2},\ldots]$.

- What could the current value of the variance of the errors plausibly depend upon?
  - Previous squared error terms.
- This leads to the AutoRegressive Conditionally Heteroscedastic Model for the variance of the errors:

  $$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$$

- This is known as an ARCH(1) model.

Autoregressive Conditionally Heteroscedastic (ARCH) Models (cont'd)

- The full model would be

  $y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t, \; u_t \sim \mathrm{N}(0, \sigma_t^2)$

  where $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$
- We can easily extend this to the general case where the error variance depends on q lags of squared errors:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_q u_{t-q}^2$$

- This is an ARCH(q) model.

- Instead of calling the variance σ_t^2, in the literature it is usually called h_t, so the model is

$$y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t, \; u_t \sim \mathrm{N}(0, h_t)$$

  where $h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_q u_{t-q}^2$

Another Way of Writing ARCH Models

- For illustration, consider an ARCH(1). Instead of the above, we can write

$$y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t, \qquad u_t = \varepsilon_t \sigma_t$$

$$\sigma_t = \sqrt{\alpha_0 + \alpha_1 u_{t-1}^2}, \qquad \varepsilon_t \sim \mathrm{N}(0,1)$$

- The two are different ways of expressing exactly the same model. The first form is easier to understand while the second form is required for simulating from an ARCH model, for example.

Testing for "ARCH Effects"

1. First, run any postulated linear regression of the form given in the equation above, e.g. $y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t$ saving the residuals, $\hat{u}_t$.

2. Then square the residuals, and regress them on q own lags to test for ARCH of order q, i.e. run the regression

$$\hat{u}_t^2 = \gamma_0 + \gamma_1 \hat{u}_{t-1}^2 + \gamma_2 \hat{u}_{t-2}^2 + \ldots + \gamma_q \hat{u}_{t-q}^2 + v_t$$

where v_t is an error term.
Obtain R^2 from this regression.

3. The test statistic is defined as TR^2 (the number of observations multiplied by the coefficient of multiple correlation) from the last regression, and is distributed as a $\chi^2(q)$.

Testing for "ARCH Effects" (cont'd)

4. The null and alternative hypotheses are

$H_0 : \gamma_1 = 0$ and $\gamma_2 = 0$ and $\gamma_3 = 0$ and ... and $\gamma_q = 0$

$H_1 : \gamma_1 \neq 0$ or $\gamma_2 \neq 0$ or $\gamma_3 \neq 0$ or ... or $\gamma_q \neq 0$.

If the value of the test statistic is greater than the critical value from the χ^2 distribution, then reject the null hypothesis (proof of autocorrelation of volatility).

Problems with ARCH(q) Models

- How do we decide on q?
- The required value of q might be very large
- Non-negativity constraints might be violated.
  - When we estimate an ARCH model, we require $\alpha_i > 0 \ \forall \ i=1,2,...,q$ (since variance cannot be negative)

- A natural extension of an ARCH(q) model which gets around some of these problems is a GARCH model.

8. Generalised ARCH (GARCH) Models

- Due to Bollerslev (1986): allow the conditional variance to be dependent upon previous own lags
- The variance equation is now

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta\sigma_{t-1}^2 \qquad (1)$$

- This is a GARCH(1,1) model, which is like an ARMA(1,1) model for the variance equation.
- We could also write

$$\sigma_{t-1}^2 = \alpha_0 + \alpha_1 u_{t-2}^2 + \beta\sigma_{t-2}^2$$

$$\sigma_{t-2}^2 = \alpha_0 + \alpha_1 u_{t-3}^2 + \beta\sigma_{t-3}^2$$

- Substituting into (1) for σ_{t-1}^2 :

$$\boldsymbol{\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta(\alpha_0 + \alpha_1 u_{t-2}^2 + \beta\sigma_{t-2}^2)}$$

$$\boldsymbol{= \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_0\beta + \alpha_1\beta u_{t-2}^2 + \beta\sigma_{t-2}^2}$$

Generalised ARCH (GARCH) Models (cont'd)

- Now substituting into (2) for σ_{t-2}^2

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_0\beta + \alpha_1\beta u_{t-2}^2 + \beta^2(\alpha_0 + \alpha_1 u_{t-3}^2 + \beta\sigma_{t-3}^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_0\beta + \alpha_1\beta u_{t-2}^2 + \alpha_0\beta^2 + \alpha_1\beta^2 u_{t-3}^2 + \beta^3\sigma_{t-3}^2$$

$$\sigma_t^2 = \alpha_0(1+\beta+\beta^2) + \alpha_1 u_{t-1}^2(1+\beta L+\beta^2 L^2) + \beta^3\sigma_{t-3}^2$$

- An infinite number of successive substitutions would yield

$$\sigma_t^2 = \alpha_0(1+\beta+\beta^2+\ldots) + \alpha_1 u_{t-1}^2(1+\beta L+\beta^2 L^2+\ldots) + \beta^\infty\sigma_0^2$$

- So the GARCH(1,1) model can be written as an infinite order ARCH model.

- We can again extend the GARCH(1,1) model to a GARCH(p,q):

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_q u_{t-q}^2 + \beta_1\sigma_{t-1}^2 + \beta_2\sigma_{t-2}^2 + \ldots + \beta_p\sigma_{t-p}^2$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q}\alpha_i u_{t-i}^2 + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2$$

Generalised ARCH (GARCH) Models (cont'd)

- But in general a GARCH(1,1) model will be sufficient to capture the volatility clustering in the data.

- Why is GARCH Better than ARCH?

  - more parsimonious - avoids overfitting

  - then less likely to breech non-negativity constraints

The Unconditional Variance under the GARCH Specification

- The unconditional variance of u_t is given by

$$\text{Var}(u_t) = \frac{\alpha_0}{1-(\alpha_1+\beta)}$$

when $\alpha_1 + \beta < 1$

- $\alpha_1 + \beta \geq 1$ is termed "non-stationarity" in variance
- $\alpha_1 + \beta = 1$ is termed Intergrated GARCH
- For stationarity in variance, the conditional variance forecasts will converge to the long-term average of variance
- For non-stationarity in variance, the conditional variance forecasts will go to infinity
- For IGARCH, the conditional variance forecasts will not converge to their unconditional value as the horizon increases.

9. Estimation of ARCH / GARCH Models

- Since the model is no longer of the usual linear form, RSS does not depend on the conditional variance, we cannot use OLS.

- We use another technique known as maximum likelihood.

- The method works by finding the most likely values of the parameters given the actual data.

- More specifically, we form a log-likelihood function and maximise it.

Estimation of ARCH / GARCH Models (cont'd)

- The steps involved in actually estimating an ARCH or GARCH model are as follows

1. Specify the appropriate equations for the mean and the variance - e.g. an AR(1)- GARCH(1,1) model:

$$y_t = \mu + \phi y_{t-1} + u_t \quad , u_t \sim N(0,\sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta\sigma_{t-1}^2$$

2. Specify the log-likelihood function to maximise:

$$L = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log(\sigma_t^2) - \frac{1}{2}\sum_{t=1}^{T}(y_t - \mu - \phi y_{t-1})^2 / \sigma_t^2$$

3. The computer will maximise the function and give parameter values and their standard errors

Example: Parameter Estimation using Maximum Likelihood for a linear model

- Consider the bivariate regression case with homoscedastic errors for simplicity: $y_t = \beta_1 + \beta_2 x_t + u_t$

- Assuming that $u_t \sim N(0,\sigma^2)$, then $y_t \sim N(\beta_1 + \beta_2 x_t, \sigma^2)$ so that the probability density function for a normally distributed random variable with this mean and variance is given by

$$f(y_t \mid \beta_1 + \beta_2 x_t, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2} \right\} \qquad (1)$$

- Successive values of y_t would trace out the familiar bell-shaped curve.

- Assuming that u_t are iid, then y_t will also be iid.

Parameter Estimation using Maximum Likelihood (cont'd)

- Then the joint pdf for all the y's can be expressed as a product of the individual density functions

$$f(y_1, y_2, \ldots, y_T \left| \beta_1 + \beta_2 X_t, \sigma^2 \right.) = f(y_1 \left| \beta_1 + \beta_2 X_1, \sigma^2 \right.) f(y_2 \left| \beta_1 + \beta_2 X_2, \sigma^2 \right.) \ldots$$

$$f(y_T \left| \beta_1 + \beta_2 X_4, \sigma^2 \right.) \qquad (2)$$

$$= \prod_{t=1}^{T} f(y_t \left| \beta_1 + \beta_2 X_t, \sigma^2 \right.)$$

- Substituting into equation (2) for every y_t from equation (1),

$$f(y_1, y_2, \ldots, y_T \left| \beta_1 + \beta_2 x_t, \sigma^2 \right.) = \frac{1}{\sigma^T (\sqrt{2\pi})^T} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} \frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2} \right\} \qquad (3)$$

Parameter Estimation using Maximum Likelihood (cont'd)

- The typical situation we have is that the x_t and y_t are given and we want to estimate β_1, β_2, σ^2. If this is the case, then f(•) is known as the likelihood function, denoted $LF(\beta_1, \beta_2, \sigma^2)$, so we write

$$LF(\beta_1, \beta_2, \sigma^2) = \frac{1}{\sigma^T (\sqrt{2\pi})^T} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} \frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2} \right\} \qquad (4)$$

- Maximum likelihood estimation involves choosing parameter values (β_1, β_2,σ^2) that maximise this function.

- We want to differentiate (4) w.r.t. β_1, β_2,σ^2, but (4) is a product containing T terms.

Parameter Estimation using Maximum Likelihood (cont'd)

- Since $\max_x f(x) = \max_x \log(f(x))$, we can take logs of (4).
- Then, using the various laws for transforming functions containing logarithms, we obtain the log-likelihood function, *LLF*:

$$LLF = -T\log\sigma - \frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2}$$

- which is equivalent to

$$LLF = -\frac{T}{2}\log\sigma^2 - \frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^2} \qquad (5)$$

- Differentiating (5) w.r.t. $\beta_1, \beta_2, \sigma^2$, we obtain

$$\frac{\partial LLF}{\partial \beta_1} = -\frac{1}{2}\sum\frac{(y_t - \beta_1 - \beta_2 x_t).2.-1}{\sigma^2} \qquad (6)$$

Parameter Estimation using Maximum Likelihood (cont'd)

$$\frac{\partial LLF}{\partial \beta_2} = -\frac{1}{2}\sum \frac{(y_t - \beta_1 - \beta_2 x_t).2. - x_t}{\sigma^2} \quad (7)$$

$$\frac{\partial LLF}{\partial \sigma^2} = -\frac{T}{2}\frac{1}{\sigma^2} + \frac{1}{2}\sum \frac{(y_t - \beta_1 - \beta_2 x_t)^2}{\sigma^4} \quad (8)$$

- Setting (6)-(8) to zero to minimise the functions, and putting hats above the parameters to denote the maximum likelihood estimators,

- From (6), $\sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t) = 0$

$$\sum y_t - \sum \hat{\beta}_1 - \sum \hat{\beta}_2 x_t = 0$$

$$\sum y_t - T\hat{\beta}_1 - \hat{\beta}_2 \sum x_t = 0$$

$$\frac{1}{T}\sum y_t - \hat{\beta}_1 - \hat{\beta}_2 \frac{1}{T}\sum x_t = 0$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x} \quad (9)$$

Parameter Estimation using Maximum Likelihood (cont'd)

- From (7), $$\sum(y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t)x_t = 0$$

$$\sum y_t x_t - \sum \hat{\beta}_1 x_t - \sum \hat{\beta}_2 x_t^2 = 0$$

$$\sum y_t x_t - \hat{\beta}_1 \sum x_t - \hat{\beta}_2 \sum x_t^2 = 0$$

$$\hat{\beta}_2 \sum x_t^2 = \sum y_t x_t - (\bar{y} - \hat{\beta}_2 \bar{x}) \sum x_t$$

$$\hat{\beta}_2 \sum x_t^2 = \sum y_t x_t - T\bar{x}\bar{y} - \hat{\beta}_2 T\bar{x}^2$$

$$\hat{\beta}_2 (\sum x_t^2 - T\bar{x}^2) = \sum y_t x_t - T\bar{x}\bar{y}$$

$$\hat{\beta}_2 = \frac{\sum y_t x_t - T\bar{x}\bar{y}}{(\sum x_t^2 - T\bar{x}^2)} \qquad (10)$$

- From (8), $$\frac{T}{\hat{\sigma}^2} = \frac{1}{\hat{\sigma}^4} \sum(y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t)^2$$

Parameter Estimation using Maximum Likelihood (cont'd)

- Rearranging, $\hat{\sigma}^2 = \frac{1}{T}\sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t)^2$

$$\hat{\sigma}^2 = \frac{1}{T}\sum \hat{u}_t^2 \qquad (11)$$

- How do these formulae compare with the OLS estimators?

  (9) & (10) are identical to OLS

  (11) is different. The OLS estimator was

$$\hat{\sigma}^2 = \frac{1}{T-k}\sum \hat{u}_t^2$$

- Therefore the ML estimator of the variance of the disturbances is biased, although it is consistent.
- Q: But how to use ML in estimating heteroscedastic models?

Estimation of GARCH Models Using Maximum Likelihood

- Now we have $y_t = \mu + \phi y_{t-1} + u_t \;,\; u_t \sim \mathrm{N}(0, \sigma_t^2)$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2$$

$$L = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log(\sigma_t^2) - \frac{1}{2}\sum_{t=1}^{T}(y_t - \mu - \phi y_{t-1})^2 / \sigma_t^2$$

- Unfortunately, the LLF for a model with time-varying variances cannot be maximised analytically, except in the simplest cases. So a numerical procedure is used to maximise the log-likelihood function. A potential problem: local optima or multimodalities in the likelihood surface.

- The way we do the optimisation is:

1. Set up LLF.
2. Use regression to get initial guesses for the mean parameters.
3. Choose some initial guesses for the conditional variance parameters.
4. Specify a convergence criterion - either by criterion or by value.

Non-Normality and Maximum Likelihood

- Recall that the conditional normality assumption for u_t is essential.

- We can test for normality using the following representation

$$u_t = v_t \sigma_t \qquad v_t \sim \mathrm{N}(0,1)$$

$$\sigma_t = \sqrt{\alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 \sigma_{t-1}^2} \qquad v_t = \frac{u_t}{\sigma_t}$$

- The sample counterpart is $\hat{v}_t = \dfrac{\hat{u}_t}{\hat{\sigma}_t}$

- Are the $\hat{v}_t$ normal? Typically $\hat{v}_t$ are still **leptokurtic**, although less so than the $\hat{u}_t$. Is this a problem? Not really, as we can use the ML with a robust variance/covariance estimator. **ML with robust standard errors** is called **Quasi-Maximum Likelihood or QML**.

10. Extensions to the Basic GARCH Model

- Since the GARCH model was developed, a huge number of extensions and variants have been proposed. Three of the most important examples are **EGARCH, GJR, and GARCH-M** models.

- Problems with GARCH(p,q) Models:

  - Non-negativity constraints may still be violated

  - GARCH models cannot account for **leverage effects**

- Possible solutions: the exponential GARCH (EGARCH) model or the GJR model, which are asymmetric GARCH models.

11. The EGARCH Model

- Suggested by Nelson (1991). The variance equation is given by

$$\log(\sigma_t^2) = \omega + \beta \log(\sigma_{t-1}^2) + \gamma \frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha \left[\frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right]$$

- Advantages of the model

- Since we model the $\log(\sigma_t^2)$, then even if the parameters are negative, σ_t^2 will be positive.

- We can account for the **leverage effect**: if the relationship between volatility and returns is negative, γ **will be negative**.

12. The GJR Model

- Due to Glosten, Jaganathan and Runkle

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma u_{t-1}^2 I_{t-1}$$

where $I_{t-1} = 1$ if $u_{t-1} < 0$
$= 0$ otherwise

- For a **leverage effect**, we would see $\gamma > 0$**.**

- We require $\alpha_1 + \gamma \geq 0$ and $\alpha_1 \geq 0$ for non-negativity.

An Example of the use of a GJR Model

- Using monthly S&P 500 returns, December 1979- June 1998
- Estimating a GJR model, we obtain the following results.

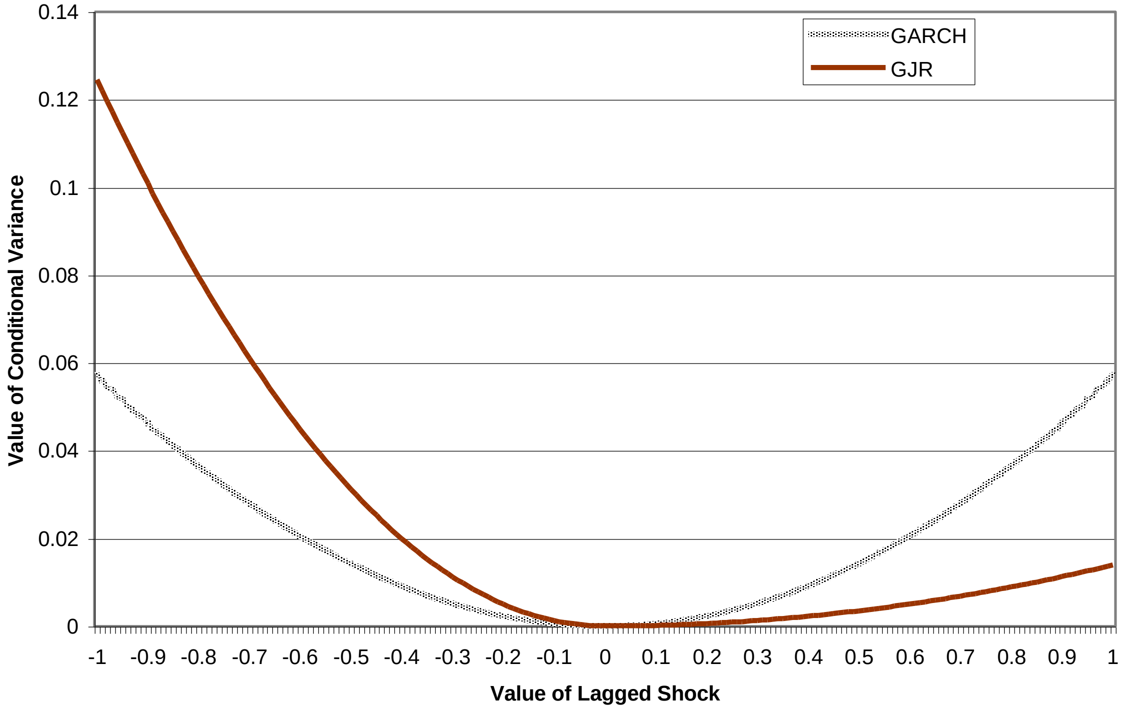
$$\underset{(3.198)}{y_t = 0.172}$$

$$\sigma_t^2 = \underset{(16.372)}{1.243} + \underset{(0.437)}{0.015}u_{t-1}^2 + \underset{(14.999)}{0.498}\sigma_{t-1}^2 + \underset{(5.772)}{0.604}u_{t-1}^2 I_{t-1}$$

13. News Impact Curves

The news impact curve plots the next period volatility (h_t) that would arise from various positive and negative values of u_{t-1}, given an estimated model.

News Impact Curves for S&P 500 Returns using Coefficients from GARCH and GJR Model Estimates:



14. GARCH-in Mean

- We expect a **risk to be compensated by a higher return**. So why not let the return of a security be partly determined by its risk?

- Engle, Lilien and Robins (1987) suggested the ARCH-M specification. A GARCH-M model would be

$$y_t = \mu + \delta\sigma_{t-1} + u_t \quad , u_t \sim \mathrm{N}(0,\sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta\sigma_{t-1}^2$$

- δ can be interpreted as a sort of **risk premium**.

- It is possible to combine all or some of these models together to get more complex "**hybrid**" models - e.g. an ARMA-EGARCH(1,1)-M model.

15. What Use Are GARCH-type Models?

- GARCH can model the volatility clustering effect since the conditional variance is autoregressive. Such models can be used to forecast volatility.

- We could show that

$$\text{Var}(y_t \mid y_{t-1}, y_{t-2}, \ldots) = \text{Var}(u_t \mid u_{t-1}, u_{t-2}, \ldots)$$

- So modelling σ_t^2 will give us models and forecasts for y_t as well.

Forecasting Variances using GARCH Models

- Producing conditional variance forecasts from GARCH models uses a very similar approach to producing forecasts from ARMA models.
- It is again an exercise in iterating with the conditional expectations operator.
- Consider the following GARCH(1,1) model:

$$y_t = \mu + u_t \quad , u_t \sim \mathrm{N}(0,\sigma_t^2),\ \sigma_t^{\,2} = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta\sigma_{t-1}^{\,2}$$

- What is needed is to generate all forecasts of $\sigma_{T+1}^{\,2} \mid \Omega_T$, $\sigma_{T+2}^{\,2} \mid \Omega_T$, ..., $\sigma_{T+s}^{\,2} \mid \Omega_T$ where Ω_T denotes all information available up to and including observation T.
- Adding one to each of the time subscripts of the above conditional variance equation, and then two, and then three would yield the following equations

$$\sigma_{T+1}^{\,2} = \alpha_0 + \alpha_1 u_T^{\,2} + \beta\sigma_T^{\,2} \ ,$$
$$\sigma_{T+2}^{\,2} = \alpha_0 + \alpha_1\, u_{T+1}^{\,2} + \beta\, \sigma_{T+1}^{\,2} \ ,$$
$$\sigma_{T+3}^{\,2} = \alpha_0 + \alpha_1\, u_{T+2}^{\,2} + \beta\, \sigma_{T+2}^{\,2}$$

Forecasting Variances using GARCH Models (Cont'd)

- Let $\sigma_{1,T}^{f\ 2}$ be the one step ahead forecast for σ^2 made at time T. This is easy to calculate since, at time T, the values of all the terms on the RHS are known.
- $\sigma_{1,T}^{f\ 2}$ would be obtained by taking the conditional expectation of the first equation at the bottom of slide 36:

$$\sigma_{1,T}^{f\ 2} = \alpha_0 + \alpha_1 u_T^2 + \beta\sigma_T^2$$

- Given, $\sigma_{1,T}^{f\ 2}$ how is $\sigma_{2,T}^{f\ 2}$, the 2-step ahead forecast for σ^2 made at time T, calculated? Taking the conditional expectation of the second equation at the bottom of slide 36:

$$\sigma_{2,T}^{f\ 2} = \alpha_0 + \alpha_1 \mathrm{E}(u_{T+1}^2 \mid \Omega_T) + \beta\,\sigma_{1,T}^{f\ 2}$$

- where $\mathrm{E}(u_{T+1}^2 \mid \Omega_T)$ is the expectation, made at time T, of u_{T+1}^2, which is the squared disturbance term.

Forecasting Variances using GARCH Models (Cont'd)

- We can write

$$E(u_{T+1}^2 \mid \Omega_t) = \sigma_{T+1}^2$$

- But σ_{T+1}^2 is not known at time T, so it is replaced with the forecast for it, $\sigma_{1,T}^{f\ 2}$, so that the 2-step ahead forecast is given by

$$\sigma_{2,T}^{f\ 2} = \alpha_0 + \alpha_1 \sigma_{1,T}^{f\ 2} + \beta \sigma_{1,T}^{f\ 2}$$

$$\sigma_{2,T}^{f\ 2} = \alpha_0 + (\alpha_1 + \beta)\sigma_{1,T}^{f\ 2}$$

- By similar arguments, the 3-step ahead forecast will be given by

$$\begin{aligned}\sigma_{3,T}^{f\ 2} &= E_T(\alpha_0 + \alpha_1 + \beta\sigma_{T+2}^2) \\ &= \alpha_0 + (\alpha_1 + \beta)\sigma_{2,T}^{f\ 2} \\ &= \alpha_0 + (\alpha_1 + \beta)[\ \alpha_0 + (\alpha_1 + \beta)\sigma_{1,T}^{f\ 2}\ ] \\ &= \alpha_0 + \alpha_0(\alpha_1 + \beta) + (\alpha_1 + \beta)^2\sigma_{1,T}^{f\ 2}\end{aligned}$$

- Any s-step ahead forecast ($s \geq 2$) would be produced by

$$h_{s,T}^f = \alpha_0 \sum_{i=1}^{s-1} (\alpha_1 + \beta)^{i-1} + (\alpha_1 + \beta)^{s-1} h_{1,T}^f$$

17. What Use Are Volatility Forecasts?

1. Option pricing

$$C = f(S, X, \sigma^2, T, r_f)$$

2. Conditional betas

$$\beta_{i,t} = \frac{\sigma_{im,t}}{\sigma^2_{m,t}}$$

3. Dynamic hedge ratios (page 546)

   The Hedge Ratio - the size of the futures position to the size of the underlying exposure, i.e. the number of futures contracts to buy or sell per unit of the spot good.

What Use Are Volatility Forecasts? (Cont'd)

- What is the optimal value of the hedge ratio?
- Assuming that the objective of hedging is to minimise the variance of the hedged portfolio, the optimal hedge ratio will be given by

$$h = p\frac{\sigma_S}{\sigma_F}$$

where h = hedge ratio

p = correlation coefficient between change in spot price (ΔS) and change in futures price (ΔF)

σ_S = standard deviation of ΔS

σ_F = standard deviation of ΔF

- What if the standard deviations and correlation are changing over time?

Use

$$h_t = p_t\frac{\sigma_{s,t}}{\sigma_{F,t}}$$