Time Varying Volatility Models

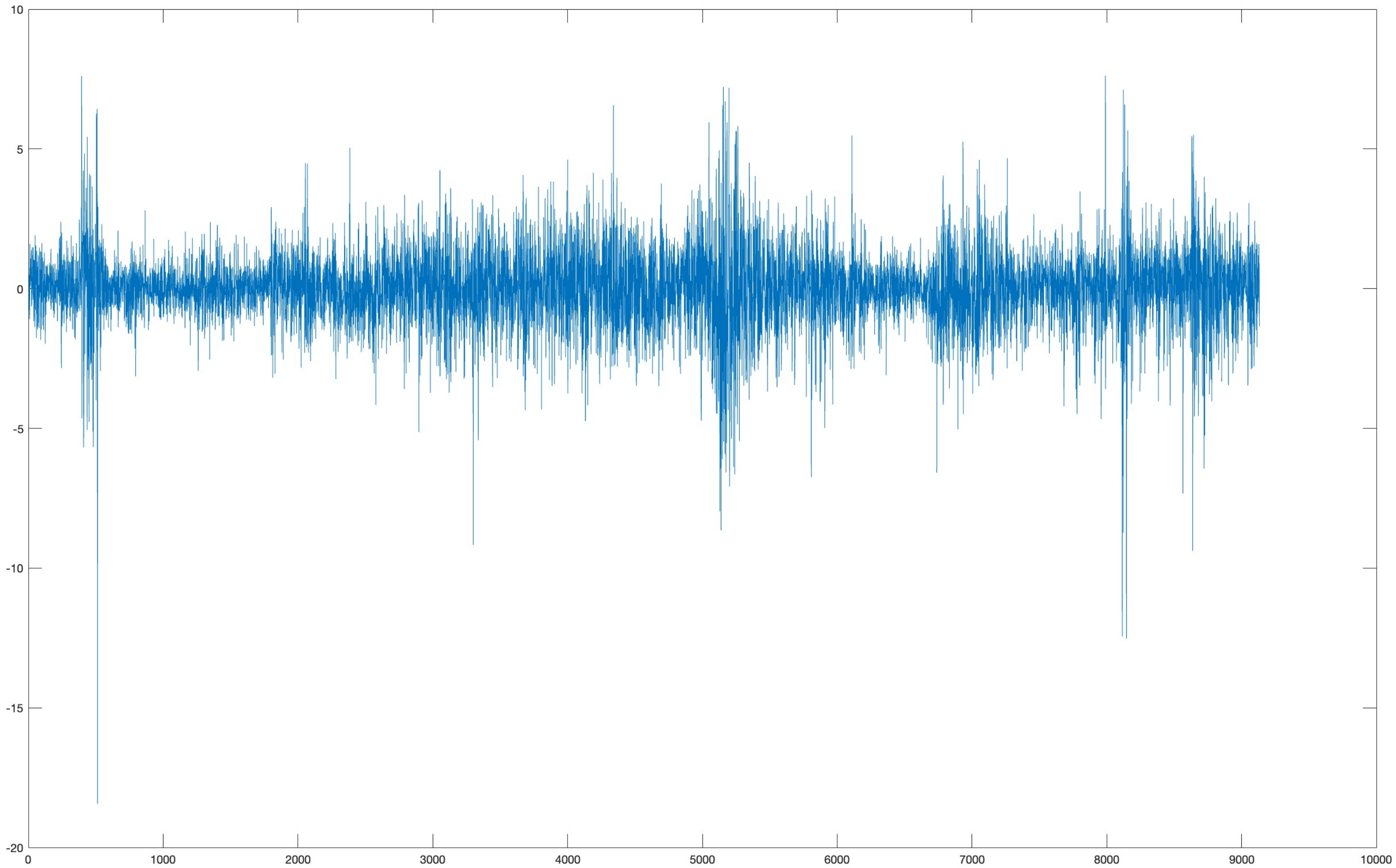
An Excursion into Non-linearity Land

- Motivation: the linear structural (and time series) models cannot explain a number of important features common to much financial data
  - leptokurtosis
  - volatility clustering or volatility pooling
  - leverage effects

- Our "traditional" structural model could be something like:
  $y_t = \beta_1 + \beta_2 x_{2t} + ... + \beta_k x_{kt} + u_t$, or more compactly $y = X\beta + u$.

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

Figure 4: Example Commodity Price Index - 02/02/1989 - 01/02/2024



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
  - switching models
  - bilinear models

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:

$$\hat{u}_t = \beta_0 + \beta_1 \hat{y}_t^2 + \beta_2 \hat{y}_t^3 + \ldots + \beta_{p-1} \hat{y}_t^p + v_t$$

- 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.

Autoregressive Conditionally Heteroscedastic (ARCH) Models

- So use a model which does not assume that the variance is constant.
- Recall the definition of the 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

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

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

- 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 iid.
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.

- Note that the ARCH test is also sometimes applied directly to returns instead of the residuals from Stage 1 above.

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.

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

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

- 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):

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

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

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

  - less likely to breech non-negativity constraints

The Unconditional Variance under the GARCH Specification

- The unconditional variance of u_t is given by

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

  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 non-stationarity in variance, the conditional variance forecasts will not converge on their unconditional value as the horizon increases.

Estimation of ARCH / GARCH Models

- Since the model is no longer of the usual linear form, 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:

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

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

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

Parameter Estimation using Maximum Likelihood

- 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

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

$$f(y_T \mid \beta_1 + \beta_2 X_4, \sigma^2) \qquad (2)$$

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

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

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

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

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

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

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

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

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

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

- 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.
- But how does this help us in estimating heteroscedastic models?

Estimation of GARCH Models Using Maximum Likelihood

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

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

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

- Unfortunately, the LLF for a model with time-varying variances cannot be maximised analytically, except in the simplest of 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

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

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

• 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.

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.

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.

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 $u_{t-1} < 0$
$\quad = 0$ 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.

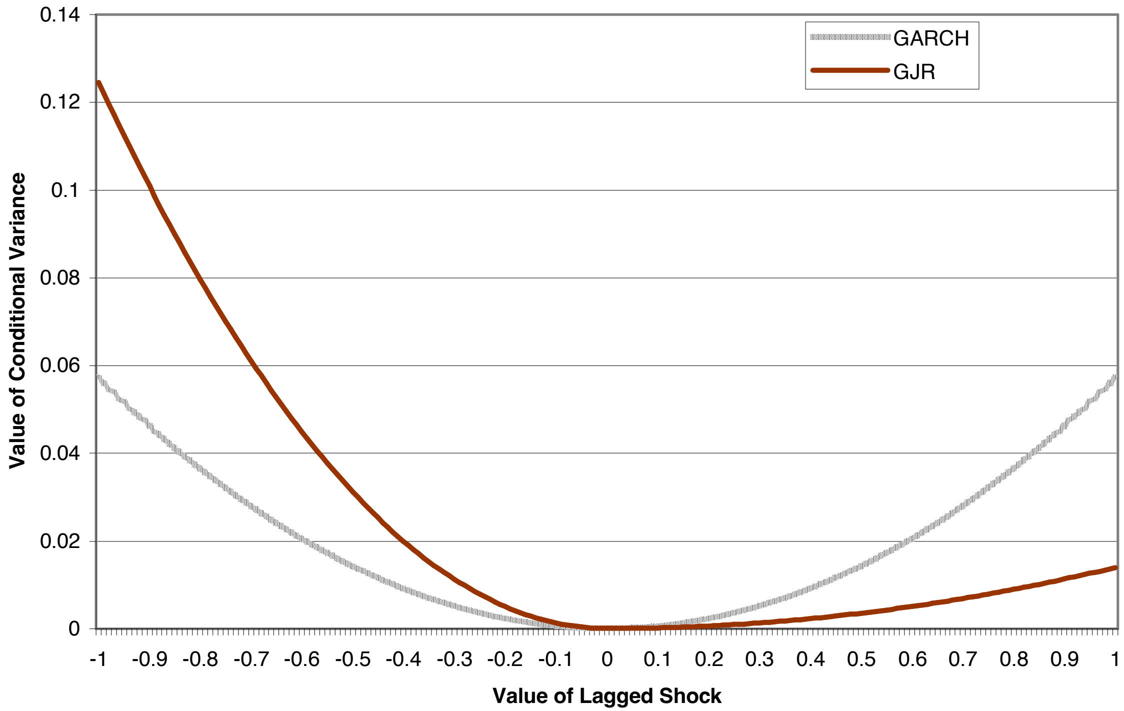
$$y_t = 0.172$$
$$(3.198)$$

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

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:



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.

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.

- Variance forecasts are additive over time.

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 are 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

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

Forecasting Variances using GARCH Models (Cont'd)

- Let ${\sigma^f_{1,T}}^2$ 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.
- ${\sigma^f_{1,T}}^2$ would be obtained by taking the conditional expectation of the first equation at the bottom of slide 36:

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

- Given, ${\sigma^f_{1,T}}^2$ how is ${\sigma^f_{2,T}}^2$, 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:

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

- 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

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

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

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

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

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

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

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

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

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

   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 (S) and change in futures price (F)

σ_S = standard deviation of S

σ_F = standard deviation of F

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

Use

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

Testing Non-linear Restrictions or

Testing Hypotheses about Non-linear Models

- Usual t- and F-tests are still valid in non-linear models, but they are not flexible enough.

- There are three hypothesis testing procedures based on maximum likelihood principles: Wald, Likelihood Ratio, Lagrange Multiplier.

- Consider a single parameter, θ to be estimated, Denote the MLE as $\hat{\theta}$ and a restricted estimate as $\tilde{\theta}$.

Likelihood Ratio Tests

- Estimate under the null hypothesis and under the alternative.
- Then compare the maximised values of the LLF.
- So we estimate the unconstrained model and achieve a given maximised value of the LLF, denoted L_u
- Then estimate the model imposing the constraint(s) and get a new value of the LLF denoted L_r.
- Which will be bigger?
- $L_r \leq L_u$ comparable to RRSS $\geq$ URSS

- The LR test statistic is given by

$$LR = -2(L_r - L_u) \sim \chi^2(m)$$

where m = number of restrictions

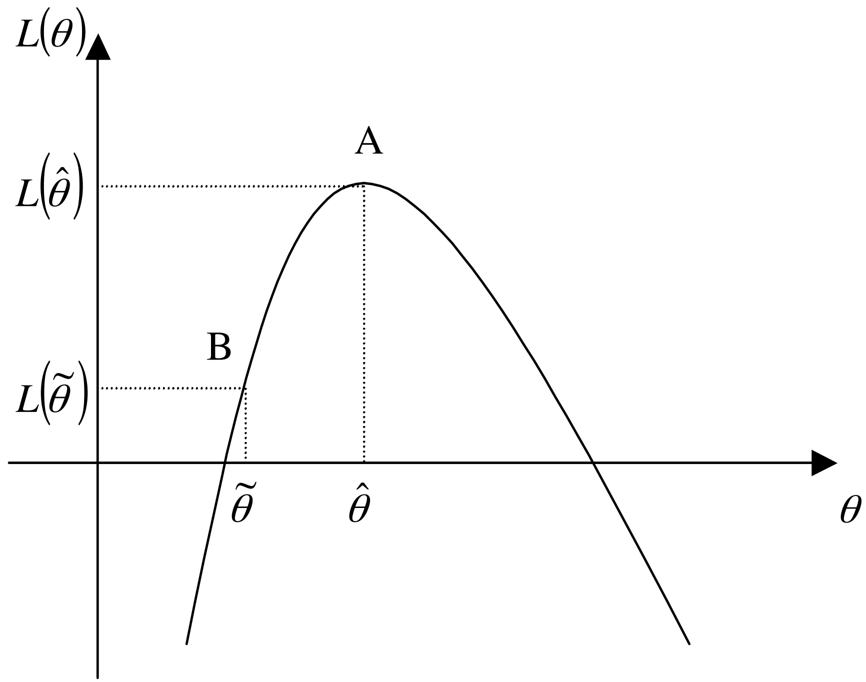
Likelihood Ratio Tests (cont'd)

- Example: We estimate a GARCH model and obtain a maximised LLF of 66.85. We are interested in testing whether $\beta = 0$ in the following equation.

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

- We estimate the model imposing the restriction and observe the maximised LLF falls to 64.54. Can we accept the restriction?
- LR = -2(64.54-66.85) = 4.62.
- The test follows a $\chi^2(1) = 3.84$ at 5%, so reject the null.
- Denoting the maximised value of the LLF by unconstrained ML as $L(\hat{\theta})$ and the constrained optimum as $L(\tilde{\theta})$. Then we can illustrate the 3 testing procedures in the following diagram:

Comparison of Testing Procedures under Maximum Likelihood: Diagramatic Representation



Hypothesis Testing under Maximum Likelihood

- The vertical distance forms the basis of the LR test.
- The Wald test is based on a comparison of the horizontal distance.
- The LM test compares the slopes of the curve at A and B.
- We know at the unrestricted MLE, $L(\hat{\theta})$, the slope of the curve is zero.
- But is it “significantly steep” at $L(\tilde{\theta})$?
- This formulation of the test is usually easiest to estimate.

An Example of the Application of GARCH Models - Day & Lewis (1992)

- <u>Purpose</u>
- To consider the out of sample forecasting performance of GARCH and EGARCH Models for predicting stock index volatility.

- Implied volatility is the markets expectation of the "average" level of volatility of an option:

- Which is better, GARCH or implied volatility?

- <u>Data</u>
- Weekly closing prices (Wednesday to Wednesday, and Friday to Friday) for the S&P100 Index option and the underlying 11 March 83 - 31 Dec. 89

- Implied volatility is calculated using a non-linear iterative procedure.

The Models

- <u>The "Base" Models</u>

  For the conditional mean

  $$R_{Mt} - R_{Ft} = \lambda_0 + \lambda_1 \sqrt{h_t} + u_t \qquad (1)$$

  And for the variance
  $$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} \qquad (2)$$

  or $$\ln(h_t) = \alpha_0 + \beta_1 \ln(h_{t-1}) + \alpha_1 \left(\theta \frac{u_{t-1}}{\sqrt{h_{t-1}}} + \gamma \left[\left| \frac{u_{t-1}}{\sqrt{h_{t-1}}} \right| - \left(\frac{2}{\pi} \right)^{1/2} \right] \right) \qquad (3)$$

  where

  R_{Mt} denotes the return on the market portfolio

  R_{Ft} denotes the risk-free rate

  h_t denotes the conditional variance from the GARCH-type models while σ_t^2 denotes the implied variance from option prices.

The Models (cont'd)

- Add in a lagged value of the implied volatility parameter to equations (2) and (3).

  (2) becomes

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} + \delta \sigma_{t-1}^2 \qquad (4)$$

  and (3) becomes

$$\ln(h_t) = \alpha_0 + \beta_1 \ln(h_{t-1}) + \alpha_1 \left(\theta \frac{u_{t-1}}{\sqrt{h_{t-1}}} + \gamma \left[\left| \frac{u_{t-1}}{\sqrt{h_{t-1}}} \right| - \left(\frac{2}{\pi} \right)^{1/2} \right] \right) + \delta \ln(\sigma_{t-1}^2) \qquad (5)$$

- We are interested in testing $H_0 : \delta = 0$ in (4) or (5).
- Also, we want to test $H_0 : \alpha_1 = 0$ and $\beta_1 = 0$ in (4),
- and $H_0 : \alpha_1 = 0$ and $\beta_1 = 0$ and $\theta = 0$ and $\gamma = 0$ in (5).

The Models (cont'd)

- If this second set of restrictions holds, then (4) & (5) collapse to

$$h_t^2 = \alpha_0 + \delta\sigma_{t-1}^2 \qquad (4')$$

- and (3) becomes

$$\ln(h_t^2) = \alpha_0 + \delta \ln(\sigma_{t-1}^2) \qquad (5')$$

- We can test all of these restrictions using a likelihood ratio test.

In-sample Likelihood Ratio Test Results: GARCH Versus Implied Volatility

$$R_{Mt} - R_{Ft} = \lambda_0 + \lambda_1 \sqrt{h_t} + u_t \qquad (8.78)$$

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} \qquad (8.79)$$

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} + \delta\sigma_{t-1}^2 \qquad (8.81)$$

$$h_t^2 = \alpha_0 + \delta\sigma_{t-1}^2 \qquad (8.81')$$

Equation for Variance specification	λ_0	λ_1	$\alpha_0 \times 10^{-4}$	α_1	β_1	δ	Log-L	χ^2
(8.79)	0.0072	0.071	5.428	0.093	0.854	-	767.321	17.77
	(0.005)	(0.01)	(1.65)	(0.84)	(8.17)			
(8.81)	0.0015	0.043	2.065	0.266	-0.068	0.318	776.204	-
	(0.028)	(0.02)	(2.98)	(1.17)	(-0.59)	(3.00)		
(8.81′)	0.0056	-0.184	0.993	-	-	0.581	764.394	23.62
	(0.001)	(-0.001)	(1.50)			(2.94)		

Notes: t-ratios in parentheses, Log-L denotes the maximised value of the log-likelihood function in each case. χ^2 denotes the value of the test statistic, which follows a $\chi^2(1)$ in the case of (8.81) restricted to (8.79), and a $\chi^2(2)$ in the case of (8.81) restricted to (8.81′). Source: Day and Lewis (1992). Reprinted with the permission of Elsevier Science.

In-sample Likelihood Ratio Test Results: EGARCH Versus Implied Volatility

$$R_{Mt} - R_{Ft} = \lambda_0 + \lambda_1 \sqrt{h_t} + u_t \qquad (8.78)$$

$$\ln(h_t) = \alpha_0 + \beta_1 \ln(h_{t-1}) + \alpha_1 \left(\theta \frac{u_{t-1}}{\sqrt{h_{t-1}}} + \gamma \left[\left| \frac{u_{t-1}}{\sqrt{h_{t-1}}} \right| - \left(\frac{2}{\pi} \right)^{1/2} \right] \right) \qquad (8.80)$$

$$\ln(h_t) = \alpha_0 + \beta_1 \ln(h_{t-1}) + \alpha_1 \left(\theta \frac{u_{t-1}}{\sqrt{h_{t-1}}} + \gamma \left[\left| \frac{u_{t-1}}{\sqrt{h_{t-1}}} \right| - \left(\frac{2}{\pi} \right)^{1/2} \right] \right) + \delta \ln(\sigma_{t-1}^2) \qquad (8.82)$$

$$\ln(h_t^2) = \alpha_0 + \delta \ln(\sigma_{t-1}^2) \qquad (8.82')$$

ıation for 'ariance cification	λ_0	λ_1	$\alpha_0 \times 10^{-4}$	β_1	θ	γ	δ	Log-L	χ^2
(c)	-0.0026	0.094	-3.62	0.529	-0.273	0.357	-	776.436	8.09
	(-0.03)	(0.25)	(-2.90)	(3.26)	(-4.13)	(3.17)			
(e)	0.0035	-0.076	-2.28	0.373	-0.282	0.210	0.351	780.480	-
	(0.56)	(-0.24)	(-1.82)	(1.48)	(-4.34)	(1.89)	(1.82)		
(e′)	0.0047	-0.139	-2.76	-	-	-	0.667	765.034	30.89
	(0.71)	(-0.43)	(-2.30)				(4.01)		

Notes: t-ratios in parentheses, Log-L denotes the maximised value of the log-likelihood function in each case. χ^2 denotes the value of the test statistic, which follows a $\chi^2(1)$ in the case of (8.82) restricted to (8.80), and a $\chi^2(2)$ in the case of (8.82) restricted to (8.82′). Source: Day and Lewis (1992).

Conclusions for In-sample Model Comparisons & Out-of-Sample Procedure

- IV has extra incremental power for modelling stock volatility beyond GARCH.
- But the models do not represent a true test of the predictive ability of IV.
- So the authors conduct an out of sample forecasting test.
- There are 729 data points. They use the first 410 to estimate the models, and then make a 1-step ahead forecast of the following week's volatility.
- Then they roll the sample forward one observation at a time, constructing a new one step ahead forecast at each step.

Out-of-Sample Forecast Evaluation

- They evaluate the forecasts in two ways:
- The first is by regressing the realised volatility series on the forecasts plus a constant:

$$\sigma_{t+1}^2 = b_0 + b_1\sigma_{ft}^2 + \xi_{t+1} \qquad (7)$$

where σ_{t+1}^2 is the "actual" value of volatility, and σ_{ft}^2 is the value forecasted for it during period t.

- Perfectly accurate forecasts imply $b_0 = 0$ and $b_1 = 1$.
- But what is the "true" value of volatility at time t ?

Day & Lewis use 2 measures

1. The square of the weekly return on the index, which they call SR.
2. The variance of the week's daily returns multiplied by the number of trading days in that week.

Out-of Sample Model Comparisons

$$\sigma_{t+1}^2 = b_0 + b_1\sigma_{ft}^2 + \xi_{t+1} \qquad (8.83)$$

Forecasting Model	Proxy for *ex post* volatility	b_0	b_1	R^2
Historic	SR	0.0004	0.129	0.094
		(5.60)	(21.18)	
Historic	WV	0.0005	0.154	0.024
		(2.90)	(7.58)	
GARCH	SR	0.0002	0.671	0.039
		(1.02)	(2.10)	
GARCH	WV	0.0002	1.074	0.018
		(1.07)	(3.34)	
EGARCH	SR	0.0000	1.075	0.022
		(0.05)	(2.06)	
EGARCH	WV	-0.0001	1.529	0.008
		(-0.48)	(2.58)	
Implied Volatility	SR	0.0022	0.357	0.037
		(2.22)	(1.82)	
Implied Volatility	WV	0.0005	0.718	0.026
		(0.389)	(1.95)	

Notes: Historic refers to the use of a simple historical average of the squared returns to forecast volatility; t-ratios in parentheses; SR and WV refer to the square of the weekly return on the S&P 100, and the variance of the week's daily returns multiplied by the number of trading days in that week, respectively. Source: Day and Lewis (1992). Reprinted with the permission of Elsevier Science.

Encompassing Test Results: Do the IV Forecasts Encompass those of the GARCH Models?

$$\sigma_{t+1}^2 = b_0 + b_1\sigma_{It}^2 + b_2\sigma_{Gt}^2 + b_3\sigma_{Et}^2 + b_4\sigma_{Ht}^2 + \xi_{t+1} \qquad (8.86)$$

Forecast comparison	b_0	b_1	b_2	b_3	b_4	R^2
Implied vs. GARCH	-0.00010 (-0.09)	0.601 (1.03)	0.298 (0.42)	-	-	0.027
Implied vs. GARCH vs. Historical	0.00018 (1.15)	0.632 (1.02)	-0.243 (-0.28)	-	0.123 (7.01)	0.038
Implied vs. EGARCH	-0.00001 (-0.07)	0.695 (1.62)	-	0.176 (0.27)	-	0.026
Implied vs. EGARCH vs. Historical	0.00026 (1.37)	0.590 (1.45)	-0.374 (-0.57)	-	0.118 (7.74)	0.038
GARCH vs. EGARCH	0.00005 (0.37)	-	1.070 (2.78)	-0.001 (-0.00)	-	0.018

Notes: *t*-ratios in parentheses; the ex post measure used in this table is the variance of the week's daily returns multiplied by the number of trading days in that week. Source: Day and Lewis (1992). Reprinted with the permission of Elsevier Science.

Conclusions of Paper

- Within sample results suggest that IV contains extra information not contained in the GARCH / EGARCH specifications.

- Out of sample results suggest that nothing can accurately predict volatility!

Multivariate GARCH Models

- Multivariate GARCH models are used to estimate and to forecast covariances and correlations. The basic formulation is similar to that of the GARCH model, but where the covariances as well as the variances are permitted to be time-varying.
- There are 3 main classes of multivariate GARCH formulation that are widely used: VECH, diagonal VECH and BEKK.

<u>VECH and Diagonal VECH</u>

- e.g. suppose that there are two variables used in the model. The conditional covariance matrix is denoted H_t, and would be 2×2. H_t and VECH(H_t) are

$$H_t = \begin{bmatrix} h_{11t} & h_{12t} \\ h_{21t} & h_{22t} \end{bmatrix} \qquad VECH(H_t) = \begin{bmatrix} h_{11t} \\ h_{22t} \\ h_{12t} \end{bmatrix}$$

VECH and Diagonal VECH

- In the case of the VECH, the conditional variances and covariances would each depend upon lagged values of all of the variances and covariances and on lags of the squares of both error terms and their cross products.
- In matrix form, it would be written

$$VECH(H_t) = C + A\,VECH(\Xi_{t-1}\Xi'_{t-1}) + B\,VECH(H_{t-1}) \qquad \Xi_t|\psi_{t-1} \sim N(0, H_t)$$

- Writing out all of the elements gives the 3 equations as

$$h_{11t} = c_{11} + a_{11}u_{1t}^2 + a_{12}u_{2t}^2 + a_{13}u_{1t}u_{2t} + b_{11}h_{11t-1} + b_{12}h_{22t-1} + b_{13}h_{12t-1}$$
$$h_{22t} = c_{21} + a_{21}u_{1t}^2 + a_{22}u_{2t}^2 + a_{23}u_{1t}u_{2t} + b_{21}h_{11t-1} + b_{22}h_{22t-1} + b_{23}h_{12t-1}$$
$$h_{12t} = c_{31} + a_{31}u_{1t}^2 + a_{32}u_{2t}^2 + a_{33}u_{1t}u_{2t} + b_{31}h_{11t-1} + b_{32}h_{22t-1} + b_{33}h_{12t-1}$$

- Such a model would be hard to estimate. The diagonal VECH is much simpler and is specified, in the 2 variable case, as follows:

$$h_{11t} = \alpha_0 + \alpha_1 u_{1t-1}^2 + \alpha_2 h_{11t-1}$$
$$h_{22t} = \beta_0 + \beta_1 u_{2t-1}^2 + \beta_2 h_{22t-1}$$
$$h_{12t} = \gamma_0 + \gamma_1 u_{1t-1}u_{2t-1} + \gamma_2 h_{12t-1}$$

BEKK and Model Estimation for M-GARCH

- Neither the VECH nor the diagonal VECH ensure a positive definite variance-covariance matrix.
- An alternative approach is the BEKK model (Engle & Kroner, 1995).
- In matrix form, the BEKK model is

$$H_t = W'W + A'H_{t-1}A + B'\Xi_{t-1}\Xi'_{t-1}B$$

- Model estimation for all classes of multivariate GARCH model is again performed using maximum likelihood with the following *LLF*:

$$\ell(\theta) = -\frac{TN}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\left(\log\left|H_t\right| + \Xi_t' H_t^{-1}\Xi_t\right)$$

where N is the number of variables in the system (assumed 2 above), θ is a vector containing all of the parameters to be estimated, and T is the number of observations.

An Example: Estimating a Time-Varying Hedge Ratio for FTSE Stock Index Returns (Brooks, Henry and Persand, 2002).

- Data comprises 3580 daily observations on the FTSE 100 stock index and stock index futures contract spanning the period 1 January 1985 - 9 April 1999.
- Several competing models for determining the optimal hedge ratio are constructed. Define the hedge ratio as β.
  - No hedge (β=0)
  - Naïve hedge (β=1)
  - Multivariate GARCH hedges:
    - Symmetric BEKK
    - Asymmetric BEKK

    In both cases, estimating the OHR involves forming a 1-step ahead forecast and computing

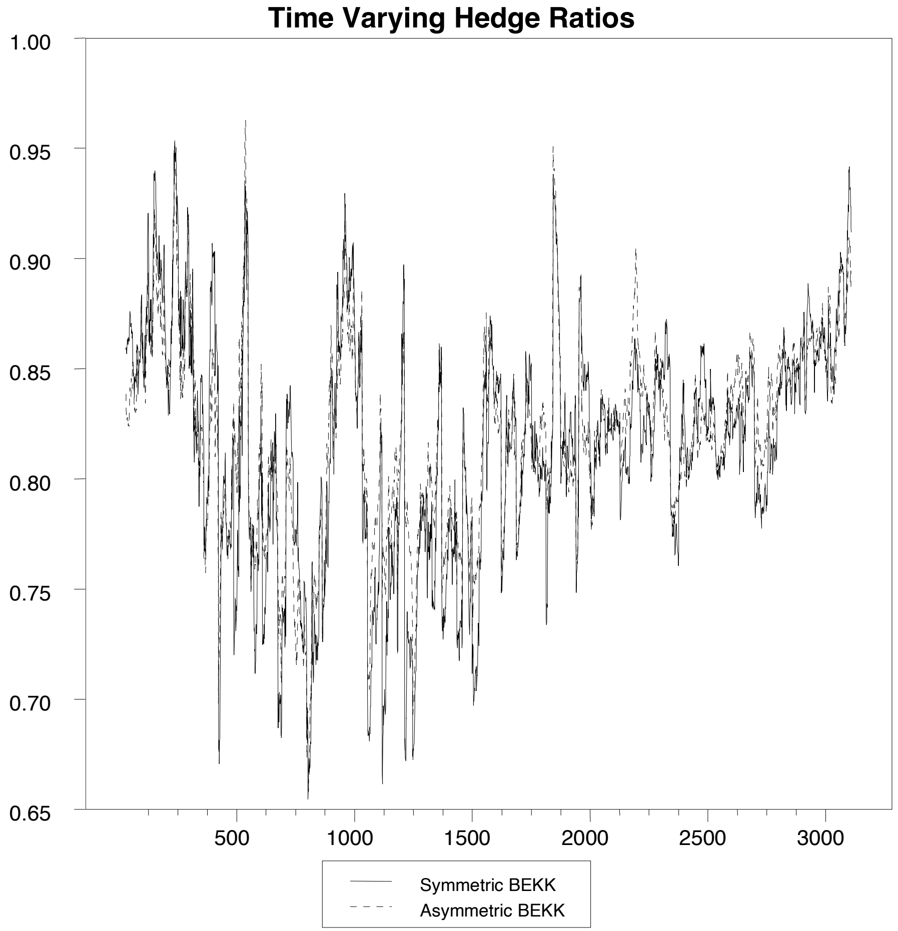
$$OHR_{t+1} = -\left.\frac{h_{CF,t+1}}{h_{F,t+1}}\right| \quad \Omega_t$$

OHR Results

	In Sample			
	Unhedged β = 0	Naïve Hedge β = 1	Symmetric Time Varying Hedge $\beta_t = \frac{h_{FC,t}}{h_{F,t}}$	Asymmetric Time Varying Hedge $\beta_t = \frac{h_{FC,t}}{h_{F,t}}$
Return	0.0389 {2.3713}	-0.0003 {-0.0351}	0.0061 {0.9562}	0.0060 {0.9580}
Variance	0.8286	0.1718	0.1240	0.1211

	Out of Sample			
	Unhedged β = 0	Naïve Hedge β = 1	Symmetric Time Varying Hedge $\beta_t = \frac{h_{FC,t}}{h_{F,t}}$	Asymmetric Time Varying Hedge $\beta_t = \frac{h_{FC,t}}{h_{F,t}}$
Return	0.0819 {1.4958}	-0.0004 {0.0216}	0.0120 {0.7761}	0.0140 {0.9083}
Variance	1.4972	0.1696	0.1186	0.1188

Plot of the OHR from Multivariate GARCH



<u>Conclusions</u>

- OHR is time-varying and less than 1
- M-GARCH OHR provides a better hedge, both in-sample and out-of-sample.
- No role in calculating OHR for asymmetries