Module 3: Conditionally Heteroskedastic Models

References:

- *Chapter 3 in Tsay*
- *Chapter 8 in Brooks*
- *Chapter 4 in McNeil, Frey and Embrechts*
- *Chapter 12 (sections 12.1 and 12.2) in Campbell, Lo and MacKinlay*

Section 2: GARCH models
Definition

- The ARCH model is a simple dynamic volatility model.

- However, it often requires many parameters (lags) to adequately describe the volatility process.

- Plus, . . . ??

- Rob Engle's student Tim Bollerslev introduced the Generalised Auto-Regressive Conditionally Heteroscedastic process in his seminal 1986 paper.

- Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, **31**, 307-327.

- Bollerslev (1986) proposed a useful extension called the generalised ARCH (or GARCH) model.

- Let $a_t = r_t - \mu_t$ be the mean-corrected return process.

- Then, a_t follows a GARCH(p, q) model if

$$\begin{aligned} a_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \end{aligned}$$

- Here it is common to assume that:

  (i) $\{\epsilon_t\}$ is a sequence of iid random variables with mean 0 and variance 1

  (ii) $\alpha_0 > 0$

  (iii) $\alpha_i \geq 0$, $\beta_j \geq 0$

  (iv) $\sum_{i-1}^{\max(p,q)}(\alpha_i + \beta_i) < 1$,

  where $\alpha_i \equiv 0$ for $i > p$ and $\beta_i \equiv 0$ for $i > q$.

- We shall see why, in detail, shortly.

- Constraint (iv) is required to ensure the variance process of $\{a_t\}$ is stationary and finite (see below).

- In general the constraints (ii)-(iii) are to **ensure** that each σ_t^2 is positive. However they are NOT all strictly **necessary** in general, see Nelson and Cao (1992, JBES).

- The distributional shape of the series $\{\epsilon_t\}$ can be varied as required.

- The GARCH(p, q) model collapses to an ARCH(p) model if $q = 0$.

- Let

$$\eta_t = a_t^2 - \sigma_t^2 = a_t^2 - E[a_t^2|\mathcal{F}_{t-1}]$$

be the *innovation* of the squared process.

- Note that

$$\begin{aligned} E[\eta_t] &= E[a_t^2] - E[E[a_t^2|\mathcal{F}_{t-1}]] \\ &= 0 \end{aligned}$$

- and

$$\text{Cov}(\eta_t, \eta_{t-1}) = 0$$

- so that $\{\eta_t\}$ is an uncorrelated series (also called a Martingale difference series if $\text{Cov}(\eta_t, \eta_{t-j}) = 0$ for all $j > 0$).

- But it is not independent. Why?

- By plugging in $\sigma^2_{t-j} = a^2_{t-j} - \eta_{t-j}$ and $\sigma^2_t = a^2_t - \eta_t$ into the model definition, the GARCH model can be rewritten as

$$a_t^2 = \alpha_0 + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) a^2_{t-i} + \eta_t - \sum_{j=1}^{q} \beta_j \eta_{t-j}$$

- This is *like* an ARMA($\max(p, q), q$) model for the series a_t^2, with a positivity constraint.

- Thus, a GARCH model is an application of the ARMA model to the squared series a_t^2.

- From ARMA theory:

$$\mathrm{Var}(a_t) = E[a_t^2] = \frac{\alpha_0}{1 - \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i)}$$

- Hence the requirement that $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$

GARCH(1,1) MODEL

- The GARCH(1,1) model is by far the most popular and well-used GARCH specification:
$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$
where it is usual to set
$$\alpha_1, \beta_1 \geq 0\,,\ \alpha_0 > 0\,,\ \alpha_1 + \beta_1 < 1$$
and where $\{\epsilon_t\}$ is an iid sequence, mean 0, variance 1.

- This model captures volatility clustering since large a_{t-1}^2 or σ_{t-1}^2 result in large σ_t^2.

- Can this model produce smoother volatility estimates than an ARCH model? How?

- The error distribution is assumed to have moments
  - $E(\epsilon_t) = 0$
  - $\text{Var}(\epsilon_t) = 1$
  - $E(\epsilon_t^4) = K_\epsilon + 3$

- Here, K_ϵ is the excess kurtosis of ϵ_t, i.e. $K_\epsilon = \kappa_\epsilon - 3$

- Then, it is possible to calculate the unconditional kurtosis of the mean-corrected errors as below.

- First, note that:

  (i)

$$\text{Var}(a_t) = E[a_t^2] = \frac{\alpha_0}{(1 - (\alpha_1 + \beta_1))}$$

where

$$E[a_t^2] = E[\sigma_t^2]E[\epsilon_t^2] = E[\sigma_t^2]$$

(ii)

$$E[a_t^4] = E[\epsilon_t^4]E[\sigma_t^4]$$

- Result (i) follows because:

$$\begin{aligned} a_t^2 &= \sigma_t^2 + \eta_t \\ &= \alpha_0 + (\alpha_1 + \beta_1)a_{t-1}^2 + \eta_t - \beta_1\eta_{t-1} \end{aligned}$$

as shown now ...

- Taking the square of the volatility equation, then taking expectations and substituting in the results at (i) and (ii) above, using simple, but fairly lengthy, algebra, leads to:

- The excess kurtosis of $\{a_t\}$ is:

$$
\begin{aligned}
K_a &= \frac{E[a_t^4]}{(E[a_t^2])^2} - 3 \\
&= (K_\epsilon + 3)E[\sigma_t^4]\frac{(1-(\alpha_1+\beta_1))^2}{\alpha_0^2} - 3 \\
&= \frac{(K_\epsilon + 3)(1-(\alpha_1+\beta_1)^2)}{1-2\alpha_1^2-(\alpha_1+\beta_1)^2 - K_\epsilon\alpha_1^2} - 3
\end{aligned}
$$

where K_ϵ is the excess kurtosis for the specific distribution assumed for ϵ_t.

- While the algebra is tedious, this is a very informative expression.

- Consider the case where $\epsilon_t \sim N(0,1)$, so that $K_\epsilon = 0$, then we could use the formula above to show that:

$$
\begin{aligned}
K_a^{\phi} &= \frac{3 - 3(\alpha_1 + \beta_1)^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2} - 3 \\
&= \frac{6\alpha_1^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2}
\end{aligned}
$$

where K_a^{ϕ} is notation for the *excess* kurtosis for a Gaussian GARCH(1,1) model.

- This has many implications for a conditionally Gaussian GARCH(1,1) model, two are:

  (a) the 4th moment (and hence kurtosis) of $\{a_t\}$ exists iff

  $$2\alpha_1^2 + (\alpha_1 + \beta_1)^2 < 1$$

  (b) if $\alpha_1 = 0$ then $K_a^{\phi} = 0$.

- The tails of a GARCH(1,1) model are heavier than a normal iid process when ...

- Implication (b) indicates that the tails of the distribution are not heavy compared to a Gaussian, in fact the distribution of $\{a_t\}$ is Gaussian, if $\alpha_1 = 0$.

- If $\beta_1 = 0$, then the GARCH(1,1) reduces to a conditionally Gaussian ARCH(1) model with excess kurtosis

$$K_a^{\phi} = \frac{6\alpha_1^2}{1 - 3\alpha_1^2}$$

and kurtosis

$$K_a^{\phi} + 3 = \frac{3(1 - 3\alpha_1^2) + 6\alpha_1^2}{(1 - 3\alpha_1^2)} = 3\frac{(1 - \alpha_1^2)}{(1 - 3\alpha_1^2)}$$

exactly as shown for the Gaussian ARCH(1) model in section 2 of this module.

- In the case when $\{\epsilon_t\}$ is not Gaussian, then

$$K_a = \frac{K_\epsilon(1 - 2\alpha^2 - (\alpha_1 + \beta_1)^2) + 6\alpha_1^2 + 5K_\epsilon\alpha_1^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 - K_\epsilon\alpha_1^2}$$

GARCH Estimation

- ML estimation is by far the most popular method for GARCH models.

- Estimation of the GARCH(p, q) model using MLE follows the same principle as for the ARCH(p) model.

- Again, consider the mean to be constant (so that $\mu_t = \mu$ for all t) and that

$$\epsilon_t \sim N(0, 1)$$

- Take the special case of a GARCH(1,1) model. Then, the parameter space is

$$\theta = \{\mu, \alpha_0, \alpha_1, \beta_1\}$$

  with

- Then the conditional *quasi*-likelihood function is

$$p(r_2, \ldots, r_T | r_1, \theta) = \prod_{t=2}^{T} p(r_t | \mathcal{F}_{t-1}, \theta)$$

- The conditional quasi-log-likelihood is therefore

$$\begin{aligned} l_c(\theta) &= -\frac{T-1}{2}\ln(2\pi) \\ &- \frac{1}{2}\sum_{t=2}^{T}\left\{\ln(\sigma_t^2) + (r_t-\mu)^2/\sigma_t^2\right\} \\ &\propto -\frac{1}{2}\sum_{t=2}^{T}\left\{\ln(\sigma_t^2) + (r_t-\mu)^2/\sigma_t^2\right\} \end{aligned}$$

- Here,

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

can be computed recursively.

- A numerical/computational MLE estimator can be computed by maximimsing l_c above with respect to θ.

- Constrained numerical optimisation methods are used (see IMSL or Matlab optimisation toolbox).

- Note that:

$$\sigma_2^2 = \alpha_0 + \alpha_1 a_1^2 + \beta_1 \sigma_1^2$$

- Value needs to be chosen for: σ_1^2.

- This is often set to equal either the sample variance (s^2) OR the unconditional variance

$$\frac{\alpha_0}{1 - (\alpha_1 + \beta_1)}$$

- Any other logical choices?

- If we assume a Student-t error distribution, we can add the degrees of freedom parameter ν to the sample space and estimate it simultaneously with the other parameters.

- Care needs to be taken since $\nu = \infty$ is a sensible value!

- The exact or full likelihood expression for GARCH is an **open** research question.

- i.e. the unconditional distribution of r_1 is not known under a GARCH(1,1) model.

Examples

- Figure 1 displays the daily log returns for Commonwealth Bank (CBA) and News Corp (NWS) stock.

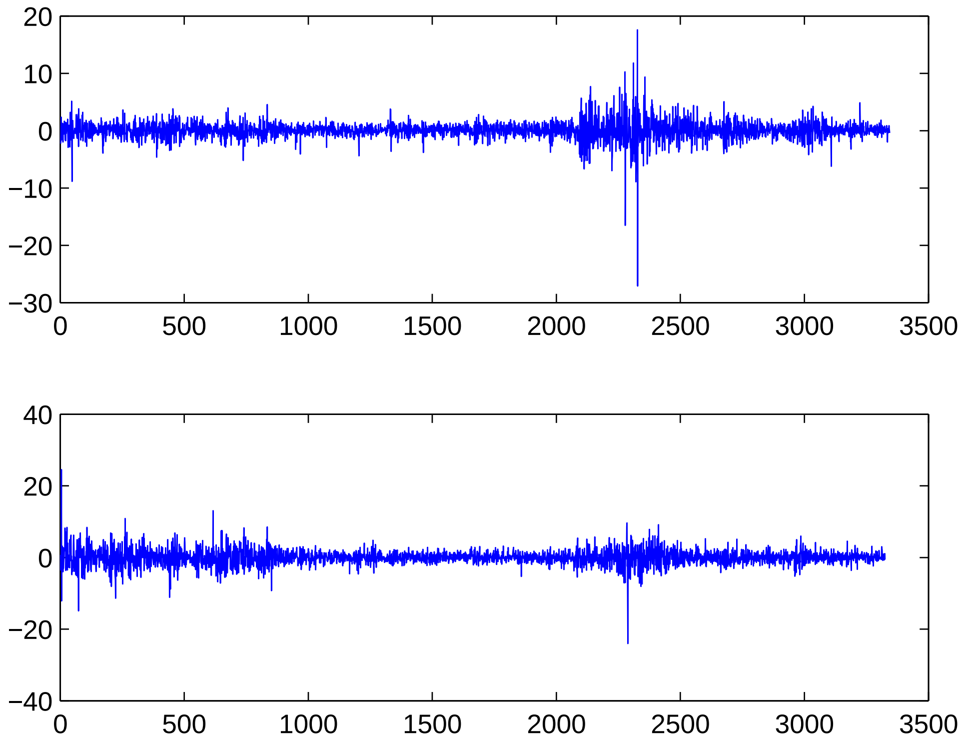

Figure 1: Log returns for CBA (top) and NWS from 2000 to 2013.

- The data are percentage log returns calculated from daily closing prices from January, 2000 to February, 2011.

- We fit a GARCH(1,1) model to each of the two series.

- The results for CBA are:

$$\begin{aligned} r_t &= \underset{(0.018)}{0.066} + a_t \\ \sigma_t^2 &= \underset{(0.003)}{0.026} + \underset{(0.005)}{0.090} a_{t-1}^2 + \underset{(0.005)}{0.896} \sigma_{t-1}^2 \end{aligned}$$

with average volatility estimated as:

$$\frac{\hat{\alpha}_0}{1 - \hat{\alpha}_1 - \hat{\beta}_1} = 2.01 \ ,$$

unconditional kurtosis estimated as:

$$\frac{6\hat{\alpha}_1^2}{1 - 2\hat{\alpha}_1^2 - (\hat{\alpha}_1 + \hat{\beta}_1)^2} + 3 = 7.98$$

and estimated volatility persistence of $\alpha_1 + \beta_1 = 0.987$.

- The sample variance and kurtosis for CBA returns are $s_C^2 = 2.50$ and $\hat{\kappa} = 40.64$.

- A summary of the GARCH(1,1) results for CBA.

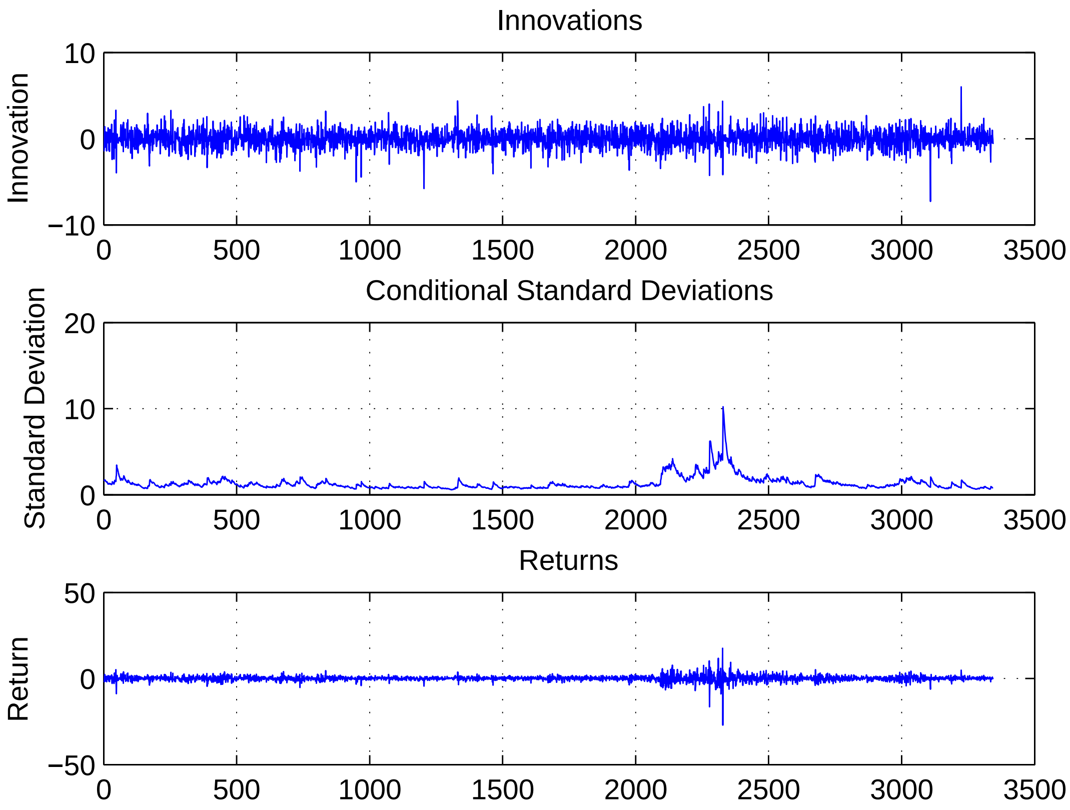

Figure 2: A summary of the GARCH(1,1) results for CBA.

- For NWS:

$$\begin{aligned} r_t &= 0.034 + a_t \\ &\quad (0.025) \\ \sigma_t^2 &= 0.024 + 0.070 a_{t-1}^2 + 0.925 \sigma_{t-1}^2 \\ &\quad (0.006)\ (0.006)\ (0.007) \end{aligned}$$

with average volatility estimated as:

$$\frac{\hat{\alpha}_0}{1 - \hat{\alpha}_1 - \hat{\beta}_1} = 5.21 \ ,$$

unconditional kurtosis estimated as:

$$\frac{6\hat{\alpha}_1^2}{1 - 2\hat{\alpha}_1^2 - (\hat{\alpha}_1 + \hat{\beta}_1)^2} + 3 = -33.62 \ ??$$

and estimated volatility persistence of 0.995.

- What has happened here?

- A summary of the GARCH(1,1) results for NWS.

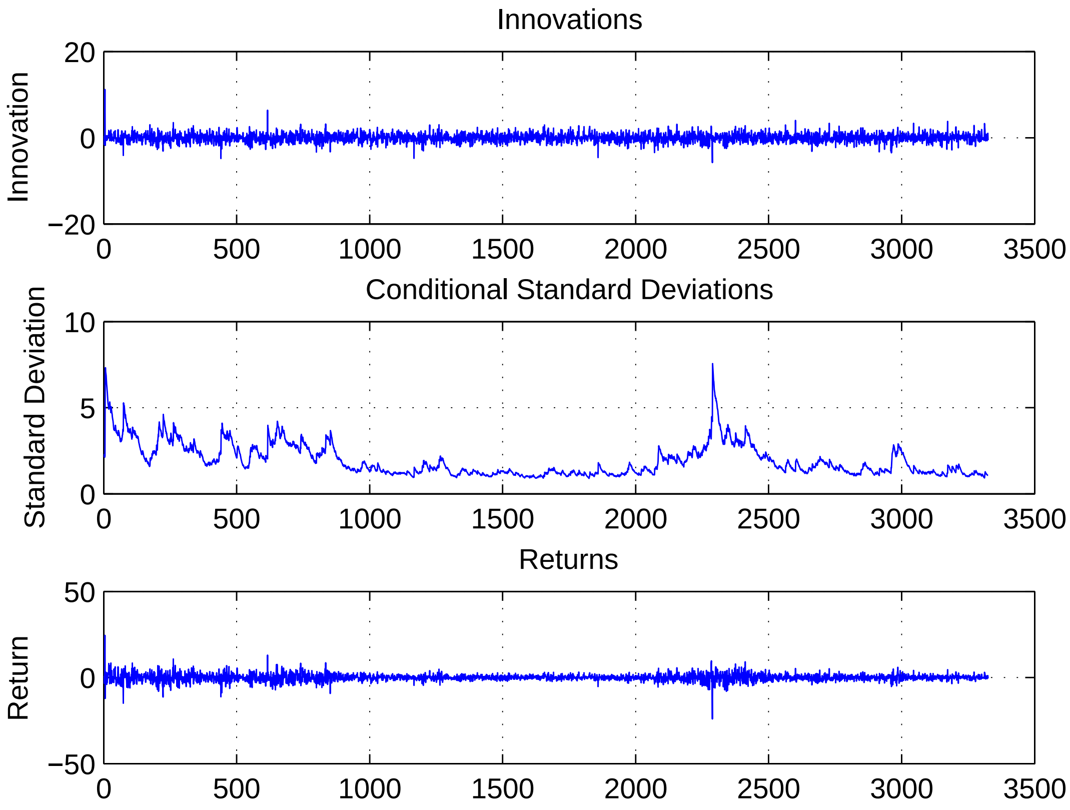

Figure 3: A summary of the GARCH(1,1) results for NWS.

- The term $2\hat{\alpha}_1^2 + (\hat{\alpha}_1 + \hat{\beta}_1)^2 = 1.001$ for NWS

- For CBA it is 0.990.

- CBA is very close to the region where 4th moments are infinite! NWS is estimated to have infinite 4th moments!

- NWS (now) has a higher estimated persistence in volatility, as measured by $\hat{\alpha}_1+\hat{\beta}_1$, and a higher unconditional variance estimate than CBA.

- Note that both conditional volatility series estimates form a smoother series compared to those for the ARCH models used previously.

- This is mainly due to the large estimated persistence coefficients for the GARCH(1, 1) model here.

Examples (ctd)

- It is important to examine the standardised residuals $\hat{\epsilon}_t = \frac{\hat{a}_t}{\hat{\sigma}_t}$ after model fitting.

- Examining figure 4 the standardised residuals for CBA do not appear to cluster,

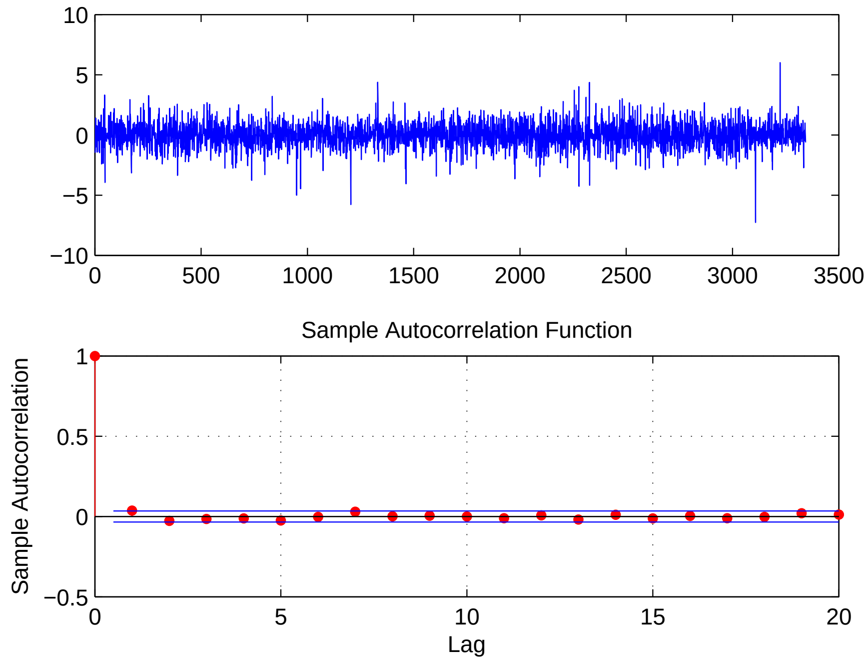

Figure 4: The standardised residuals and their ACF from the GARCH(1,1) model for CBA.

- but still show some possible outliers; See figure 5.

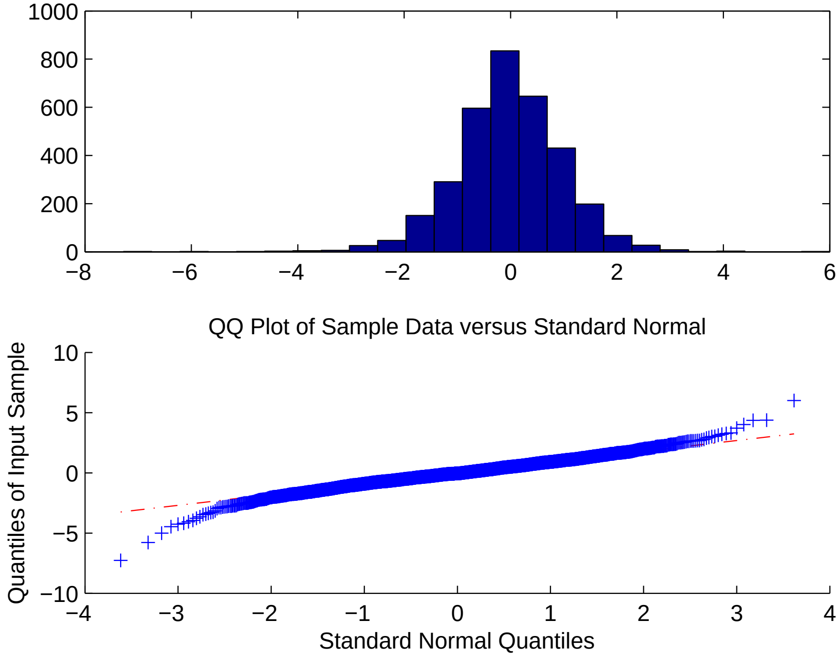

Figure 5: The standardised residual histogram and qq-plot from the GARCH(1,1) model for CBA.

- The ACF perhaps shows a weakly significant 1st order auto-correlation.

- Figure 6 displays the ACF of the squares of these standardised residuals, again showing a significant 1st order auto-correlation.

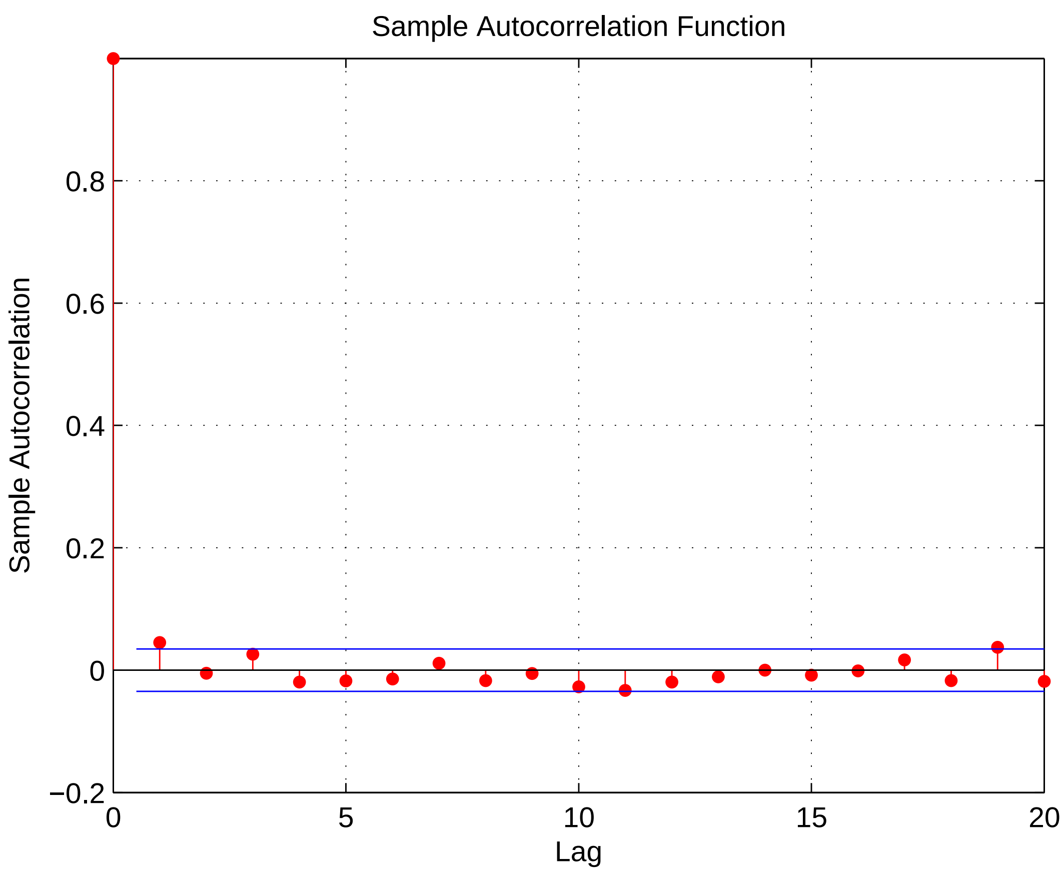

Figure 6: Autocorrelation function of the squared standardised residuals from the GARCH(1,1) model for CBA.

- Ljung-Box tests, with $m = 7, 12$ and thus $5, 10$ df respectively are conducted as follows:

- We obtain p-values of $0.021, 0.18$ for the standardised residuals $\hat{\epsilon}_t$,

- and p-values of $0.028, 0.020$ for the squared standardised residuals $\hat{\epsilon}_t^2$.

- These tests (mostly) suggest weakly significant remaining ARCH effects in the residuals at the 5% level but not at the 1% level; i.e. the volatility equation seems close to adequate, but could still be improved.

- Further, the residuals display mildly significant auto-correlation in the 1st seven lags (only at 5% level, not 1% level), indicating the (constant) mean equation is also close to reasonably well-specified.

- How can we improve these mildly rejected mean and variance equations?

- The sample skewness and kurtosis for the standardised residuals are -0.21 and 5.48 respectively.

- The residuals display some outliers and/or fat-tails and the JB test has p-value < 0.001

- How might we model these outliers and lack of Gaussianity?

GARCH properties

- GARCH models allow a local estimate of volatility, i.e. an estimate of $\sigma_t^2 = Var(r_t|\mathcal{F}_{t-1})$ that uses $a_{t-1}^2, \sigma_{t-1}^2$ as inputs to σ_t^2.

- But σ_{t-1}^2 is then a function of $a_{t-2}^2, \sigma_{t-2}^2$, ...

- i.e.

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
&= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1(\alpha_0 + \alpha_1 a_{t-2}^2 + \beta_1 \sigma_{t-2}^2) \\
&= \alpha_0(1 + \beta_1) + \alpha_1 a_{t-1}^2 + \alpha_1 \beta_1 a_{t-2}^2 + \beta_1^2 \sigma_{t-2}^2 \\
&\vdots
\end{aligned}$$

- If we continue this process we find that:

$$\begin{aligned}\sigma_t^2 &= \alpha_0(1+\beta_1+\beta_1^2+\ldots+\beta_1^{t-1}) \\ &+ \alpha_1 a_{t-1}^2+\alpha_1\beta_1 a_{t-2}^2+\alpha_1\beta_1^2 a_{t-3}^2+\ldots+\alpha_1\beta_1^{t-1}a_1^2 \\ &+ \beta_1^t\sigma_1^2\end{aligned}$$

- Since both $\alpha_1, \beta_1 > 0$ and $\alpha_1 + \beta_1 < 1$, the weights on the squared shocks a_{t-k}^2 decrease as k increases.

- The weight on a_{t-k}^2 in the formula for σ_t^2 is $\alpha_1\beta_1^{k-1}$

- This gives an estimate of $\sigma_t^2 = Var(r_t|\mathcal{F}_{t-1})$ that is 'local' and weights the most recent shocks more highly than shocks further back in the data history.

- The GARCH(1,1) model hence has a few attractive features.

- It has only 3 parameters, yet allows for diminishing weight, smooth and local volatility estimation.

- It also uses ALL previous data points to estimate the current volatility σ_t^2

Student-t distribution

- The Student-t distribution is an extension of a Gaussian, that can allow higher kurtosis and hence fatter tails, to capture outliers.

- It is symmetric and has three parameters.

- If $X \sim t_\nu(\mu, h)$ then:

$$E(X) = \mu \,;\ \mathrm{Var}(X) = \frac{h\nu}{\nu - 2}$$

  and

$$E[(X-\mu)^4] = \frac{3\nu^2 h^2}{(\nu-2)(\nu-4)}$$

- ν is the degrees of freedom parameter. It combines with the scale parameter h to control the even moments of the Student-t distribution.

- As ν tends to ∞ the Student-t becomes *exactly* a Gaussian distribution.

- Figure 7 compares the normal to various Student-t densities with differing degrees of freedom. All have $\mu = 0$, $h = 1$ (i.e. only Gaussian has variance 1 here).

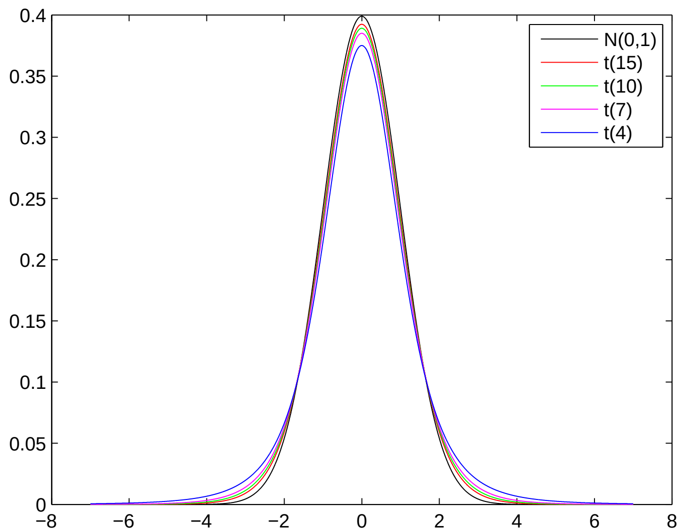

Figure 7: Normal and Student-t density functions.

- $\mu = 0$, $h = 1$ indicates a *standard* Student-t distribution.

- The density function for a Student-t distribution follows:

- If $X \sim t_\nu(\mu, h)$ then:

$$p(X|\mu, h, \nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi h}} \left[1 + \frac{1}{\nu}\left(\frac{X-\mu}{h}\right)^2\right]^{-(\nu+1)/2}$$

- Bollerslev (1987) was the first to use a Student-t error distribution in a GARCH model.

- The AR(1)-GARCH(1,1)-t model is written:

$$\begin{aligned} r_t &= \phi_0 + \phi_1 r_{t-1} + a_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned}$$

where $a_t = \sigma_t \eta_t$ and

$$\eta_t \equiv \sqrt{\frac{\nu-2}{\nu}} \times t_\nu(0,1)$$

- Thus $E(\eta_t) = 0$ and $\text{Var}(\eta_t) = 1$, as required.

- η_t has a *standardised* Student-t distribution as follows. If $X \sim t^*_\nu(\mu, h)$ then:

$$
\begin{aligned}
p(X|\mu, h, \nu) \; &= \; \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{(\nu-2)\pi h}} \\
&\quad \times \left[1 + \frac{1}{\nu-2}\left(\frac{X-\mu}{h}\right)^2\right]^{-(\nu+1)/2}
\end{aligned}
$$

- A standardised error distribution ensures that

$$\text{Var}(a_t|\mathcal{F}_{t-1}) = \sigma_t^2$$

as required in a GARCH type model.

- Figure 8 compares the normal to various *standardised* Student-t densities with differing degrees of freedom. All have mean 0 and variance 1.

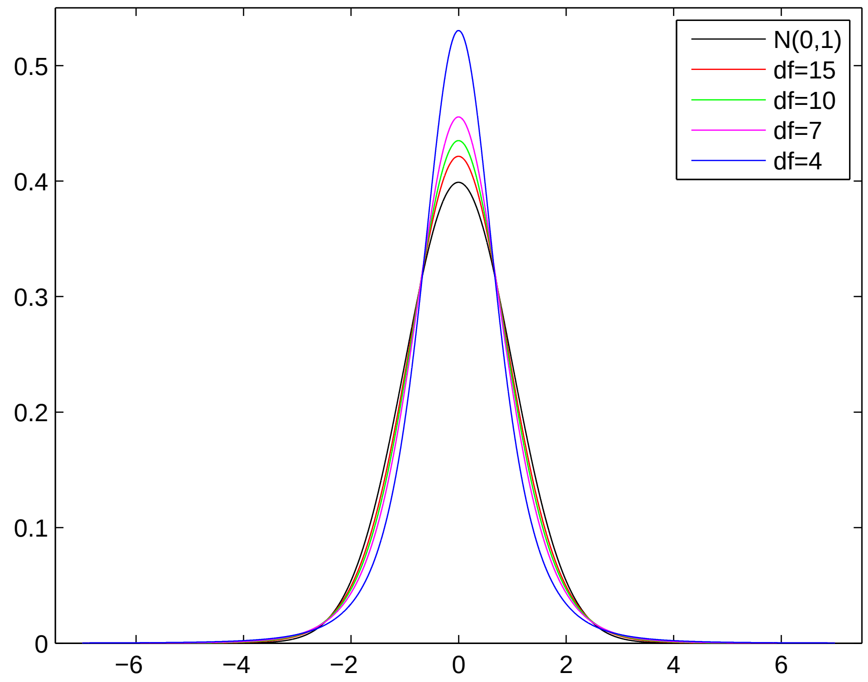

Figure 8: Normal and standardized Student-t density functions.

- Notice anything strange ...??

- Figure 9 compares the normal to various *standardised* Student-t log-densities with differing degrees of freedom.

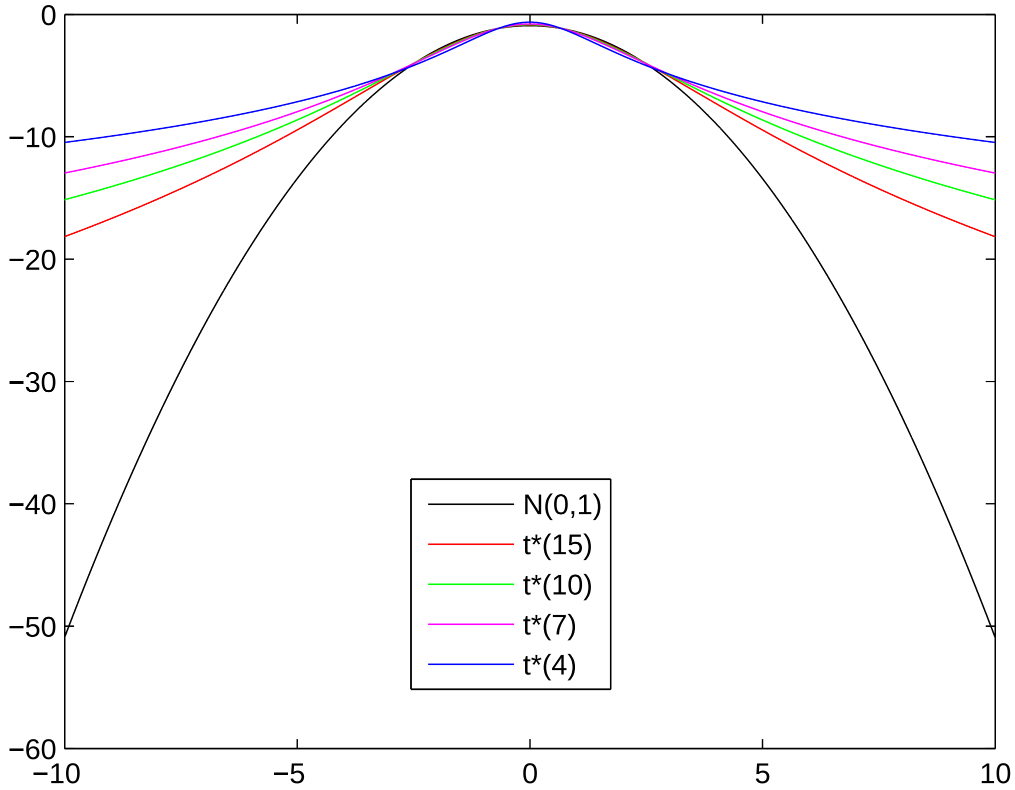

Figure 9: Normal and standardized Student-t log-density functions.

- The conditional likelihood function is

$$p(\mathbf{r}^{m+1,T}|r_1, \ldots, r_m, \theta) = \prod_{t=m+1}^{T} p(r_t|\mathcal{F}_{t-1}, \theta)$$

- The conditional log-likelihood is therefore

$$\begin{aligned} l_c(\theta) &= (T-m)\left[\log(\Gamma((\nu+1)/2)) - \log(\Gamma(\nu/2))\right] \\ &- \frac{T-m}{2}[\log(\pi(\nu-2))] \\ -\frac{1}{2} &\sum_{t=m+1}^{T} \left\{\log(\sigma_t^2) + (\nu+1)\log\left(1 + \frac{(r_t-\mu_t)^2}{(\nu-2)\sigma_t^2}\right)\right\} \end{aligned}$$

- A conditional MLE estimator can be computed by maximimsing l_c above with respect to θ.

- The usual restrictions on the GARCH and AR parameters are enforced, as also is $\nu > 2$, and sometimes $\nu > 4$ *why ??*

- The results for CBA are:

$$
\begin{aligned}
r_t &= 0.068 + 0.043 r_{t-1} + a_t \\
&\quad (0.017)\,(0.018) \\
\sigma_t^2 &= 0.021 + 0.093 a_{t-1}^2 + 0.898 \sigma_{t-1}^2 \\
&\quad (0.005)\,(0.011)\,(0.011)
\end{aligned}
$$

with average volatility estimated as:

$$
\frac{\hat{\alpha}_0}{(1-\hat{\phi}_1^{\,2})(1-\hat{\alpha}_1-\hat{\beta}_1)} = 2.375
$$

and estimated volatility persistence of $\hat{\alpha}_1 + \hat{\beta}_1 = 0.991$.

- Note that this is now much closer to the sample variance of 2.50.

- The df parameter estimate is $\hat{\nu} = 6.85$, with SE of 0.69.

- All parameter estimates are significant at a 5% level.

- Figure 10 summarises the results.

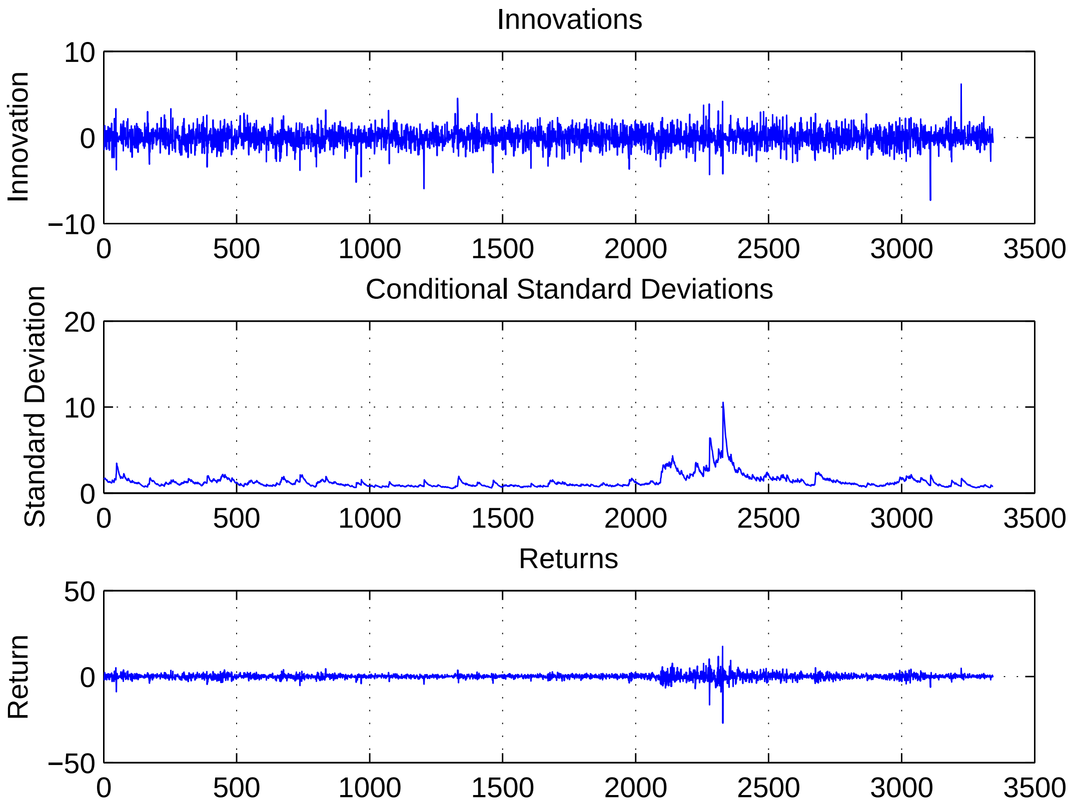

Figure 10: A summary of the AR-GARCH(1,1)-t results for CBA.

- Examining figure 11 the standardised residuals still show some possible outliers and poor fit in the tails.

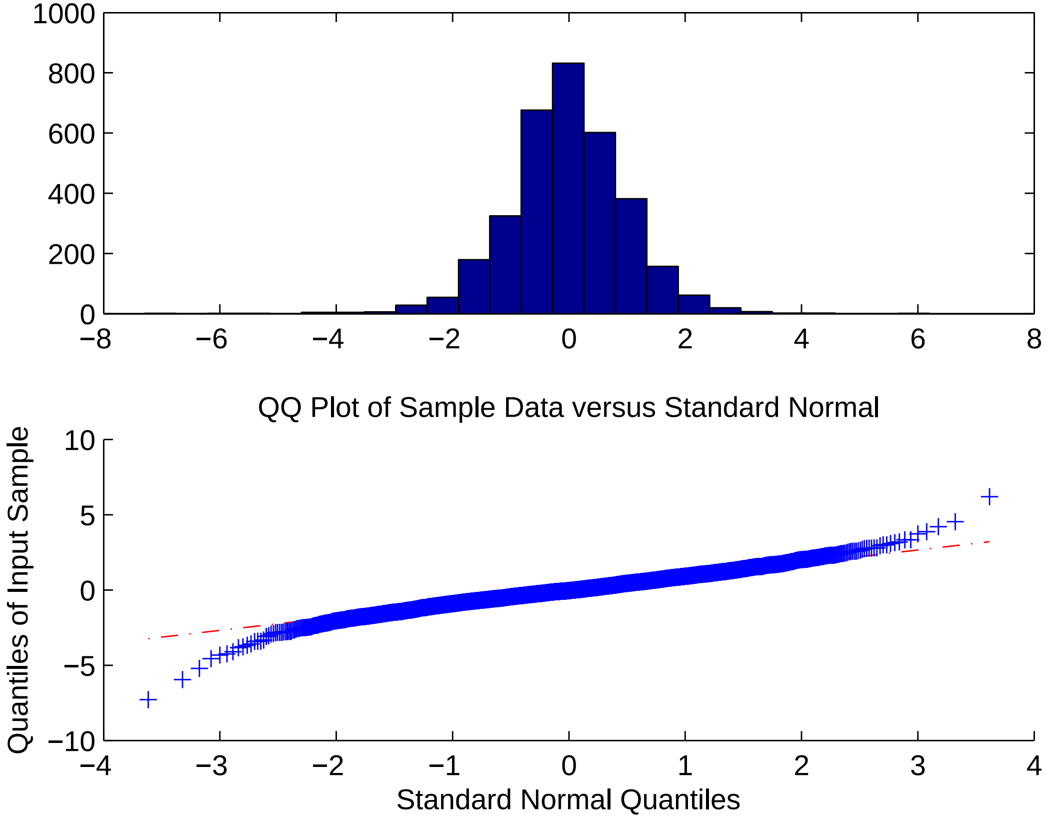

Figure 11: The standardised residual histogram and qq-plot from the AR-GARCH(1,1)-t model for CBA.

- **We expect this**, since the errors are estimated to follow:
$$\hat{\epsilon}_t \sim t^*_{6.85}(0,1)$$
which has kurtosis of:
$$\kappa = 3\frac{6.85-2}{6.85-4} = 5.10$$
but the qq-plot compares to a Gaussian distribution with $\kappa = 3$!

- There is a result in probability theory allowing us to transform between ANY two probability distributions, as follows:

- If $X \sim F$ and $Y \sim G$ are two rvs with cdfs F, G, then:
$$Y \equiv G^{-1}\left(F(X)\right) \;\; X \equiv F^{-1}\left(G(Y)\right)$$

- This is because a cdf, like $F(X)$ generates a set of probabilities.

- An inverse cdf (like G^{-1}), take a set of probabilities and changes them back into a r.v. (in this case Y).

- We can transform between ANY two distributions in this manner

- We thus transform our Student-t residuals back to Gaussian residuals via:

$$e_t = \Phi^{-1} F_{t_{6.85}} \left(\sqrt{\frac{6.85}{6.85 - 2}} \hat{\epsilon}_t \right)$$

- Here $\hat{\epsilon}_t$ should have a standardised Student-t, while $\frac{6.85}{6.85-2}\hat{\epsilon}_t$ has a usual Student-t, with 6.85 df.

- Matlab only has the cdf for a usual t, not a standardised t.

- Examining figure 12, the histogram and qq-plot shows that the transformed residuals now seem quite close to a standard normal $N(0,1)$.

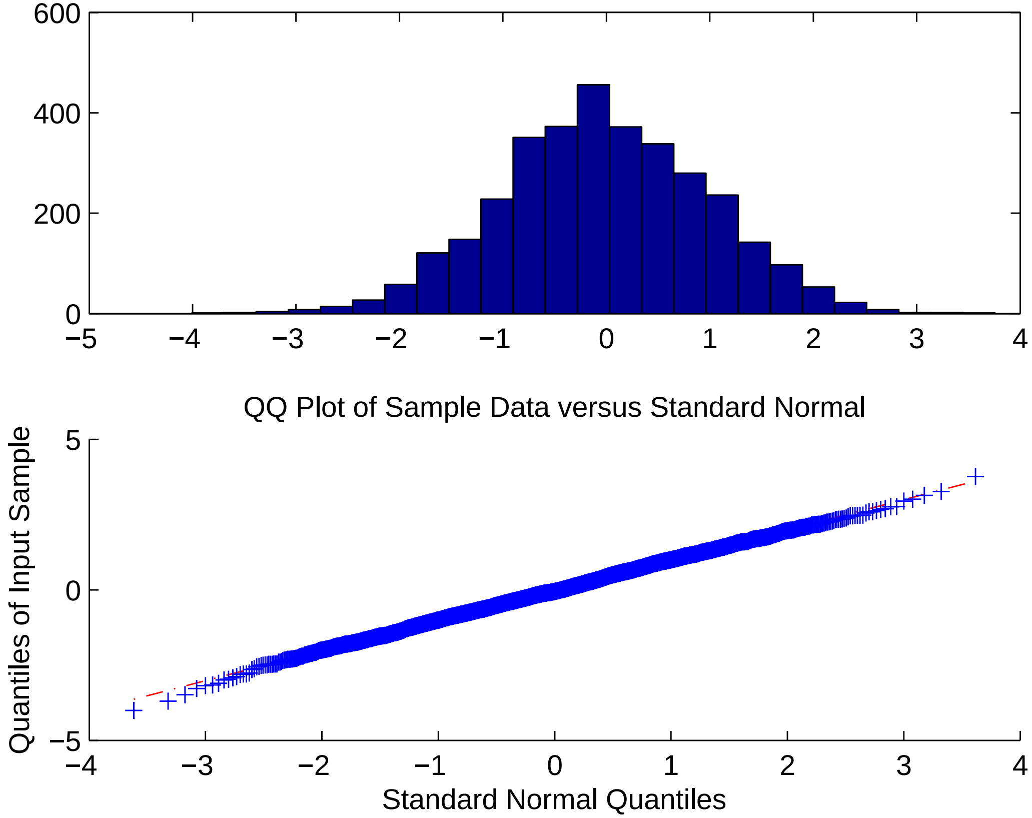

Figure 12: The transformed standardised residual histogram and qq-plot from the AR-GARCH(1,1)-t model for CBA.

- Some possible outliers close to -4 standard deviations may still be problematic, but this is the best fit we have seen so far!

- The JB test gave a p-value of 0.5 with sample skewness and kurtosis being $-0.04, 3.02$ respectively. We cannot reject a Student-t as the conditional distribution for CBA returns!

- The ACF plots for the transformed residuals and their squares, in figure 13 show no significant auto-correlations at low lags in either plot.

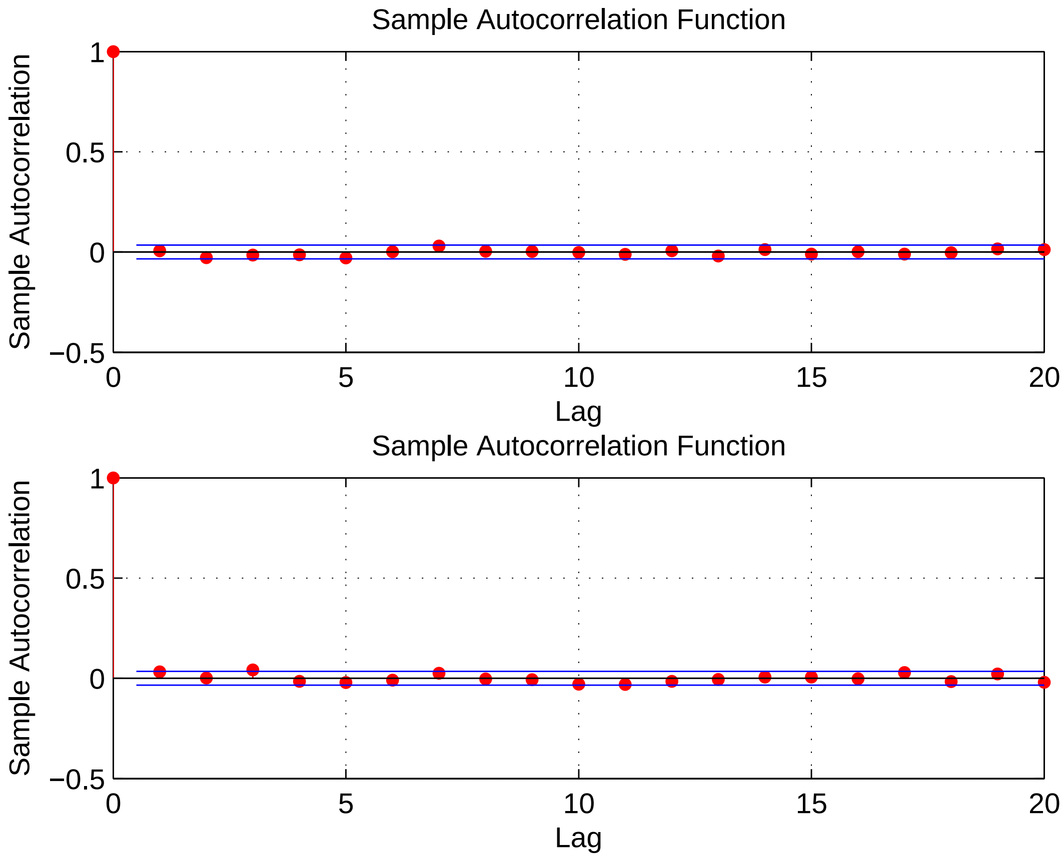

Figure 13: Autocorrelation function of the transformed standardised residuals (top) and their squares (bottom) from the AR-GARCH(1,1)-t model for CBA.

- Ljung-Box tests, with $m = 9, 14$ and $5, 10$ df were conducted as follows:

- We obtained p-values of $0.071, 0.25$ for the standardised transformed residuals, indicating the AR(1) mean equation is well specified in this case: no remaining significant autocorrelation exists.

- For the squared transformed residuals, LB tests found p-values of $0.013, 0.017$, indicating that the GARCH(1,1) volatility equation can still be improved and the residuals contain remaining significant ARCH effects.

- Finally, we cannot reject the normality for the transformed standardised residuals, suggesting the Student-t is a good choice of error distribution for this data.

GARCH order selection

- The SIC and AIC can again be used to select (p, q) in a GARCH(p,q) model:

$$\begin{aligned} \text{AIC} &= -2 * l_c(\theta) + 2 * (p + q) \\ \text{SIC} &= -2 * l_c(\theta) + (p + q) * log(n) \end{aligned}$$

  The model order (p, q) that minimises the AIC and/or SIC is chosen.

- Usually only $(p, q) \in \{1, 2, 3, 4, 5\}$ are considered, giving 25 models to consider.

- When using Student-t errors an extra penalty (i.e. +1) should be applied to the AIC, SIC penalties, when comparing to Gaussian error models.

- Similarly, when fitting an AR(l) mean equation, an extra penalty of l should be applied when comparing models with different orders.

- We'll try this out in one of the lab sessions.

- For this data, among constant mean models with Gaussian errors, a GARCH(2,2) is favoured by AIC and a GARCH(1,2) is favoured by SIC.

- When allowing an AR(1) with Gaussian errors, the preferred model is an AR(1)-GARCH(2,2) by AIC, but a constant mean GARCH(1,2) by SIC.

- When allowing a constant mean and Student-t errors, the preferred model is an GARCH(1,2) by AIC, but a GARCH(1,1) by SIC.

- When allowing an AR(1) and Student-t errors, the preferred model is an AR(1)-GARCH(1,2) by AIC, but a constant mean GARCH(1,1) by SIC.

- The results of fitting an AR(1)-GARCH(1,2) with Student-t errors for CBA are:

$$
\begin{aligned}
r_t &= 0.067 + 0.043 r_{t-1} + a_t \\
&\ (0.017)\,(0.018) \\
\sigma_t^2 &= 0.028 + 0.129 a_{t-1}^2 + 0.324 \sigma_{t-1}^2 + 0.535 \sigma_{t-2}^2 \\
&\ (0.016)\,(0.137)\,(0.130)
\end{aligned}
$$

- The results of the tests are almost exactly the same as before, except that:

- Ljung-Box tests, with $m = 10, 15$ and $5, 10$ df on the squared transformed standardised residuals had p-values of 0.017 and 0.070 respectively.

- Thus, there still seems to be significant ARCH effects in the residuals, but the GARCH(1,2) has made a minor improvement in fit.

Forecasting single period Value-at-Risk and Expected Shortfall

- The relevant forecast distribution is $r_{t+1}|\mathcal{F}_t$.

- Any parametric volatility model assumes: $r_{t+1}|\mathcal{F}_t \sim D(\mu_{t+1}, \sigma^2_{t+1})$.

- If μ_{t+1} and σ^2_{t+1} are available using $\mathcal{F}_t$ we can do VaR forecasting directly

- $$\begin{aligned} \mathrm{VaR}_p &= D^{-1}_{(\mu_{t+1},\sigma^2_{t+1})}(p) \\ &= \mu_{t+1} + D^{-1}(p)\sigma_{t+1} \end{aligned}$$

- If $p = 0.01$ and $D \equiv N(0,1)$ then $D^{-1}(p) = \Phi^{-1}(0.01) = -2.326$

- If $p = 0.05$ and $D \equiv N(0,1)$ then $D^{-1}(p) = \Phi^{-1}(0.05) = -1.645$

- For level p, when $D \equiv \sqrt{\frac{\nu-2}{\nu}} \times t_\nu(0,1)$ then $D^{-1}(p) = \sqrt{\frac{\nu-2}{\nu}} \times T_\nu^{-1}(p)$.

- Matlab calculates all these for us.

- For CBA and NWS, the following table shows VaR forecasts for the next day in the sample:

Table 1: The 1-step-ahead forecast VaRs for all models for CBA for p=0.05 and $p = 0.01$.

Model	ARCH(9)-N	GARCH(1,1)-N	AR-GARCH(1,1)-N	GARCH(1,1)-t	AR-GARCH(1,2)-t
CBA $p = 0.05$	-1.114	-1.194	-1.202	-1.119	-1.132
NWS $p = 0.05$	-1.751	-1.708	-1.707	-1.699	-1.684
CBA $p = 0.01$	-1.605	-1.716	-1.722	-1.815	-1.825
NWS $p = 0.01$	-2.493	-2.430	-2.429	-2.656	-2.566

- For expected shortfall we again use the distribution of $r_{t+1}|\mathcal{F}_t$

- It is not hard to show that, if $r_{t+1}|\mathcal{F}_t \sim N(\mu_{t+1}, \sigma^2_{t+1})$ then:

$$ES_p = \mu_{t+1} - \sigma_{t+1}\frac{\phi\left(\Phi^{-1}(p)\right)}{p}$$

where $\phi()$ is the standard normal probability density function (pdf) and $\Phi^{-1}(p)$ is the inverse standard normal cdf at probability p.

- If $r_{t+1}|\mathcal{F}_t \sim t^*_\nu(\mu_{t+1}, \sigma^2_{t+1})$ then (it is MUCH harder to show that):

$$ES_p(h) = \mu_{t+1} - \sigma_{t+1}\frac{f\left(T_\nu^{-1}(p)\right)}{p}\left(\frac{\nu + (T_\nu^{-1})^2}{\nu - 1}\right)\sqrt{\frac{\nu-2}{\nu}}$$

where $f()$ is the standard Student-t density function (pdf) and $T^{-1}(p)$ is the inverse standard Student-t cdf at probability p.

- For CBA and NWS, the following table shows ES forecasts for the next day in the sample:

Table 2: The 1-step-ahead forecast ES for all models for CBA, for p=0.05 and p=0.01.

Model	ARCH(9)-N	GARCH(1,1)-N	AR-GARCH(1,1)-N	GARCH(1,1)-t	AR-GARCH(1,2)-t
CBA $p = 0.05$	-1.415	-1.514	-1.521	-1.864	-1.872
NWS $p = 0.05$	-2.206	-2.150	-2.150	-2.661	-2.477
CBA $p = 0.01$	-1.849	-1.975	-1.980	-2.744	-2.746
NWS $p = 0.01$	-2.862	-2.789	-2.788	-3.809	-3.446

- In module 4 we will compare these parametric model forecasts over time and with other methods.