Certificate in Quantitative Finance

Volatility Models: the ARCH framework

Lecture notes provided by:

Stephen Taylor

Lancaster University Management School, UK

s.taylor@lancaster.ac.uk

10 March 2016

Asset Price Dynamics, Volatility, and Prediction

Yesterday's lecture	Covered in Chapters
Volatility clustering	2
Properties of daily returns	4
Properties of high-frequency returns	12
Today's lecture	
Volatility concepts	8
GARCH models	9 & 10
Background material	
Random variables and stochastic processes	3

After today's class you will:

- Know there are many ways to define volatility,
- Know why ARCH models are often estimated from daily price series,
- Appreciate that estimating ARCH models and volatilities is straightforward,
- Have seen simple equations for predicting volatility,
- Realize that stock index volatility generally moves in the opposite direction to the index.

Volatility modelled using time series information

Volatility is simply the variability of prices, defined by the standard deviation of returns. There are several definitions of volatility, each of which has constructive applications.

5 ways to refer to volatility:

- A *parameter* which appears in special stochastic differential equations, such as geometric Brownian motion,

- A *realized* quantity, calculated as the standard deviation of a set of returns,
- A *conditional* quantity, which depends on a model and a history of returns,
- A *stochastic* variable, which can have its own s.d.e.,
- An *implied* quantity, determined by one or more option prices and often by applying an option pricing formula.

Recommended reading: Section 8.1.

1. Why does volatility change ?

There is **no** simple answer to this fundamental question, although several factors are known to be relevant:

- Volatility increases during crises –

  Great Depression (economic, 1929 ..),

  Watergate (political, 1973-4),

  Crashes (financial, 1987).

- Volatility is more likely to increase when the stock market falls, which is when leverage (the debt/equity ratio) increases.
- Volatility does depend on some macroeconomic variables.
- Volatility increases when the amount of relevant information per unit time increases.
- Volatility correlates with trading volume, but it is inappropriate to say that one variable causes the other.

Recommended reading: Sections 8.2 and 8.3.

2. The family of ARCH models

These are models (stochastic processes) which successfully describe time series of asset returns, recorded weekly, daily or hourly.

There are applications for all types of assets - equities, interest rates, currencies, etc.

AR = *autoregressive*,

CH = *conditional heteroscedastic*.

Working within the ARCH framework allows:

- Estimation of parameters as accurately as possible, using the maximum likelihood principle.
- Tests which compare simple models with more complicated specifications.
- Multi-asset calculation of covariances.
- Calculation of forecasts for future volatility, which might be used to
  - Manage risk,
  - Value derivatives,
  - Optimize portfolio weights.

ARCH models are defined by conditional density functions. A specific model is given by defining the

- Conditional mean,
- Conditional variance, and
- Shape of the density.

A popular example is first discussed, followed by the general framework and then further examples. These examples are all for *daily* returns.

3. GARCH(1, 1) - the simplest credible example

A simple and very popular choice defines conditional distributions for random variables r_t, that represent returns, by using previous returns r_{t-1} etc, :

$$r_t | r_{t-1}, r_{t-2}, \ldots \sim N(\mu, h_t)$$

with the conditional variances defined recursively by

$$h_t = \omega + \alpha(r_{t-1} - \mu)^2 + \beta h_{t-1}.$$

- The conditional means are constant and equal to μ.
- The conditional densities are all normal.

- Parameter vector $\theta = (\mu, \omega, \alpha, \beta)$, so that four numbers have to be estimated from returns data.
- To ensure h_t is positive, require $\omega \geq 0, \alpha \geq 0, \beta \geq 0$.
- To ensure a stationary process, need $\alpha + \beta < 1$.

For daily returns, typically

- α is near zero, often < 0.1,
- $\alpha + \beta$ is near one, often > 0.95.

GARCH(1, 1) states that conditional variances are a weighted combination of all previous squared excess returns :

$$h_t = \frac{\omega}{1-\beta} + \alpha\left(r_{t-1} - \mu\right)^2$$

$$+ \alpha\beta\left(r_{t-2} - \mu\right)^2 + \alpha\beta^2\left(r_{t-3} - \mu\right)^2 + \ldots$$

Note: the conditional variance has a positive lower bound.

Stylized facts

GARCH(1, 1) and many other ARCH specifications are consistent with the three major stylized facts for asset returns.

For the GARCH(1, 1) stochastic process, with $\alpha+\beta<1$:

- Returns have kurtosis which exceeds 3 and may even be infinite,
- Returns are uncorrelated, because the conditional mean does not depend on previous returns.
- The autocorrelations of squared excess returns, $s_t=(r_t-\mu)^2$, exist when the kurtosis is finite and then they are all positive.

Positive dependence in the series of conditional variances h_t creates positive dependence in s_t.

The autocorrelations of $s_t = (r_t - \mu)^2$ for GARCH(1, 1) have the same shape as those of an ARMA(1, 1) process:

$$\rho_{\tau,s} = C(\alpha + \beta)^{\tau}, \quad \tau > 0,$$

with C a positive function of α and β alone, providing the autocorrelations exist.

Recommended reading: Section 9.3.

Textbook example

DM/$ daily returns, 1991-2000.

Parameters are estimated by maximising the log-likelihood function, as explained later. The estimates include:

$$\hat{\alpha} = 0.035 \text{ (standard error 0.008)}$$

$$\hat{\beta} = 0.955 \text{ (standard error 0.010)}$$

$$\hat{\alpha} + \hat{\beta} = 0.991 \text{ (standard error 0.005)}$$

Volatility σ_t is shown on the following Figures as an annualised quantity, obtained from trading day standard deviations $\sqrt{h_t}$ by $\sigma_t = \sqrt{259 h_t}$.

Figure 9.1 shows σ_t for all ten years, with :

Mean 10.6%

Interquartile range 8.8% to 11.7%

Range 6.5% to 19.1%.

Figure 9.2 shows σ_t for 1992, a high volatility year.

Figure 9.3 shows σ_t for 1996, a low volatility year.

Figure 9.1 DM/$ volatility 1991-2000

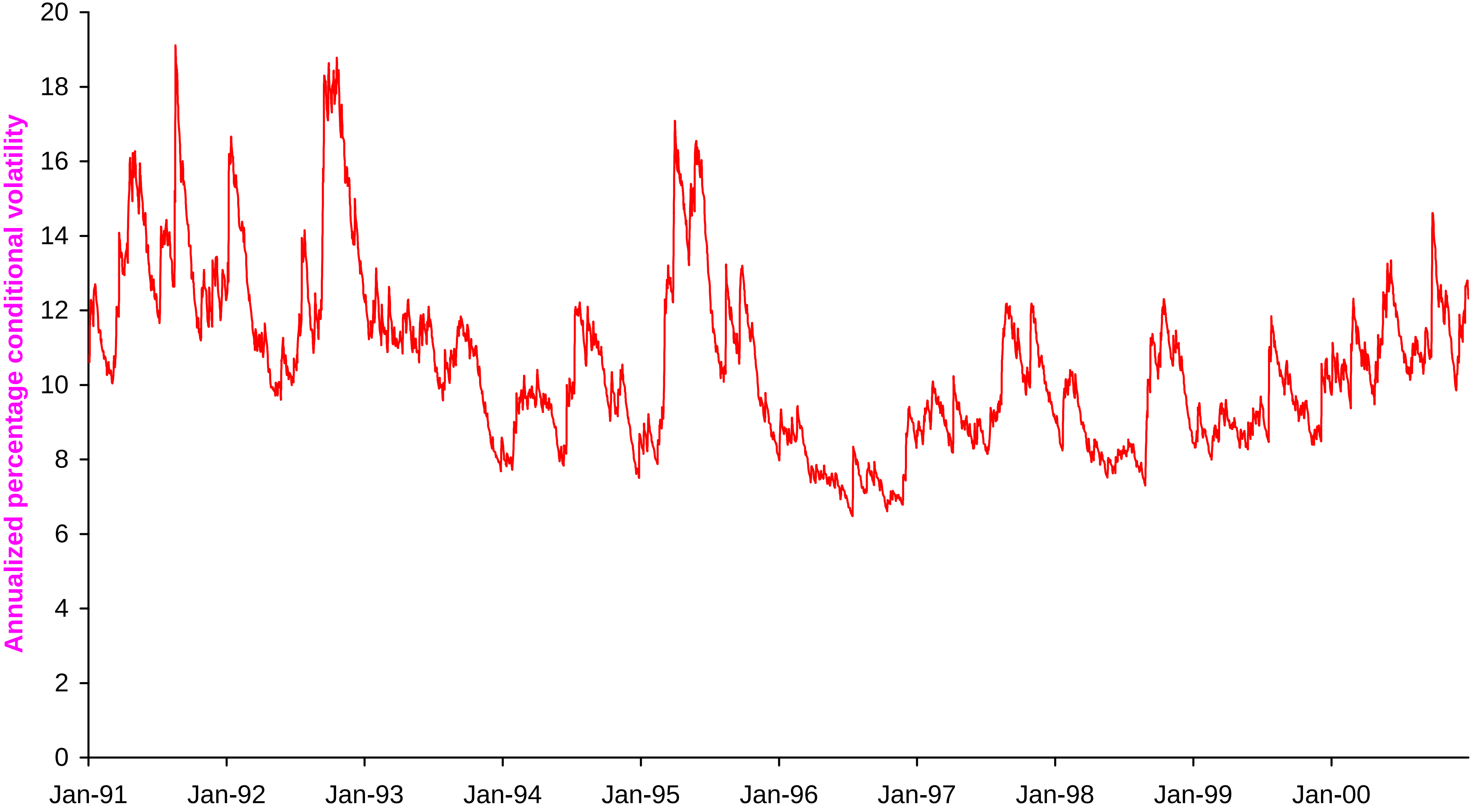


Figure 9.2 DM/$ volatility and returns in 1992

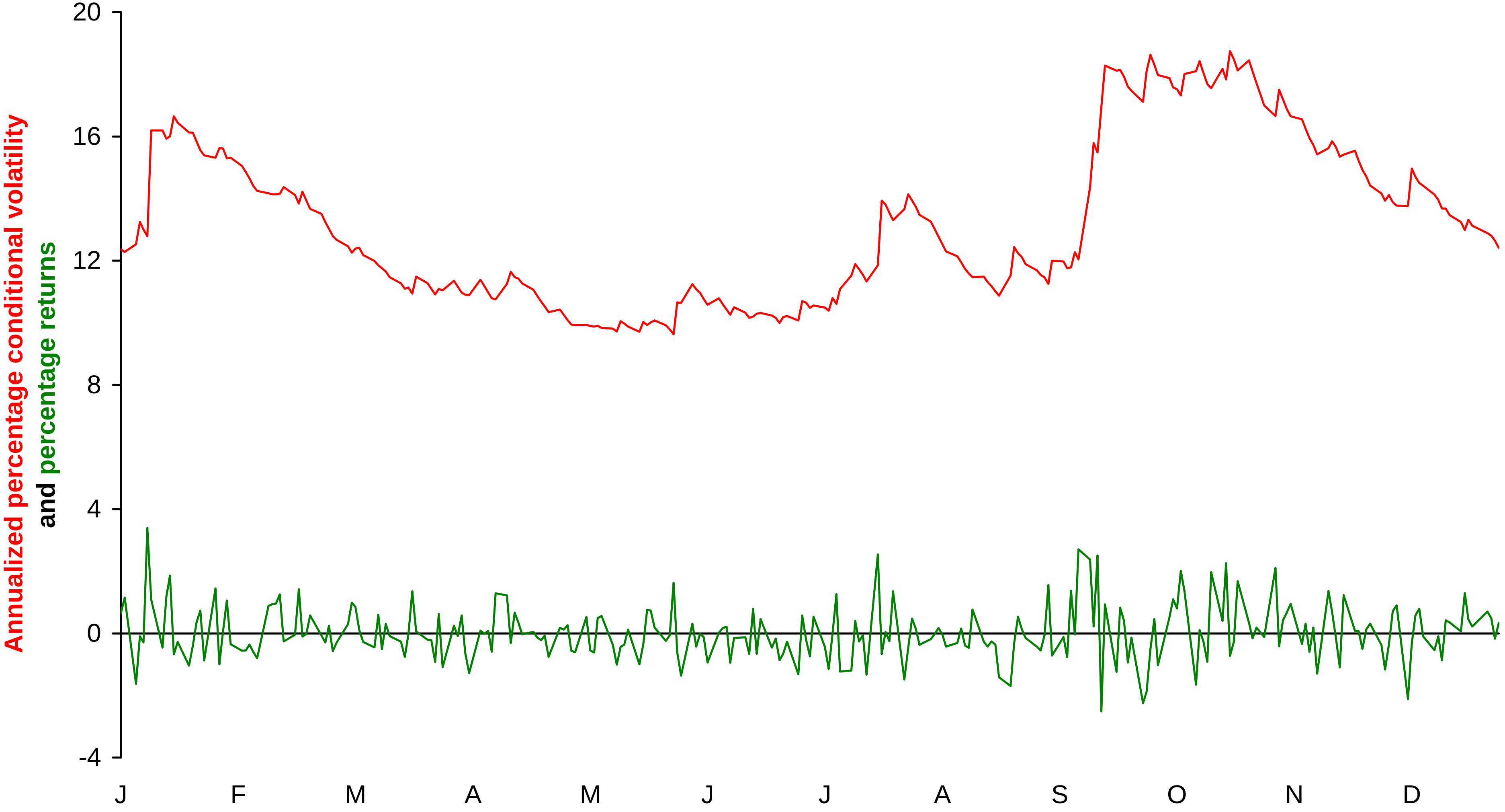
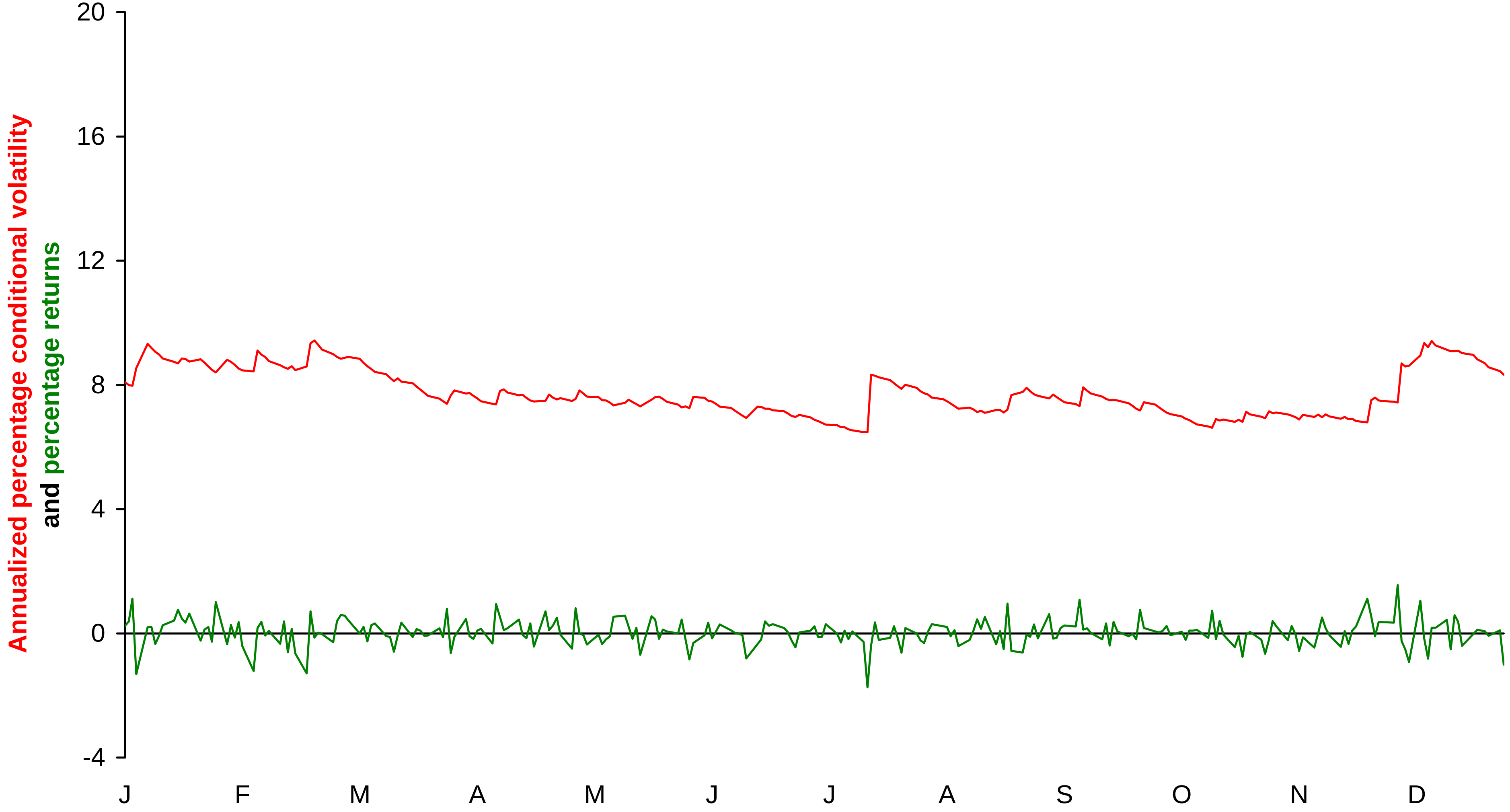


Figure 9.3 DM/$ volatility and returns in 1996



Variance forecasts

For the stationary GARCH(1,1) model, these are obtained at time $t-1$ from information

$$I_{t-1} = \{r_{t-1}, r_{t-2}, r_{t-3}, \ldots\}$$

and the *conditional* variances

$$\text{var}(r_t | I_{t-1}) = h_t,$$

$$\text{var}(r_{t+n} | I_{t-1}) = V + (\alpha + \beta)^n (h_t - V), \quad n = 1,2,..$$

which depend on the *unconditional* variance

$$V = \text{var}(r_t) = \frac{\omega}{1-\alpha-\beta}.$$

Figure 9.4 illustrates forecasts of annualised volatility for the DM/$ estimates. They converge to the unconditional annualised standard deviation $\sqrt{259V}$ as n increases.

The rate of convergence is determined by the *persistence* parameter, $\alpha + \beta$. The *half-life* is the time H for which the forecast is half-way between the current level of volatility and the unconditional level. It solves

$$(\alpha + \beta)^H = \frac{1}{2}$$

and equals 75 trading days (3.5 months) for the DM/$ series.

Recommended reading: Section 9.4.

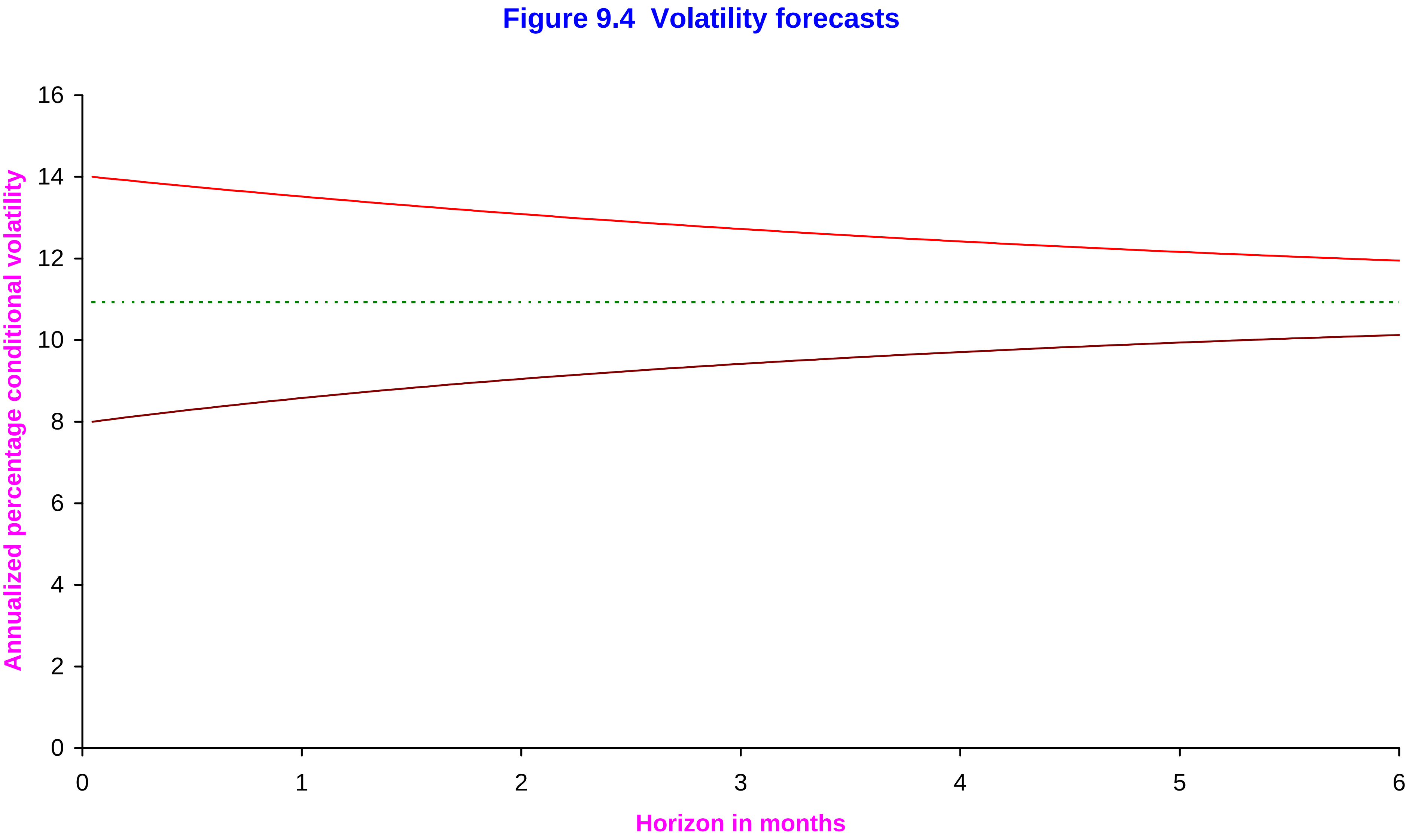


Figure 9.4 Volatility forecasts
Annualized percentage conditional volatility
16
14
12
10
8
6
4
2
0
0
1
2
3
4
5
6
Horizon in months

4. The general framework

ARCH models have six components:

a) Returns r_t, with t counting trading periods. Often t counts trading days.

b) Information sets I_{t-1} whose contents are known before subsequent times t. The simplest example is the returns history $I_{t-1} = \{r_{t-1}, r_{t-2}, r_{t-3}, \ldots\}$. More complicated examples include options information.

c) A function that defines conditional means:

$$\mu_t = E[r_t | I_{t-1}].$$

This function might be a constant, or a function of previous returns, or dependent on the day-of-the-week, etc. It can also depend on the conditional variance.

d) A function that defines conditional variances:

$$h_t = \text{var}(r_t | I_{t-1}).$$

This is the key feature of the model. It is usually a function of previous returns, may include calendar terms and may use information that is additional to the returns price history.

e) A family of conditional distributions D, for which

$$r_t | I_{t-1} \sim D(\mu_t, h_t).$$

The simplest example is provided by conditional normal distributions. For any family D it is required that the standardised variables

$$z_t = \frac{r_t - \mu_t}{\sqrt{h_t}}$$

have the property

$$z_t \mid I_{t-1} \sim D(0,1)$$

whatever the information I_{t-1}. The standardised variables are then independent and identically distributed (i.i.d.).

f) A parameter vector θ. This contains all the parameters in the functions μ_t and h_t, and also any needed to define the density of $D(0,1)$.

Recommended reading: Section 9.5.

5. GJR(1, 1) - more credible for equity returns

For any ARCH model, the residuals are: $e_t = r_t - \mu_t$.

GARCH has the limitation that positive and negative residuals have the same impact on the next variance. This is very often inappropriate for equity markets.

Asymmetric volatility models contain a variable which separates positive and negative residuals:

$$\begin{aligned} S_t &= 1 \text{ if } e_t < 0, \\ &= 0 \text{ if } e_t \geq 0. \end{aligned}$$

GARCH(1, 1) can be changed to incorporate asymmetry by changing the conditional variances to:

$$h_t = \omega + (\alpha + \alpha^- S_{t-1})(r_{t-1} - \mu)^2 + \beta h_{t-1}.$$

This specification:

- Has **one** additional parameter, α^-, which is almost always **positive** for equity index returns.
- Multiplies squared residuals by either $\alpha + \alpha^-$ or α, when returns r_{t-1} are respectively below and above expectations μ_{t-1}.

- Has *persistence* parameter $\phi = \alpha + \frac{1}{2}\alpha^{-} + \beta$, which appears in forecasting equations. Now

$$\mathrm{var}\left(r_{t+n} \middle| I_{t-1}\right) = V + \phi^{n}\left(h_t - V\right), \quad n = 1,2,..$$

$$V = \mathrm{var}\left(r_t\right) = \frac{\omega}{1-\phi}.$$

- GJR is named after Glosten, Jagannathan & Runkle.

Recommended reading: Section 9.7.

ARCH-M, textbook example

Data are daily S & P 100 index returns, 1991 to 2000.

Suppose the conditional mean has MA(1) and conditional variance components ("M"):

$$r_t = \mu_t + e_t,$$

$$\mu_t = \mu + \lambda\sqrt{h_t} + \Theta e_{t-1}.$$

Also suppose the conditional variance follows GJR(1,1):

$$h_t = \omega + (\alpha + \alpha^{-} S_{t-1}) e_{t-1}^2 + \beta h_{t-1}.$$

We discuss results here when the conditional distributions are normal.

The parameter vector is now: $\theta = \left(\mu, \lambda, \omega, \Theta, \alpha, \alpha^{-}, \beta\right).$

The estimates include:

- A positive value for λ, which is probably significantly different from zero (a test is described later).
- A value of Θ very near zero.
- $\alpha = 0.011$, $\alpha^{-} = 0.087$, so there is a **very strong** asymmetric effect. Squares of negative residuals are multiplied by 0.098, which is 9 times the multiplier for squares of positive residuals.
- Persistence $\phi = 0.987$, forecast half-life $H = 52$ trading days (2.5 months).

There is clear evidence that the true conditional distributions are not normal – but ignoring this fact is acceptable when estimating the parameters.

Figure 9.6 shows the annualised volatility $\sigma_t = \sqrt{253 h_t}$ for all ten years, with:

Mean 14.6%

Interquartile range 10.5% to 17.6%

Range 7.3% to 47.1%

The mean is 17.9% for 1996-2000, compared with 11.3% for 1991-1995.

Figure 9.7 shows σ_t for 2000, a high volatility year.

Figure 9.8 shows σ_t for 1995, a low volatility year.

Recommended reading: Section 9.8.

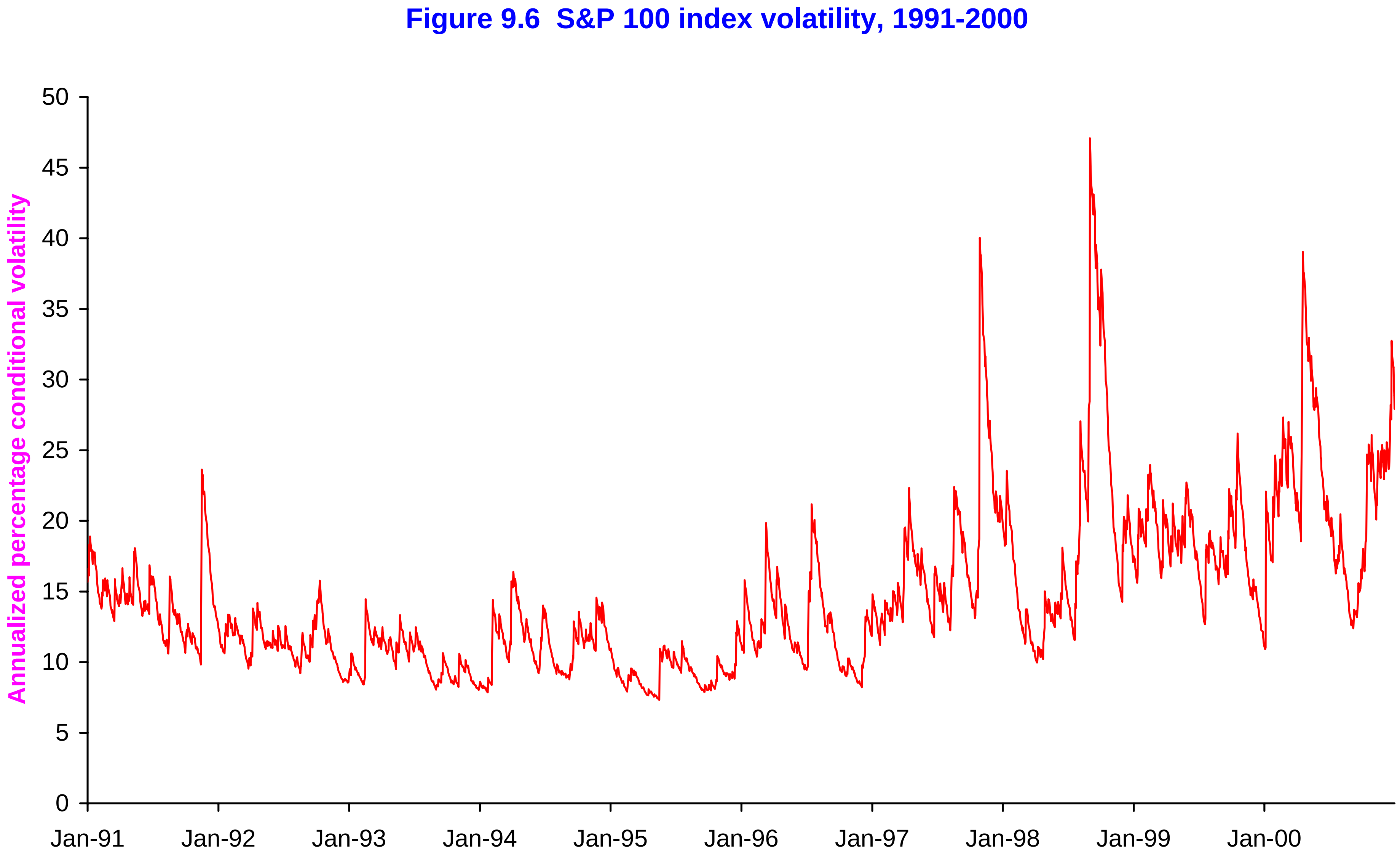
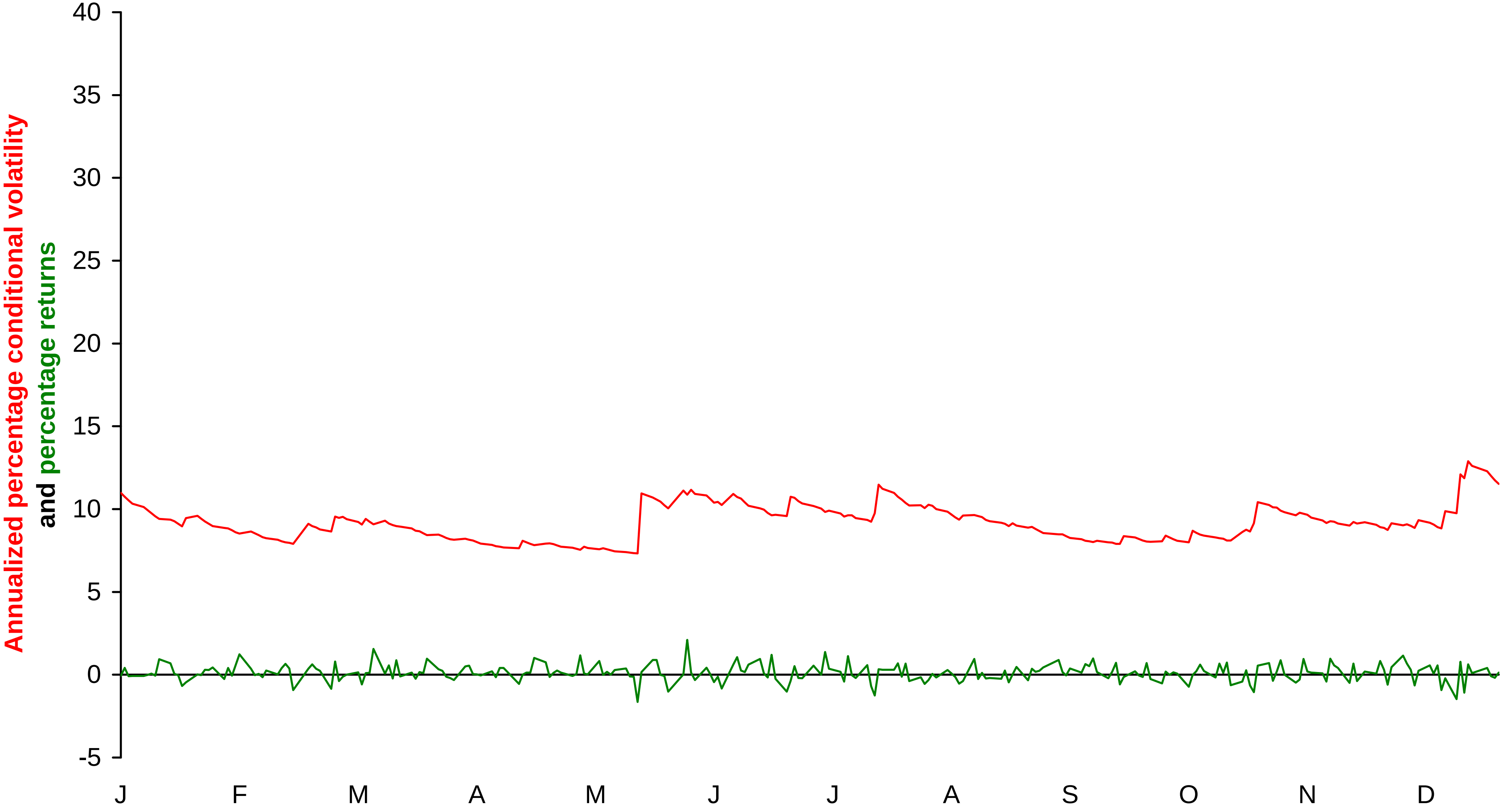


Figure 9.6 S&P 100 index volatility, 1991-2000
Annualized percentage conditional volatility
50
45
40
35
30
25
20
15
10
5
0
Jan-91
Jan-92
Jan-93
Jan-94
Jan-95
Jan-96
Jan-97
Jan-98
Jan-99
Jan-00

Figure 9.7 SP 100 index volatility and returns in 2000

Figure 9.8 S&P 100 index volatility and returns in 1995



A second ARCH-M example

Data are daily FTSE 100 index returns, from November 2005 to October 2015.

For the same ARCH-M specification as before, the estimates include:

- $\alpha = 0.000$, $\alpha^{-} = 0.191$, so there is an **exceptional** asymmetric effect. In this case, volatility only increases after a market fall.
- Persistence $\phi = 0.974$.

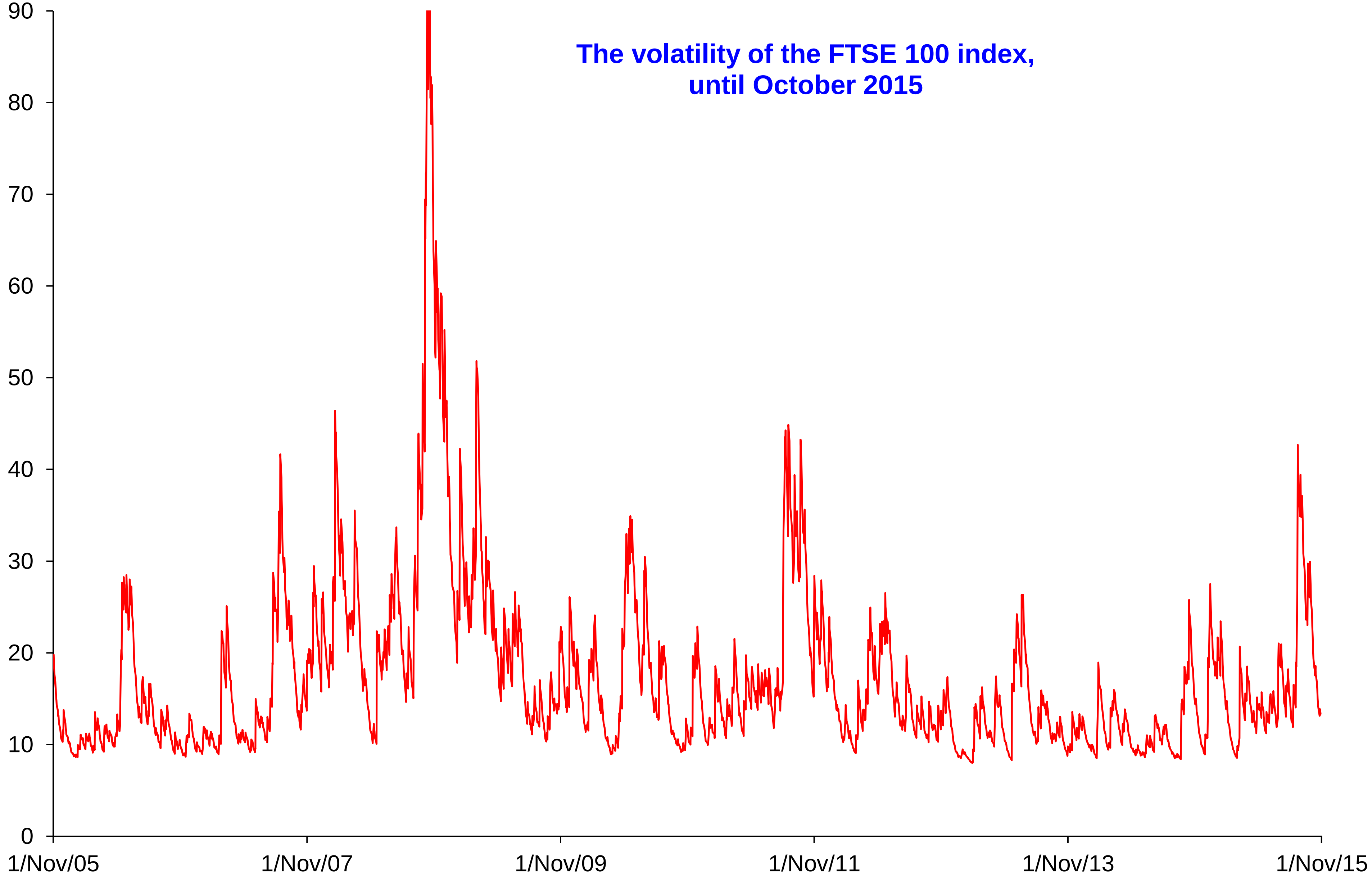
The next Figure shows the annualised volatility for the 10-year period, which

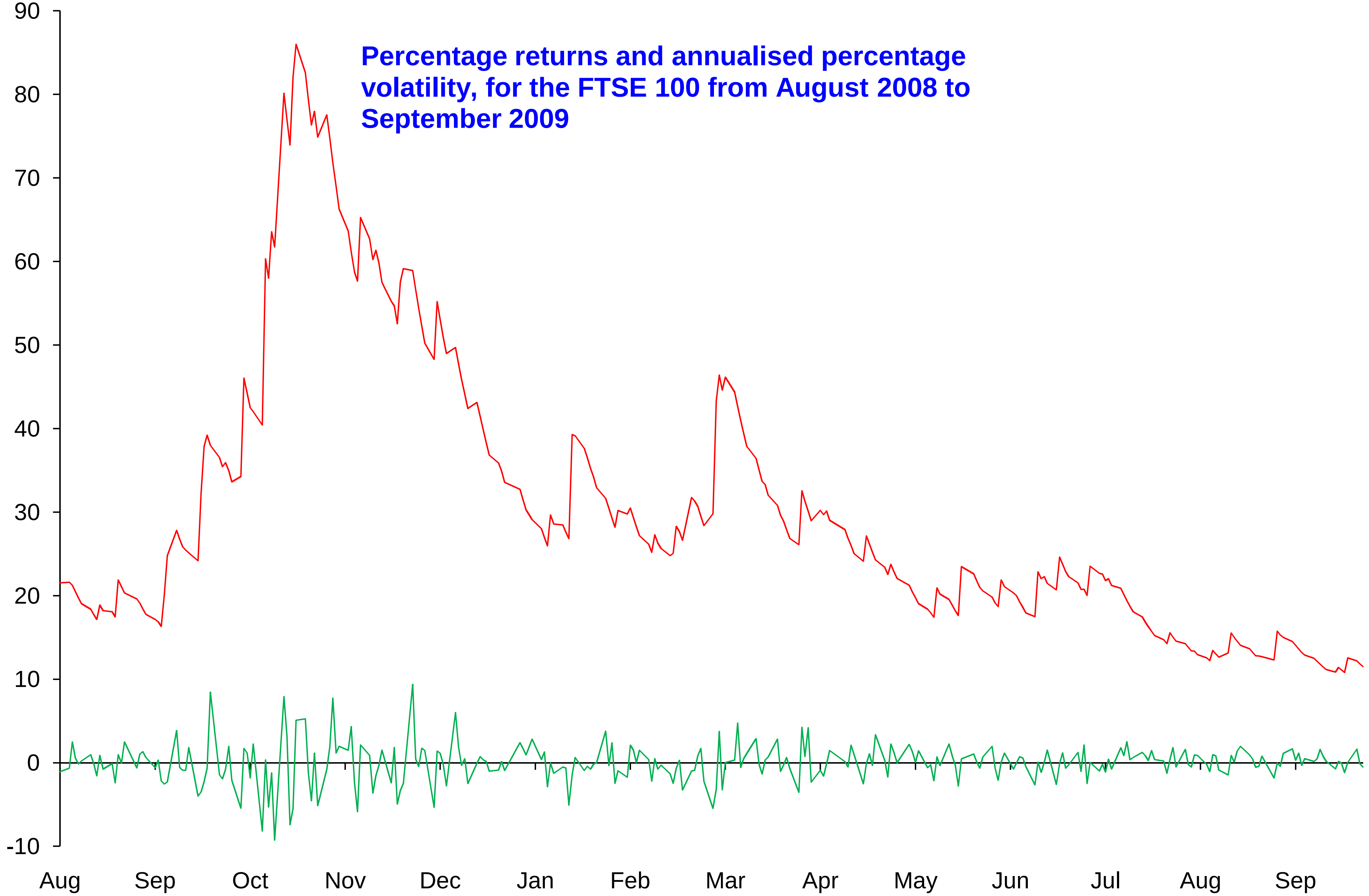
- commences at 20%,
- falls below 10% on some days in 2006, 2007, 2010, 2012, 2013, 2014 & 2015,
- rises above 40% on some days in 2007, 2008, 2009, 2011 & 2015.

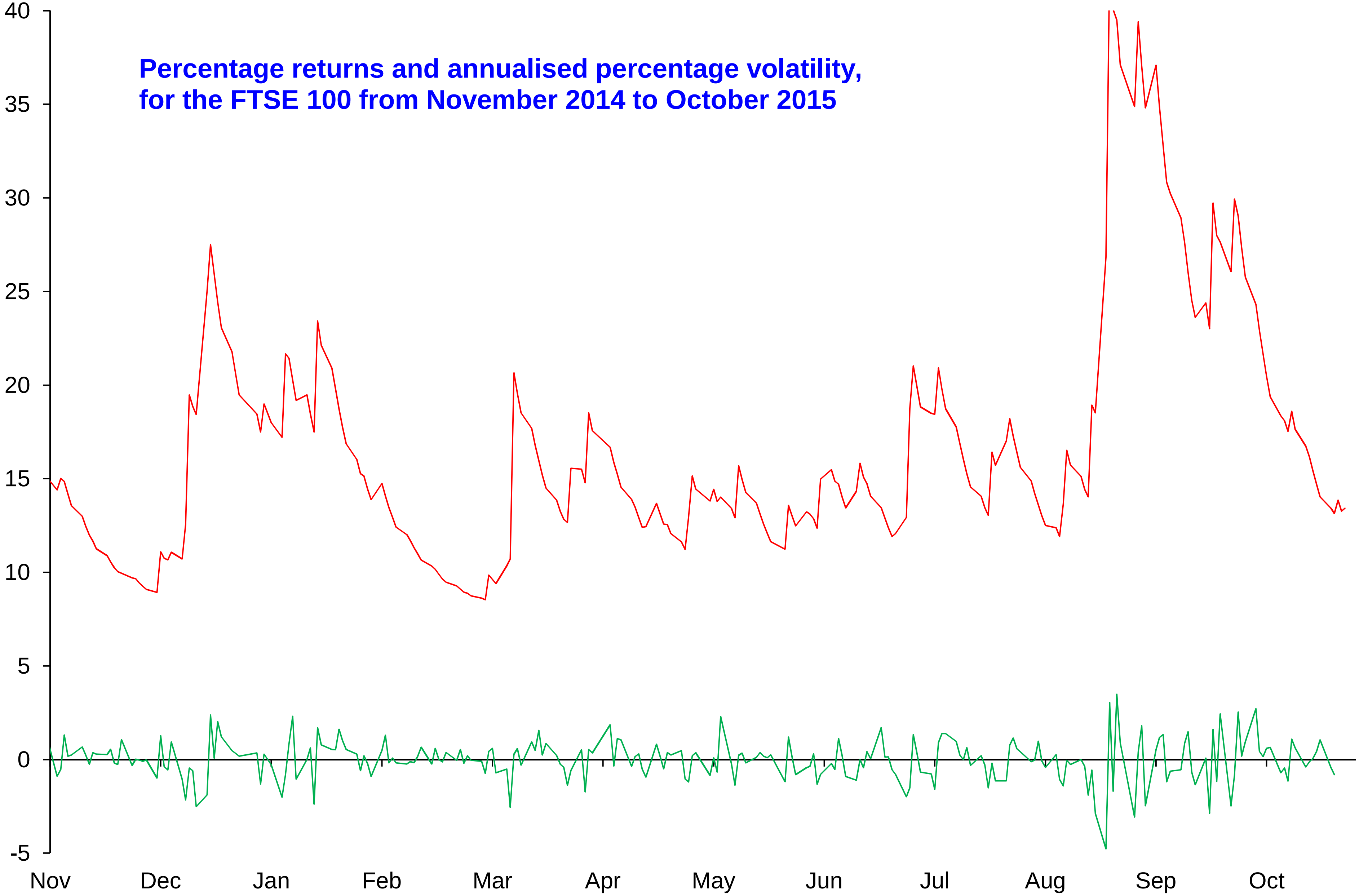
The first detailed Figure covers the 2008 crisis:

- volatility soars to approximately 90% after the index falls by 9.3% on 10 October, followed by consecutive losses of 7.4% and 5.5% on 15 & 16 October.

The final Figure has a range for volatility from 9% to 43% during the last year of the data.


The volatility of the FTSE 100 index,
until October 2015
90
80
70
60
50
40
30
20
10
0
1/Nov/05
1/Nov/07
1/Nov/09
1/Nov/11
1/Nov/13
1/Nov/15


Percentage returns and annualised percentage volatility, for the FTSE 100 from August 2008 to September 2009
90
80
70
60
50
40
30
20
10
0
-10
Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep


Percentage returns and annualised percentage volatility,
for the FTSE 100 from November 2014 to October 2015
40
35
30
25
20
15
10
5
0
-5
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct

Why is index volatility asymmetric?

Asymmetric effects are found in U.S. indices throughout the 1900s. The magnitude of the effect is time-varying and it has become stronger in recent years.

The effects are weaker for individual firms than for indices. *Correlations between firm returns increase when the market falls*, which explains the higher index asymmetry.

An older opinion is that debt/equity ratios increase in a falling market and hence equity is more risky and so more volatile. But the asymmetric effect, (1) is very large compared with changes in d/e ratios, (2) occurs for firms which are debt-free.

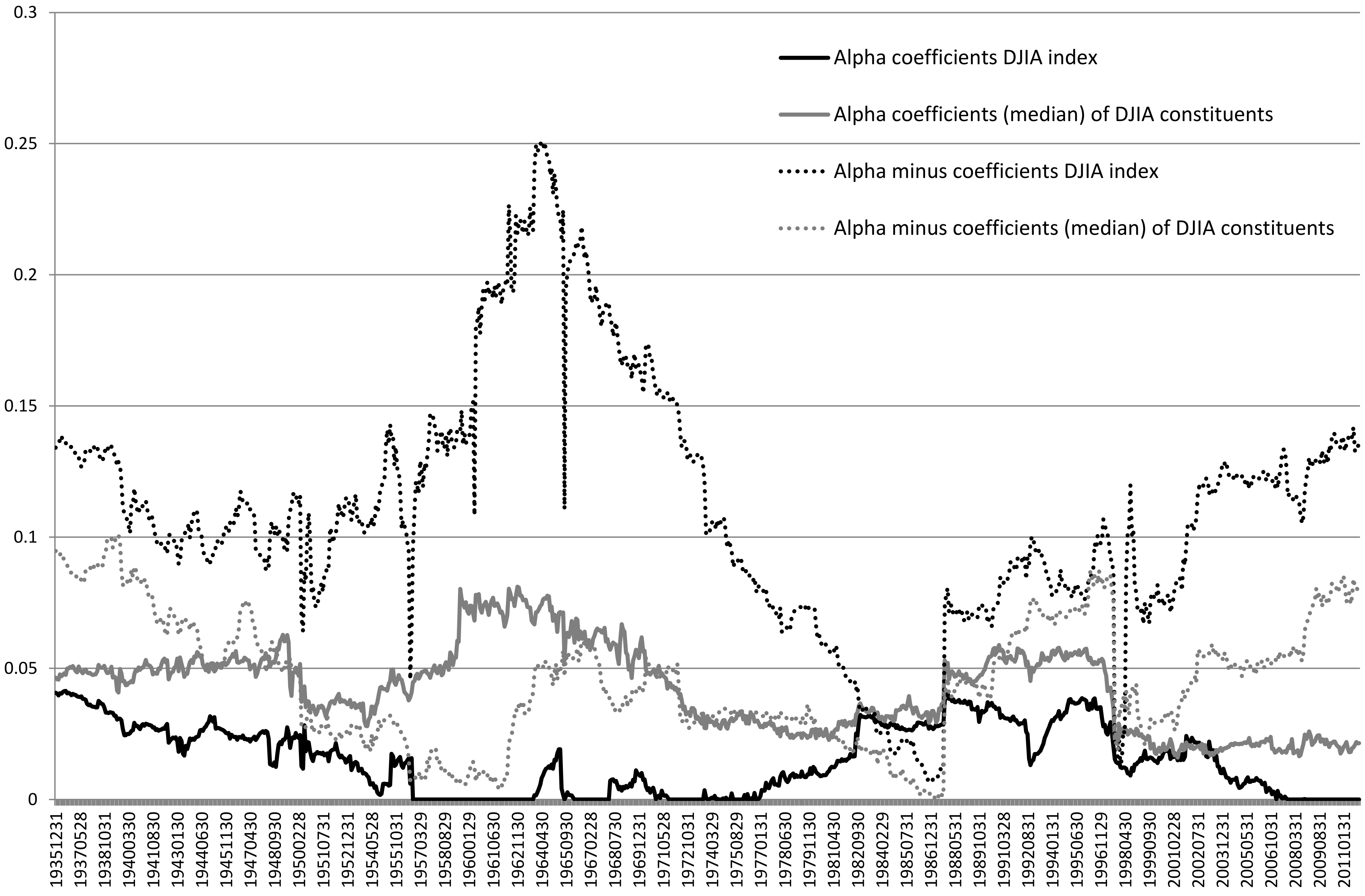
The next page shows estimates of α and α^- calculated by Tristan Linke:

- For 10-year windows, of *weekly* returns, ending in years from 1936 to 2011,
- For the Dow Jones index,
- For the DJIA firm having the median estimate.

It is seen that:

- α is usually higher for firms than the index,
- α^- is usually higher for the index than for firms,
- Thus the index has more asymmetry than the firms.

Asymmetry has increased since 1990, both for the index and the median firm.



Alpha coefficients DJIA index
Alpha coefficients (median) of DJIA constituents
Alpha minus coefficients DJIA index
Alpha minus coefficients (median) of DJIA constituents
0.3
0.25
0.2
0.15
0.1
0.05
0
19351231
19370528
19381031
19400330
19410830
19430130
19440630
19451130
19470430
19480930
19500228
19510731
19521231
19540528
19551031
19570329
19580829
19600129
19610630
19621130
19640430
19650930
19670228
19680730
19691231
19710528
19721031
19740329
19750829
19770131
19780630
19791130
19810430
19820930
19840229
19850731
19861231
19880531
19891031
19910328
19920831
19940131
19950630
19961129
19980430
19990930
20010228
20020731
20031231
20050531
20061031
20080331
20090831
20110131

6. Parameter estimation

The parameters θ are obtained by selecting the values that are "most likely" for the observed data. This requires maximisation of the likelihood function. For large datasets the parameter estimates will be as accurate as possible.

The *likelihood* of a dataset is the density function evaluated using the observed data and values of the model's parameters. The *likelihood function* $L(\theta)$ is a function of the parameters θ.

For n returns and conditional information restricted to the history of returns, and with f representing the conditional density of the variable in brackets,

$$\begin{aligned} L(\theta) &= f\left(r_1, r_2, \ldots, r_n \middle| \theta\right) \\ &= f\left(r_1 \middle| \theta\right) f\left(r_2 \middle| I_1, \theta\right) \ldots f\left(r_n \middle| I_{n-1}, \theta\right). \end{aligned}$$

When the conditional densities are normal,

$$r_t \middle| I_{t-1}, \theta \sim N\left(\mu_t(\theta), h_t(\theta)\right).$$

Here μ_t and/or h_t are functions of θ and

$$f\left(r_t \middle| I_{t-1}, \theta\right) = \frac{1}{\sqrt{2\pi h_t(\theta)}} e^{-\frac{1}{2}\left(r_t - \mu_t(\theta)\right)^2 / h_t(\theta)}.$$

The *log-likelihood function* is then

$$\log L(\theta) = -\tfrac{1}{2}\sum_{t=1}^{n}\left[\log(2\pi) + \log(h_t(\theta)) + z_t^2(\theta)\right]$$

with

$$z_t(\theta) = \frac{r_t - \mu_t(\theta)}{\sqrt{h_t(\theta)}}.$$

In calculations it is convenient to make use of:

$$l_t(\theta) = -\tfrac{1}{2}[\log(2\pi) + \log(h_t(\theta)) + z_t^2(\theta)],$$

so $\log L(\theta) = \sum_{t=1}^{n} l_t(\theta)$.

A single value of the log-likelihood can be calculated as follows:

- Select values for all the parameters in θ,
- Select μ_1, for example the sample mean of all n returns,
- Select h_1, usually the sample variance of all n returns,
- Use r_1 to get
    - $z_1 = (r_1 - \mu_1)/\sqrt{h_1}$,
    - $l_1 = -\frac{1}{2}[\log(2\pi) + \log(h_1) + z_1^2]$,
    - μ_2,
    - h_2,

- Use r_2 to get
  - $z_2 = (r_2 - \mu_2)/\sqrt{h_2}$,
  - $l_2 = -\frac{1}{2}[\log(2\pi) + \log(h_2) + z_2^2]$,
  - μ_3,
  - h_3,
- Use r_3 to get z_3, l_3, μ_4, h_4,
- etc. etc.

Finally $\log L(\theta)$ is the sum $l_1 + l_2 + ... + l_n$.

The *maximum likelihood estimate* $\hat{\theta}$ is that set of parameter values that maximises $L(\theta)$ or, equivalently, that maximises log $L(\theta)$.

Software is required for the maximisation. The Solver tool can be used in Excel, although it has limitations. A logical way to organise Excel calculations is to have the terms $r_t, \mu_t, h_t, z_t, l_t$ in the same row.

Recommended reading: The likelihood function is defined in Section 9.5.

Excel calculations are described in Sections 9.4 and 9.8.

File **FTSE_GJR for 14MAR2016.XLSX**:

- Contains the FTSE-100 data discussed previously in these notes.
- Illustrates the GJR calculations.
- Excludes MA(1) and ARCH-M effects.
- Relies on simple Excel calculations.
- Of course, more advanced software can achieve more with less effort.

7. Hypothesis tests

Selection of an ARCH model can be guided by hypothesis tests based upon the likelihood function.

Is the conditional distribution normal?

There are many alternatives to the normal family. Consider a family of t-distributions. These have three parameters: mean, variance and degrees-of-freedom ν. The t-distribution is fat tailed. The kurtosis is finite when $\nu > 4$ and then equals $3 + \dfrac{6}{\nu - 4}$.

As $\nu \to \infty$, the t-distribution converges to a normal distribution.

Let $\zeta = 1/v$. The null hypothesis of a normal distribution can be evaluated by testing

$$H_0 : \zeta = 0 \quad \text{against} \quad H_1 : \zeta > 0.$$

To test the null,

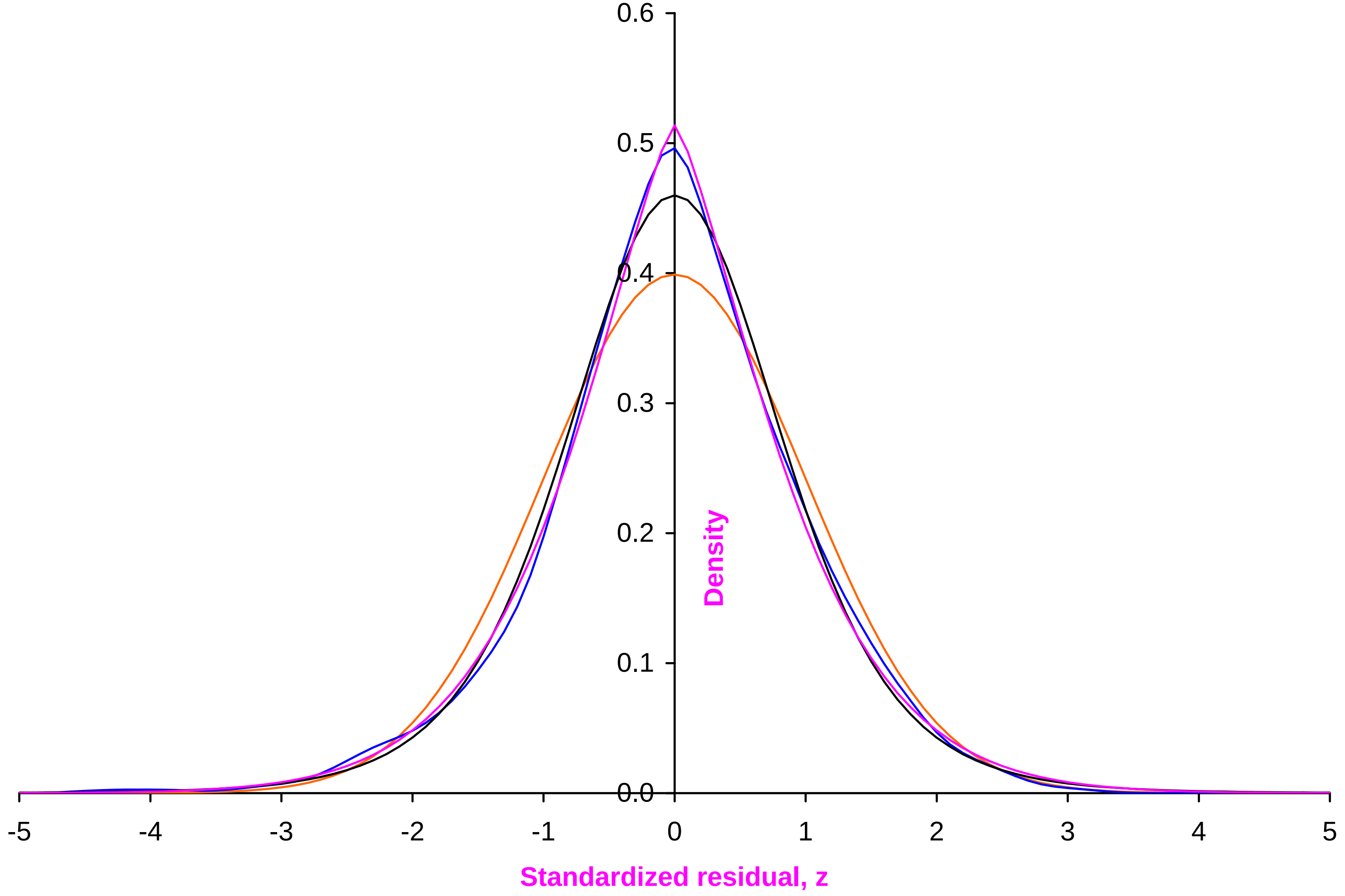
- Estimate the model twice, first for the null hypothesis and second for the alternative hypothesis.
- Let L_0 and L_1 be the maximum values of the log-likelihood function for the two hypotheses.
- Compare $2(L_1 - L_0)$ with the χ^2 distribution, with one degree-of-freedom; one d.o.f. because the alternative has one more parameter than the null.

- As $2(L_1 - L_0) = 257$ for the S & P 100 data, reject the null at very low significance levels. Therefore the best conditional distribution is not normal.
- An alternative test uses $\hat{\zeta}$ divided by its standard error and reaches the same conclusion.

Figure 9.5 shows the empirical density of the standardised residuals z_t (blue curve, peak value 0.50) and three theoretical densities (normal, t and GED, using orange, black and pink curves).

If interested, Sections 9.6 and 9.8 cover t and GED distributions.

Figure 9.5 Density estimates for standardized residuals



Do expected returns depend on volatility?

Consider this ARCH-M specification of the conditional mean :

$$\mu_t = \mu + \lambda \sqrt{h_t} \, .$$

Then the question can be answered by testing

$$H_0 : \lambda = 0 \quad \text{against} \quad H_1 : \lambda \neq 0 .$$

- One method is as before, so let L_0 and L_1 be the maximum values of the log-likelihood function for the two hypotheses and compare $2(L_1 - L_0)$ with the χ^2 distribution, with one degree-of-freedom.

- Obtain $2(L_1 - L_0) = 3.64$ for the S & P 100 data, when t-distributions are used. This is less than 3.84 (the one-tail 5% point of χ_1^2) and suggests the null should be accepted.
- A second method uses λ divided by its standard error, and compares the result with $N(0,1)$. This gives a test value of 2.00 when normal distributions and robust standard errors are used. Then the null is rejected at the 5% level.

The results from several research studies show that any relationship between μ_t and h_t is weak. It appears to vary through time for the U.S. market (if it exists at all) with either positive or negative λ being possible.

Do volatility shocks persist for ever?

The persistence of the GJR(1,1) model is $\phi = \alpha + \frac{1}{2}\alpha^{-} + \beta$. The impact of a return on future volatility eventually disappears when the persistence is less than one. However, if ϕ is one the ARCH model is *integrated*, volatility has a *unit root* and returns have a permanent impact on future volatility.

Tests of

$$H_0 : \phi = 1 \text{ against } H_1 : \phi < 1$$

produce mixed conclusions :

- The null is usually rejected for foreign exchange returns. The robust t-ratio for the illustrative DM/$ returns is $(0.9908-1)/0.00482=-1.91$, which rejects the null at the 5% level (one-tail test).
- The null is sometimes accepted for U.S. stock indices, although the robust t-ratio is –2.48 for the illustrative S & P returns.

Does option trading increase volatility?

Hypothesis tests can also be used to decide if particular events have a permanent impact on volatility.

An often asked question is "Does option trading increase volatility?". We could define an event as the introduction of option trading. Let

$$\begin{aligned} \omega &= \omega_0 \quad \text{before the event,} \\ &= \omega_1 \quad \text{after the event,} \end{aligned}$$

and then test

$$H_0 : \omega_0 = \omega_1 \quad \text{against} \quad H_1 : \omega_0 \neq \omega_1 \ .$$

This requires estimating ARCH models from a history of returns that starts before the event and finishes after it. For U.K. firms, the evidence favours the null hypothesis.

Recommended reading: Section 10.5.

Likelihood theory is a very technical subject, which is covered in Section 10.4.

8. High-frequency ARCH

There are issues which arise when ARCH models are estimated from intraday returns.

For GARCH(1,1) and other simple specifications, estimates of the persistence parameter change a lot as the data frequency increases, which is unsatisfactory.

Appropriate models need to take account of:

- The intraday volatility pattern.
- Multiple volatility components, which have different persistence levels.

9. Multivariate ARCH

Conditional variances and covariances can be estimated simultaneously for two or more assets. A choice then has to be made about the number of estimated parameters, which can be substantial when many assets are modelled.

A relatively simple methodology for estimating large, dynamic correlation matrices has been developed by Engle and is described in his 2002 paper on *dynamic conditional correlation* and in his 2009 book “Anticipating correlations: a new paradigm for risk management”.