Lecture Notes (4)

Modelling Volatility

1

1

Motivations: An Excursion into Non-linearity Land

- All of the models that have been discussed so far have been linear in nature- that is, the model is linear in the parameters. The "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 $u_t \sim N(0, \sigma^2)$.
- Motivation: the linear structural (and time series) models cannot explain a number of important features common to much financial data
  - Leptokurtosis: the tendency for financial asset returns to have distributions that exhibit fat tails and excess peakedness at the mean
  - Volatility clustering or volatility pooling: the tendency for volatility in financial markets to appear in bunches. Thus the large returns (of either sign) are expected to follow large returns, and small returns (of either sign) to follow small returns.
  - Leverage effects: the tendency for volatility to rise more following a large price fall than following a price rise of the same magnitude.

2

Non-linear Models: A Definition

- Campbell, Lo and MacKinlay (1997) define a non-linear data generating process as one that can be written
  $y_t = f(u_t, u_{t-1}, u_{t-2}, \ldots)$
  where u_t is an iid error term and f is a non-linear function.
- They also give a slightly more specific definition as
  $y_t = g(u_{t-1}, u_{t-2}, \ldots) + u_t\sigma^2(u_{t-1}, u_{t-2}, \ldots)$
  where g is a function of past error terms only and σ^2 is a variance term.
- Models with nonlinear g(•) are "non-linear in mean", while those with nonlinear σ^2(•) are "non-linear in variance". - ARCH / GARCH.

Models for Volatility

- Historical volatility models
- Implied volatility models
- Exponentially weighted moving average models
- Autoregressive volatility models

(Details refer to pages 420 ~423)

- Autoregressive conditionally heteroscedastic (ARCH) models

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

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

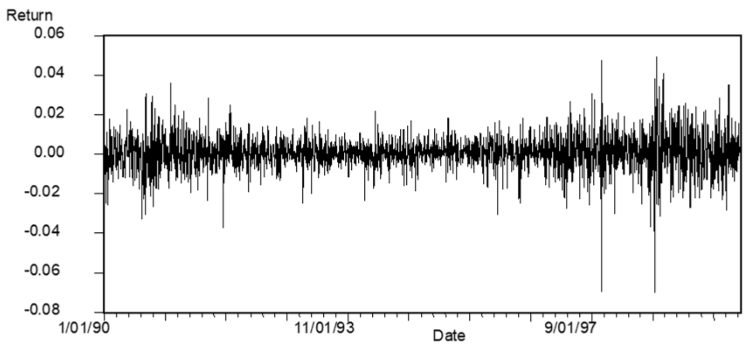
- What if the variance of the errors is not constant?
  - heteroscedasticity
  - would imply that standard error estimates could be wrong.

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

5

A Sample Financial Asset Returns Time Series

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



6

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:
  σ_t^2 = Var($u_t \mid u_{t-1}, u_{t-2}$,...) = E[($u_t$-E($u_t$))$^2 \mid u_{t-1}, u_{t-2}$,...]
  We usually assume that E(u_t) = 0
  so σ_t^2 = Var($u_t \mid u_{t-1}, u_{t-2}$,...) = E[$u_t^2 \mid u_{t-1}, u_{t-2}$,...].
- 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
  $y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t\,,\; u_t \sim N(0, \sigma_t^2)$
  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
  $$y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t\,,\; u_t \sim N(0, h_t)$$
  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

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

$\sigma_t = \sqrt{\alpha_0 + \alpha_1 u_{t-1}^2}$, $v_t \sim$ 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

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

- 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:

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

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

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

Estimation of GARCH Models Using Maximum Likelihood

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

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

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

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

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

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

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

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

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 $u_{t-1} < 0$
$= 0$ otherwise

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

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

An Example of the use of a GJR Model

- Using monthly S&P 500 returns, December 1979- June 1998

- Estimating a GJR model, we obtain the following results.

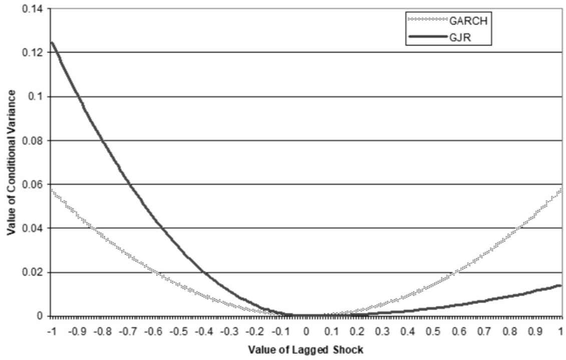
$$\underset{(3.198)}{y_t = 0.172}$$

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

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:

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

- 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

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

Forecasting Variances using GARCH Models (Cont'd)

- Let $\sigma_{1,T}^{f\ 2}$ be the one step ahead forecast for σ^2 made at time T. This is easy to calculate since, at time T, the values of all the terms on the RHS are known.
- $\sigma_{1,T}^{f\ 2}$ would be obtained by taking the conditional expectation of the first equation:
  $\sigma_{1,T}^{f\ 2} = \alpha_0 + \alpha_1 u_T^2 + \beta\sigma_T^2$
- Given, $\sigma_{1,T}^{f\ 2}$ how is $\sigma_{2,T}^{f\ 2}$, the 2-step ahead forecast for σ^2 made at time T, calculated? Taking the conditional expectation of the second equation:
  $\sigma_{2,T}^{f\ 2} = \alpha_0 + \alpha_1 \mathrm{E}(u_{T+1}^2 \mid \Omega_T) + \beta\, \sigma_{1,T}^{f\ 2}$
- where $\mathrm{E}(u_{T+1}^2 \mid \Omega_T)$ is the expectation, made at time T, of u_{T+1}^2 which is the squared disturbance term.