Lecture notes: Financial time series, ARCH and GARCH models

Piotr Fryzlewicz
Department of Mathematics
University of Bristol
Bristol BS8 1TW
UK
p.z.fryzlewicz@bristol.ac.uk
http://www.maths.bris.ac.uk/~mapzf/

July 18, 2007

1 Financial time series

Let P_k, $k = 0, \ldots, n$, be a time series of prices of a financial asset, e.g. daily quotes on a share, stock index, currency exchange rate or a commodity. Instead of analysing P_k, which often displays unit-root behaviour and thus cannot be modelled as stationary, we often analyse *log-returns* on P_k, i.e. the series

$$y_k = \log P_k - \log P_{k-1} = \log\left(\frac{P_k}{P_{k-1}}\right) = \log\left(1 + \frac{P_k - P_{k-1}}{P_{k-1}}\right).$$

By Taylor-expanding the above, we can see that y_k is almost equivalent to the relative return $(P_k - P_{k-1})/P_{k-1}$. The reason we typically consider log-returns instead of relative returns is the additivity property of log-returns, which is not shared by relative returns.

As an illustration, consider the time series of daily closing values of the 'WIG' index, which is the main summary index of Warsaw Stock Exchange, running from 16 April 1991 to 5 January 2007. The data are available from `http://bossa.pl/notowania/daneatech/metastock/` (page in Polish). The top left plot in Figure 1 shows the actual series P_k. The top right plot shows y_k.

The series y_k displays many of the typical 'stylised facts' present in financial log-return series. As shown in the middle left plot of Figure 1, the series y_k is uncorrelated, here with the exception of lag 1 (typically, log-return series are uncorrelated with the exception of the first few lags). However, as shown in the middle right plot, the squared series y_k^2 is strongly auto-correlated even for very large lags. In fact, in this example it is not obvious that the auto-correlation of y_k^2 decays to zero at all.

It is also typical of financial log-return series to be heavy-tailed, as illustrated in the bottom left plot of Figure 1, which shows the Q-Q plot of the marginal distribution of y_k against the standard normal.

Finally, the bottom right plot illustrates the so-called *leverage effect*: the series y_k responds differently to its own positive and negative movements, or in other words the conditional distribution of $|y_k| \Big| \{y_{k-1} > 0\}$ is different from that of $|y_k| \Big| \{y_{k-1} < 0\}$. The bottom right plot of Figure 1 shows the sample quantiles of the two conditional distributions plotted against each other. The explanation is that the market responds differently to "good" and "bad" news, which is only too natural.

Statisticians "like" stationarity as it enables them to estimate parameters globally, using the entire available data set. However, to propose a stationary model for y_k which captures the above "stylised facts" is not easy, as the series does not "look stationary": the local variance (volatility) is clearly clustered in bunches of low/high values. If we were to fit a linear time series model (such as ARMA) to y_k, the estimated parameters would come out

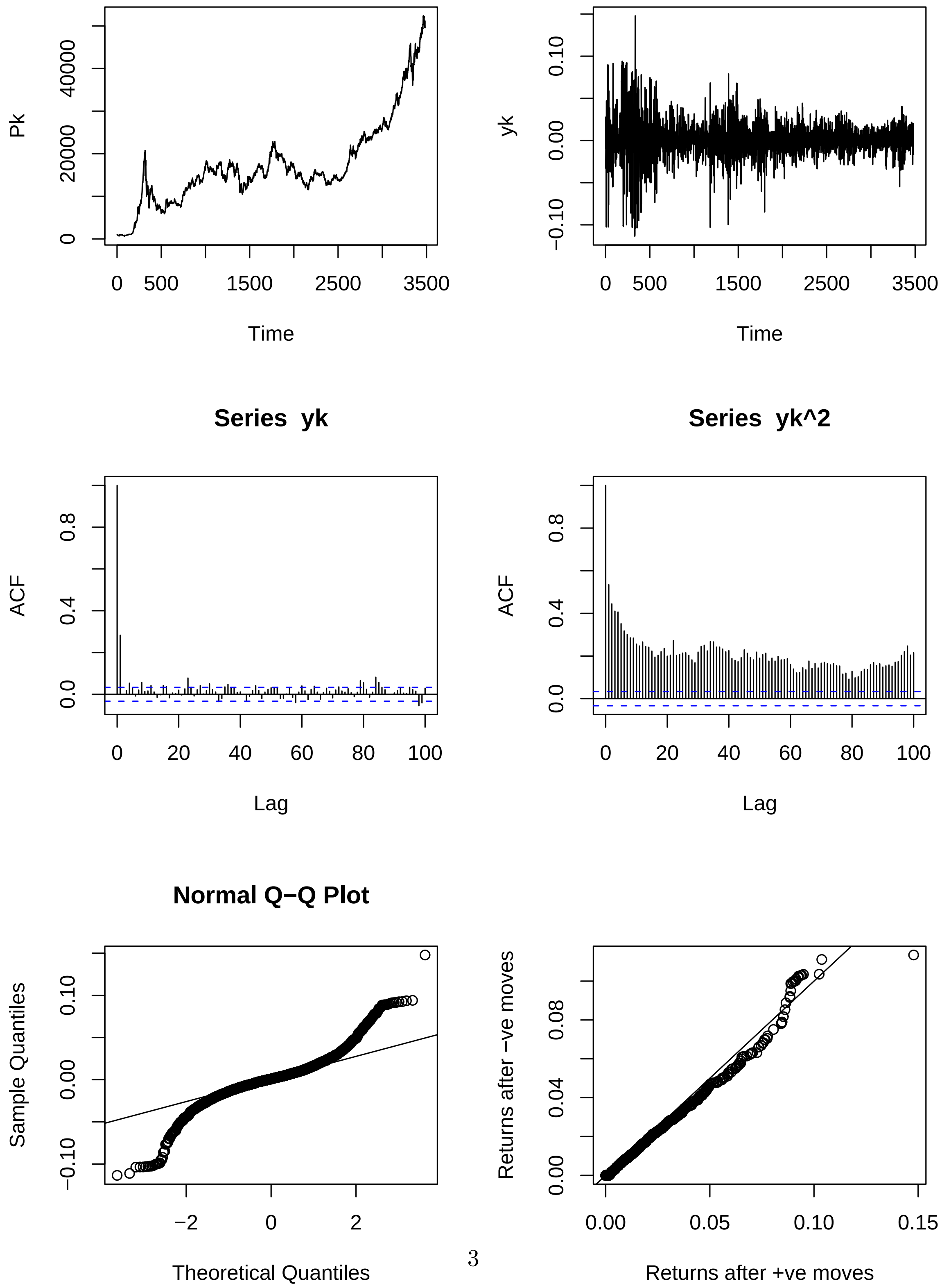
Pk
Time
yk
Time
Series yk
ACF
Lag
Series yk^2
ACF
Lag
Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
Returns after −ve moves
Returns after +ve moves

Figure 1:

as zero because of the lack of serial correlation property, which clearly would not be what we wanted.

2 (G)ARCH models

2.1 Introduction

Noting the above difficulties, Engle (1982) was the first to propose a stationary *non-linear* model for y_k, which he termed ARCH (Auto-Regressive Conditionally Heteroscedastic; it means that the conditional variance of y_k evolves according to an autoregressive-type process). Bollerslev (1986) and Taylor (1986) independently generalised Engle's model to make it more realistic; the generalisation was called "GARCH". GARCH is probably the most commonly used financial time series model and has inspired dozens of more sophisticated models.

Literature. Literature on GARCH is massive. My favourites are: Giraitis et al. (2005), Bera and Higgins (1993), Berkes et al. (2003), and the book by Straumann (2005). This chapter is based on the latter three.

Definition. The GARCH(p, q) model is defined by

$$\begin{aligned} y_k &= \sigma_k \varepsilon_k \\ \sigma_k^2 &= \omega + \sum_{i=1}^{p} \alpha_i y_{k-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{k-j}^2, \end{aligned} \tag{1}$$

where $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and the innovation sequence $\{\varepsilon_i\}_{i=-\infty}^{\infty}$ is independent and identically distributed with $E(\varepsilon_0) = 0$ and $E(\varepsilon_0^2) = 1$.

The main idea is that σ_k^2, the conditional variance of y_k given information available up to time $k-1$, has an autoregressive structure and is positively correlated to its own recent past and to recent values of the squared returns y^2. This captures the idea of volatility (=

conditional variance) being "persistent": large (small) values of y_k^2 are likely to be followed by large (small) values.

2.2 Basic properties

2.2.1 Uniqueness and stationarity

The answer to the question of whether and when the system of equations (1) admits a unique and stationary solution is not straightforward and involves the so-called *top Lyapunov exponent* associated with a certain sequence of random matrices. This is beyond the scope of this course. See Bougerol and Picard (1992) for a theorem which gives a necessary and sufficient condition for (1) to have a unique and stationary solution. Proposition 3.3.3 (b) in Straumann (2005) implies that it is the case if

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1. \tag{2}$$

2.2.2 Mean zero

Define the information set

$$\mathcal{F}_{k-1} = \sigma\{\varepsilon_i, \quad -\infty < i \leq k-1\}.$$

In any model in which σ_k is measurable with respect to $\mathcal{F}_{k-1}$ (which is the case in the GARCH model (1)), the mean of y_k is zero:

$$E(y_k) = E(\sigma_k \varepsilon_k) = E[E(\sigma_k \varepsilon_k | \mathcal{F}_{k-1})] = E[\sigma_k E(\varepsilon_k | \mathcal{F}_{k-1})] = E[\sigma_k E(\varepsilon_k)] = E[\sigma_k \cdot 0] = 0.$$

2.2.3 Lack of serial correlation

In the same way, we can show that y_k is not correlated to y_{k+h} for $h > 0$:

$$E(y_k y_{k+h}) = E(y_k \sigma_{k+h} \varepsilon_{k+h}) = E[E(y_k \sigma_{k+h} \varepsilon_{k+h} | \mathcal{F}_{k+h-1})] = E[y_k \sigma_{k+h} E(\varepsilon_{k+h} | \mathcal{F}_{k+h-1})] = 0.$$

2.2.4 Unconditional variance

In order to compute $E(y_k^2)$, it is useful to consider an alternative representation of y_k^2. First define the sequence

$$Z_k = y_k^2 - \sigma_k^2 = \sigma_k^2(\varepsilon_k^2 - 1).$$

Proceeding exactly like in Sections 2.2.2 and 2.2.3, we can show that Z_k is a martingale difference and therefore has mean zero (the "lack of serial correlation" property is more tricky as it would require $E(y_k^4) < \infty$, which is not always true). The main point of this definition is that for many purposes, Z_k can be treated as a "white noise" sequence: there are a lot of results about martingale difference sequences which extend results on independent sequences. See for example Davidson (1994).

We now proceed with our alternative representation. Write

$$\begin{aligned} y_k^2 &= \sigma_k^2 + Z_k \\ &= \omega + \sum_{i=1}^{p} \alpha_i y_{k-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{k-j}^2 + Z_k \\ &= \omega + \sum_{i=1}^{p} \alpha_i y_{k-i}^2 + \sum_{j=1}^{q} \beta_j y_{k-j}^2 - \sum_{j=1}^{q} \beta_j Z_{k-j} + Z_k. \end{aligned}$$

If we denote $R = \max(p, q)$, $\alpha_i = 0$ for $i > p$ and $\beta_j = 0$ for $j > q$, then the above can be written as

$$y_k^2 = \omega + \sum_{i=1}^{R} (\alpha_i + \beta_i) y_{k-i}^2 - \sum_{j=1}^{q} \beta_j Z_{k-j} + Z_k. \tag{3}$$

In other words, y_k^2 is an ARMA process with martingale difference innovations.

Using stationarity (which implies $E(y_k^2) = E(y_{k+h}^2)$), the unconditional variance is now easy to obtain:

$$\begin{aligned} E(y_k^2) &= \omega + \sum_{i=1}^{R}(\alpha_i + \beta_i)E(y_{k-i}^2) - \sum_{j=1}^{q}\beta_j E(Z_{k-j}) + E(Z_k) \\ &= \omega + E(y_k^2)\sum_{i=1}^{R}\alpha_i + \beta_i, \end{aligned}$$

which gives

$$E(y_k^2) = \frac{\omega}{1 - \sum_{i=1}^{R}\alpha_i + \beta_i}.$$

This result shows again the importance of condition (2).

2.2.5 Heavy tails of y_k

In this section, we argue that the GARCH model (1) can easily be heavy-tailed. For ease of presentation, we only show it for the GARCH(1,1) model. We first assume the following condition:

$$E(\alpha_1\varepsilon_k^2 + \beta_1)^{q/2} > 1 \tag{4}$$

for some $q > 0$. This condition is, for example, satisfied if $\varepsilon_k \sim N(0,1)$ (but not only in this case, obviously).

We write

$$\sigma_{k+1}^2 = \omega + \alpha_1 y_k^2 + \beta_1\sigma_k^2 = \omega + (\alpha_1\varepsilon_k^2 + \beta_1)\sigma_k^2,$$

which, using the independence of ε_k of $\mathcal{F}_{k-1}$, gives

$$E(\sigma_{k+1}^q) = E[\omega + (\alpha_1\varepsilon_k^2 + \beta_1)\sigma_k^2]^{q/2} \geq E[(\alpha_1\varepsilon_k^2 + \beta_1)\sigma_k^2]^{q/2} = E(\alpha_1\varepsilon_k^2 + \beta_1)^{q/2}E(\sigma_k^q).$$

If $E(\sigma_k^q)$ were finite, then by stationarity, it would be equal to $E(\sigma_{k+1}^q)$. Then, simplifying, we would obtain

$$1 \geq E(\alpha_1 \varepsilon_k^2 + \beta_1)^{q/2},$$

which contradicted assumption (4). Thus $E(\sigma_k^q)$ is infinite, which also implies that $E(y_k^2)$ is infinite. Thus y_k does not have all finite moments, so it is heavy-tailed.

2.3 Estimation

2.3.1 Some extra notation

The parameter vector will be denoted by θ, that is

$$\theta = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q).$$

We assume that there is a unique stationary solution to the set of equations (1). By expanding σ_{k-j}^2 in (1), σ_k^2 can be represented as

$$\sigma_k^2 = c_0 + \sum_{i=1}^{\infty} c_i y_{k-i}^2, \tag{5}$$

(given that $E(\log \sigma_0^2) < \infty$, which is a kind of a "stability condition", but this is not important for now). How to compute the coefficients c_k in (5)? They obviously depend on the parameter θ. More generally, if the parameter is $u = (x, s_1, \ldots, s_p, t_1, \ldots, t_q)$, then:

- If $q \geq p$, then

$$\begin{aligned} c_0(u) &= \frac{x}{1 - t_1 - \ldots - t_q} \\ c_1(u) &= s_1 \\ c_2(u) &= s_2 + t_1 c_1(u) \\ \ldots \end{aligned}$$

$$
\begin{aligned}
c_p(u) &= s_p + t_1 c_{p-1}(u) + \ldots + t_{p-1} c_1(u) \\
c_{p+1}(u) &= t_1 c_p(u) + \ldots + t_p c_1(u) \\
&\ldots \\
c_q(u) &= t_1 c_{q-1}(u) + \ldots + t_{q-1} c_1(u)
\end{aligned}
$$

- If $q < p$, then

$$
\begin{aligned}
c_0(u) &= \frac{x}{1 - t_1 - \ldots - t_q} \\
c_1(u) &= s_1 \\
c_2(u) &= s_2 + t_1 c_1(u) \\
&\ldots \\
c_{q+1}(u) &= s_{q+1} + t_1 c_q(u) + \ldots + t_q c_1(u) \\
&\ldots \\
c_p(u) &= s_p + t_1 c_{p-1}(u) + \ldots + t_q c_{p-q}(u)
\end{aligned}
$$

For $i > \max(p, q)$,

$$
c_i(u) = t_1 c_{i-1}(u) + t_2 c_{i-2}(u) + \ldots + t_q c_{i-q}(u). \tag{6}
$$

2.3.2 Estimation via Quasi Maximum Likelihood

Define the parameter space

$$
U = \{u : t_1 + \ldots + t_q \leq \rho_0, \min(x, s_1, \ldots, s_p, t_1, \ldots, t_q) \geq \underline{u}, \max(x, s_1, \ldots, s_p, t_1, \ldots, t_q) \leq \overline{u}\},
$$

where ρ_0, $\underline{u}$, $\overline{u}$ are arbitrary but such that $0 < \underline{u} < \overline{u}$, $0 < \rho_0 < 1$, $q\underline{u} < \rho_0$. Clearly, U is a compact set (it is a closed and bounded subset of $\mathbb{R}^{p+q+1}$).

Quasi-likelihood for GARCH(p,q) is defined as

$$L_n(u) = \sum_{1 \le k \le n} -\frac{1}{2}\left\{\log\, w_k(u) + \frac{y_k^2}{w_k(u)}\right\},$$

where

$$w_k(u) = c_0(u) + \sum_{1 \le i < \infty} c_i(u) y_{k-i}^2$$

(in practice we truncate this sum, but let us not worry about it for now). Note that $w_k(\theta) = \sigma_k^2$.

If ε_k is standard normal, then conditionally on $\mathcal{F}_{k-1}$, y_k is normal with mean zero and variance $w_k(\theta)$. Normal log-likelihood for $N(0, \sigma^2)$ is

$$\log \sigma^{-1} - \frac{x^2}{2\sigma^2} + C \sim -\frac{1}{2}\left(\log \sigma^2 + \frac{x^2}{\sigma^2}\right).$$

Thus, if ε_k are $N(0,1)$ and the parameter is a general u, then the conditional likelihood of y_k given $\mathcal{F}_{k-1}$ is

$$-\frac{1}{2}\left\{\log\, w_k(u) + \frac{y_k^2}{w_k(u)}\right\}.$$

Summing over k, we get exactly $L_n(u)$. The quasi-likelihood estimator $\hat{\theta}_n$ is defined as

$$\hat{\theta}_n = \mathrm{argmax}_{u \in U} L_n(u).$$

The estimator is called a "quasi"-likelihood estimator as it will also be consistent if the variables ε_k are not normally distributed.

In order to prove consistency, we first need to list a few assumptions.

Assumption 2.1 *(i) ε_0^2 is a non-degenerate random variable.*

(ii) The polynomials

$$
\begin{aligned}
A(x) &= \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_p x^p \\
B(x) &= 1 - \beta_1 x - \beta_2 x^2 - \ldots - \beta_q x^q
\end{aligned}
$$

do not have common roots.

(iii) θ lies in the interior of U. (This implies that there are no zero coefficients and that the orders p and q are known!)

(iv)

$$E\left\{\varepsilon_0^{2(1+\delta)}\right\} < \infty$$

for some $\delta > 0$.

(v)

$$\lim_{t \to 0} t^{-\mu} P(\varepsilon_0^2 \leq t) = 0$$

for some $\mu > 0$.

(vi)

$$E(\varepsilon_0^2) = 1.$$

(vii)

$$E\left\{|\varepsilon_0|^{2\gamma}\right\} < \infty$$

for some $\gamma > 0$.

(viii)

$$E|y_0^2|^\delta < \infty$$

for some $\delta > 0$.

We are now ready to formulate a consistency theorem for the quasi-likelihood estimator $\hat{\theta}_n$.

Theorem 2.1 *Under Assumptions 2.1 (i)–(vi), we have*

$$\hat{\theta}_n \to \theta \quad \text{a.s.},$$

as $n \to \infty$.

2.3.3 Consistency proof for the Quasi Maximum Likelihood estimator

Before we prove Theorem 2.1, we need a number of technical lemmas. These lemmas use, in an ingenious and "beautiful" way, a number of *simple* but extremely useful mathematical techniques/tricks, such as: mathematical induction, the Hölder inequality, an alternative representation of the expectation of a nonnegative random variable, a method for proving implications, the Bonferroni inequality, the Markov inequality, the mean-value theorem, ergodicity, the ergodic theorem and the concept of equicontinuity.

Lemma 2.1 *For any $u \in U$, we have*

$$\begin{aligned} C_1 \underline{u}^i &\leq c_i(u) \quad 0 \leq i < \infty && (7) \\ c_i(u) &\leq C_2 \rho_0^{i/q} \quad 0 \leq i < \infty, && (8) \end{aligned}$$

for some positive constants C_1, C_2.

Proof. By induction. (7) and (8) hold for $0 \leq i \leq \max(p, q)$ if C_1 (C_2) is chosen small (large) enough. Let $j > R$ and assume (7), (8) hold for all $i < j$. By (6),

$$c_j(u) \geq \underline{u} \min_{k=1,\ldots,q} c_{j-k}(u) \geq \underline{u} C_1 \underline{u}^{j-1} = C_1 \underline{u}^j.$$

Also by (6),

$$c_j(u) \le (t_1 + \ldots + t_q) \max_{k=1,\ldots,q} c_{j-k}(u) \le \rho_0 C_2 \rho_0^{(j-q)/q} = C_2 \rho_0^{j/q},$$

which completes the proof. $\square$

Lemma 2.2 *Suppose Assumptions 2.1 (iii), (v) and (vii) hold. Then for any* $0 < \nu < \gamma$ *we have*

$$E\left\{ \left(\sup_{u \in U} \frac{\sigma_k^2}{w_k(u)} \right)^{\nu} \right\} < \infty.$$

Proof. Take $M \ge 1$. We have

$$\frac{\sigma_k^2}{w_k(u)} \le \frac{\sigma_k^2}{\sum_{i=1}^{M} c_i(u) y_{k-i}^2} = \frac{\sigma_k^2}{\sum_{i=1}^{M} c_i(u) \varepsilon_{k-i}^2 \sigma_{k-i}^2}. \tag{9}$$

By definition of σ_k^2,

$$\begin{aligned} \sigma_{k-1}^2 &> \beta_i \sigma_{k-i-1}^2, \quad 1 \le i \le q \\ \sigma_{k-1}^2 &> \alpha_i y_{k-i-1}^2, \quad 1 \le i \le p \\ \sigma_{k-1}^2 &> \omega. \end{aligned}$$

Hence

$$\begin{aligned} \frac{\sigma_k^2}{\sigma_{k-1}^2} &= \frac{\omega + \alpha_1 y_{k-1}^2 + \ldots + \alpha_p y_{k-p}^2 + \beta_1 \sigma_{k-1}^2 + \ldots + \beta_q \sigma_{k-q}^2}{\sigma_{k-1}^2} \\ &\le 1 + \frac{\alpha_1 \varepsilon_{k-1}^2 \sigma_{k-1}^2}{\sigma_{k-1}^2} + \frac{\alpha_2 y_{k-2}^2}{\alpha_1 y_{k-1-1}^2} + \frac{\alpha_3}{\alpha_2} + \ldots + \frac{\alpha_p}{\alpha_{p-1}} + \beta_1 + \frac{\beta_2}{\beta_1} + \ldots + \frac{\beta_q}{\beta_{q-1}} \\ &\le K_1(1 + \varepsilon_{k-1}^2), \end{aligned}$$

for some $K_1 > 1$. This leads to

$$\frac{\sigma_k^2}{\sigma_{k-i}^2} \leq K_1^M \prod_{1 \leq j \leq M} (1 + \varepsilon_{k-j}^2),$$

for all $i = 1, \ldots, M$. Thus, using Lemma 2.1, (9) gives

$$\frac{\sigma_k^2}{w_k(u)} \leq \frac{\sigma_k^2}{\sum_{i=1}^M c_i(u)\varepsilon_{k-i}^2 \sigma_{k-i}^2} \leq K_1^M \prod_{1 \leq j \leq M} (1+\varepsilon_{k-j}^2) \frac{1}{\sum_{i=1}^M c_i(u)\varepsilon_{k-i}^2} \leq K_2^M \frac{\prod_{1 \leq j \leq M}(1 + \varepsilon_{k-j}^2)}{\sum_{i=1}^M \varepsilon_{k-i}^2},$$

for some $K_2 > 0$, which is now independent of u (a uniform bound)! Thus, using the Hölder inequality,

$$\begin{aligned} E\left\{\left(\sup_{u \in U} \frac{\sigma_k^2}{w_k(u)}\right)^\nu\right\} &\leq E\left\{\left(K_2^M \frac{\prod_{1 \leq j \leq M}(1 + \varepsilon_{k-j}^2)}{\sum_{i=1}^M \varepsilon_{k-i}^2}\right)^\nu\right\} \\ &\leq K_2^{M\nu} \left\{E\left(\prod_{1 \leq j \leq M} (1 + \varepsilon_{k-j}^2)\right)^\gamma\right\}^{\nu/\gamma} \left\{E\left(\sum_{i=1}^M \varepsilon_{k-i}^2\right)^{-\frac{\nu\gamma}{\gamma-\nu}}\right\}^{\frac{\gamma-\nu}{\gamma}} \end{aligned}$$

By the independence of the ε_i's, and by Assumption 2.1 (vii),

$$E\left(\prod_{1 \leq j \leq M} (1 + \varepsilon_{k-j}^2)\right)^\gamma = \left\{E(1 + \varepsilon_0^2)^\gamma\right\}^M < \infty.$$

It is now enough to show

$$E\left(\sum_{i=1}^M \varepsilon_{k-i}^2\right)^{-\frac{\nu\gamma}{\gamma-\nu}} < \infty.$$

Dealing with moments of sums can be difficult, but we will use a "trick" to avoid having to do so. From 1st year probability, we know that for a nonnegative random variable Y,

$$EY = \int P(Y > t)dt.$$

Thus if $P(Y > t) \leq K_3 t^{-2}$ for t large enough, then $EY < \infty$. We will show

$$P\left(\left(\sum_{i=1}^{M} \varepsilon_{k-i}^2\right)^{-\frac{\nu\gamma}{\gamma-\nu}} > t\right) \leq K_3 t^{-2}$$

for t large enough. We have

$$P\left(\left(\sum_{i=1}^{M} \varepsilon_{k-i}^2\right)^{-\frac{\nu\gamma}{\gamma-\nu}} > t\right) = P\left(\sum_{i=1}^{M} \varepsilon_{k-i}^2 \leq t^{-\frac{\gamma-\nu}{\nu\gamma}}\right) \tag{10}$$

Obviously the following implication holds:

$$\sum_{i=1}^{M} \varepsilon_{k-i}^2 \leq t^{-\frac{\gamma-\nu}{\nu\gamma}} \quad \Rightarrow \quad \forall\, i, \quad \varepsilon_{k-i}^2 \leq t^{-\frac{\gamma-\nu}{\nu\gamma}},$$

so (10) can be bounded from above by

$$\left\{P\left(\varepsilon_0^2 \leq t^{-\frac{\gamma-\nu}{\nu\gamma}}\right)\right\}^M. \tag{11}$$

Now, Assumption 2.1 (vii) implies

$$P(\varepsilon_0^2 \leq s) \leq \tilde{C} s^{\mu},$$

for all s, for some $\tilde{C}$. Thus, we bound (11) from above by

$$\tilde{C}^M t^{-\frac{\mu(\gamma-\nu)M}{\gamma\nu}} \leq K_3 t^{-2},$$

if M is chosen large enough, which completes the proof. $\square$

Lemma 2.3 *Suppose Assumptions 2.1 (iii) and (viii) hold. Then for any* $\nu > 0$,

$$E\left(\sup_{u\in U} \frac{\sum_{i=1}^{\infty} i c_i(u) y_{k-i}^2}{1 + \sum_{i=1}^{\infty} c_i(u) y_{k-i}^2}\right)^{\nu} < \infty.$$

Proof. For any $M \geq 1$, we have

$$\begin{aligned}
\frac{\sum_{i=1}^{\infty} ic_i(u)y_{k-i}^2}{1+\sum_{i=1}^{\infty} c_i(u)y_{k-i}^2} &= \frac{\sum_{i=1}^{M} ic_i(u)y_{k-i}^2}{1+\sum_{i=1}^{\infty} c_i(u)y_{k-i}^2} + \frac{\sum_{i=M+1}^{\infty} ic_i(u)y_{k-i}^2}{1+\sum_{i=1}^{\infty} c_i(u)y_{k-i}^2} \\
&\leq \frac{\sum_{i=1}^{M} ic_i(u)y_{k-i}^2}{\sum_{i=1}^{M} c_i(u)y_{k-i}^2} + \sum_{i=M+1}^{\infty} ic_i(u)y_{k-i}^2 \\
&\leq M + \sum_{i=M+1}^{\infty} ic_i(u)y_{k-i}^2. \qquad (12)
\end{aligned}$$

We now recall another basic fact of probability. For a nonnegative variable Y, if $P(Y > t) \sim e^{-\beta t^{\alpha}}$, then all moments of Y are finite, i.e. $E|Y|^{\nu} < \infty$ for all ν. Explanation:

$$E|Y|^{\nu} = \int P(Y^{\nu} > t)dt = \int P(Y > t^{1/\nu})dt \sim \int e^{-\beta t^{\alpha/\nu}} dt < \infty.$$

We will show

$$P\left(\sup_{u\in U} \sum_{i=M+1}^{\infty} ic_i(u)y_{k-i}^2 > t\right) \sim e^{-\beta t^{\alpha}}.$$

Choose constants $\rho_0^{1/q} < \rho_* < 1$, $\rho_{**} > 1$ such that $\rho_*\rho_{**} < 1$ and take $M \geq M_0(C_2, \rho_*, \rho_{**})$ "large enough". Then by (8) we have

$$P\left(\sup_{u\in U} \sum_{i=M+1}^{\infty} ic_i(u)y_{k-i}^2 > t\right) \leq P\left(C_2 \sum_{i=M+1}^{\infty} i\rho_0^{i/q} y_{k-i}^2 > t\right) \leq P\left(\sum_{i=M+1}^{\infty} \rho_*^i y_{k-i}^2 > t\right). \qquad (13)$$

Now,

$$\sum_{i=M+1}^{\infty} \rho_*^i y_{k-i}^2 > t \quad \Rightarrow \quad \exists\, i = M+1, \ldots, \infty \quad y_{k-i}^2 > t\rho_*^{-i} \left(\frac{\rho_{**}}{\rho_{**}-1}\right)^{-1} \rho_{**}^{-i}$$

To see that this implication of the form $p \Rightarrow q$ is true, it is easy to show that $\neg q \Rightarrow \neg p$.

Thus, by the Bonferroni inequality, (13) can be bounded from above by

$$\sum_{i=M+1}^{\infty} P\left(y_{k-i}^2 > t\rho_*^{-i}\left(\frac{\rho_{**}}{\rho_{**}-1}\right)^{-1} \rho_{**}^{-i}\right) = \sum_{i=M+1}^{\infty} P\left(y_{k-i}^{2\delta} > t^{\delta}\left(\frac{\rho_{**}}{\rho_{**}-1}\right)^{-\delta} (\rho_*\rho_{**})^{-i\delta}\right).$$

Now using the Markov inequality, we bound the above by

$$\sum_{i=M+1}^{\infty} E|y_0|^{2\delta} t^{-\delta} \left(\frac{\rho_{**}}{\rho_{**}-1}\right)^{\delta} (\rho_* \rho_{**})^{i\delta} = E|y_0|^{2\delta} t^{-\delta} \left(\frac{\rho_{**}}{\rho_{**}-1}\right)^{\delta} \frac{(\rho_* \rho_{**})^{M\delta}}{1-(\rho_*\rho_{**})^{\delta}} \tag{14}$$

We now take $t > 2\max(M_0, 1)$ (it is enough to show the above for "large" t) and $M = t/2$. Combining (12) and (14), we get

$$\begin{aligned} P\left(M + \sup_{u \in U} \sum_{i=M+1}^{\infty} i c_i(u) y_{k-i}^2 > t\right) &= P\left(\sup_{u \in U} \sum_{i=M+1}^{\infty} i c_i(u) y_{k-i}^2 > t/2\right) \\ &\leq E|y_0|^{2\delta} (t/2)^{-\delta} \left(\frac{\rho_{**}}{\rho_{**}-1}\right)^{\delta} \frac{(\rho_*\rho_{**})^{t\delta/2}}{1-(\rho_*\rho_{**})^{\delta}} \leq K_4 e^{-K_5 t}, \end{aligned}$$

which completes the proof. □

Lemma 2.4 *Let $|\cdot|$ denote the maximum norm of a vector. Suppose that Assumptions 2.1 (iii), (iv), (v), (viii) hold. Then*

$$\begin{aligned} E \sup_{u,v \in U} \frac{1}{|u-v|} |\log w_k(u) - \log w_k(v)| &< \infty \\ E \sup_{u,v \in U} \frac{1}{|u-v|} \left| \frac{y_k^2}{w_k(u)} - \frac{y_k^2}{w_k(v)} \right| &< \infty \end{aligned}$$

Proof. The mean value theorem says that for a function f, we have

$$\frac{|f(u) - f(v)|}{|u-v|} = |f'(\xi)|,$$

where $\max(|\xi - u|, |\xi - v|) \leq |u - v|$. Applying it to $f(u) = y_k^2 / w_k(u)$, we get

$$\left| \frac{y_k^2}{w_k(u)} - \frac{y_k^2}{w_k(v)} \right| = |u-v| \left| \frac{y_k^2}{w_k(\xi)} \right| \left| \frac{w_k'(\xi)}{w_k(\xi)} \right|.$$

Clearly

$$w_k'(u) = c_0'(u) + \sum_{i=1}^{\infty} c_i'(u) y_{k-i}^2.$$

We now use a fact which we accept without proof. The proof is easy but long and is again done by induction. If you are interested in the details, see Berkes et al. (2003), Lemma 3.2.

$$\begin{aligned} |c_0'(u)| &< \tilde{C} \\ |c_i'(u)| &< \bar{C} i c_i(u). \end{aligned}$$

Using the above, we get

$$\sup_{u \in U} \left| \frac{w_k'(u)}{w_k(u)} \right| \leq K_6 \frac{1 + \sum_{i=1}^{\infty} i c_i(u) y_{k-i}^2}{1 + \sum_{i=1}^{\infty} c_i(u) y_{k-i}^2}.$$

Given the above, Lemma 2.3 implies that

$$E \left(\sup_{u \in U} \left| \frac{w_k'(u)}{w_k(u)} \right| \right)^{\frac{2+\delta}{\delta}} < \infty.$$

On the other hand, by Lemma 2.2 and the assumptions of this lemma,

$$E \left(\sup_{u \in U} \frac{y_k^2}{w_k(u)} \right)^{1+\delta/2} = E(\varepsilon_k^2)^{1+\delta/2} E \left(\sup_{u \in U} \frac{\sigma_k^2}{w_k(u)} \right)^{1+\delta/2} < \infty.$$

The Hölder inequality completes the proof. The proof for log is very similar. $\square$

Lemma 2.5 *Suppose Assumptions 2.1 (iii), (iv) and (v) hold. Then*

$$\sup_{u \in U} \left| \frac{1}{n} L_n(u) - L(u) \right| \to 0$$

almost surely, as $n \to \infty$*, where*

$$L(u) = -1/2 E \left(\log w_0(u) + \frac{y_0^2}{w_0(u)} \right).$$

Proof. We start by stating the fact that if $E|\varepsilon_0^2|^\delta < \infty$ for some $\delta > 0$, then there exists a $\delta^* > 0$ such that $E|y_0^2|^{\delta^*} < \infty$ and $E|\sigma_0^2|^{\delta^*} < \infty$. The proof is beyond the scope of this course. See Berkes et al. (2003), Lemma 2.3, for details.

Thus, the assumptions of this lemma mean that

$$E|y_0^2|^{\delta^*} < \infty \tag{15}$$

for some δ^*. Using Lemma 2.1,

$$0 < C_1 \leq w_k(u) \leq C_2 \left(1 + \sum_{1 \leq i < \infty} \rho_0^{i/q} y_{k-i}^2 \right),$$

which implies

$$|\log w_0(u)| \leq \log C_2 + \log \left(1 + \sum_{1 \leq i < \infty} \rho_0^{i/q} y_{k-i}^2 \right) \leq A + B \left| \sum_{1 \leq i < \infty} \rho_0^{i/q} y_{k-i}^2 \right|^{\delta^*},$$

which implies $E|\log w_0(u)| < \infty$ by (15).

By Lemma 2.2,

$$E\left(\frac{y_0^2}{w_0(u)}\right) = E\varepsilon_0^2 E\left(\frac{\sigma_0^2}{w_0(u)}\right) < \infty.$$

Clearly, there exists a function g such that

$$y_k = g(\varepsilon_k, \varepsilon_{k-1}, \ldots),$$

and therefore y_k is stationary and ergodic by Theorem 3.5.8 of Stout (1974) (since $\{\varepsilon_k\}_k$ is stationary and ergodic as it is independent).

As $E|L(u)| < \infty$, we can use the ergodic theorem, which says that for any $u \in U$,

$$\frac{1}{n} L_n(u) \to L(u)$$

almost surely.

Also,

$$\sup_{u,v\in U} |L_n(u) - L_n(v)| \frac{1}{|u-v|} \leq 1/2 \sum_{1\leq k\leq n} \eta_k,$$

where

$$\eta_k = \sup_{u,v\in U} \frac{1}{|u-v|} \left\{ |\log\, w_k(u) - \log\, w_k(v)| + \left| \frac{y_k^2}{w_k(u)} - \frac{y_k^2}{w_k(v)} \right| \right\}.$$

Again by Theorem 3.5.8 of Stout (1974), η_k is stationary and ergodic. By Lemma 2.4, $E\eta_0 < \infty$. Using the ergodic theorem,

$$\frac{1}{n} \sum_{i=1}^{n} \eta_i = O(1)$$

almost surely, showing that

$$\sup_{u,v\in U} \left| \frac{1}{n} L_n(u) - \frac{1}{n} L_n(v) \right| \frac{1}{|u-v|} = O(1).$$

Thus the sequence of functions $L_n(u)/n$ is <u>equicontinuous</u>. Also as shown earlier it converges almost surely to $L(u)$ for all $u \in U$. This, along with the fact that U is a compact set, implies that the convergence is <u>uniform</u>, which completes the proof. (Recall a well-known fact of mathematical analysis: let f_n be an equicontinuous sequence of functions from a compact set to R. If $f_n(x) \to f(x)$ for all x, then $f_n \to f$ uniformly in x.) $\square$

Lemma 2.6 *Suppose the conditions of Theorem 2.1 are satisfied. Then $L(u)$ has a unique maximum at θ.*

Proof. $w_0(\theta) = \sigma_k^2$. As $E\varepsilon_0^2 = 1$,

$$E\left(\frac{y_0^2}{w_0(u)}\right) = E\left(\frac{\sigma_0^2}{w_0(u)}\right).$$

We have

$$\begin{aligned}
L(\theta) - L(u) &= -\frac{1}{2}E\left(\log\,\sigma_0^2 + \frac{y_0^2}{\sigma_0^2}\right) + \frac{1}{2}E\left(\log\,w_0(u) + \frac{y_0^2}{w_0(u)}\right) \\
&= -\frac{1}{2}\left(E(\log\,\sigma_0^2) + 1 - E(\log\,w_0(u)) - E\left(\frac{\sigma_0^2}{w_0(u)}\right)\right) \\
&= -\frac{1}{2}\left(E\left(\log\frac{\sigma_0^2}{w_0(u)} - \frac{\sigma_0^2}{w_0(u)}\right) + 1\right) \\
&= -\frac{1}{2} + \frac{1}{2}E\left(\frac{\sigma_0^2}{w_0(u)} - \log\frac{\sigma_0^2}{w_0(u)}\right).
\end{aligned}$$

The function $x - \log x$ is positive for all $x > 0$ and attains its minimum value (of 1) for $x = 1$. Thus $L(u)$ has a global maximum at θ.

Is the maximum unique?

Assume $L(u^*) = L(\theta)$ for some $u^* \in U$.

$$0 = L(\theta) - L(u^*) = -\frac{1}{2} + \frac{1}{2}E\left(\frac{\sigma_0^2}{w_0(u^*)} - \log\frac{\sigma_0^2}{w_0(u^*)}\right).$$

When is it possible that $E(X - \log\,X) = 1$ if $X > 0$? $X - \log\,X \geq 1$, so it is only possible if $X = 1$ almost surely. Thus $\sigma_0^2 = w_0(u^*)$ almost surely, so we must also have $c_i(\theta) = c_i(u^*)$ for all i (we accept this "intuitively obvious" statement here without proof; see Berkes et al. (2003) for details). So we also have $\sigma_k^2 = w_k(u^*)$ for all k. Let

$$u^* = (x^*, s_1^*, \ldots, s_p^*, t_1^*, \ldots, t_q^*).$$

On the one hand, by definition,

$$\sigma_k^2 = w_k(\theta) = \omega + \alpha_1 y_{k-1}^2 + \ldots + \alpha_p y_{k-p}^2 + \beta_1 \sigma_{k-1}^2 + \ldots + \beta_q \sigma_{k-q}^2.$$

On the other hand, by the above discussion,

$$\sigma_k^2 = w_k(u^*) = x^* + s_1^* y_{k-1}^2 + \ldots + s_p^* y_{k-p}^2 + t_1^* \sigma_{k-1}^2 + \ldots + t_q^* \sigma_{k-q}^2.$$

Equating coefficients (using the "uniqueness of GARCH representation", also without proof: see Berkes et al. (2003) for details), we have $u^* = \theta$, which completes the proof. $\square$

We are finally ready to prove Theorem 2.1.

Proof of Theorem 2.1. U is a compact set. L_n/n converges uniformly to L on U with probability one (Lemma 2.5) and L has a unique maximum at $u = \theta$ (Lemma 2.6). Thus by standard arguments (best seen graphically!) the locations of the maxima of L_n/n converge to that of L. This completes the proof of the Theorem. $\square$

Exercise: try to think why we need *uniform* convergence for this reasoning to be valid. Would it not be enough if $L_n(u)/n$ converged pointwise to $L(u)$ for all u?

2.4 Forecasting

By standard Hilbert space theory, the best point forecasts of y_k under the L_2 norm are given by $E(y_{k+h}|\mathcal{F}_k)$ and are equal to zero if $h > 0$ by the martingale difference property of y_k.

The equation (3) is a convenient starting point for the analysis of the optimal forecasts for y_k^2. Again under the L_2 norm, they are given by $E(y_{k+h}^2|\mathcal{F}_k)$. Formally, this only makes sense if $E(y_k^4) < \infty$, which is not always the case. However, many authors take the above as their forecasting statistic of choice. It might be more correct (and interesting) to consider $\text{Median}(y_{k+h}^2|\mathcal{F}_k)$, which is the optimal forecast under the L_1 norm. This always makes sense as $E(y_k^2) < \infty$ as we saw before. However, it is mathematically far more tractable to look at $E(y_{k+h}^2|\mathcal{F}_k)$, which is what we are going to do below.

Take $h > 0$. Recall that $E(Z_{k+h}|\mathcal{F}_k) = 0$. From (3), we get

$$\begin{aligned}
y_{k+h}^2 &= \omega + \sum_{i=1}^{R}(\alpha_i + \beta_i)y_{k+h-i}^2 - \sum_{j=1}^{q}\beta_j Z_{k+h-j} + Z_{k+h} \\
E(y_{k+h}^2|\mathcal{F}_k) &= \omega + \sum_{i=1}^{R}(\alpha_i + \beta_i)E(y_{k+h-i}^2|\mathcal{F}_k) - \sum_{j=1}^{q}\beta_j E(Z_{k+h-j}|\mathcal{F}_k) + E(Z_{k+h}|\mathcal{F}_k) \\
E(y_{k+h}^2|\mathcal{F}_k) &= \omega + \sum_{i=1}^{R}(\alpha_i + \beta_i)E(y_{k+h-i}^2|\mathcal{F}_k) - \sum_{j=1}^{q}\beta_j E(Z_{k+h-j}|\mathcal{F}_k). \qquad (16)
\end{aligned}$$

The recursive formula (16) is used to compute the forecasts, with the following boundary conditions:

- $E(y_{k+h-i}^2|\mathcal{F}_k)$ is given recursively by (16) if $h > i$,
- $E(y_{k+h-i}^2|\mathcal{F}_k) = y_{k+h-i}^2$ if $h \le i$,
- $E(Z_{k+h-j}|\mathcal{F}_k) = 0$ if $h > j$,
- $E(Z_{k+h-j}|\mathcal{F}_k) = Z_{k+h-j}$ if $h \le j$.

2.4.1 The asymptotic forecast

For $h > p$, (16) becomes

$$E(y_{k+h}^2|\mathcal{F}_k) = \omega + \sum_{i=1}^{R}(\alpha_i + \beta_i)E(y_{k+h-i}^2|\mathcal{F}_k), \qquad (17)$$

which is a difference equation for the sequence $\{E(y_{k+h}^2|\mathcal{F}_k)\}_{h=p+1}^{\infty}$. Standard theory of difference equations says that if the roots of the polynomial

$$p(z) = 1 - (\alpha_1 + \beta_1)z - \ldots - (\alpha_R + \beta_R)z^R$$

lie outside the unit circle, then the solution of (17) converges to

$$\frac{\omega}{1-\sum_{i=1}^{R}(\alpha_i+\beta_i)},$$

which is the unconditional expectation of y_k^2! In other words, as the forecasting horizon gets longer and longer, the conditioning set $\mathcal{F}_k$ has less and less impact on the forecast and asymptotically, it "does not matter" at all.

2.4.2 Example: GARCH(1,1)

In this section, we obtain explicit formulae for forecasts in the GARCH(1,1) model. Using formula (16) and the definition of Z_k, we get

$$E(y_{k+1}^2|\mathcal{F}_k)=\omega+(\alpha_1+\beta_1)y_k^2-\beta_1 Z_k=\omega+\alpha_1 y_k^2+\beta_1\sigma_k^2.$$

Substituting recursively into (17), we obtain

$$\begin{aligned}
E(y_{k+2}^2|\mathcal{F}_k) &= \omega[1+(\alpha_1+\beta_1)^1]+\alpha_1(\alpha_1+\beta_1)y_k^2+\beta_1(\alpha_1+\beta_1)\sigma_k^2\\
E(y_{k+3}^2|\mathcal{F}_k) &= \omega[1+(\alpha_1+\beta_1)^1+(\alpha_1+\beta_1)^2]+\alpha_1(\alpha_1+\beta_1)^2y_k^2+\beta_1(\alpha_1+\beta_1)^2\sigma_k^2\\
&\dots\\
E(y_{k+h}^2|\mathcal{F}_k) &= \omega\sum_{i=0}^{h-1}(\alpha_1+\beta_1)^i+\alpha_1(\alpha_1+\beta_1)^{h-1}y_k^2+\beta_1(\alpha_1+\beta_1)^{h-1}\sigma_k^2,
\end{aligned}$$

which clearly converges to $\omega/(1-\alpha_1-\beta_1)$ as $h\to\infty$, as expected.

2.5 Extensions of GARCH

There are many extensions of the GARCH model. Two of them, EGARCH and IGARCH are probably the most popular and are covered in Straumann (2005). The Exponential

GARCH (EGARCH) model reads

$$\log\,\sigma_k^2 = \alpha + \beta \log\,\sigma_{k-1}^2 + \gamma \varepsilon_{k-1} + \delta|\varepsilon_{k-1}|.$$

The Integrated GARCH (IGARCH) process is a GARCH process for which $\sum_{i=1}^{R} \alpha_i + \beta_i = 1$.

2.6 Software for fitting GARCH models

Both S-Plus and R have their own packages containing routines for fitting and forecasting GARCH models. The S-Plus module is called FinMetrics, is described on

`http://www.insightful.com/products/finmetrics/`

and is a commerical product. Sadly, it is *much* better than its (free) R counterpart, the `tseries` package, available from

`http://cran.r-project.org/src/contrib/Descriptions/tseries.html`

The R package is only able to fit GARCH models, while the S-Plus module can fit GARCH, EGARCH and a number of other models.

2.7 Relevance of GARCH models

Are GARCH models really used in practice? The answer is YES. Only recently, a big UK bank was looking for a time series analyst to work on portfolio construction (risk management). One of the job requirements was familiarity with GARCH models!

References

A. K. Bera and M. L. Higgins. ARCH models: properties, estimation and testing. *J. Economic Surveys*, 7:305–366, 1993.

I. Berkes, L. Horváth, and P. Kokoszka. GARCH processes: structure and estimation. *Bernoulli*, 9:201–227, 2003.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31:307–327, 1986.

P. Bougerol and N. Picard. Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics*, 52:115–127, 1992.

J. Davidson. *Stochastic Limit Theory*. Oxford University Press, 1994.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987–1007, 1982.

L. Giraitis, R. Leipus, and D. Surgailis. Recent advances in ARCH modelling. In *Long Memory in Economics*, pages 3–39. Springer, Berlin, 2005.

W. F. Stout. *Almost Sure Convergence*. Academic Press, New York-London, 1974. Probability and Mathematical Statistics, Vol. 24.

D. Straumann. *Estimation in Conditionally Heteroscedastic Time Series Models*, volume 181 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, 2005.

S. J. Taylor. *Modelling Financial Time Series*. Wiley, Chichester, 1986.