FM437: Financial Econometrics*

Lecture notes #1

Volatility modeling

For more information on this topic see Chapter 3 of Tsay (2005). Additional information can be found in the review of ARCH models given by Bollerslev, Engle and Nelson (1994) and Andersen, *et al.* (2005). Other text book treatments include Taylor (2005, Chapters 8-10), Brooks (2002, Chapter 8), Enders (2004, Chapter 3) and Christoffersen (2003, Chapter 2).

Risk plays a central role in financial decision making, and it is thus no surprise that a great deal of effort has been devoted to the study of the volatility of asset returns. This effort has paid large dividends: volatility modelling and forecasting methods have been shown to be very useful in many economic applications. In this chapter we will cover some of the most widely-used models for modelling volatility and discuss the estimation of these models, testing for correct specification of the models, and evaluating the forecasts produced by these models.

1. Predictability of asset return volatilities

Consider a standard ARMA(1,1) model for an asset return:

$$Y_{t+1} = \phi_0 + \phi_1 Y_t + \varepsilon_{t+1} + \theta\varepsilon_t$$
$$\varepsilon_{t+1} \sim WN\left(0, \sigma^2\right)$$

which implies that

$$\begin{aligned}
V_t\left[Y_{t+1}\right] &= V_t\left[\phi_0 + \phi_1 Y_t + \varepsilon_{t+1} + \theta\varepsilon_t\right] \\
&= V_t\left[\varepsilon_{t+1}\right], \text{ since the other terms are known at time } t \\
&= \sigma^2 \text{ by the assumption } \varepsilon_{t+1} \sim WN\left(0, \sigma^2\right)
\end{aligned}$$

So in standard models $V_t\left[Y_{t+1}\right] = \sigma^2$, a constant.

The squared residual can reveal information about volatility, and just as we considered the time series Y_t we can consider the series of squared residuals, ε_t^2, as a time series. So let

$$\begin{aligned}
\varepsilon_{t+1}^2 &= E_t\left[\varepsilon_{t+1}^2\right] + \eta_{t+1}, \eta_{t+1} \sim WN(0) \\
&= V_t\left[\varepsilon_{t+1}\right] + E_t\left[\varepsilon_{t+1}\right]^2 + \eta_{t+1}, \text{ by the definition of variance} \\
&= V_t\left[\varepsilon_{t+1}\right] + \eta_{t+1}, \text{ since } E_t\left[\varepsilon_{t+1}\right] = 0 \\
&= \sigma^2 + \eta_{t+1}, \text{ since } \varepsilon_{t+1} \sim WN\left(0, \sigma^2\right)
\end{aligned}$$

That is, the time series ε_{t+1}^2 can be decomposed into two parts: the conditional mean and a mean-zero white noise innovation series, denoted here η_{t+1}. If the conditional variance truly was constant, what would the ACF of ε_{t+1}^2 look like?

$$\begin{aligned}
\gamma_j &= Cov\left[\varepsilon_{t+1}^2, \varepsilon_{t+1-j}^2\right] \\
&= E\left[\left(\varepsilon_{t+1}^2 - E\left[\varepsilon_{t+1}^2\right]\right) \cdot \left(\varepsilon_{t+1-j}^2 - E\left[\varepsilon_{t+1-j}^2\right]\right)\right] \\
&= E\left[\left(\varepsilon_{t+1}^2 - \sigma^2\right) \cdot \left(\varepsilon_{t+1-j}^2 - \sigma^2\right)\right] \text{ since } E\left[\varepsilon_{t+1}^2\right] = E\left[E_t\left[\varepsilon_{t+1}^2\right]\right] = \sigma^2 \\
&= E\left[\eta_{t+1} \cdot \eta_{t+1-j}\right] \\
&= 0 \ \forall j \neq 0 \text{ since } \eta_{t+1} \text{ is white noise.}
\end{aligned}$$

Thus if Y_t has constant conditional variance, the ACF of ε_{t+1}^2 would be zero for all lags. Let us now check whether this is true empirically.

1.1. Empirical predictability of asset return volatilities

For illustration purposes we will consider the continuously compounded returns on a few example financial time series:

Name	**Frequency**	**Sample period**
DM/USD exchange rate	daily	2 January, 1991 to 15 November, 2000.
FTSE 100 index	daily	1 January, 1988 to 28 July, 2000.
US 3-month T-bill rate	daily	3 July, 1989 to 14 September, 2000.

Before examining the conditional variance of these returns we must first capture any dynamics in the conditional mean. (Above I considered an ARMA$(1,1)$ but this might not be the best model for all assets.) Let us consider ARMA(p,q) models, allowing p and q to range between 0 and 5. For each model I compute the Akaike and Bayesian Information Criteria (AIC and BIC). The 'optimal model' is the one that minimises the information criterion. The optimal choices of (p,q) and the R^2s corresponding the the optimal choices are presented below:

	Optimal model			
	AIC		**BIC**	
Data	(p,q)	R^2	(p,q)	R^2
Daily exchange rates	$(0,0)$	0	$(0,0)$	0
Daily stock returns	$(2,5)$	0.0128	$(0,1)$	0.0052
Daily interest rates	$(3,4)$	0.0084	$(0,0)$	0

For the exchange rate both the AIC and the BIC suggest $p=q=0$, which corresponds to assuming that the conditional mean is constant. For the stock return the AIC suggests an ARMA(2,5), which is quite a large model (8 parameters in total), and yielding a R^2 of 1.28%. The BIC chose a more parsimonious MA(1) model. For the interest rate the AIC selected an ARMA(3,4) model while the BIC suggested a model with just a constant term.

In Figure 1.1 I present the sample autocorrelation functions of the squared residuals from the optimal ARMA(p,q) model for each of these three series according to the BIC.

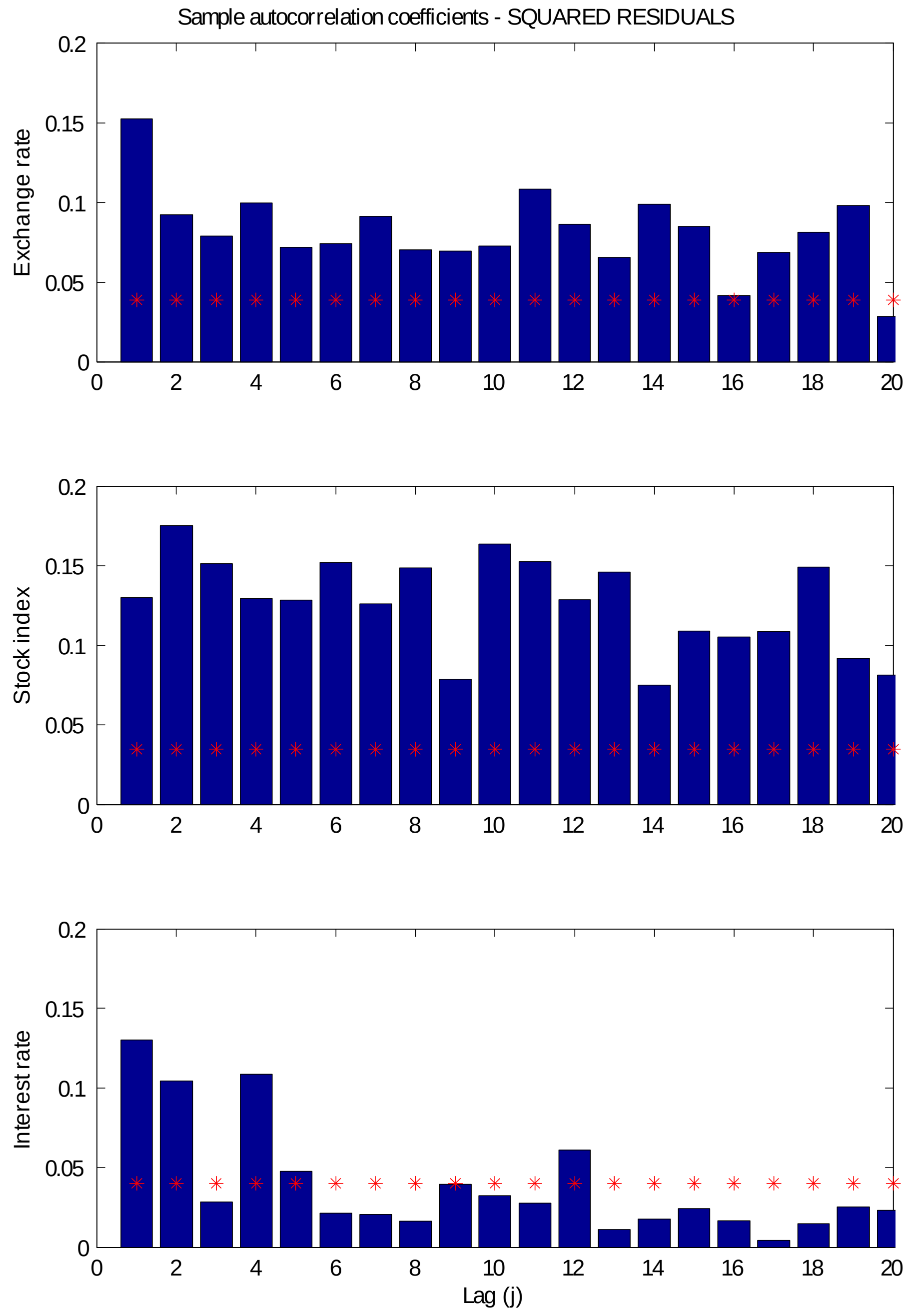

Figure 1.1: *Sample autocorrelation functions of the squared returns on the exchange rate, stock index and T-bill.*

The SACFs of the squared returns on the three assets clearly indicate significant serial correlation. This evidence is particularly strong for the stock index, and weakest for the interest rate (though still significant). This feature of financial time series gained attention with 2003 Nobel laureate Robert Engle's 1982 article, and is now one of the generally accepted stylised facts about asset returns: there is a substantial amount of predictability in return *volatility*. Studies of conditional volatility have formed a large part of the financial econometrics literature.

If we can somehow capture the predictability in volatility, we may be able to improve our portfolio decisions, risk management decisions, option pricing, amongst other things. We will now turn to a very popular and successful model for conditional variance: the ARCH model.

2. Autoregressive conditional heteroscedasticity (ARCH) processes

From our SACFs for squared returns we saw strong evidence of serial dependence. This suggests that the assumption of constant conditional variance of Y_{t+1}, or equivalently, the assumption of a constant conditional mean for ε_{t+1}^2, is false. If the conditional variance is not constant, how might we model it? A place to start might be an AR model for ε_{t+1}^2:

$$\varepsilon_{t+1}^2 = \omega + \alpha\varepsilon_t^2 + \eta_{t+1},\ \ \eta_{t+1} \sim WN\left(0\right)$$

What would this specification imply for the conditional variance function?

$$\begin{aligned}
V_t\left[Y_{t+1}\right] &= E_t\left[\varepsilon_{t+1}^2\right] \equiv \sigma_{t+1}^2, \text{ so} \\
\sigma_{t+1}^2 &= E_t\left[\omega + \alpha\varepsilon_t^2 + \eta_{t+1}\right] \\
&= \omega + \alpha\varepsilon_t^2 + 0, \text{ because } \eta_{t+1} \text{ is white noise} \\
\sigma_{t+1}^2 &= \omega + \alpha\varepsilon_t^2
\end{aligned}$$

The equation above is the famous ARCH(1) model of Engle (1982). It states that the conditional variance of tomorrow's return is equal to a constant, plus today's residual squared. This a simple and powerful model for capturing the predictability in volatility.

If an AR(1) model for ε_{t+1}^2 leads to an ARCH(1) model for the conditional variance, what would an ARMA(1,1) model for ε_{t+1}^2 (which is a more flexible specification) lead to?

$$\varepsilon_{t+1}^2 = \omega + \gamma\varepsilon_t^2 + \eta_{t+1} + \lambda\eta_t, \eta_{t+1} \sim WN(0)$$

What would this specification imply for the conditional variance function?

$$\begin{aligned}
V_t[Y_{t+1}] &= E_t\left[\varepsilon_{t+1}^2\right] \equiv \sigma_{t+1}^2, \text{ so} \\
\sigma_{t+1}^2 &= E_t\left[\omega + \gamma\varepsilon_t^2 + \eta_{t+1} + \lambda\eta_t\right] \\
&= \omega + \gamma\varepsilon_t^2 + \lambda\eta_t + 0, \text{ because } \eta_{t+1} \text{ is white noise} \\
&= \omega + \gamma\varepsilon_t^2 + \lambda\left(\varepsilon_t^2 - E_{t-1}\left[\varepsilon_t^2\right]\right), \text{ substituting in for } \eta_t \\
&= \omega + \gamma\varepsilon_t^2 + \lambda\left(\varepsilon_t^2 - \sigma_t^2\right) \\
&= \omega + (\gamma + \lambda)\varepsilon_t^2 - \lambda\sigma_t^2, \text{ re-defining the coefficients we obtain} \\
&= \omega + \alpha\varepsilon_t^2 + \beta\sigma_t^2
\end{aligned}$$

where we set $\alpha = (\gamma + \lambda)$ and $\beta = -\lambda$.

This is the famous GARCH(1,1) model due to Bollerslev (1986). The ARMA(1,1)-GARCH(1,1) model is a work-horse in financial time series analysis, which we can now write as:

$$\begin{aligned}
Y_{t+1} &= \mu_{t+1} + \varepsilon_{t+1},\ \varepsilon_{t+1} \sim WN\left(0, \sigma_{t+1}^2\right) \\
\mu_{t+1} &= E_t[Y_{t+1}] = \phi_0 + \phi_1 Y_t + \lambda\varepsilon_t \\
\sigma_{t+1}^2 &= V_t[Y_{t+1}] = \omega + \alpha\varepsilon_t^2 + \beta\sigma_t^2
\end{aligned}$$

We will cover the estimation of the parameters of this model in Section 5 below.

3. Stationarity, moments, and restrictions on parameters

In this course we will generally focus on "covariance stationary" time series, which are those where:

$$\begin{aligned} E[Y_t] &= \mu \ \forall \ t \\ V[Y_t] &= \sigma_y^2 \ \forall \ t \\ Cov[Y_t, Y_{t+j}] &= \gamma_j \ \forall \ t \end{aligned}$$

and so unconditional second moments are assumed to be constant through time. GARCH models are used for the conditional variance, but without further restrictions they can lead to a violation of covariance stationarity. The requirements for a GARCH(1,1) process to be covariance stationary are given below:

$$\begin{aligned} \text{Let } \ Y_{t+1} &= \mu_{t+1} + \varepsilon_{t+1} \\ \text{and } \ \varepsilon_{t+1} &= \sigma_{t+1}\nu_{t+1} \\ \nu_{t+1}|\mathcal{F}_t &\sim F(0,1) \\ \sigma_{t+1}^2 &= \omega + \beta\sigma_t^2 + \alpha\varepsilon_t^2 \end{aligned}$$

where $F(0,1)$ is some distribution with mean zero and variance one. Then we need:

$$\begin{aligned} &\text{Condition 1}: \omega > 0, \ \ \alpha, \beta \geq 0, \text{ for positive variance} \\ &\text{Condition 2}: \beta = 0 \text{ if } \alpha = 0, \text{ for identification} \\ &\text{Condition 3}: \alpha + \beta < 1, \text{ for covariance stationarity} \end{aligned}$$

If $\alpha + \beta < 1$ then

$$\begin{aligned} E\left[\sigma_{t+1}^2\right] &= \omega + \beta E\left[\sigma_t^2\right] + \alpha E\left[\varepsilon_t^2\right] \\ &= \omega + \beta E\left[\sigma_t^2\right] + \alpha E\left[E_{t-1}\left[\varepsilon_t^2\right]\right] \\ &= \omega + \beta E\left[\sigma_t^2\right] + \alpha E\left[\sigma_t^2\right], \text{ so} \\ E\left[\sigma_t^2\right] &= \frac{\omega}{1-\alpha-\beta} \end{aligned}$$

This quantity is interesting because:

$$E\left[\sigma_t^2\right] = E\left[E_{t-1}\left[\varepsilon_t^2\right]\right] = E\left[\varepsilon_t^2\right] = V\left[\varepsilon_t\right]$$

the unconditional variance of the residuals, ε_t. Note that this is only one part of the unconditional variance of Y_t :

$$\begin{aligned}
\mu &= E\left[Y_t\right] = E\left[E_{t-1}\left[Y_t\right]\right] = E\left[\mu_t\right] \\
\sigma_y^2 &= V\left[Y_t\right] \\
&= V\left[\mu_t + \varepsilon_t\right] \\
&= V\left[\mu_t\right] + V\left[\varepsilon_t\right] + 2Cov\left[\mu_t, \varepsilon_t\right] \\
&= V\left[\mu_t\right] + E\left[\sigma_t^2\right]
\end{aligned}$$

and so the unconditional variance of the returns is equal to the sum of the unconditional variance of the conditional mean term and the unconditional variance of the innovation term (which was computed above for the GARCH(1,1) case). If the conditional mean is constant, then the first term is zero and the unconditional variance of returns is equal to the unconditional variance of the residuals. In practise, the conditional mean of asset returns varies much less than the conditional variance of asset returns, and so the first term is close to zero. (The ratio of the second term to the first term has been estimated at something between 100 and 700.)

4. The "memory" of a GARCH model

The "memory" of a GARCH model measures how long a shock to the process takes to die out. We can get a measure of this by looking at the multi-step ahead estimate of the conditional variance. It is helpful firstly to show that we can re-write the GARCH(1,1) as a function of the unconditional variance, σ_y^2, rather than the constant ω.

$$\begin{aligned}
\sigma_{t+1,t}^2 &\equiv E_t\left[\varepsilon_{t+1}^2\right] \equiv \sigma_{t+1}^2 = \omega + \beta\sigma_{t,t-1}^2 + \alpha\varepsilon_t^2 \\
\text{Note that } \sigma_{t+1}^2 &= \omega + (\alpha+\beta)\,\sigma_y^2 + \beta\left(\sigma_{t,t-1}^2 - \sigma_y^2\right) + \alpha\left(\varepsilon_t^2 - \sigma_y^2\right) \\
&= \sigma_y^2 + \beta\left(\sigma_{t,t-1}^2 - \sigma_y^2\right) + \alpha\left(\varepsilon_t^2 - \sigma_y^2\right) \\
\text{Recalling that } \sigma_y^2 &= \frac{\omega}{1-\alpha-\beta}
\end{aligned}$$

This shows that the GARCH(1,1) forecast can be thought of as a weighted average of the unconditional variance, the deviation of last period's forecast from the unconditional variance, and the deviation of last period's squared residual from the unconditional variance. Next we work out the two-step ahead forecast:

$$\begin{aligned}
\sigma_{t+2,t}^2 &\equiv E_t\left[\varepsilon_{t+2}^2\right] \\
&= E_t\left[E_{t+1}\left[\varepsilon_{t+2}^2\right]\right] \text{ by the LIE} \\
&= E_t\left[\sigma_y^2 + \beta\left(\sigma_{t+1,t}^2 - \sigma_y^2\right) + \alpha\left(\varepsilon_{t+1}^2 - \sigma_y^2\right)\right] \\
&= \sigma_y^2 + \beta\left(\sigma_{t+1,t}^2 - \sigma_y^2\right) + \alpha\left(E_t\left[\varepsilon_{t+1}^2\right] - \sigma_y^2\right) \\
&= \sigma_y^2 + (\alpha+\beta)\left(\sigma_{t+1,t}^2 - \sigma_y^2\right)
\end{aligned}$$

Similarly, one can derive that

$$\begin{aligned}
\sigma_{t+3,t}^2 &\equiv E_t\left[\varepsilon_{t+2}^2\right] \\
&= E_t\left[E_{t+1}\left[\varepsilon_{t+3}^2\right]\right] \text{ by the LIE} \\
&= E_t\left[\sigma_y^2 + (\alpha+\beta)\left(\sigma_{t+2,t+1}^2 - \sigma_y^2\right)\right] \\
&= \sigma_y^2 + (\alpha+\beta)\left(E_t\left[\sigma_{t+2,t+1}^2\right] - \sigma_y^2\right) \\
&= \sigma_y^2 + (\alpha+\beta)\left(E_t\left[\sigma_y^2 + \beta\left(\sigma_{t+1,t}^2 - \sigma_y^2\right) + \alpha\left(\varepsilon_{t+1}^2 - \sigma_y^2\right)\right] - \sigma_y^2\right) \\
&= \sigma_y^2 + (\alpha+\beta)\left(\sigma_y^2 + \beta\left(\sigma_{t+1,t}^2 - \sigma_y^2\right) + \alpha\left(E_t\left[\varepsilon_{t+1}^2\right] - \sigma_y^2\right) - \sigma_y^2\right) \\
&= \sigma_y^2 + (\alpha+\beta)^2\left(\sigma_{t+1,t}^2 - \sigma_y^2\right)
\end{aligned}$$

Following similar arguments we can then show the general formula:

$$\sigma_{t+h,t}^2 = \sigma_y^2 + (\alpha+\beta)^{h-1}\left(\sigma_{t+1,t}^2 - \sigma_y^2\right),\ h \geq 1$$

The above expression shows that the forecast of the one-period volatility h periods from now is a weighted average of the unconditional variance and the deviation of the one-step forecast from the unconditional variance. If $\alpha + \beta < 1$ then the second term above goes to zero as $h \to \infty$, which implies that the longer our forecast horizon, the closer our forecast will get to the unconditional forecast, i.e. the unconditional variance. The size of $(\alpha + \beta)$ determines how quickly the predictability of the process dies out: if $(\alpha + \beta)$ is close to 0 then predictability will die out very quickly. If $(\alpha + \beta)$ is close to 1 then predictability will die out slowly.

An alternative way to think about volatility predictability is the "half-life" of a deviation of the conditional variance from the unconditional variance. The half-life is the number of periods, h^*, it takes for the conditional variance to revert back half-way towards the unconditional variance.

$$\sigma^2_{t+h^*,t} - \sigma^2_y = \frac{1}{2}\left(\sigma^2_{t+1,t} - \sigma^2_y\right)$$

This definition is applicable to all types of volatility models. For the GARCH(1,1) process in particular we have:

$$\begin{aligned} (\alpha+\beta)^{h^*-1}\left(\sigma^2_{t+1,t} - \sigma^2_y\right) &= \frac{1}{2}\left(\sigma^2_{t+1,t} - \sigma^2_y\right) \\ \text{so} \quad h^* &= 1 + \frac{\log(1/2)}{\log(\alpha+\beta)} \end{aligned}$$

In daily and monthly financial returns we usually see $(\alpha + \beta)$ between 0.8 and 0.99, leading to half-lives of

Half-life of a GARCH(1,1) process

$(\alpha+\beta)$	h^*
0.80	4.11
0.85	5.27
0.90	7.58
0.95	14.51
0.99	69.97
0.999	693.80

As $(\alpha + \beta) \to 1^-$ the process approaches a non-covariance stationary process, and the half-life diverges to infinity[1]. With daily asset returns we commonly see values of $(\alpha + \beta)$

[1] It is interesting to note, however, that GARCH(1,1) processes with $\alpha + \beta = 1$ (known as IGARCH processes) are strictly stationary even though they are not covariance stationary.

near one, and this prompted the development of so-called "long memory" volatility models. (A popular example of such a model is the fractionally integrated GARCH, or FIGARCH, model. We will not cover this class of models in this course.)

5. Maximum likelihood and quasi-maximum likelihood estimation

GARCH models are relatively easily estimated if we are willing to make an assumption about the distribution of ε_{t+1}. Making such an assumption means we can employ maximum likelihood. We could alternatively avoid making this assumption and estimate the model by the generalised method of moments (GMM) but this is not a common approach.

The most common distributional assumption is that of normality:

$$\varepsilon_{t+1}|\mathcal{F}_t \sim N\left(0, \sigma_{t+1}^2\right)$$

so the residuals are normally distributed. Combined with some model for the conditional mean, say an ARMA(1,1), this implies that the time series is conditionally normally distributed:

$$\begin{aligned}
Y_{t+1} &= \mu_{t+1} + \varepsilon_{t+1} \\
\mu_{t+1} &= E_t\left[Y_{t+1}\right] = \phi_0 + \phi_1 Y_t + \lambda\varepsilon_t \\
\sigma_{t+1}^2 &= V_t\left[Y_{t+1}\right] = \omega + \alpha\varepsilon_t^2 + \beta\sigma_t^2 \\
Y_{t+1}|\mathcal{F}_t &\sim N\left(\phi_0 + \phi_1 Y_t + \lambda\varepsilon_t, \omega + \alpha\varepsilon_t^2 + \beta\sigma_t^2\right)
\end{aligned}$$

We can estimate the unknown parameters $\boldsymbol{\theta} \equiv [\phi_0, \phi_1, \lambda, \omega, \alpha, \beta]'$ by maximum likelihood.

The likelihood for this model is:

$$L\left(Y_1, Y_2, ..., Y_t|\boldsymbol{\theta}\right) = \prod_{t=2}^{T} f\left(y_t|y_{t-1};\boldsymbol{\theta}\right)$$

$$= \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \cdot \exp\left\{-\frac{\varepsilon_t^2}{2\sigma_t^2}\right\}$$

$$\text{where } \varepsilon_t = Y_t - \phi_0 - \phi_1 Y_{t-1} - \lambda\varepsilon_{t-1}$$

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

$$\text{and we assume } \ \varepsilon_1 = 0 \text{ and } \sigma_1^2 = \frac{1}{T}\sum_{t=1}^{T}\left(Y_t - \bar{Y}\right)^2,$$

$$\text{where } \bar{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t, \text{ then}$$

$$\frac{1}{T-1}\log L\left(Y_1, Y_2, ..., Y_t|\boldsymbol{\theta}\right) = -\frac{1}{2}\log\left(2\pi\right) - \frac{1}{2\left(T-1\right)}\sum_{t=2}^{T}\log\sigma_t^2 - \frac{1}{2\left(T-1\right)}\sum_{t=2}^{T}\frac{\varepsilon_t^2}{\sigma_t^2}$$

No analytical solution for the maximum likelihood estimates is available and so we instead us numerical methods to maximise the likelihood. This can be done using Matlab, or one of many other standard econometric software packages such as EViews.

It should be pointed out that there are many different ways we could deal with ε_1 and σ_1^2. Above we simply set them equal to their unconditional sample means. Alternatively we could estimate them as additional parameters, though these parameters are not identified asymptotically. We could also instead set them equal to their expected values for given parameters: $\varepsilon_1 = 0$ and $\sigma_1^2 = \omega/\left(1 - \alpha - \beta\right)$. For long time series the impact of the method used to deal with observation $t = 1$ tends not to be large.

Numerous researchers have noted that asset returns are not normally distributed. Part of this non-normality comes from the fact that volatility is time-varying, but that turns out not to explain all of the non-normality. This has prompted many researchers to consider alternative distributional specifications for ε_t:

$$\varepsilon_{t+1} \equiv \sigma_{t+1}\nu_{t+1}$$

$$\nu_{t+1}|\mathcal{F}_t \sim iid\ F\left(0, 1; \kappa\right)$$

where $F(0,1;\kappa)$ is some distribution with mean zero, variance one, and "shape" parameters κ. Common choices in finance for this density are the Student's t, the generalized error disitribution (GED) and the skewed Student's t. To estimate such a model we can again use maximum likelihood:

$$L(y_1, y_2, ..., y_t|\boldsymbol{\theta},\kappa) = \prod_{t=2}^{T} f(y_t|y_{t-1};\boldsymbol{\theta},\kappa)$$

$$\log L(y_1, y_2, ..., y_t|\boldsymbol{\theta},\kappa) = \sum_{t=2}^{T} \log f(y_t|y_{t-1};\boldsymbol{\theta},\kappa)$$

However, if we are *only* interested in estimating the conditional variance it turns out that using the normal distribution in maximum likelihood estimation will give us consistent parameter estimates *even if the true density is non-normal.* This is a very useful result; it was shown by Bollerslev and Wooldridge (1992) and is based on work by Gourieroux, *et al.* (1984). This estimator is known as the quasi (or pseudo) maximum likelihood estimator (QMLE or PMLE).

The "no free lunch" rule arises through the fact that QMLEs are not efficient unless the true density actually *is* normal. If, for example, we know that the true density is a Student's t with some unknown degrees of freedom κ, then we can get a more efficient estimate than QMLE by using the MLE. However, if the true density turns out to something other than a Student's t, then the MLE based on the assumption of a Student's t distribution is not consistent (and hence efficiency is not considered) whereas the QMLE will at least be consistent.

If we are interested in quantities other than just the mean and the variance, for example Value-at-Risk, tail probabilities, skewness or kurtosis, then the choice of density is important and we should try hard to find the best fitting density. But if not, QML results suggest that using the normal density for estimation is a good choice.

6. Standard errors for GARCH models

Most software packages provide standard errors for the parameters of volatility models, and here we will review how they are obtained for MLEs and QMLEs. Consider the generic maximum likelihood estimator:

$$\underset{(k\times 1)}{\hat{\boldsymbol{\theta}}_T} \equiv \arg\max_{\boldsymbol{\theta}\in\Theta} \frac{1}{T}\sum_{t=1}^{T}\log f\left(y_t|y_{t-1},y_{t-2},...;\boldsymbol{\theta}\right)$$

Under standard regularity conditions, see Hamilton (1994) or Hayashi (2000) for details, we know

$$\hat{V}_T^{-1/2}\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\right)\to^{D} N\left(0,I\right)$$

and so we may write

$$\underset{(k\times 1)}{\hat{\boldsymbol{\theta}}_T} \overset{a}{\sim} N\left(\underset{(k\times 1)}{\boldsymbol{\theta}_0}, \frac{1}{T}\cdot\underset{(k\times k)}{\hat{V}_T}\right)$$

where $\boldsymbol{\theta}_0$ is the true parameter vector and $\hat{V}_T$ is a consistent estimator of the true asymptotic covariance matrix, denoted V_0, of the estimator. In general estimation, V_0 has a "sandwich" form:

$$\begin{aligned}
V_0 &= A_0^{-1}B_0A_0^{-1} \\
\text{where}\quad A_0 &\equiv -E\left[H_t\left(\boldsymbol{\theta}_0\right)\right]\equiv -E\left[\underset{(k\times k)}{\frac{\partial^2\log f\left(y_t;\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}}\right] \\
B_0 &\equiv E\left[\underset{(k\times 1)}{\frac{\partial\log f\left(y_t;\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}}}\underset{(1\times k)}{\frac{\partial\log f\left(y_t;\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}'}}\right]
\end{aligned}$$

assuming that $\partial\log f\left(y_t;\boldsymbol{\theta}_0\right)/\partial\boldsymbol{\theta}$ is serially uncorrelated (which is a common assumption). The matrix A_0 is also known as the "information matrix" (or "Fisher's information matrix"), and H_t is the "hessian" matrix. The matrix B_0 is sometimes called the OPG matrix (as it is the expectation of the outer-product of the gradients). These quantities are estimated using their sample counterparts:

$$\begin{aligned}
\hat{A}_T &\equiv -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\log f\left(y_t;\hat{\boldsymbol{\theta}}_T\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \\
\hat{B}_T &\equiv \frac{1}{T}\sum_{t=1}^{T}\frac{\partial\log f\left(y_t;\hat{\boldsymbol{\theta}}_T\right)}{\partial\boldsymbol{\theta}}\frac{\partial\log f\left(y_t;\hat{\boldsymbol{\theta}}_T\right)}{\partial\boldsymbol{\theta}'} \\
\hat{V}_T &\equiv \hat{A}_T^{-1}\hat{B}_T^{-1}\hat{A}_T^{-1}
\end{aligned}$$

If the density selected for the estimation is correct, then it is possible to show that the "information matrix equality" holds:

$$A_0 = -B_0$$

and so the asymptotic covariance matrix simplifies to

$$V_0 = -A_0^{-1} A_0 A_0^{-1} = -A_0^{-1} = B_0^{-1}$$
$$\text{and } \hat{V}_T = -\hat{A}_T^{-1} = \hat{B}_T^{-1}$$

This implies that, under the assumption that the density is correct, we can estimate the covariance matrix of the parameters either using the (negative of) the inverse of the expected Hessian matrix, $-\hat{A}_T^{-1}$, or using the mean of the outer-product of the scores, $\hat{B}_T$, or using the sandwich estimator.

However, in quasi-maximum likelihood estimation we do not assume that the normal density is correct, and in that case the information matrix equality does not hold. This means we must use the sandwich estimator of the covariance matrix. This is sometimes called the "robust" covariance matrix (because it is a consistent estimator even if the true density is non-normal) or, for volatility models, the Bollerslev-Wooldridge (1992) covariance matrix estimator.

7. Dealing with the conditional mean

It is important to capture conditional mean dynamics before testing for volatility clustering or estimating a conditional variance model. A mis-specified conditional mean model (for example, one that does not capture all the dynamics in the conditional mean) will usually lead to a mis-specified conditional variance model, see Lumsdaine and Ng (1999) for example. We will show this for a particular example below.

In some cases, however, such as intra-daily returns, the conditional mean is often assumed constant and near zero. If this is true, then analysing the squared residual from an ARMA model for daily asset returns, for example, is approximately the same as analysing the squared return itself. That is:

$$\begin{aligned} V_t\left[Y_{t+1}\right] &= E_t\left[Y_{t+1}^2\right] - E_t\left[Y_{t+1}\right]^2 \\ &\approx E_t\left[Y_{t+1}^2\right] \text{ if } E_t\left[Y_{t+1}\right] \approx 0 \end{aligned}$$

The squared asset return is commonly used as a proxy for the true volatility of the asset return. For lower frequency returns (daily, weekly, monthly) the assumption of a constant and zero conditional mean is generally less palatable. For some, but definitely not all, assets it may be reasonable at the daily frequency: for example, the BIC selected a constant conditional mean for our exchange rate and interest rate data sets, but not for our stock return data set. At the weekly or monthly frequency the conditional mean should always be allowed to be non-zero.

Now let us consider a specific example analytically, to emphasise this point. Let the true process for Y_t be:

$$\begin{aligned} Y_{t+1} &= \phi Y_t + \varepsilon_{t+1}, \text{ where } |\phi| < 1 \\ \varepsilon_t &\sim iid\ N\left(0, \sigma^2\right) \end{aligned}$$

The sample autocorrelation function would indicate that a good model for the conditional mean is:

$$Y_{t+1} = \beta_0 + \beta_1 Y_t + e_{t+1}$$

With enough data the OLS would yield $\hat{\beta}_0 = 0$ and $\hat{\beta}_1 = \phi$. But what would happen if instead the researcher observed that the unconditional mean of this variable was zero, and then assumed that the *conditional* mean of this variable is also zero? Such a researcher would effectively impose that $\beta_0 = \beta_1 = 0$, and so the residuals from the regression would simply be:

$$e_{t+1} = Y_{t+1}$$

What would be the first-order autocorrelation in the squared residuals?

$$
\begin{aligned}
e_{t+1}^2 &= Y_{t+1}^2 = (\phi Y_t + \varepsilon_{t+1})^2 \\
&= \phi^2 Y_t^2 + \varepsilon_{t+1}^2 + 2\phi Y_t \varepsilon_{t+1} \\
E\left[e_{t+1}^2\right] &= E\left[Y_{t+1}^2\right] \\
&= E\left[\phi^2 Y_t^2 + \varepsilon_{t+1}^2 + 2\phi Y_t \varepsilon_{t+1}\right] \\
&= \phi^2 E\left[Y_t^2\right] + E\left[\varepsilon_{t+1}^2\right] + 2\phi E\left[Y_t \varepsilon_{t+1}\right] \\
&= \phi^2 E\left[Y_t^2\right] + \sigma^2, \text{ since } \varepsilon_{t+1} \text{ is } iid\ N\left(0, \sigma^2\right) \\
\text{so } E\left[e_{t+1}^2\right] &= \frac{\sigma^2}{1-\phi^2} \\
E\left[e_{t+1}^2 e_t^2\right] &= E\left[Y_t^2 \left(\phi Y_t + \varepsilon_{t+1}\right)^2\right] \\
&= E\left[Y_t^2 \left(\phi^2 Y_t^2 + \varepsilon_{t+1}^2 + 2\phi Y_t \varepsilon_{t+1}\right)\right] \\
&= \phi^2 E\left[Y_t^4\right] + E\left[Y_t^2 \varepsilon_{t+1}^2\right] + 2\phi E\left[Y_t^3 \varepsilon_{t+1}\right] \\
&= \phi^2 \kappa_4 + E\left[Y_t^2\right] E\left[\varepsilon_{t+1}^2\right] + 2\phi E\left[Y_t^3\right] E\left[\varepsilon_{t+1}\right] \\
&= \frac{\phi^2 \kappa_4 \left(1-\phi^2\right) + \sigma^4}{1-\phi^2}
\end{aligned}
$$

where $\kappa_4 \equiv E\left[Y_t^4\right]$. (It turns out that we do not need to work this moment out.)

$$
\begin{aligned}
\text{so } Cov\left[e_{t+1}^2, e_t^2\right] &= E\left[e_{t+1}^2 e_t^2\right] - E\left[e_{t+1}^2\right]^2 \\
&= \frac{\phi^2\kappa_4\left(1-\phi^2\right)+\sigma^4}{1-\phi^2} - \left(\frac{\sigma^2}{1-\phi^2}\right)^2 \\
&= \frac{\phi^2\kappa_4\left(1-\phi^2\right)^2+\sigma^4\left(1-\phi^2\right)-\sigma^4}{\left(1-\phi^2\right)^2} \\
&= \frac{\phi^2\kappa_4\left(1-\phi^2\right)^2-\phi^2\sigma^4}{\left(1-\phi^2\right)^2} \\
V\left[e_{t+1}^2\right] &= V\left[Y_{t+1}^2\right] \\
&= E\left[Y_{t+1}^4\right] - E\left[Y_{t+1}^2\right]^2 \\
&= \frac{\kappa_4\left(1-\phi^2\right)^2-\sigma^4}{\left(1-\phi^2\right)^2} \\
Corr\left[e_{t+1}^2, e_t^2\right] &= \frac{Cov\left[e_{t+1}^2, e_t^2\right]}{V\left[e_{t+1}^2\right]} \\
&= \frac{\phi^2\kappa_4\left(1-\phi^2\right)^2-\phi^2\sigma^4}{\left(1-\phi^2\right)^2} . \frac{\left(1-\phi^2\right)^2}{\kappa_4\left(1-\phi^2\right)^2-\sigma^4} \\
&= \phi^2
\end{aligned}
$$

Thus although the variable Y_{t+1} is truly homoscedastic, if we use a constant mean model when it is truly an AR(1) (with $\phi \neq 0$) we will find evidence of volatility clustering. By similar arguments it is possible that the wrong *type* of volatility model is selected if the conditional mean model is mis-specified: for example, the true conditional variance process might be a GARCH(1,1) but the use of a mis-specified conditional mean model may lead to an EGARCH model being selected. This should emphasise the importance of correct modelling of the conditional mean *prior* to modelling the conditional variance or testing for volatility clustering.

8. Testing for volatility clustering

Before going to the trouble of specifying and estimating a volatility model it is a good idea to test for the presence of volatility clustering the data. Two simple tests are available. McLeod and Li (1983) suggest using the Ljung-Box (LB) test on the squared residuals (or squared returns, if they have conditional mean zero) to test jointly for evidence of serial correlation, while an alternative, similar, test is the Lagrange multiplier test of Engle (1982).

The Ljung-Box test is a joint test that all autocorrelation coefficients of a particular series out to lag L are zero. This the Ljung-Box test is for the following null and alternative hypotheses:

$$\begin{aligned} & H_0 : \rho_1 = \rho_2 = ...\rho_L = 0 \\ \text{vs.} \quad & H_a : \rho_j \neq 0 \text{ for some } j = 1, 2, ..., , L \end{aligned}$$

The Ljung-Box Q-statistic, denoted $Q_{LB}(L)$, is:

$$\begin{aligned} Q_{LB}(L) &= T(T+2)\sum_{j=1}^{L}\left(\frac{1}{T-j}\right)\hat{\rho}_j^2 \\ \text{where} \quad \hat{\rho}_j &= \frac{\widehat{Cov}[Y_t, Y_{t-j}]}{\hat{V}[Y_t]} \\ &= \frac{\frac{1}{T-j}\sum_{t=j+1}^{T}(y_t - \bar{y})(y_{t-j} - \bar{y})}{\frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2} \end{aligned}$$

The $Q_{LB}(L)$ statistic is simply a weighted sum of the squared autocorrelation coefficients, with j ranging from 1 to L. Under the null hypothesis, the $Q_{LB}(L)$ statistic is distributed as χ_L^2. With a chi-squared test we reject the null hypothesis if the test statistic (in our case $Q_{LB}(L)$) is larger than the 95% critical value of a χ_L^2 random variable.

The Lagrange multiplier test of Engle (1982) (known as the ARCH LM test), which involves running the regression:

$$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + ... + \alpha_L e_{t-L}^2 + u_t$$

and then performing a χ^2_L test (or an F-test) of the hypothesis:

$$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_L = 0$$

The ARCH LM test has the advantage that robust standard errors can be used to test the above null hypothesis, whereas the McLeod-Li/Ljung-Box test is based on the assumption that the process is *iid* (which is generally not realistic for squared residuals).

If we apply the ARCH LM test to the three data series considered above, using the ARMA(p,q) models suggested by the BIC model selection criterion for the conditional mean, we obtain the following test statistics:

ARCH LM tests for volatility clustering			
	Lag length, L		
Series	5	10	20
95% Critical value	*11.07*	*18.31*	*31.41*
DM/USD exchange rate	95.70	118.95	159.72
FTSE 100 index	211.61	291.80	345.10
US 3-month T-bill rate	82.47	85.69	94.12

Thus we have significant statistical evidence of volatility clustering in all three time series. Having found significant evidence of volatility clustering, we now estimate a GARCH(1,1) model on each of these series. The results are presented below:

GARCH(1,1) parameter estimates			
	ω	α	β
DM/USD exchange rate	0.005	0.039	0.949
FTSE 100 index	0.002	0.047	0.943
US 3-month T-bill rate	0.022	0.162	0.762

9. Extensions of the ARCH model

We have already considered the most widely used (and the first, it turns out) extension of the ARCH model, that is, the Generalized ARCH, or GARCH model of Bollerslev (1986).

9.1. Models with a "leverage effect"

Black (1976) was perhaps the first to observe that stock returns are negatively correlated with changes in volatility: that is, volatility tends to rise (or rise more) following bad news (a negative return) and fall (or rise less) following good news (a positive return). This is called the "leverage effect", as firms' use of leverage can provide an explanation for this correlation: if a firm uses both debt and equity then as the stock price of the firm falls its debt-to-equity ratio rises. This will raise equity return volatility if the firm's cashflows are constant. The leverage effect has since been shown to provide only a partial explanation to observed correlation, but the name persists.

Given the above correlation we might expect that negative returns today lead to higher volatility tomorrow than do positive returns. This behaviour cannot be captured by a standard GARCH model:

$$\sigma_{t+1}^2 = \omega + \beta\sigma_t^2 + \alpha\varepsilon_t^2$$

which shows that tomorrow's volatility is quadratic in today's return, so the sign of today's return does not matter. The simplest extension to accommodate this relation is the model of Glosten, Jagannathan and Runkle (1993) (so-called GJR-GARCH, sometimes known as Threshold-GARCH):

$$GJR\text{-}GARCH : \sigma_{t+1}^2 = \omega + \beta\sigma_t^2 + \alpha\varepsilon_t^2 + \delta\varepsilon_t^2\mathbf{1}\left\{\varepsilon_t < 0\right\}$$

If $\delta > 0$ then the impact on tomorrow's volatility of today's return is greater if today's return is negative.

One way of illustrating the difference between a volatility models is via their "*news impact curves*", see Engle and Ng (1993). This curve plots σ_{t+1}^2 as the value of ε_t varies, leaving everything else in the model fixed, and normalising the function to equal zero when $\varepsilon_t = 0$. This is illustrated in Figure 9.1. The news impact curve for a standard GARCH model is simply the function $\alpha\varepsilon_t^2$, while for the GJR-GARCH model it is $\alpha\varepsilon_t^2+\delta\varepsilon_t^2\mathbf{1}\left\{\varepsilon_t < 0\right\}$.

The results from estimating a GJR-GARCH(1,1) on the three series analysed above are presented below. Parameters that are significantly different from zero at the 5% level

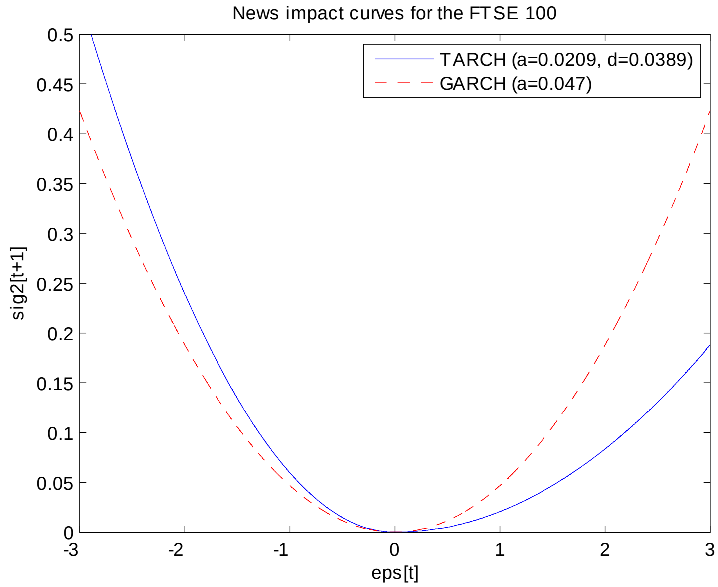

Figure 9.1: *The news impact curves of a threshold GARCH and a standard GARCH model.*

are denoted with an asterisk. From this table we see that the asymmetry paramter, δ, is significant for the stock return and the interest rate, but not for the exchange rate. Notice that the significant δ estimates are both positive, indicating that negative shocks lead to higher future volatility than do positive shocks of the same magnitude.

GJR-GARCH(1,1) parameter estimates				
	ω	α	β	δ
DM/USD exchange rate	0.0054*	0.0397*	0.9490*	−0.0014
FTSE 100 index	0.0015*	0.0209*	0.9496*	0.0389*
US 3-month T-bill rate	0.0233*	0.0940*	0.7498*	0.1535*

The other widely-used GARCH model that allows for leverage effects is the "exponential GARCH", or "EGARCH" model of Nelson (1991). Nelson proposed this model to remedy two shortcomings of the standard GARCH model. The first is its inability to capture the leverage effect, and the second is the conditions have to be imposed on the parameters of the GARCH model to ensure a positive volatility estimate. The EGARCH model deals with both:

$$\ln \sigma_{t+1}^2 = \omega + \beta \ln \sigma_t^2 + \alpha \left| \frac{\varepsilon_t}{\sigma_t} \right| + \gamma \frac{\varepsilon_t}{\sigma_t}$$

By modelling $\ln \sigma_{t+1}^2$ rather than σ_{t+1}^2 we are ensured a positive estimate of σ_{t+1}^2. Further by allowing γ to differ from zero the leverage effect can be captured. The parameter estimates for EGARCH(1,1) models estimated on our three data sets are presented below.

EGARCH(1,1) parameter estimates				
	ω	α	β	γ
DM/USD exchange rate	-0.0838*	0.0934*	0.9848*	-0.0047
FTSE 100 index	-0.0944*	0.0923*	0.9891*	-0.0335*
US 3-month T-bill rate	-0.2276*	0.2195*	0.9582*	-0.0754*

Again, parameters that are significantly different from zero are denoted with an asterisk. Similar to the results we obtained using the TARCH model, the EGARCH model results indicate that the asymmetry parameter, γ in this case, is significant for the stock return and the interest rate but not for the exchange rate. Here notice that all γ estimates are negative, which again implies that negative shocks lead to higher future volatility than do positive shocks of the same magnitude.

9.2. ARCH-in-mean model

A key idea in finance is that the return on a risky security should be positively related to its risk. This lead Engle, Lilien and Robins (1987) to develop the "ARCH in mean", or "ARCH-M", model which posits that the conditional mean of a return is dependent on some function of its conditional variance or conditional standard deviation. The particular function of conditional variance that enters the conditional mean (ie, level of variance, standard deviation, log-variance, etc) is left to the researcher. Using the standard deviation has the benefit that it is in the same units as returns, and so parameters are unaffected by scale (such as multiplying returns by 100).

If we thought that the mean had an AR(1) term, for example, then we might use one of the following models:

$$r_{t+1} = \phi_0 + \phi_1 r_t + \gamma\sigma_{t+1}^2 + \varepsilon_{t+1}, \text{ or}$$
$$r_{t+1} = \phi_0 + \phi_1 r_t + \gamma\sigma_{t+1} + \varepsilon_{t+1}, \text{ or}$$
$$r_{t+1} = \phi_0 + \phi_1 r_t + \gamma \ln \sigma_{t+1} + \varepsilon_{t+1}, \text{ with}$$
$$\sigma_{t+1}^2 = \omega + \beta\sigma_t^2 + \alpha\varepsilon_t^2$$

Estimating the ARCH-M models above (with the conditional standard deviation added to the mean equation) for the three series studied so far yielded the following parameter estimates. From this table we see that the "ARCH-in-mean" term is rarely significant: γ is only significantly different from zero for the T-bill series. This is not uncommon; many applications of the ARCH-M model find that the volatility term in the mean equation is not significant, perhaps due to the imprecision with which the ARCH model estimates the conditional variance.

AR(1)-GARCH-M(1,1) parameter estimates						
	ϕ_0	ϕ_1	γ	ω	α	β
DM/USD exchange rate	0.000	0.034	0.025	0.006*	0.040*	0.948*
FTSE 100 index	-0.012	0.069*	0.094	0.002*	0.048*	0.940*
US 3-month T-bill rate	-0.057*	0.026	0.159*	0.013*	0.137*	0.814*

9.3. NARCH, SQARCH, PARCH, QARCH, STARCH, APARCH...

The line of research that started with a humble ARCH model and the plain vanilla GARCH(1,1), has since been extended in numerous directions. Some of the many flavours of GARCH models are given below. The notation we will use is:

$$Y_{t+1} = \mu_{t+1} + \varepsilon_{t+1}$$
$$\varepsilon_{t+1} = \sigma_{t+1}\nu_{t+1}$$
$$\nu_{t+1}|\mathcal{F}_t \sim iid\ F(0,1)$$

IGARCH (Engle and Bollerslev, 1986)

$$\sigma_{t+1}^2 = \omega + \beta\sigma_t^2 + (1-\beta)\,\varepsilon_t^2$$

PARCH (Ding, Granger and Engle, 1993)

$$\sigma_{t+1}^{\gamma} = \omega + \beta\sigma_t^{\gamma} + \alpha\varepsilon_t^{2\gamma}$$

A**PARCH** (Ding, Granger and Engle, 1993)

$$\sigma_{t+1}^{\gamma} = \omega + \beta\sigma_t^{\gamma} + \alpha\gamma\left(|\varepsilon_t| - \delta\varepsilon_t\right)^{\gamma}$$

PARCH (Engle and Bollerslev, 1986)

$$\sigma_{t+1}^{2} = \omega + \beta\sigma_t^{2} + \alpha\varepsilon_t^{\delta}$$

SQR-GARCH (Taylor, 1986 and Schwert, 1989)

$$\sigma_{t+1} = \omega + \beta\sigma_t + \alpha\left|\varepsilon_t\right|$$

QGARCH (Sentana, 1991)

$$\sigma_{t+1}^{2} = \omega + \beta\sigma_t^{2} + \alpha\varepsilon_t^{2} + \delta\varepsilon_t$$

NARCH (Higgins and Bera, 1992)

$$\sigma_{t+1}^{2} = \left(\phi_0\omega^{\delta} + \phi_1\varepsilon_t^{2\delta} + \phi_2\varepsilon_{t-1}^{2\delta} + ... + \phi_p\varepsilon_{t-p+1}^{2\delta}\right)^{1/\delta}$$

All-in-the-family GARCH (Hentschel, 1995)

$$\frac{\sigma_{t+1}^{\gamma} - 1}{\gamma} = \omega + \alpha\sigma_t^{\gamma}\left[\left|\nu_t - \delta\right| - \lambda\left(\nu_t - \delta\right)\right] + \beta\frac{\sigma_t^{\gamma} - 1}{\gamma}$$

SQARCH (Ishida and Engle, 2001)

$$\sigma_{t+1}^{2} = \omega + \beta\sigma_t^{2} + \alpha\sigma_t\left(\nu_t^{2} - 1\right)$$

9.4. Does anything beat a GARCH(1,1)?

In the two decades or so since the first ARCH model was proposed over a dozen extensions have been published. With increased computing power has come increasingly complex volatility models. A very reasonable question to ask is "does anything beat the benchmark GARCH(1,1) volatility model". A recent paper by Hansen and Lunde (2005) addresses this question. They consider a total of 330 different ARCH-type models for the Deutsche mark-U.S. dollar exchange rate and for IBM equity returns. For the exchange rate they find no evidence against the simple GARCH(1,1).

For the equity return they find that the "Asymmetric Power GARCH(2,2)", or APARCH(2,2), model performed best. This model is:

$$\sigma_{t+1}^{\delta} = \omega + \sum_{i=1}^{2} \alpha_i \left(|\varepsilon_t| - \gamma_i \varepsilon_t \right)^{\delta} + \sum_{j=1}^{2} \beta_i \sigma_t^{\delta}$$

The APARCH model is one of the most complicated in use. It allows for a leverage effect (when $\gamma \neq 0$). Allowing δ to differ from 2 enables the model to use the fact that serial correlation in the absolute value of returns to the power δ (with $\delta < 2$) tends to be stronger than that in squared returns.

We will discuss the ways Hansen and Lunde (2005) measured the performance of these volatilty models below.

10. Choosing a volatility model

We've shown above that there are numerous volatility models available to a financial forecaster. How can we choose the "best" one? Choosing a volatility model should depend on what we intend to do with it. For example, if we intend to use it for out-of-sample forecasting of volatility, which is perhaps the most common use for a volatility model, then the correct way to choose a volatility model should be according to some measure of its out-of-sample forecast performance. If we instead want to determine whether there is statistical evidence of a leverage effect, then we should pick the model that gives the best in-sample fit and thus (hopefully) the most precise parameter estimates. Here let us focus on choosing a model by in-sample goodness-of-fit.

Below we will again use the notation:

$$
\begin{aligned}
Y_{t+1} &= \mu_{t+1} + \varepsilon_{t+1} \\
\varepsilon_{t+1} &= \sigma_{t+1}\nu_{t+1} \\
\nu_{t+1}|\mathcal{F}_t &\sim iid\ F\left(0,1\right)
\end{aligned}
$$

where F is some unspecified distribution with mean zero and variance one.

10.1. Comparing nested models via tests on parameters

If the two models being compared are "nested", in the sense that for certain choices of paramters the models are identical, then we can conduct statistical tests to see if the models are significantly different. The simplest example of this is comparing a GJR-GARCH model with a GARCH model:

$$
\begin{aligned}
\textit{GJR-GARCH} : \sigma_{t+1}^2 &= \omega + \beta\sigma_t^2 + \alpha\varepsilon_t^2 + \delta\varepsilon_t^2 \mathbf{1}\left\{\varepsilon_t < 0\right\} \\
\textit{GARCH} : \sigma_{t+1}^2 &= \omega + \beta\sigma_t^2 + \alpha\varepsilon_t^2
\end{aligned}
$$

When $\delta = 0$ the GJR-GARCH is simply the GARCH model, and so the GJR-GARCH "nests" the GARCH model. Using the formulas in Section 6 to compute the standard errors for the GJR-GARCH parameters we can test:

$$
\begin{aligned}
& H_0 : \delta = 0 \\
\text{vs.} \quad & H_a : \delta \neq 0
\end{aligned}
$$

If we reject the null hypothesis then we conclude that the GJR-GARCH is significantly better than the GARCH model, at least in-sample. In out-of-sample forecast comparisons it is often the case that more parsimonious models perform best, even if a more flexible model is significantly better in-sample. If the more flexible model is not significantly better in-sample (e.g. if we fail to reject H_0) then it is very unlikely to do better out-of-sample.

10.2. Using information criteria

As we noted earlier, measures of performance that do not account for the number of parameters in the model will generally just pick the largest model. But in forecasting extra parameters can lead to increased estimation error and worsened forecast performance. Thus it is important to incorporate some sort of trade-off between increased goodness-of-fit and increased estimation error. The Akaike information criterion (AIC), Hannan-Quinn information criterion (HQIC) and Schwarz's Bayesian information criterion (BIC) may be used to accomplish this. The formulas for these can be expressed either as a function of the sum of squared residuals (for conditional mean modelling) or more generally as a function of the <u>mean</u> log-likelihood at the optimum, $\mathcal{L}$, the sample size, T, and the number of parameters, k. For volatility forecasting the general formulas are the correct ones to use:

$$\begin{aligned} AIC &= -2\mathcal{L} + \frac{2}{T}k \\ HQIC &= -2\mathcal{L} + \frac{2\log(\log(T))}{T}k \\ BIC &= -2\mathcal{L} + \frac{\log(T)}{T}k \end{aligned}$$

As before, the first terms in these expressions represent the goodness-of-fit, and the second terms represent a penalty for extra parameters. We want to choose the volatility model that yields the smallest information criterion.

Of course, as in forecasts of the variable itself, it is possible (and indeed preferable) to instead compute the pseudo out-of-sample forecasting performance of the competing models. This can be done using a rolling, expanding or fixed window of data, as in the case previously considered. Information criteria can be used to obtain a fast, approximate estimate of out-of-sample forecast performance, but an actual out-of-sample test is preferable if time and computational power are available.

10.3. Using statistical goodness-of-fit measures

When estimating standard econometric models we would usually evaluate goodness-of-fit by the R^2, which corresponds to ranking competing models by their MSE:

$$MSE = \frac{1}{T}\sum_{t=1}^{T} e_t^2$$

where e_t is the residual from the model for the conditional mean. We would choose the model with the highest R^2 , or the lowest MSE. What would be the right way to measure the accuracy of a volatility model?

Evaluating accuracy requires some knowledge of the realised value of the variable of interest, in this case the conditional variance. But the conditional variance is not observable even *ex-post*, and so we must instead rely on *volatility proxies*. The simplest volatility proxy is the squared residual (or squared return, if we assume that returns have zero mean)

$$\hat{\sigma}_t^2 = \varepsilon_t^2$$

A volatility proxy is some variable that is useful for estimating the value of the true volatility. The squared residual can be justified as a volatility proxy because it is a conditionally unbiased estimator of the true conditional variance:

$$E_{t-1}\left[\varepsilon_t^2\right] = E_{t-1}\left[\sigma_t^2 \nu_t^2\right] = \sigma_t^2 E_{t-1}\left[\nu_t^2\right] = \sigma_t^2$$

and so, on average, the squared residual will correctly estimate the true conditional variance. Note, importantly, that while it will be correct on average it will estimate the conditional variance with error. That is, the squared residual is a "noisy" volatility proxy. Other, less noisy, volatility proxies have gained attention recently, see Andersen, *et al.* (2001, 2003).

If we denote a volatility forecast by h_t (where the h stands for "heteroscedasticity") and a volatility proxy by $\hat{\sigma}_t^2$ then two possible goodness-of-fit measures for a volatility forecast are:

$$Squared\ error : L\left(\hat{\sigma}_t^2, h_t\right) = \left(\hat{\sigma}_t^2 - h_t\right)^2$$
$$QLIKE : L\left(\hat{\sigma}_t^2, h_t\right) = \log h_t + \frac{\hat{\sigma}_t^2}{h_t}$$

and we would rank two forecasts according to their average loss over the forecast period:

$$\frac{1}{T}\sum_{t=1}^{T} L\left(\hat{\sigma}_t^2, h_t\right)$$

You may recognise the "QLIKE" loss function as the central part of the normal log-likelihood estimating a volatility model. Thus ranking by the average QLIKE loss function is equivalent to ranking by the average (normal) log-likelihood. Hansen and Lunde (2005) used these two goodness-of-fit measures to compare the numerous ARCH-type models they considered out-of-sample.

Some authors have used other loss functions, such as absolute error, or squared error on standard deviations or squared error on log-variances. These loss functions can be shown to lead to serious problems for evaluating volatility models, see Patton (2006) and Patton and Sheppard (2006). I recommend only using the squared error or the QLIKE loss function.

10.4. Using economic goodness-of-fit measures

An alternative to statistical measures of goodness-of-fit is some *economic* measure of goodness-of-fit. Good examples of this are in West, *et al.* (1993) and Fleming, *et al.* (2001), who find that the value of modelling conditional volatility to a risk-averse investor is between 5 and 200 basis points per year. If a conditional variance forecast is going to be used to make some economic decision, such as an investment decision, then the right way to compare the performance of competing models is to compare the profits that are generated by each model. Ideally, we would *always* compare forecasting models by their performance in economic decision-making. Our use of statistical measures of goodness-of-fit is motivated by the simple fact that we generally don't know the various economic uses of the forecasts, and so we instead use general measures of goodness-of-fit.

Consider an investor with utility function that is quadratic in future wealth, generating mean-variance preferences. Let's consider speculating in a risky asset, using borrowings at the risk-free rate to finance it:

$$\begin{aligned}
\mathcal{U}\left(W_{t+1}\right) &= W_{t+1} - 0.5\gamma W_{t+1}^2 \\
W_{t+1} &= W_t\left(\omega_{t+1}\left(R_{t+1} - R_{t+1}^f\right) + 1\right) \\
W_t &= \text{wealth at time } t \\
\gamma &= \text{risk aversion parameter} \\
R_t &= \text{gross return on risky asset} \\
R_t^f &= \text{gross return on risk-free asset} \\
\omega_{t+1} &= \text{portfolio weight in risky asset}
\end{aligned}$$

If we set $W_t = 1$ we get the following rule for the optimal portfolio weight:

$$\omega_{t+1}^* = \frac{E_t\left[R_{t+1} - R_{t+1}^f\right]}{V_t\left[R_{t+1} - R_{t+1}^f\right] - E_t\left[R_{t+1} - R_{t+1}^f\right]^2} \cdot \frac{1+\gamma}{\gamma}$$

Thus the optimal portfolio weight is a function of a conditional mean and a conditional variance forecast. Combined with some model for the conditional mean, we can use volatility forecasts to obtain optimal portfolio weights at each point in time. From these we can then work out the portfolio returns obtained by using a particular volatility model, and finally work out the realised utility from these returns. The better volatility forecast should yield a higher expected utility over the sample period.

Another application of volatility models is in option pricing. If we take the simple Black-Scholes formula for pricing a European option on a non-dividend paying stock we get:

$$\begin{aligned}
c_t &= S_t\Phi\left(d_1\right) - K\exp\left\{-r\left(T-t\right)\right\}\Phi\left(d_2\right) \\
p_t &= K\exp\left\{-r\left(T-t\right)\right\}\Phi\left(-d_2\right) - S_t\Phi\left(-d_1\right) \\
d_1 &= \frac{\log\left(S_t/K\right) + \left(r + \sigma^2/2\right)\left(T-t\right)}{\sigma\sqrt{T-t}} \\
d_2 &= d_1 - \sigma\sqrt{T-t}
\end{aligned}$$

where

$$
\begin{aligned}
c_t &= \text{European call option price at time } t \\
p_t &= \text{European put option price at time } t \\
S_t &= \text{value of underlying stock at time } t \\
K &= \text{strike price of option contract} \\
r &= \text{risk-free rate (annualised)} \\
T - t &= \text{time (in years) until expiry of contract} \\
\sigma &= \text{volatility of underlying stock price}
\end{aligned}
$$

The only unobservable input to the B-S option pricing formula is the stock price volatility. This is where volatility models may be used. If h_t is a forecast of the volatility of the return on the asset between time t and time T, then one way of evaluating the performance of the model is to look at the pricing errors it generates when used in the B-S pricing formula:

$$
\begin{aligned}
\hat{c}_{t,j} &= S_t \Phi\left(\hat{d}_{1,j}\right) - K_j \exp\left\{-r\left(T-t\right)\right\} \Phi\left(\hat{d}_{2,j}\right) \\
\hat{p}_{t,j} &= K_j \exp\left\{-r\left(T-t\right)\right\} \Phi\left(-\hat{d}_{2,j}\right) - S_t \Phi\left(-\hat{d}_{1,j}\right) \\
\hat{d}_{1,j} &= \frac{\log\left(S_t/K_j\right) + \left(r + h_t/2\right)\left(T-t\right)}{\sqrt{h_t\left(T-t\right)}} \\
\hat{d}_{2,j} &= \hat{d}_{1,j} - \sqrt{h_t\left(T-t\right)} \\
MSE - call &\equiv \sum_{j=1}^{n} \left(c_{t,j} - \hat{c}_{t,j}\right)^2 \\
MSE - put &\equiv \sum_{j=1}^{n} \left(p_{t,j} - \hat{p}_{t,j}\right)^2
\end{aligned}
$$

where $\{c_{t,j}\}_{j=1}^{n}$ and $\{p_{t,j}\}_{j=1}^{n}$ are the option prices available in the market at time t, for options with various strikes prices. It should be noted that there is evidence against the B-S model empirically, and so even if h_t happened to be a perfect volatility forecast it would generate pricing errors. Furthermore, time-varying conditional variance is not consistent with the B-S assumptions, and so there is an internal inconsistency with this method of evaluating volatility forecasts. Nevertheless, evaluation of volatility forecasts through their use in option pricing is a useful measure of their goodness-of-fit.

11. Alternatives to the ARCH class of models

11.1. Stochastic volatility models

Notice in the GARCH model that, conditioning on the information set $\mathcal{F}_t$, the variable σ_{t+1}^2 is known:

$$V_t\left[Y_{t+1}\right] = \sigma_{t+1}^2 = \omega + \alpha\varepsilon_t^2 + \beta\sigma_t^2$$

so if we know the parameters $[\omega, \alpha, \beta]$ and we know ε_t^2 and σ_t^2 (which we usually do) then we know the conditional volatility. An alternative type of volatility model, known as a "stochastic volatility" model, assumes that the process σ_{t+1}^2 is itself random, with some innovation term that is not known at time t :

$$\begin{aligned}
\sigma_{t+1}^2 &= \omega + \beta\sigma_t^2 + \eta_{t+1} \\
\eta_{t+1} &\sim F
\end{aligned}$$

where F is some distribution, usually assumed to be $N\left(0, \sigma_\eta^2\right)$. A normal ARMA-SV model is thus:

$$\begin{aligned}
Y_{t+1} &= \phi_0 + \phi_1 Y_t + \varepsilon_{t+1} + \theta\varepsilon_t \\
\varepsilon_{t+1}|\mathcal{F}_t &\sim N\left(0, \sigma_{t+1}^2\right) \\
\sigma_{t+1}^2 &= \omega + \beta\sigma_t^2 + \eta_{t+1} \\
\eta_{t+1} &\sim iid\ N\left(0, \sigma_\eta^2\right)
\end{aligned}$$

Notice that, unlike the ARMA-GARCH model, the ARMA-SV model has *two* innovation terms, ε_{t+1} for the return itself, and η_{t+1} for the conditional variance of the return. The presence of this additional innovation term makes both the estimation of the model and forecasting using the model more difficult than for GARCH models. SV models have proven useful in option pricing, as they can be written quite naturally in continuous-time form, which is the usual form for models in this area, see Heston (1993). For a nice summary of SV models in finance see Shephard (2005); see Taylor (2005, Chapter 11) for a text book treatment of this topic.

SV models are not as popular for forecasting as GARCH models, mainly because:

1. They are harder to estimate (due to the extra innovation term)

2. There is little evidence that they produce superior volatility forecasts to a simple GARCH model.

11.2. Regime switching volatility models

Regime switching volatility models capture the intuitively appealing idea that markets sometimes switch between volatile and tranquil periods[2]. As an example, let us consider the following model:

$$\begin{aligned} \textit{State } 1 &: r_t \sim N\left(0, \sigma_1^2\right) \\ \textit{State } 2 &: r_t \sim N\left(0, \sigma_2^2\right) \\ \sigma_2^2 &> \sigma_1^2 \end{aligned}$$

So State 1 is the relatively tranquil period and State 2 is the relatively volatile period. Define a state variable S_t as being equal to 1 if we are in State 1 and equal to 2 if we are in State 2. A regime switching volatility model captures persistence in volatility by allowing the probability of being in the same state next period to be higher than the probability of

[2] The first regime switching model in economics, see Hamilton (1989), focussed on the conditional mean of macroeconomic variables. In particular, he wanted to capture another intuitively appealing idea that the economy moves between booms and recessions. These models have proven very useful in macroeconomics, but we will focus on their use for volatility modelling only.

switching between states. Let

$$\pi_{11} = \Pr\left[S_{t+1} = 1 | S_t = 1\right]$$
$$\pi_{12} = \Pr\left[S_{t+1} = 2 | S_t = 1\right] = 1 - \pi_{11}$$
$$\pi_{22} = \Pr\left[S_{t+1} = 2 | S_t = 2\right]$$
$$\pi_{21} = \Pr\left[S_{t+1} = 1 | S_t = 1\right] = 1 - \pi_{22}$$
$$P = \left[\begin{array}{cc} \pi_{11} & 1 - \pi_{11} \\ 1 - \pi_{22} & \pi_{22} \end{array}\right]$$

If there is volatility persistence then $\pi_{11} > 0.5$ and $\pi_{22} > 0.5$. When estimated on financial data we usually see estimates like $\hat{\pi}_{11} \approx 0.95$ and $\hat{\pi}_{22} \approx 0.90$. The state variable in regime switching models is unobservable. At each point in time we form estimates of which state we are in using the returns data available. This makes the estimation more difficult than ARCH-type models.

Regime switching models have proven quite useful in macroeconomic forecasting, and for certain applications in finance (density forecasting in particular). They are used by many researchers for volatility forecasting, though ARCH-type models are more widely used.

11.3. Implied volatilities

Above we noted that the only unobservable input to the Black-Scholes option pricing formula is the conditional variance of the return on the underlying asset between time t and the expiry of the option. Because the Black-Scholes value of an option is strictly increasing in the volatility of the underlying asset, we can use observed option prices to obtain the market's estimate of the volatility on the underlying asset:

$$c_t = c_{BS}\left(S_t, K, r, T - t, \sigma_t\right)$$
$$\sigma_t = c_{BS}^{-1}\left(S_t, K, r, T - t, c_t\right)$$

Practically, the implied volatility is computed numerically by searching over σ_t for the value that returns the observed option price, when plugged into the appropriate above equation. This volatility estimate is the value "implied" by observed option prices and the Black-Scholes model[3].

One of the main attractions of this type of volatility estimator is that they are based on current market prices rather than on historical data. Thus these may be termed "forward looking" estimators of volatility. The main drawback of this type of volatility estimator is that it hinges critically on the accuracy of the Black-Scholes option pricing formula, which is known to be only approximately correct. In particular, the Black-Scholes model relies on an assumption of constant conditional volatility, which is obviously not consistent with using implied volatilities for forecasting time-varying volatility. More sophisticated option pricing formulas are available, and some of these may also be used to obtain implied volatility estimates, though this is not common. In general, researchers have found that implied volatilities tend to be higher than volatilities obtained from ARCH-type models, possibly reflecting the fact that the true distribution of asset returns has fatter tails than the log-normal distribution on which the Black-Scholes formula relies.

The above discussion also relates to the implied volatility "smile" or "smirk". This refers to a plot of implied volatilities from a range of options on the same underlying asset with the same expiry, differing only by their strike prices. Each option price leads to one implied volatility, and if the Black-Scholes model was correct all of these implied volatilities would be equal. In general, researchers have found that call options that are deep out-of-the-money tend to lead to implied vols that are much higher than call options that are at-the-money. Call options that are deep in-the-money tend also to yield higher implied vols, though to a lesser extent than for out-of-the-money options, see Figure 11.1. The plot, which resembles a smile or smirk, is usually taken as being further evidence that the true distribution of asset returns has fatter tails than the log-normal, and has a fatter left tail than right tail. These fatter tails lead to higher option prices (because there is a greater

[3] Implied volatilities are quite popular in financial markets. Option trades often refer to the price of an option not in dollars, but in units of volatility. For example, "the September 2004 European call on the S&P500 with strike 1200 is trading at 25%", rather than "the September 2004 European call on the S&P500 with strike 1200 is trading at $4.50".

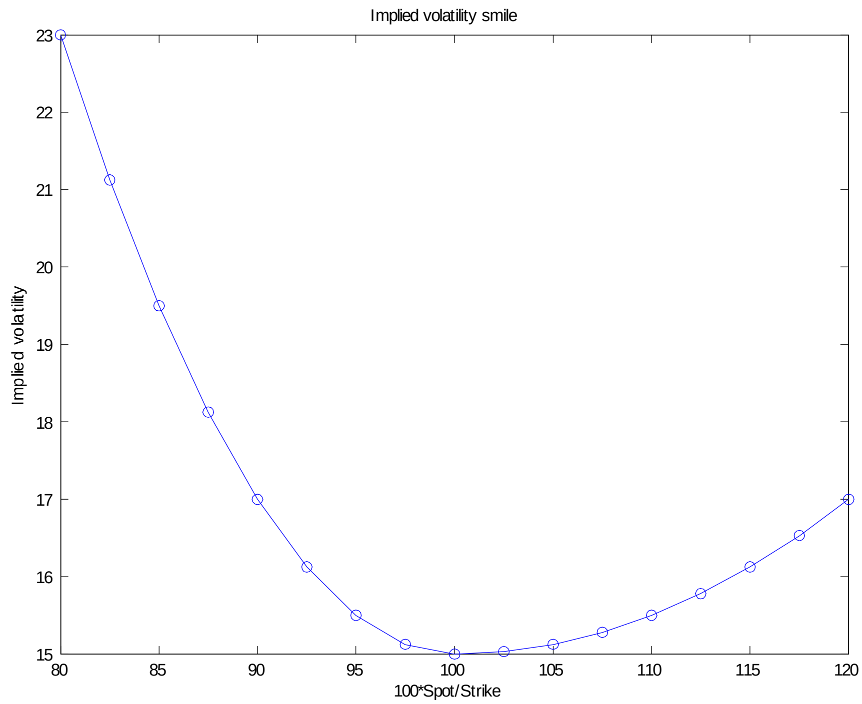

Figure 11.1: *Implied volatility as a function of the spot price.*

probability of large pay-offs) than under log-normality, and so higher volatility estimates are required to push the Black-Scholes price up to the observed option price.

Overall, in spite of the problems with the Black-Scholes model, implied volatilities represent a rich source of information on the variability of the underlying asset. The problems with the Black-Scholes model mean that they perhaps should not be used without adjustment, but they have been found to be useful when combined with other volatility estimators, such as those obtained from GARCH models, see Blair, et al. (2001) for example. The main reason implied vols are not more widely used relates to the data: options data can be hard to obtain and hard to prepare for analysis.

References

[1]Abarbanell, Jeffrey S., Bernard, Victor L., 1991, Tests of Analysts' Overreaction/Underreaction to Earnings Information as an Explanation for Anomalous Stock Price Behaviour, *Journal of Finance*, 47(3), 1181-1207.

[2]Alexander, C., and Chibumba, A., 1997, working paper, University of Sussex.

[3]Andersen, T.G., and Bollerslev, T., 1998, Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies, Journal of Finance, 53(1), 219-265.

[4]Andersen, Torben, Bollerslev, Tim, Diebold, Francis X., and Christoffersen, Peter F., 2005, Volatility and Correlation Forecasting, in G. Elliott, A. Timmermann and C.W.J. Granger (ed.s), *Handbook of Economic Forecasting*, North Holland, forthcoming.

[5]Bauwens, Luc, Laurent, Sébastien, and Rombouts, Jeroen, 2006, Multivariate GARCH Models: A Survey, *Journal of Applied Econometrics*, 21(1), 79-109.

[6]Bera, Anil K., and Yannis Bilias, 2002, The MM, ME, ML, EL, EF and GMM Approaches to Estimation: A Synthesis, *Journal of Econometrics*, 107, 51-86.

[7]Blair, B.J., Poon, S.-H., and Taylor, S.J., 2001, Forecasting S&P 100 Volatility: the Incremental Information Content of Implied Volatilities and High Frequency Index Returns, Journal of Econometrics, 105, 5-26.

[8]Bollerslev, Tim, 1986, Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31, 307-327.

[9]Bollerslev, Tim, 1990, Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach, *Review of Economics and Statistics*, 72, 498-505.

[10]Bollerslev, Tim, and Wooldridge, Jeffrey M., 1992, Quasi-Maximum Likelihood Estimation and Inference in Dynamics Models with Time-Varying Covariances, *Econometric Reviews*, 11, 143-172.

[11]Brealy, Richard A., and Myers, Stewart C., 1991, *Principles of Corporate Finance*, 4th Edition, McGraw Hill, U.S.A.

[12]Brooks, Chris, 2002, *Introductory Econometrics for Finance*, Cambridge University Press, Cambridge.

[13]Campbell, John Y., Lo, Andrew W. and Mackinlay, A. Craig, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

[14]Cappiello, Lorenzo, Engle, Robert F., and Sheppard, Kevin, 2006, Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns, Journal of Financial Econometrics, 4(4), 537-572.

[15]Casella, George, and Berger, Roger L., 1990, *Statistical Inference*, Duxbury Press, U.S.A.

[16]Christoffersen, Peter F., 1998, Evaluating Interval Forecasts, *International Economic Review*, 39(4), 841-862.

[17]Christoffersen, Peter, 2003, *Elements of Financial Risk Management*, Academic Press.

[18]Cochrane, John H., 2001, *Asset Pricing*, Princeton University Press, Princeton, New Jersey.

[19]Comte, F., and Lieberman, O., 2003, Asymptotic Theory for Multivariate GARCH Processes, Journal of Multivariate Analysis, 84, 61–84.
[20]Diebold, Francis X., 2001, *Elements of Forecasting*, 2nd Edition, South-Western College Publishing, Ohio.
[21]Diebold, Francis X., and Mariano, Roberto S., 1995, Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13(3), 253-263.
[22]Diebold, Francis X., Gunther, T., and Tay, Anthony S., 1998, Evaluating Density Forecasts, with Applications to Financial Risk Management, *International Economic Review*, 39, 863-883.
[23]Diebold, Francis X., Hahn, Jinyong, and Tay, Anthony S., 1999, Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange, *The Review of Economics and Statistics*, 81(4), 661-673.
[24]Elliott, Graham, Rothenberg, Thomas J., and Stock, James H., 1996, Efficient Tests for an Autoregressive Unit Root, *Econometrica*, 64(4), 813-836.
[25]Embrechts, Paul, Klüppelberg, Claudia, and Mikosch, Thomas, 1997, *Modelling Extremal Events*, Springer, Berlin.
[26]Enders, Walter, 2004, *Applied Econometric Time Series,* Second Edition, John Wiley & Sons, USA.
[27]Engle, Robert F., 1982, Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of U.K. Inflation, *Econometrica,* 50, 987-1008.
[28]Engle, Robert F., 2002, Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models, *Journal of Business and Economic Statistics*, 17(5).
[29]Engle, Robert R., and Bollerslev, Tim, 1986, Modelling the Persistence of Conditional Variances, *Econometric Reviews*, 5, 1-50.
[30]Engle, Robert F., and Granger, Clive W. J., 1987, Co-Integration and Error Correction: Representation, Estimation and Testing, *Econometrica*, 55, 251-276.
[31]Engle, Robert F., and Kroner, Kenneth, 1995, Multivariate Simultaneous GARCH, *Econometric Theory,* 11, 122-150.
[32]Engle, Robert F., and Ng, Victor K., 1993, Measuring and Testing the Impact of News on Volatility, *Journal of Finance*, 48(5), 1749-1778.
[33]Engle, Robert F., and Sheppard, Kevin K., 2002, Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH", Working Paper 2001-15, Department of Economics, University of California, San Diego.
[34]Fisher, R. A., 1932, *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh.
[35]Fleming, Jeffrey, Kirby, Chris, and Ostdiek, Barbara, 2001, The Economic Value of Volatility Timing, *Journal of Finance*, 56, 329-352.
[36]Fleming, Jeffrey, Kirby, Chris, and Ostdiek, Barbara, 2003, The Economic Value of Volatility Timing using "Realized" Volatility, *Journal of Financial Economics*, 67, 473-509.
[37]Froot, Kenneth A., Scharfstein David S., and Stein, Jeremy C., 1993, Risk Management: Coordinating Corporate Investment and Financing Policies, *Journal of Finance*, 48(5), 1629-1658.

[38]Gourieroux, C., Monfort, A., and Trognon, A., 1984, Pseudo Maximum Likelihood Methods: Theory, *Econometrica*, 52(3), 681-700.

[39]Granger, Clive W. J., 1969, Investigating Causal Relations by Econometric Methods and Cross-Spectral Methods, *Econometrica*, 37, 424-438.

[40]Granger, C.W.J., 1969, Prediction with a Generalized Cost Function. *OR*, 20, 199-207.

[41]Granger, C.W.J., 1999, Outline of Forecast Theory Using Generalized Cost Functions. *Spanish Economic Review*, 1, 161-173.

[42]Granger, Clive W. J., and Newbold, Paul, 1974, Spurious Regressions in Econometrics, *Journal of Econometrics*, 2, 111-120.

[43]Granger, Clive W. J. and Timmerman, Allan, 2004, Efficient Market Hypothesis and Forecasting, *International Journal of Forecasting*, 20(1), 15-27.

[44]Greene, William H., 1997, *Econometric Analysis*, Third Edition, Prentice-Hall, U.S.A.

[45]Griffiths, William E., Hill, R. Carter, and Judge, George G., 1993, *Learning and Practicing Econometrics*, John Wiley and Sons, U.S.A.

[46]Grossman, S. and Sitglitz, J., 1980, On the Impossibility of Informationally Efficient Markets, *American Economic Review*, 70, 393-408.

[47]Gourieroux, Christian, Monfort, Alain and Trognon, A., 1984, Pseudo Maximum Likelihood Methods: Theory, *Econometrica*, 52, 681-700.

[48]Hall, Alastair R., 2005, *Generalized Method of Moments*, Oxford University Press, Oxford.

[49]Hamilton, James D., 1989, A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica*, 57, 357-384.

[50]Hamilton, James D., 1994, *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

[51]Hansen, Lars Peter, 1982, Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50, 1029-1054.

[52]Hansen, Peter Reinhard, and Lunde, Asger, 2005, A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?, *Journal of Applied Econometrics*, 20, 873-889.

[53]Hansen, Peter R., and Lunde, Asger, 2006, Consistent Ranking of Volatility Models, Journal of Econometrics, 131, 97-121.

[54]Hanselman, Duane, and Littlefield, Bruce, 2001, *Mastering Matlab 6: A Comprehensive Tutorial and Reference*, Prentice-Hall, London.

[55]Heston, S. L. 1993, A Closed Form Solution for Options with Stochastic Volatility with Applications to Bonds and Currency Options, *Review of Financial Studies*, 6(2), 327-343.

[56]Hentschel, Ludger, 1995, All in the Family: Nesting Symmetric and Asymmetric GARCH Models, *Journal of Financial Economics*, 39, 71-104.

[57]Huang, Chi-fu, and Litzenberger, Robert H., 1988, *Foundations for Financial Economics*, Prentice Hall, New Jersey.

[58]Ishida, Isao, and Engle, Robert F., 2001, Modeling Variance of Variance: the Square-Root, the Affine and the CEV GARCH Models, working paper, Department of Economics, University of California, San Diego.

[59]Jarque, C.M. and A.K. Bera, 1987, A Test for Normality of Observations and Regression Residuals, *International Statistical Review*, 55, 163-172.
[60]Jeantheau, T, 1998, Strong Consistency of Estimators for Multivariate ARCH models, Econometric Theory, 14, 70–86.
[61]Judd, Kenneth L, 1998, *Numerical Methods in Economics*, MIT Press, USA.
[62]Kandel, Shmuel, and Stambaugh, Robert F., 1996, On the Predictability of Stock Returns: An Asset-Allocation Perspective, Journal of Finance, 51(2), 385-424.
[63]Kawakatsu, Hiroyuki, 2006, Matrix Exponential GARCH, *Journal of Econometrics*, 134(1), 95-128.
[64]Komunjer, I., 2005, Quasi-Maximum Likelihood Estimation for Conditional Quantiles, Journal of Econometrics, 128, 137-164.
[65]Kroner, Kenneth F., and Ng, Victor K., 1998, Modeling Asymmetric Comovements of Asset Returns, *Review of Financial Studies*, 11(4), 817-844.
[66]Lo, Andrew W., and MacKinlay, A. Craig, 1990, An Econometric Analysis of Nonsynchronous-Trading, *Journal of Econometrics*, 45, 181-212.
[67]Longin, François, and Solnik, Bruno, 2001, Extreme Correlation of International Equity Markets, *Journal of Finance*, 56(2), 649-676.
[68]Lumsdaine, Robin L., and Ng, Serena, 1999, Testing for ARCH in the Presence of a Possibly Misspecified Conditional Mean, *Journal of Econometrics*, 93(2), 257-279.
[69]Maddala, G. S., 2001, *Introduction to Econometrics*, John Wiley and Sons, U.S.A.
[70]McNeil, Alexander J., and Frey, Rüdiger, 2000, Estimation of Tail-Related Risk Measures fo Heteroscedastic Financial Time Series: An Extreme Value Approach, *Journal of Empirical Finance*, 7, 271-300.
[71]Mendenhall, William, Wackerly, Dennis D., and Scheaffer, Richard L., 1990, *Mathematical Statistics with Applications*, 4th Edition, Duxbury Press, U.S.A.
[72]Nelson, Daniel B., 1991, Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica*, 59(2), 347-370.
[73]Newey, Whitney K., and West, Kenneth D., 1987, A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55(3), 703-708.
[74]Newey, Whitney K., and McFadden, Daniel, 1994, Large Sample Estimation and Hypothesis Testing, in R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics, Volume IV*, Elsevier Science, The Netherlands.
[75]Patton, Andrew J., 2004, On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation, *Journal of Financial Econometrics*, 2(1), 130-168.
[76]Patton, Andrew J., 2006b, Modelling Asymmetric Exchange Rate Dependence, *International Economic Review*, 47(2), 527-556.
[77]Patton, Andrew J., 2006, Volatility Forecast Comparison using Imperfect Volatility Proxies, Quantitative Finance Research Centre, University of Technology Sydney, Research Paper 175.
[78]Patton, Andrew J., and Timmermann, Allan, 2003, Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity, *Journal of Econometrics*, forthcoming.

[79]Patton, Andrew J., and Timmermann, Allan, 2006, Testing Forecast Optimality under Unknown Loss, Journal of the Americal Statistical Association, forthcoming.
[80]Patton, Andrew J. and Sheppard, Kevin, 2006, Evaluating Volatility and Correlation Forecasts, in T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch (eds.) Handbook of Financial Time Series, Springer Verlag. Forthcoming.
[81]Pindyck, Robert S., and Rubinfeld, Daniel L., 1991, *Econometric Models and Economic Forecasts*, 3rd Edition, McGraw Hill, U.S.A.
[82]Poterba, J. and L. Summers, 1986, The Persistence of Volatility and Stock Market Fluctuations, *American Economic Review*, 76, 1124-1141.
[83]Ramanathan, Ramu, 2002, *Introductory Econometrics with Applications*, 5th Edition, Harcourt College Publishers, U.S.A.
[84]Roberts, H., 1967, Statistical versus Clinical Prediction of the Stock Market, manuscript, Center for Research in Securities Prices, University of Chicago.
[85]Schwert, G. William, 1989, Why Does Stock Market Volatility Change Over Time?, *Journal of Finance*, 44, 1115-1153.
[86]Sentana, Enrique, 1991, Quadratic ARCH Models, *Review of Economic Studies*, 62 (4), 639-661.
[87]Shephard, Neil, 1996, Statistical Aspects of ARCH and Stochastic Volatility, in D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen (eds.), *Time Series Models in Econometrics, Finance and Other Fields,* Chapman and Hall, London.
[88]Shephard, Neil, 2005, *Stochastic Volatility: Selected Readings*, Oxford University Press, Oxford.
[89]Strang, Gilbert, *Linear Algebra and its Applications*, Third Edition, Harcourt Brace & Co., U.S.A.
[90]Sullivan, Ryan, Timmermann, Allan and White, Halbert, 1999, Data-Snooping, Technical Trading Rules and the Bootstrap, *Journal of Finance*, 54, 1647-1692.
[91]Sullivan, Ryan, Timmermann, Allan and White, Halbert, 2001, Dangers of Data Mining: The Case of Calendar Effects in Stock Returns, *Journal of Econometrics*, 249-286.
[92]Taylor, Stephen J., 1986, *Modelling Financial Time Series*, Wiley, New York.
[93]Taylor, Stephen J., 2005, *Asset Price Dynamics, Volatility and Prediction*, Princeton University Press, Oxford.
[94]Theil, H., 1958, *Economic Forecasts and Policy*, North-Holland, Amsterdam.
[95]Tsay, Ruey S., 2002, *Analysis of Financial Time Series*, John Wiley & Sons, New York.
[96]West, Kenneth D., Edison, Hali J., and Cho, Dongchul, 1993, A Utility-Based Comparison of Some Models of Exchange Rate Volatility, *Journal of International Economics*, 35, 23-45.
[97]White, Halbert, 1982, Maximum Likelihood Estimation of Misspecified Models, *Econometrica*, 48, 817-838.
[98]White, Halbert., 2000, A Reality Check for Data Snooping, *Econometrica*, 68, 1097-1126.
[99]Wonnacott, Thomas H., and Wonnacott, Ronald J., 1990, *Introductory Statistics for*

Business and Economics, 4th Edition, John Wiley and Sons, U.S.A.