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 1: ARCH Models

- Mandelbrot (1963) was one of the first published papers to illustrate that neither price differences nor log returns were Gaussian.

- Mandelbrot, B (1963). The variation of certain speculative prices. *The Journal of Business*, **4**, 394-419.

- He posited a different iid distribution, so as to capture empirical fat tails in log returns.

- However, modern financial analysts and practitioners believe that asset return volatility changes over time. This can also cause 'fat' tails.

- Nobel Laureate Rob Engle introduced the Auto-Regressive Conditional Heteroskedastic process in his seminal 1982 paper.

- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of variance of United Kingdom inflation. *Econometrica*, **50**, 987-1008.

Definition & Moments

- Figures 1 and 2 show daily closing prices and close to close returns for CBA and NWS stock on the ASX for January, 2000 to February, 2013.

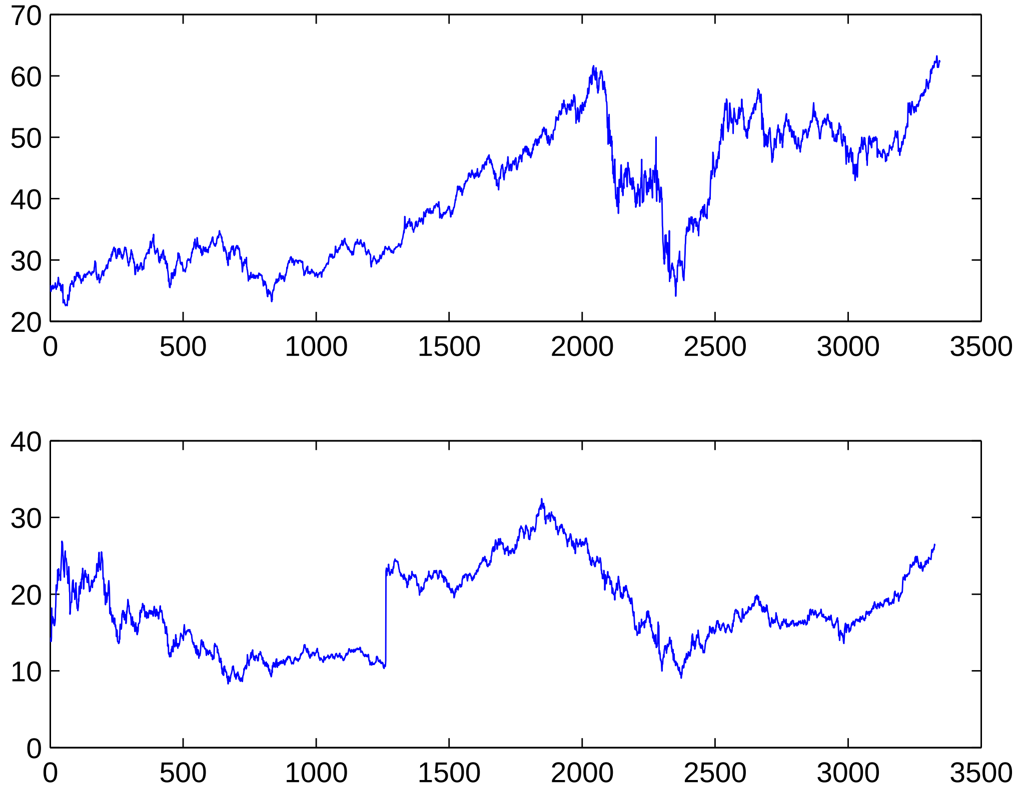

Figure 1: Daily closing prices for CBA (top) and NWS from 2000 to 2013.

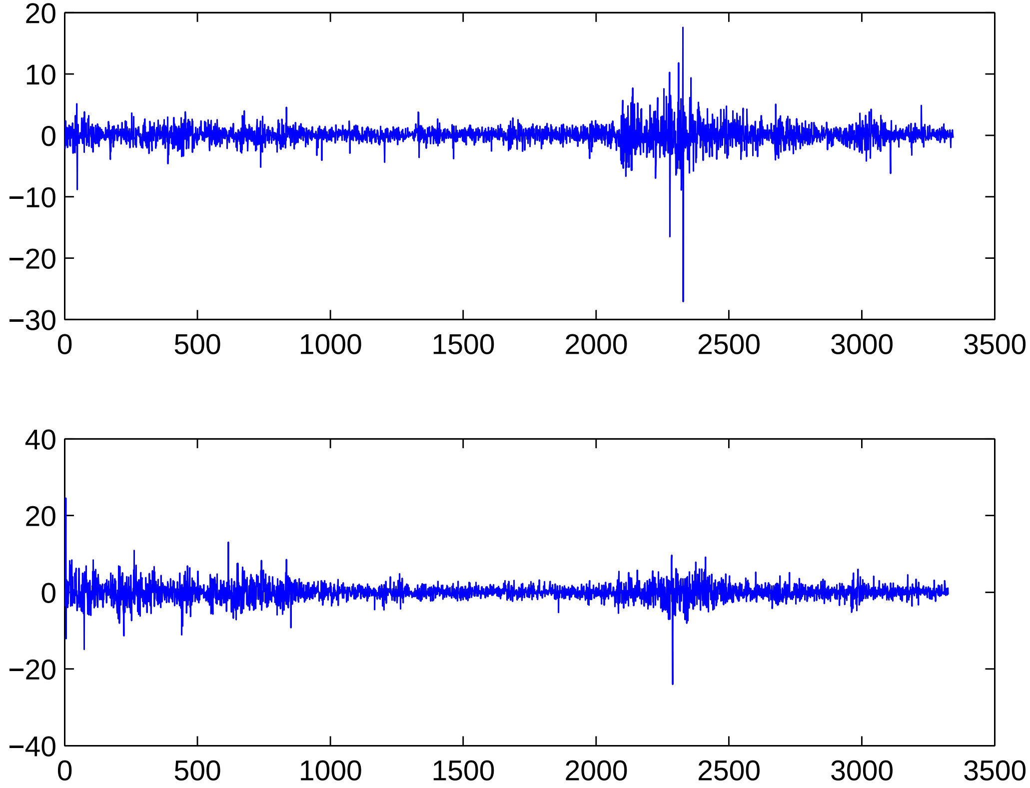

Figure 2: Daily percentage log-returns for CBA (top) and NWS from 2000 to 2013.

- What happened during late 2004 to NWS?

- I set the return on that day to the median NWS return.

- Figure 3 shows the NWS returns for analysis.

- The data suggest that $Var(r_t|\mathcal{F}_{t-1})$ might change over time, in a correlated fashion

- Modelling such persistent volatility has been prevalent since Engle (1982)

- The ARCH process for a series of returns $\{r_t\}$, is defined as:

$$
\begin{aligned}
r_t &= \mu_t + a_t = \mu_t + \sigma_t \epsilon_t \\
\sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \ldots + \alpha_p a_{t-p}^2
\end{aligned}
$$

where $\sigma_t^2 \equiv \mathrm{Var}(a_t|\mathcal{F}_{t-1})$ and $\mathcal{F}_{t-1}$ is the available information at time $t-1$.

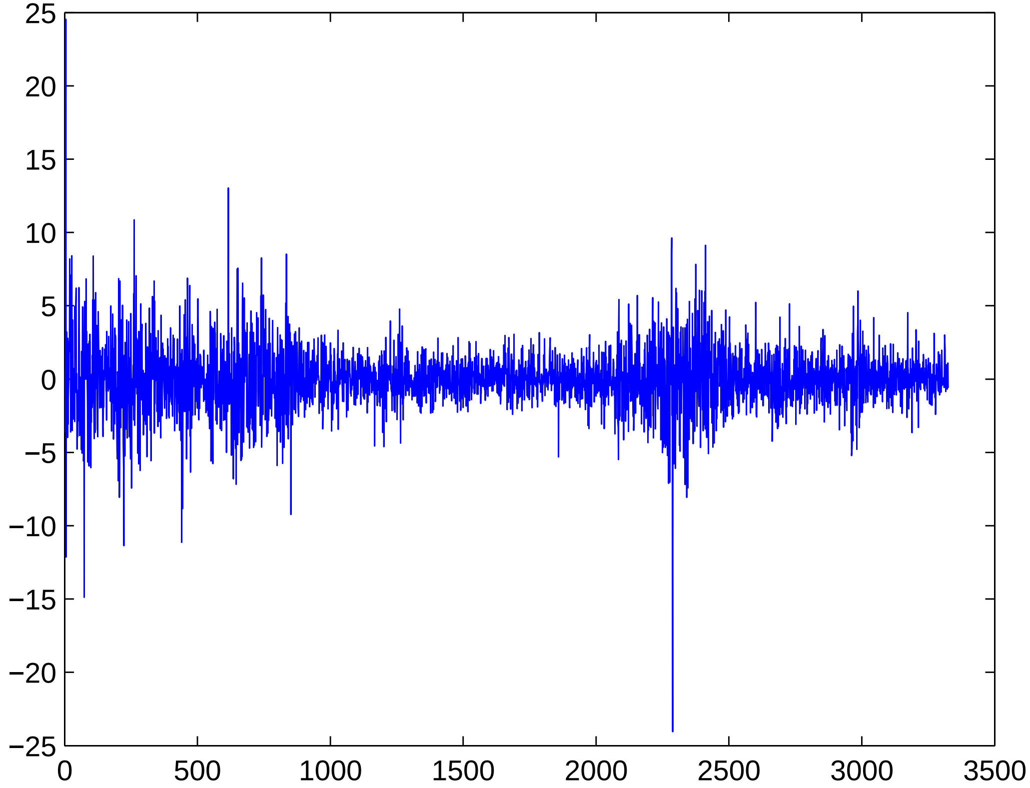

Figure 3: Daily adjusted percentage log-returns for NWS from 2000 to 2013.

- Rather than regression, or an auto-regression, where we usually specify a model **only** for the conditional expectation, $E(r_t|\mathcal{F}_{t-1})$, here we specify a model for the conditional variance, $\sigma_t^2 \equiv \text{Var}(r_t|\mathcal{F}_{t-1})$.

- Note that $E(r_t|\mathcal{F}_{t-1}) = \mu_t$, which we leave unspecified for the moment.

- The series $\{a_t\}$ are *mean-corrected* returns, and it is this series that is modeled as ARCH.

- The heteroskedasticity (i.e. σ_t^2 changes with time) implies an extra level of uncertainty over standard quantitative models.

- But there is still only one random disturbance $\{\epsilon_t\}$ for $t = 1, 2, \ldots$ in this model.

- The 1st and 2nd order moments are *assumed* to be
  - $E(\epsilon_t) = 0 = E(\epsilon_t|\mathcal{F}_{t-1})$
  - $\text{Var}(\epsilon_t) = 1 = \text{Var}(\epsilon_t|\mathcal{F}_{t-1})$

and the sequence $\{\epsilon_1, \epsilon_2, \ldots\}$ is *assumed* to be iid.

- The series $\{a_t\}$ is then uncorrelated, i.e.

$$Cov(a_t, a_{t-j}) = 0, \;\; j = 1, 2, ...$$

- However, $\{a_t\}$ is NOT an independent series.

- Why?

- The first and second unconditional moments of the mean-corrected series $a_t = r_t - \mu_t$ are computed as:

$$E[a_t|\mathcal{F}_{t-1}] = E[\sigma_t \epsilon_t|\mathcal{F}_{t-1}] = \sigma_t E[\epsilon_t|\mathcal{F}_{t-1}] = 0$$

and

$$\begin{aligned}\mathrm{Var}(a_t) &= E[a_t^2] = E[E[a_t^2|\mathcal{F}_{t-1}]] \\ &= E[\sigma_t^2 \mathrm{Var}(\epsilon_t|\mathcal{F}_{t-1})] \\ &= E[\alpha_0 + \alpha_1 a_{t-1}^2 + \ldots + \alpha_p a_{t-p}^2] \\ &= \alpha_0 + \alpha_1 E[a_{t-1}^2] + \ldots + \alpha_p E[a_{t-p}^2] \\ &= \alpha_0 + \sum_{j=1}^{p} \alpha_j \mathrm{Var}[a_{t-j}]\end{aligned}$$

- This has implications for stationarity, as we shall see.

- Recall that **strict stationarity** occurs iff

$$\{a_{t_1}, a_{t_1+1}, \ldots, a_{t_1+k}\}$$

has a distribution that is exactly equivalent for all choices of t_1, and for any fixed k.

- Note that the ARCH(p) model above is almost **equivalent** to the model

$$\begin{aligned} r_t &= \mu_t + a_t \\ a_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \ldots + \alpha_p a_{t-p}^2 + \eta_t \end{aligned}$$

  where $\{\eta_t\}$ is an uncorrelated, zero mean (but not iid) process.

- This means that $\{a_t^2\}$ evolves similarly to an ordinary linear AR(p) process, except with a positivity constraint.

ARCH(1)

- When $p = 1$:

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

- Normally, in ARCH models (as in most time series models) stationarity is *imposed* on the process.

- Why?

- Stationarity requires (e.g.) that

$$\text{Var}(a_t) = \text{Var}(a_s)$$

for all s, t.

- From the formula above, we have

$$\mathrm{Var}(a_t) = \alpha_0 + \alpha_1 \mathrm{Var}(a_{t-1})$$

- so, stationarity implies that

$$\mathrm{Var}(a_t) = \frac{\alpha_0}{(1 - \alpha_1)} > 0$$

  which is constant for all t.

- To ensure this variance exists and is positive, it is usual to assume $\alpha_0 > 0$ and $\alpha_1 < 1$.

- It is also common to enforce $\alpha_1 \geq 0$. *why?*

- In many applications, higher order moments need to exist, which can result in additional constraints

- For instance, often the fourth order moment needs to be finite.

- Recall that the kurtosis of a series **r** is:

$$\kappa = \frac{E\left((r_t - \mu)^4\right)}{Var(r_t)^2} = \frac{E\left(a_t^4\right)}{Var(a_t)^2}$$

- Consider imposing $\epsilon_t \sim N(0,1)$.

- This implies that the conditional kurtosis of $r_t|\mathcal{F}_{t-1} \equiv a_t|\mathcal{F}_{t-1}$ is 3.

- What is the unconditional kurtosis? Is it greater than 3? Can ARCH induce fat-tailed returns?

- Then, some algebra reveals that:

$$E[a_t^4] = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$$

- Since the fourth order moment of $\{a_t\}$ has to be positive, both $1-\alpha_1^2$ and $1-3\alpha_1^2$ must be of the same sign.

- Since we already have $\alpha_0 > 0, 0 \leq \alpha_1 < 1$, we now further require $\alpha_1 < \frac{1}{\sqrt{3}} \approx 0.577$.

- The unconditional kurtosis of $\{a_t\}$ is then

$$\kappa = \frac{E[a_t^4]}{(\mathrm{Var}(a_t))^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2}$$

- Since $1-\alpha_1^2 > 1-3\alpha_1^2$, the kurtosis in an ARCH model with Gaussian errors IS in excess of that of a normal distribution (i.e. > 3), whenever $\alpha_1 \neq 0$.

- Therefore, the tail behaviour in an ARCH model, with Gaussian errors, is *heavier* than that of a homoskedastic Gaussian distribution.

- Some implications for modeling financial return data are:

- To derive such formulas for higher order ARCH models is trickier (but possible).

ARCH Estimation

- Consider a LS estimator for μ, α_0 and α_1 in a constant mean ARCH(1) model

- What is the sum of squares to minimise here?

$$\sum_{t=1}^{n} a_t^2 = \sum_{t=1}^{n} (r_t - \mu)^2 \quad ??$$

OR

$$\sum_{t=1}^{n} \epsilon_t^2 = \sum_{t=1}^{n} \frac{(r_t - \mu)^2}{\sigma_t^2} \quad ??$$

- Which should be used? Any issues here?

- Note that in an ARCH(1) we have:

$$\mathrm{Var}(a_t|\mathcal{F}_{t-1}) = E(a_t^2|\mathcal{F}_{t-1}) = \alpha_0 + \alpha_1 a_{t-1}^2$$

- Engle (1982) considered the two-stage LS estimator, obtained via first minimising:

$$\sum_{t=1}^{n} (r_t - \mu)^2,$$

then using the estimates $\hat{\mu} = \bar{r}$, so as to minimise:

$$\sum_{t=1}^{n} (\hat{a}_t^2 - \alpha_0 - \alpha_1 \hat{a}_{t-1}^2)^2 \ ,$$

where $\hat{a}_t = r_t - \hat{\mu}_t$.

- This second sum of squares is simply a LS auto-regression, thus giving estimates:

$$\begin{aligned}
\hat{\mu} &= \bar{r} \\
\hat{\alpha}_0 &= \bar{a^2}(1 - \hat{\alpha}_1) \\
\hat{\alpha}_1 &= \frac{\sum_{t=2}^{n}(\hat{a}_t^2 - \bar{a^2})(\hat{a}_{t-1}^2 - \bar{a^2})}{\sum_{t=2}^{n}(\hat{a}_{t-1}^2 - \bar{a^2})^2} = \hat{\rho}_1(\mathbf{a^2})
\end{aligned}$$

- What are the assumptions of LS for AR models and when are they optimal, efficient, BLUE, etc.?

- The error series from this regression is $\eta_t = \hat{a}_t^2 - \alpha_0 - \alpha_1 \hat{a}_{t-1}^2$.

- To establish the standard properties of least squares estimators requires three assumptions:

  (i) The error series $\eta_1, \ldots, \eta_n$ are an iid random sample.

  (ii) The conditional expectation $E(\hat{a}_t^2|\mathcal{F}_{t-1}) = \alpha_0 + \alpha_1\hat{a}_{t-1}^2$

  (iii) The 4th moments of the observations, i.e. $E([\hat{a}_t^2]^4) = E(\hat{a}_t^8)$ must exist and be finite.

- Assumption (i) is dubious, actually the $\text{Var}(\eta_t|\mathcal{F}_{t-1})$ changes with time in this model.

- Assumption (ii) is somewhat doubtful (as we shall see)

- Assumption (iii) is highly doubtful and very likely untrue in real return data. Why?

- Thus, LS estimation is not efficient or optimal for ARCH models in general.

- For ARCH models, likelihood estimation is much more commonly applied.

Maximum Likelihood

- The *likelihood* function is simply the *joint* density function for a sample of data, conditional upon the parameter values.

- It is viewed as a function of the unknown parameter values (not the sample data)

- Say we observe $y = 20$ from the normal distribution $N(25, 25)$. i.e. $n = 1$.

- For particular values of μ and σ^2, the normal density function evaluated at $y = 20$ **is** the relative likelihood of observing that observation. i.e.

$$\begin{aligned} L(\mu, \sigma^2 | y = 20) &= p(y = 20 | \mu, \sigma^2) \\ &= (2\pi\sigma^2)^{-0.5} \exp\left[\frac{-(20-\mu)^2}{2\sigma^2}\right] \end{aligned}$$

- In general we don't know the true values for μ, σ^2

- But for any particular values of these parameters, say $\mu = 10, \sigma^2 = 16$ we can evaluate the likelihood and compare with any other values, say $\mu = 15, \sigma^2 = 10$.

- If $L(\mu = 10, \sigma^2 = 16|y = 20) > L(\mu = 15, \sigma^2 = 10|y = 20)$ then it is more likely to observe $y = 20$ when $\mu = 10, \sigma^2 = 16$.

- We wish to choose as estimates for μ, σ^2 those values that make the likelihood function a *maximum*.

- Figure 4 shows what are called *profile* likelihoods: the likelihood as a function of 1 variable, where the other variable is fixed at its estimated (or true) value.

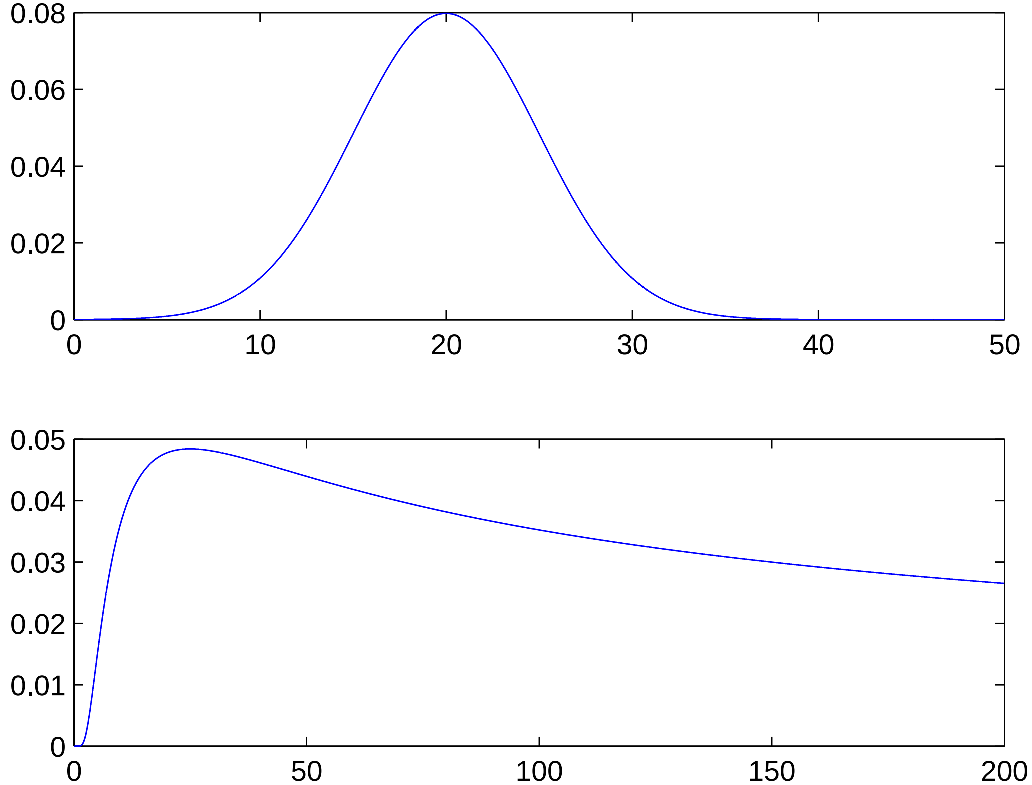

Figure 4: Profile likelihoods when $y = 20$.

- The (profile) likelihood function, for $y = 20$, for different values of μ, seen in Figure 4 (top), looks almost the same as the original normal density.

- The maximum likelihood estimate (mle) is $\hat{\mu} = 20 = y$ in this case.

- The (profile) likelihood as a function of σ^2 is reasonably flat, e.g. σ^2 as high as 200 is quite likely. Even if we know the true value of μ, one observation doesn't tell us much about σ^2. Although the maximum occurs at ...

- Note that the likelihood is really a bivariate function of the parameters μ, σ^2 here. The data are fixed.

Maximum Likelihood (ctd)

- The likelihood is the *joint* density for observing the sample obtained.

- Say we have observed $y_1, y_2, \ldots, y_n$ independently from a $N(\mu, \sigma^2)$ distribution

- The likelihood function is then:

$$\begin{aligned} L(\mu, \sigma^2 \mid y_1, \ldots, y_n) &= p(y_1, y_2, \ldots, y_n | \mu, \sigma^2) \\ &= p(y_1|\mu, \sigma^2) \ldots p(y_n|\mu, \sigma^2) \text{ by independence} \\ &= \prod_{t=1}^{n} (2\pi\sigma^2)^{-0.5} \exp\left[\frac{-(y_t - \mu)^2}{2\sigma^2}\right] \end{aligned}$$

- This is the joint likelihood of observing the sample when $y_t \sim N(\mu, \sigma^2)$.

- After some manipulation and letting $\mathbf{y} = y_1, y_2, \ldots, y_n$ this becomes:

$$L(\mu, \sigma^2|\mathbf{y}) = (2\pi\sigma^2)^{-n/2} \exp\left[\frac{-\sum_{t=1}^{n}(y_t - \mu)^2}{2\sigma^2}\right]$$

- Figure 5 displays the profile likelihoods for different values of μ (top) and σ^2, for a sample where $n = 100$. *comments?*

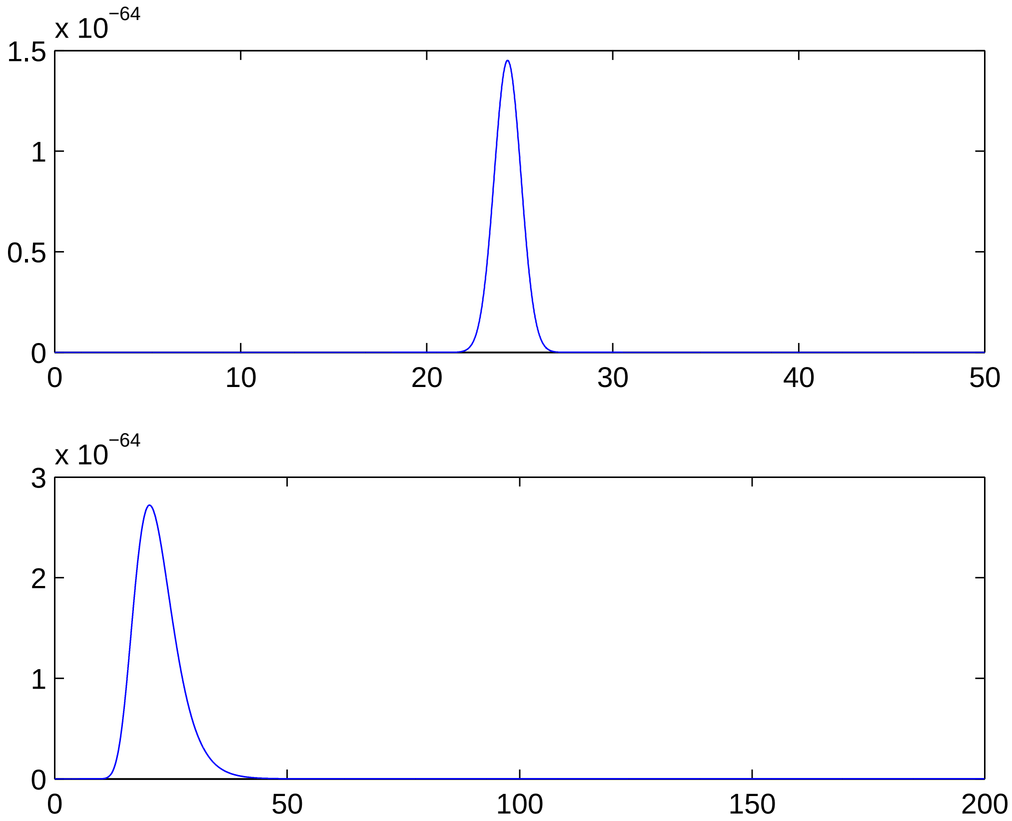

Figure 5: Profile likelihood functions when n=100.

- Figure 6 displays the full bivariate likelihood function.

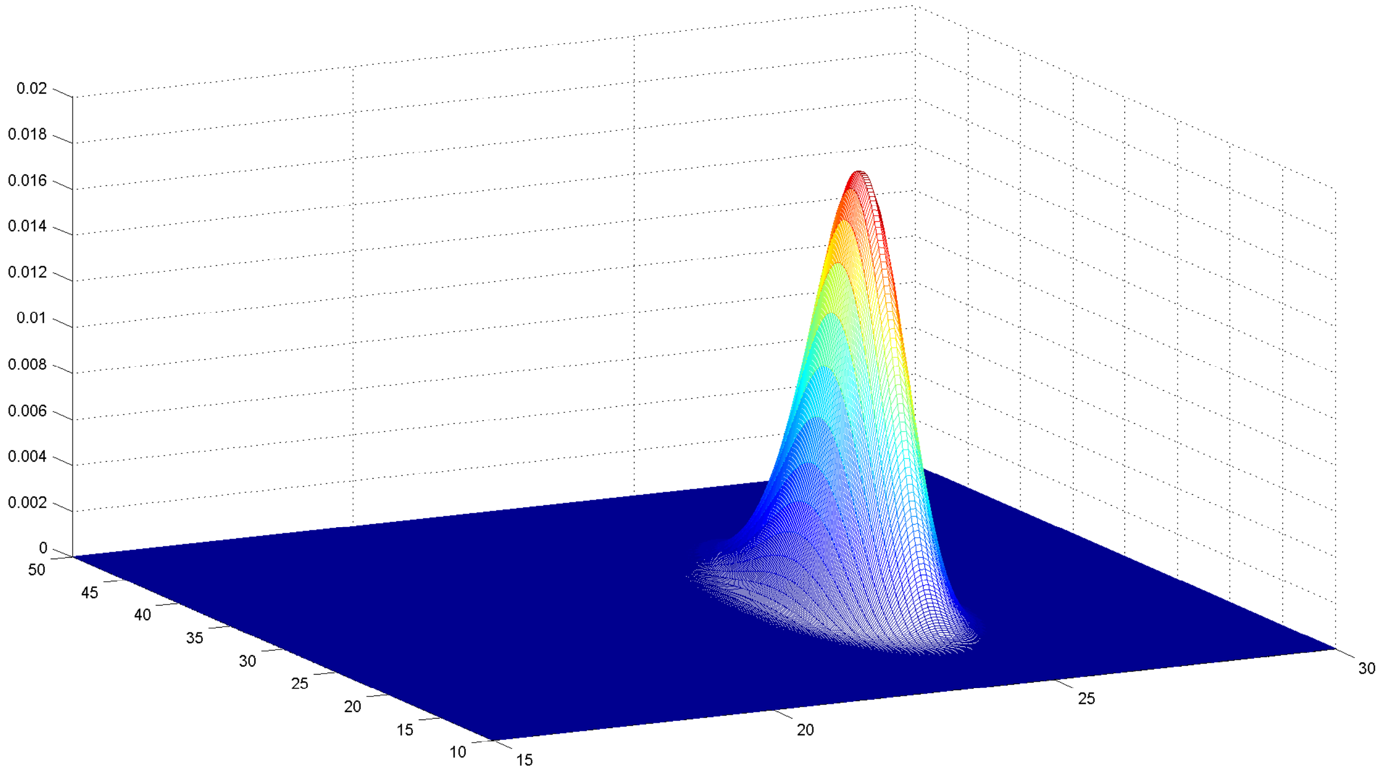

Figure 6: Bivariate likelihood function when n=100.

- We see that the values of the parameters that maximise the likelihood function are very close to the true values, $\mu = 25$ and $\sigma^2 = 25$.

- The values of μ, σ^2 that maximise the likelihood function are called the *maximum likelihood estimates* (mle).

- We can show that the mle here are $\hat{\mu} = \bar{y}$ and $\hat{\sigma}^2 = (n-1)s^2/n$ i.e. the sample mean and (almost) sample variance.

- Note that the likelihood function here contains the term

$$\frac{-\sum_{t=1}^{n}(y_t - \mu)^2}{2\sigma^2}$$

  which is a sum of squares.

- Thus, to maximise the Gaussian likelihood (for μ) is equivalent to minimising this sum of squares. What is the implication here?

- Note also that

$$\log(L(\mu, \sigma^2|\mathbf{y})) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{\sum_{t=1}^{n}(y_t - \mu)^2}{2\sigma^2}$$

- Maximising the likelihood is equivalent to maximising the *log* likelihood OR minimising $-log$ likelihood.

- Large sample theory says that maximum likelihood estimates have a CLT and are normally distributed across samples.

- Their large sample variance (beyond this course) is estimated via the Fisher information matrix, which involves the second derivatives of the log-likelihood function.

- In general the mle must be found numerically by a search algorithm.

- **Summary**
  - The likelihood function is the joint relative ’probability’ of observing the sample of data
  - The likelihood is a function of unknown parameter values
  - Parameter estimates are chosen to maximise this function, in order to allow the data observed to have the highest likelihood of occurring if the estimates were really the true parameter values.
  - We put ultimate faith in our sample of data under this method!

Maximum Likelihood for ARCH

- Consider the ARCH(1) model:

$$\begin{aligned} r_t &= \mu + a_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 \end{aligned}$$

where $\sigma_t^2 \equiv \mathrm{Var}(a_t|\mathcal{F}_{t-1})$, $a_t = \sigma_t \epsilon_t$ and $\epsilon_1, \ldots, \epsilon_n$ is an iid series.

- To find the joint density of a sample of data $r_1, \ldots, r_n$ under an ARCH(1) model, we need to specify a probability distribution for ϵ.

- First, we try $\epsilon_t \sim N(0, 1)$ for all t.

- This implies that $a_t|\mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$.

- and that $r_t|\mathcal{F}_{t-1} \sim N(\mu, \sigma_t^2)$.

- The likelihood function for a sample of returns $r_1, \ldots, r_n$ from this model is

$$\begin{aligned} L(\mu, \alpha_0, \alpha_1 \quad | \quad r_1, r_2, \ldots, r_n) &= p(r_1, r_2, \ldots, r_n|\theta) \\ &= p(r_1|\theta)p(r_2|r_1, \theta) \ldots p(r_n|r_1, \ldots, r_{n-1}, \theta) \\ &= p(r_1|\theta) \prod_{t=2}^{n} (2\pi\sigma_t^2)^{-0.5} exp\left[\frac{-(r_t - \mu)^2}{2\sigma_t^2}\right] \end{aligned}$$

- This is the joint probability density of observing the sample of returns under any ARCH model with Gaussian errors.

- The likelihood is a function of the unknown parameters $\theta = (\mu, \alpha_0, \alpha_1)$.

- By ignoring $p(r_1|\mu, \alpha)$ and taking logs we see that the *conditional* log-likelihood function is:

$$\log(L_c(\theta|\mathbf{r})) = -\frac{n-1}{2}\log(2\pi) - \frac{1}{2}\sum_{t=2}^{n}\left[\log(\sigma_t^2) + \frac{(r_t - \mu)^2}{\sigma_t^2}\right]$$

- For any sample of returns $\mathbf{r}$, Matlab will search for the values of the parameters μ, α_0, α_1 that maximise this function.

- It will also estimate the t-statistic and p-value from the hypothesis test that each parameter's true value equals 0.

- Approximate MLEs can be computed by maximising $\log(L_c(\theta|\mathbf{r}))$.

- Note that the MLEs will not (ever) be infinite.

- Matlab actually sets r_0 to a value, either:

  1. $\bar{r}$, the sample mean; OR
  2. any user-supplied value.

- If r_0 is known, then the distribution of $a_1|r_0, \mu, \theta \sim N\left(0, \alpha_0 + \alpha_1(r_0 - \mu)^2\right)$ is known and employed in the likelihood function.

- Closed form solutions do not exist for MLEs in ARCH models.

Examples

- CBA is considered a *defensive* stock, and is expected to display lower volatility.

- NWS is considered a more *aggressive* (though still blue-chip) stock, and is expected to display higher volatility.

- We first fit a Gaussian ARCH(1) model to each of these series.

- For CBA this resulted in the following fit:

$$\begin{aligned} r_t &= \underset{(0.020)}{0.057} + a_t \\ \sigma_t^2 &= \underset{(0.030)}{1.264} + \underset{(0.021)}{0.519} a_{t-1}^2 \end{aligned}$$

where standard errors are in brackets below estimates.

- with estimated average volatility of

$$\frac{\hat{\alpha_0}}{1-\hat{\alpha_1}} = \frac{1.264}{1-0.519} = 2.628$$

- Table 1 shows ML and LS estimates for the ARCH(1) models for CBA and NWS.

Table 1: ML and LS estimates from ARCH(1) model applied to CBA and NWS returns

	CBA			NWS		
	LS	ML	SE	LS	ML	SE
μ	0.026	0.058	0.018	-0.003	-0.033	0.033
α_0	1.555	1.244	0.025	4.017	3.305	0.033
α_1	0.377	0.476	0.018	0.123	0.341	0.014
$\frac{\alpha_0}{1-\alpha_1}$	2.498	2.628		4.579	5.012	

- Note that consistent with our a priori expectation the estimated average volatility is (much) larger for NWS.

- But the estimated ARCH parameter for NWS is somewhat smaller than that for CBA. What does that imply?

- Figure 7 provides summaries of the ARCH(1) estimation and output for CBA.

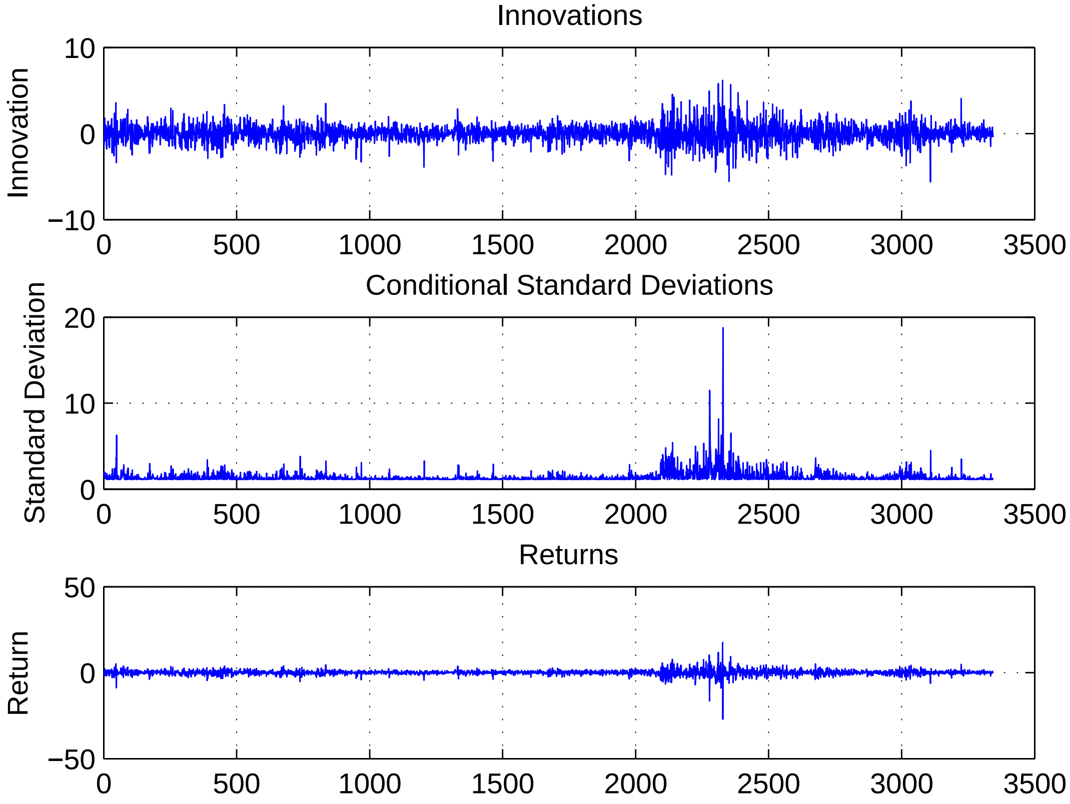

Figure 7: A summary of the ARCH(1) results for CBA.

- Figure 8 provides summaries of the ARCH(1) estimation and output for NWS.

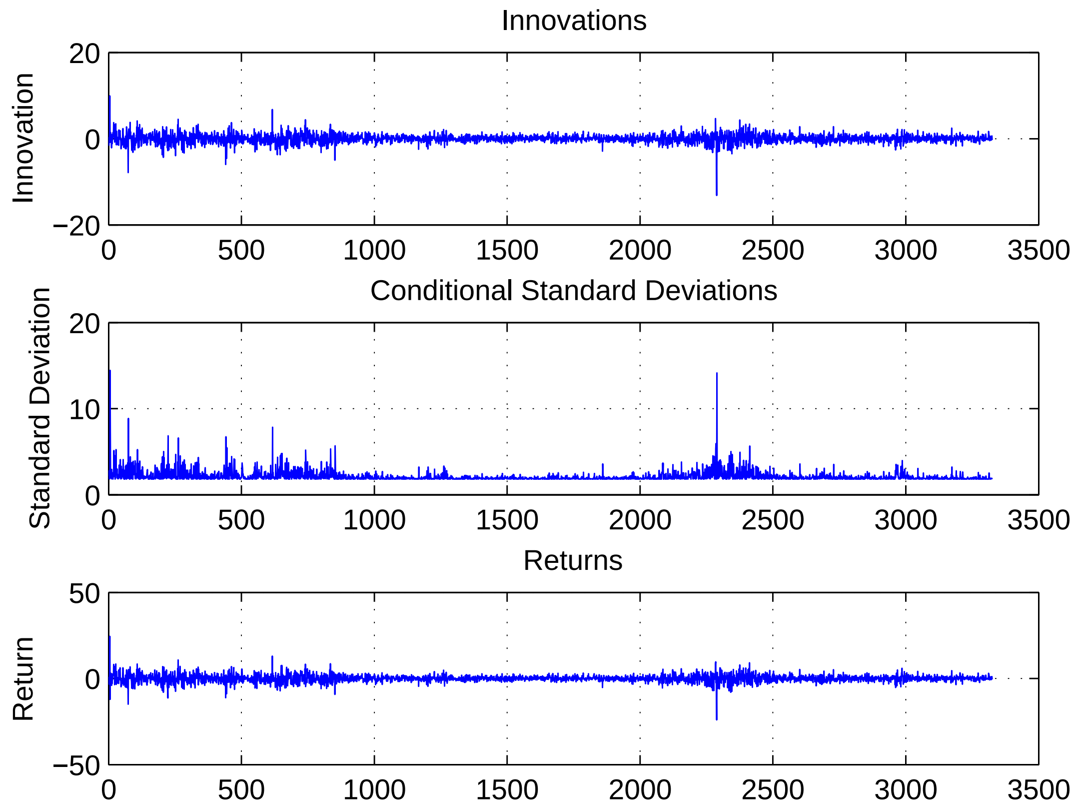

Figure 8: A summary of the ARCH(1) results for NWS.

- The summary statistics for these asset returns are:

Asset	mean	variance	std.	skewness	kurtosis
CBA	0.0262	2.498	1.581	-1.306	40.64
NWS	-0.0029	4.580	2.140	0.010	16.02

- Discussion.

ARCH Estimation (ctd)

- We now develop an MLE estimator for the case ARCH(p), $p > 1$, with Gaussian errors.

- Assuming Gaussian errors leads to the name quasi-ML estimator.

- $$\begin{aligned} r_t &= \mu_t + a_t = \mu_t + \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \ldots + \alpha_p a_{t-p}^2 \end{aligned}$$

  where $\sigma_t^2 \equiv \mathrm{Var}(a_t|\mathcal{F}_{t-1})$.

- Note that the parameter space is $\theta = \{\alpha, \mu\}$, where
  $$\alpha = (\alpha_0, \ldots, \alpha_p)$$

- Let $\mathbf{r} = (r_{p+1}, \ldots, r_T)$, then the conditional likelihood function is:
  $$\begin{aligned} p(\mathbf{r}|r_1, \ldots, r_p, \theta) &= \prod_{t=p+1}^{T} p(r_t|\mathcal{F}_{t-1}, \theta) \\ &= \prod_{t=p+1}^{T} p(a_t|\mathcal{F}_{t-1}, \theta) \end{aligned}$$

- Now,

$$a_t|\mathcal{F}_{t-1}, \theta \sim N(0, \sigma_t^2)$$

so that the *conditional* log-likelihood is

$$\begin{aligned} l_c(\theta) &= \ln[p(\mathbf{r}|r_1, \ldots, r_p, \theta)] \\ &= -\frac{(T-p)}{2} \ln(2\pi) \\ &- \frac{1}{2} \sum_{t=p+1}^{T} \left(\ln(\sigma_t^2) + (r_t - \mu)^2/\sigma_t^2 \right) \end{aligned}$$

- Matlab sets each of $r_{1-p}, \ldots, r_0$ to values, either:

  1. $\bar{r}$, the sample mean; OR
  2. any user-supplied values.

- If $r_{1-p}, \ldots, r_0$ are all known, then the distributions

$$a_1 | r_{1-p}, \ldots, r_0, \mu, \theta \sim N\left(0, \alpha_0 + \sum_{i=1}^{p} \alpha_i (r_{1-i} - \mu)^2\right)$$

, . . .,

$$a_p | r_0, \ldots, r_{p-1}, \mu, \theta \sim N\left(0, \alpha_0 + \sum_{i=1}^{p} \alpha_i (r_{i-1} - \mu)^2\right)$$

are also all known and employed in the likelihood function.

- Note that Matlab thus uses:

$$p(r_1, \ldots, r_T | \theta, r_{1-p}, \ldots, r_0) = \prod_{t=1}^{T} p(a_t | a_{t-1}, \ldots, a_{t-p}, \mu, \theta)$$

ARCH model and order selection

- When is an ARCH model useful? needed?
- Tests for zero auto-correlations in a time series could be useful here.
- The ACF for $\{a_t^2\}$ can help identify the existence of 'ARCH effects'.
- Significant auto-correlations in $\{a_t^2\}$ indicate that an ARCH model may be useful.
- We can again use the Ljung-Box test to examine the hypotheses

  H_0: $\rho_1 = \rho_2 = \ldots = \rho_m = 0$

  H_a: $\rho_i \neq 0$ for some $i \in \{1, 2, \ldots, m\}$

- Their $Q(m)$ statistic is

$$Q(m) = T(T+2)\sum_{l=1}^{m}\frac{\hat{\rho}_l^2}{T-l}$$

- $Q(m)$ is also asymptotically distributed χ^2 with m degrees of freedom.

- When testing **residuals (not observations)** for autocorrelation, the degrees of freedom of the test must be adjusted to account for parameter estimation.

- This is done by subtracting (from m) the number of terms in the mean and variance equations, **not including those for a constant mean and constant variance**.

- For example, when fitting an ARCH(p) model to data and using the Q(m) statistic above on the residuals, the correct degrees of freedom would be $m - p$.

ARCH Order Selection

- **Information criteria** (IC) are often used to choose p in an ARCH(p) model.
- The most popular are Akaike's (AIC) and Schwarz's IC (SIC or BIC).
- These methods plug MLEs into the log-likelihood function, and then penalise for the number of parameters (here p):

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

The model order p that minimises AIC and/or SIC is chosen.

Examples

- A Ljung-Box test for ARCH effects is run on CBA returns
- $m = 5, 8$ and $m = 15$ are chosen. Why?
- Hypotheses are:
- Q statistics and test results are:

m	Q	Critical value	p-val
5	442.47	11.07	0
8	512.95	15.51	0
15	520.88	18.31	0

- The consistent conclusion is that highly significant ARCH effects exist in the CBA return series.

- Figure 9 shows the ACF for the CBA and NWS squared returns.

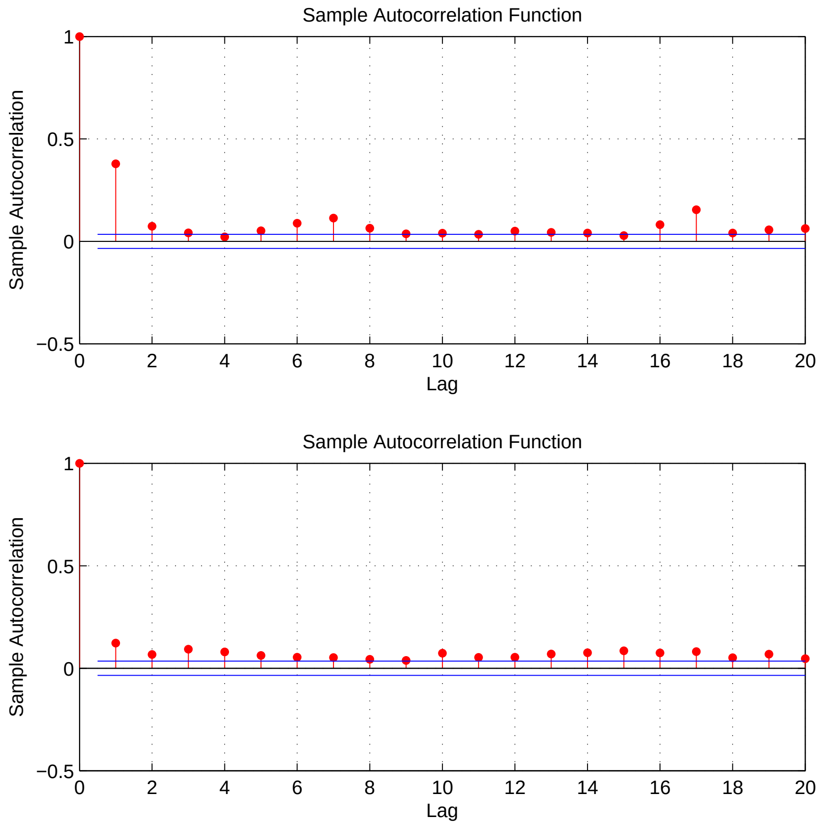

Figure 9: ACF for squared returns for CBA and NWS.

- LB test p-values for NWS squared returns are also all 0.

- The following table shows AIC and SIC figures for ARCH(p) models over $p = 1, 2, \ldots, 19$

p	1	2	3	4	5	6	7	8	9	10	11	12	19	20
AIC	11489	11249	11076	10980	10936	10886	10842	10823	10812	10813	10815	10816	**10749**	10750
SIC	11495	11261	11095	11004	10967	10922	10885	10872	10867	10875	10883	10889	**10865**	10873

- Figure 10 plots these numbers;

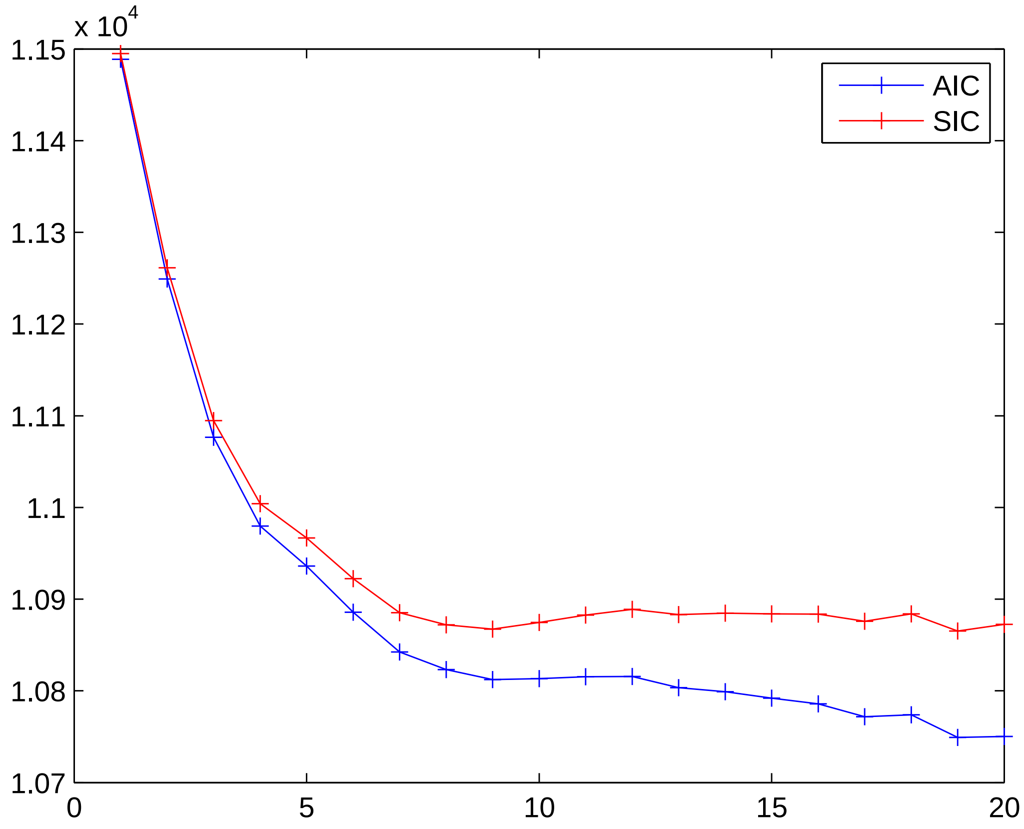

Figure 10: AIC and SIC for ARCH(p) model for CBA data.

- whilst figure 11 shows the same for NWS.

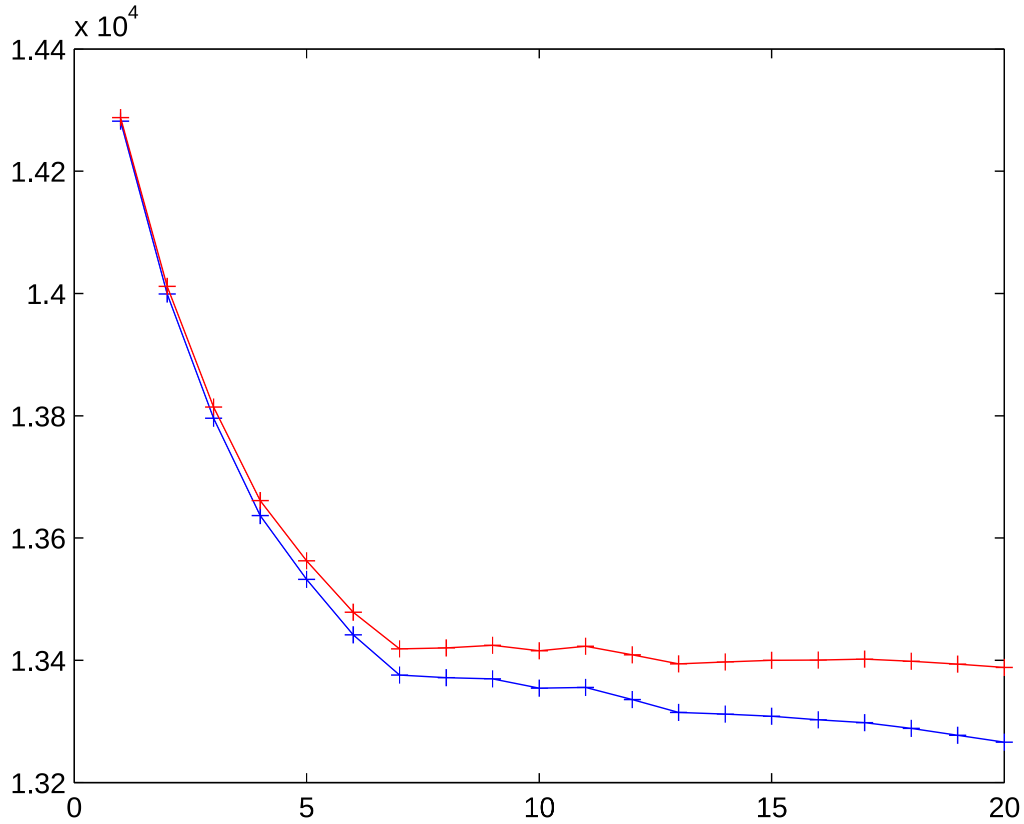

Figure 11: AIC and SIC for ARCH(p) model for NWS data.

- For CBA, the SIC suggests an ARCH(19), while the AIC suggests an ARCH(19) model, is appropriate.

- Since the SIC for q=9 and q=19 are so close, I choose ARCH(9).

- The estimated ARCH(9) model is:

$$\begin{aligned} r_t &= 0.070 + a_t \\ \sigma_t^2 &= 0.388 + 0.156a_{t-1}^2 + 0.105a_{t-2}^2 + 0.151a_{t-3}^2 \\ &\quad +0.080a_{t-4}^2 + 0.043a_{t-5}^2 + 0.078a_{t-6}^2 + 0.112a_{t-7}^2 \\ &\quad +0.064a_{t-8}^2 + 0.054a_{t-9}^2 \end{aligned}$$

- All ARCH effects are significant, having t-stats above 2 (lowest is 5th lag at 3.5)

- Stationarity requires $\sum_{i=1}^{9} \alpha_i < 1$. Here, for CBA, the sum is 0.84.

- Figure 12 provides a summary of the ARCH(9) estimation output

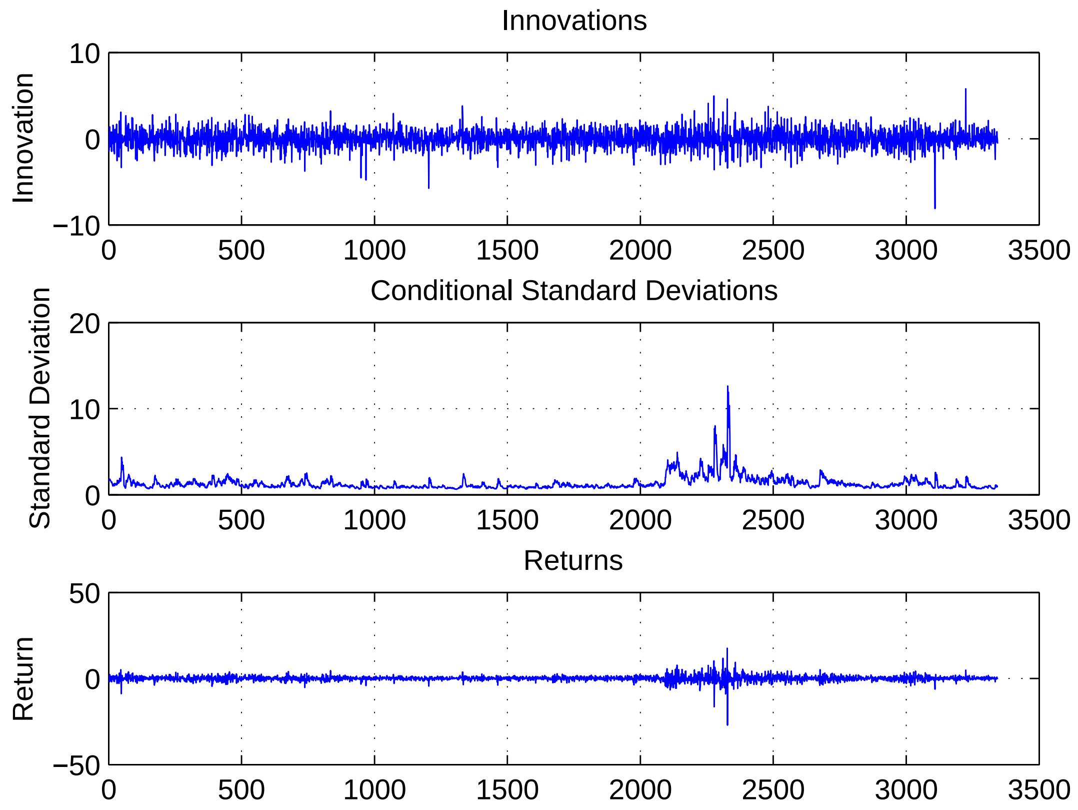

Figure 12: A summary of the results for CBA from an ARCH(9) model.

- and Figure 13 shows both the ARCH(1) and the ARCH(9) volatility estimates.

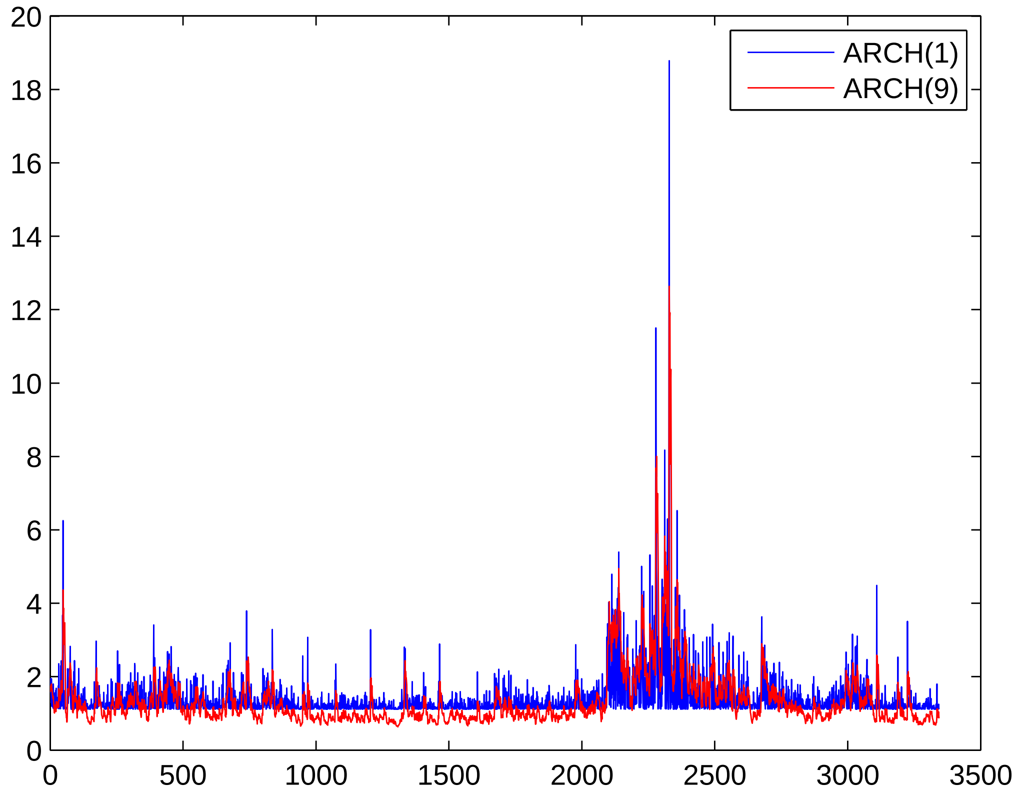

Figure 13: Volatility estimates for CBA from an ARCH(1) and and ARCH(9) model.

- What is noticeable about the estimated volatility series compared to that for the ARCH(1)?

- Table 2 shows ML estimation output for the ARCH(9) model for CBA.

Table 2: ML estimates from ARCH(9) model applied to CBA returns

	MLE	SE	t-stat
μ	0.070	0.018	3.966
α_0	0.388	0.022	17.526
α_1	0.156	0.016	9.584
α_2	0.105	0.019	5.613
α_3	0.151	0.020	7.481
α_4	0.080	0.016	4.969
α_5	0.043	0.012	3.497
α_6	0.078	0.015	5.231
α_7	0.112	0.021	5.493
α_8	0.064	0.016	4.007
α_9	0.054	0.015	3.679
$\frac{\alpha_0}{1-\sum_{i=1}^{9}\alpha_i}$	2.460		

Model checking and diagnostics

- What properties should the residuals from an ARCH model have?

- The ML estimation method assumed that standardised shocks

$$\epsilon_t = a_t/\sigma_t$$

were iid random variables, following a $N(0,1)$ distribution.

- Thus the estimated standardised shocks

$$\hat{\epsilon}_t = a_t/\hat{\sigma}_t$$

should be iid random variables, following a $N(0,1)$ distribution, **iff the model fits the data**.

- The standardised shock series can be estimated using the fitted volatilities

$$\hat{\sigma}_t = \hat{\alpha}_0 + \sum_{j=1}^{p} \hat{\alpha}_j a_{t-j}^2$$

so that

$$\hat{\epsilon}_t = \frac{(r_t - \hat{\mu})}{\hat{\sigma}_t}$$

- The Ljung-Box Q statistic on these standardised shocks $\hat{\epsilon}_t$ can be used to test the adequacy of the **mean** equation.

- If these residuals are iid then ...

- The Q statistic on the **squared** standardised shocks $\hat{\epsilon}_t^2$ can be used to test for the adequacy of the **volatility** equation.

- If these residuals have a constant variance (equal to 1), then ...

- The degrees of freedom for the LB test needs adjusting by the order of the ARCH(p) model (and any additional mean equation parameters above μ).

- e.g. To test the residuals from an ARCH(2) model using 10 lags of the ACF, the df of the LB statistic ($Q(10)$) should be 10-2 $= 8$.

- The Jarque-Bera test can be used to test for the normality of the residual series.

- Histograms and qq-plots can be used to visually assess this.

Example

- Figure 14 shows the standardised residuals, and their ACF, from the ARCH(1) model.

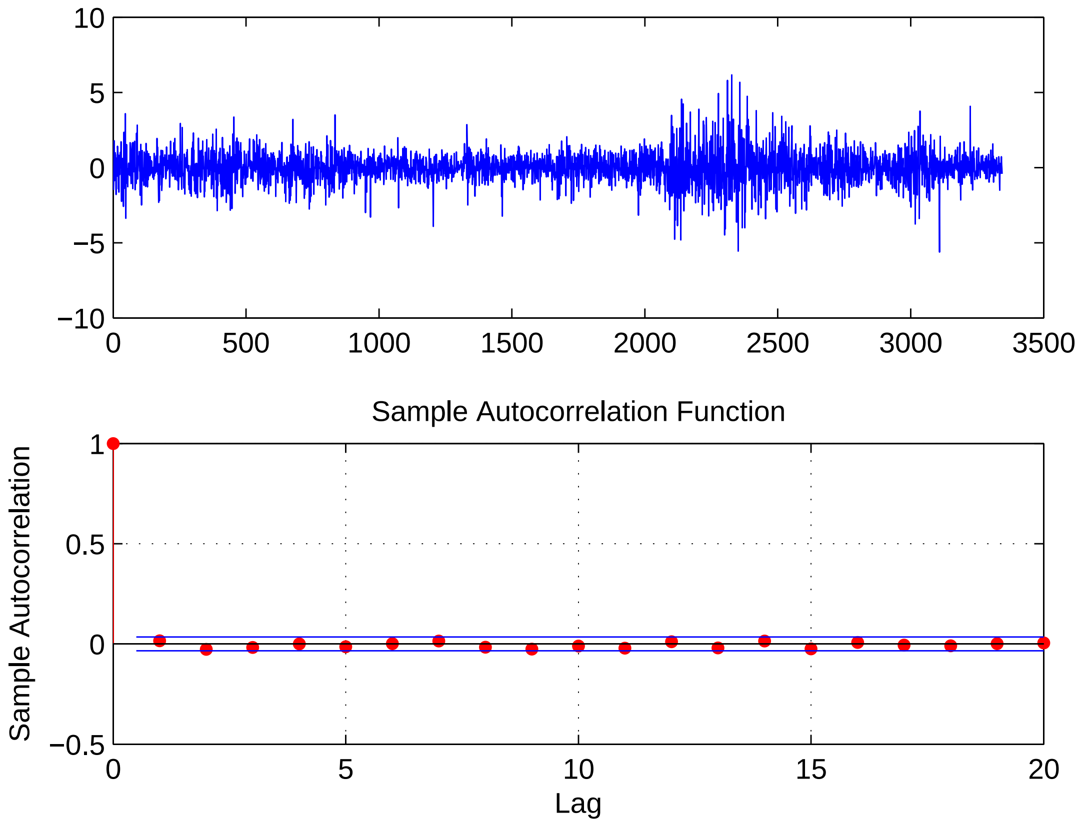

Figure 14: The standardised residuals for CBA from an ARCH(1) model.

- These show some moderate clustering and may show some possible outliers.

- Figure 15 displays the squared standardised residuals and their ACF. These clearly cluster and display positive autocorrelation.

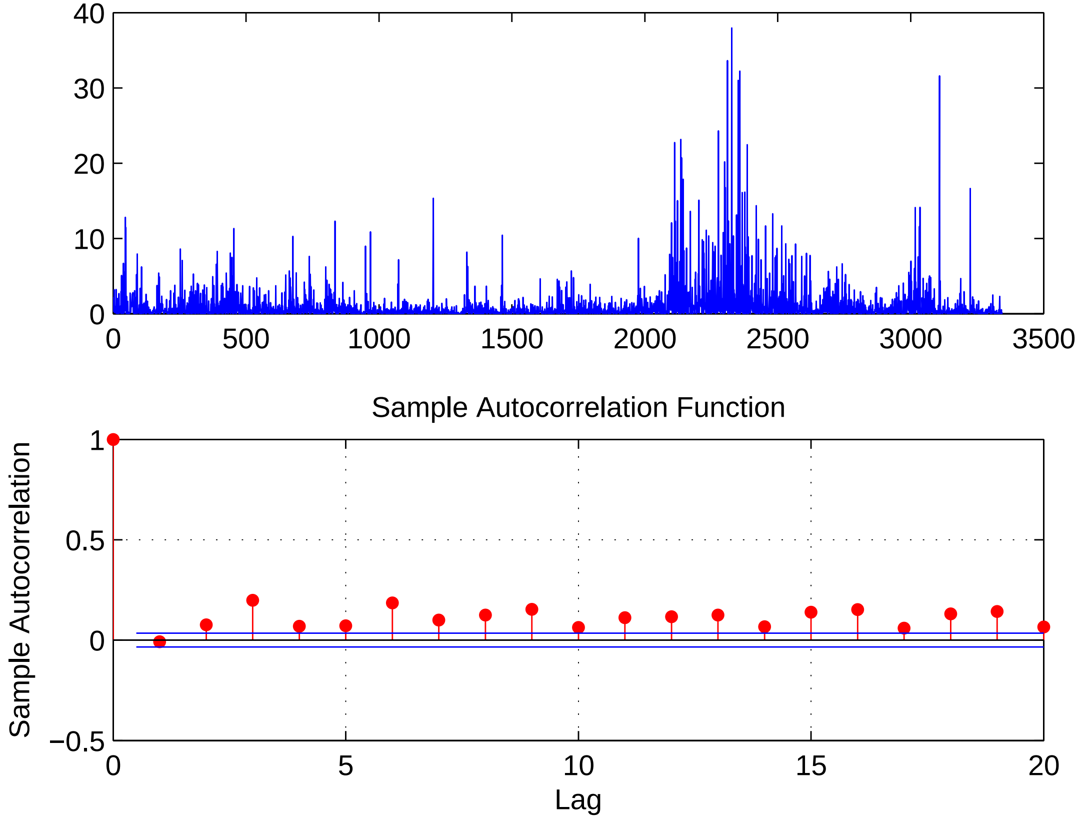

Figure 15: The squared standardised residuals for CBA from an ARCH(1) model.

- Ljung-Box statistics, using $m = 6$ and $m = 11$ (why?), show p-values of $0.39, 0.37$ for the residuals and p-values of $0, 0$ for the squared residuals.

- These suggest significant remaining significant ARCH effects in the residuals and suggest the volatility equation is inadequate for this data.

- The residuals do not display significant auto-correlation indicating the mean equation is reasonably well-specified.

- Further tests can be carried out on the standardised residuals to check the validity of the distributional assumption (i.e. standard normal). e.g. qq-plots, probability plots, etc.

- For example, see Figure 16

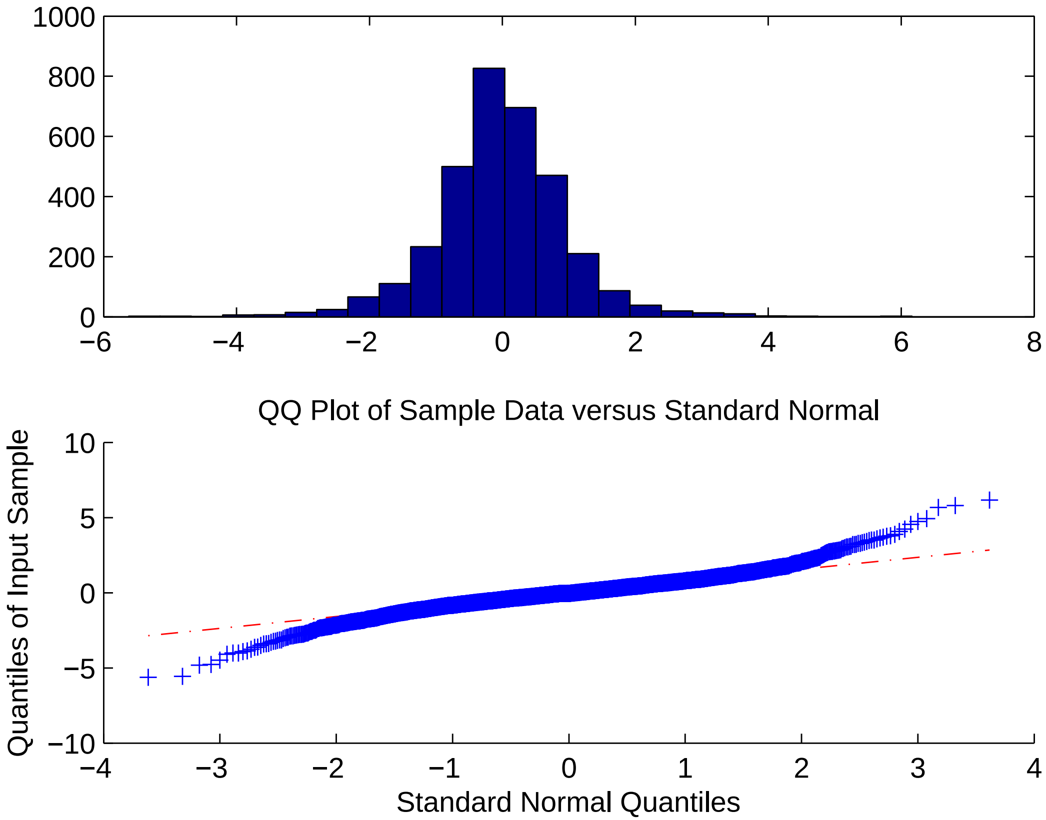

Figure 16: The standardised residuals for CBA from an ARCH(1) model: histogram and Gaussian QQ-plot.

- Clearly the standardised residuals have fatter tails than a Gaussian.

- Their skewness is 0.066 and kurtosis is 6.76.

- The Jarque-Bera test for skewness and excess kurtosis has a p-value < 0.001. Thus, the standardised residuals are not Gaussian.

- Figure 17 shows the standardised residuals, and their ACF, from the ARCH(9) model.

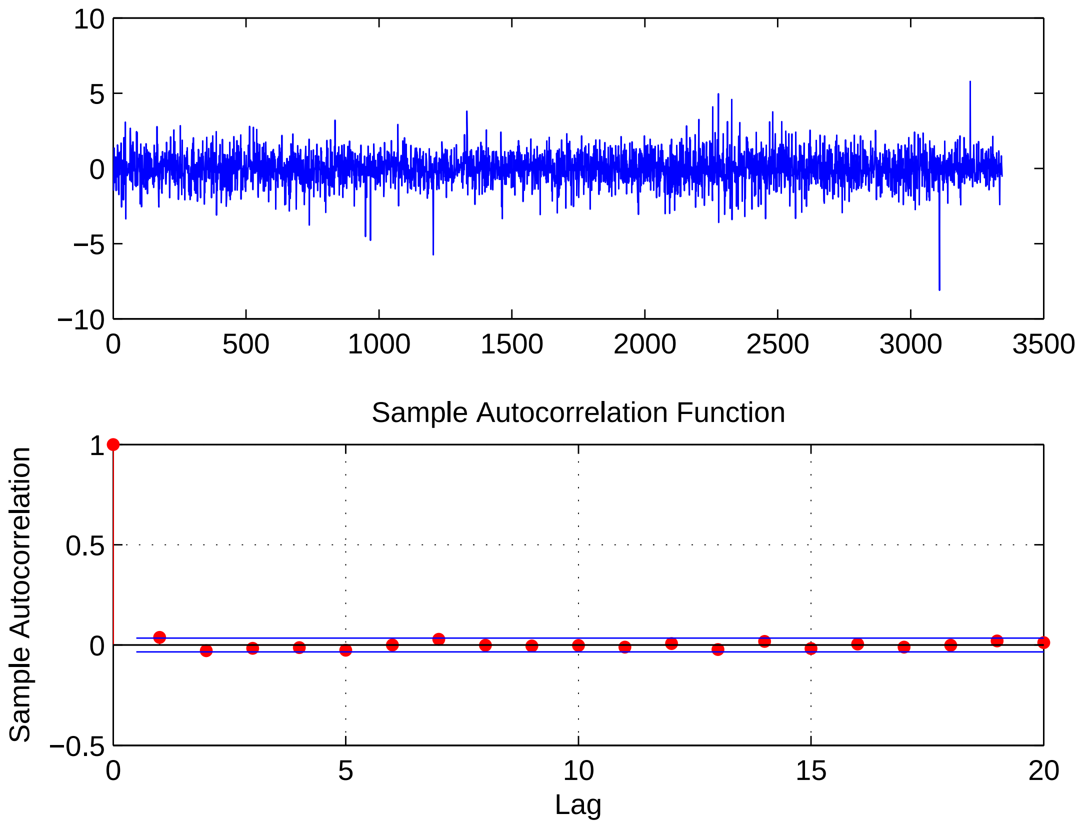

Figure 17: The standardised residuals for CBA from an ARCH(9) model.

- These do not appear to clearly cluster, but still show some possible outliers.

- Figure 18 displays the squared standardised residuals and their ACF. These do seem to cluster and show positive autocorrelation.

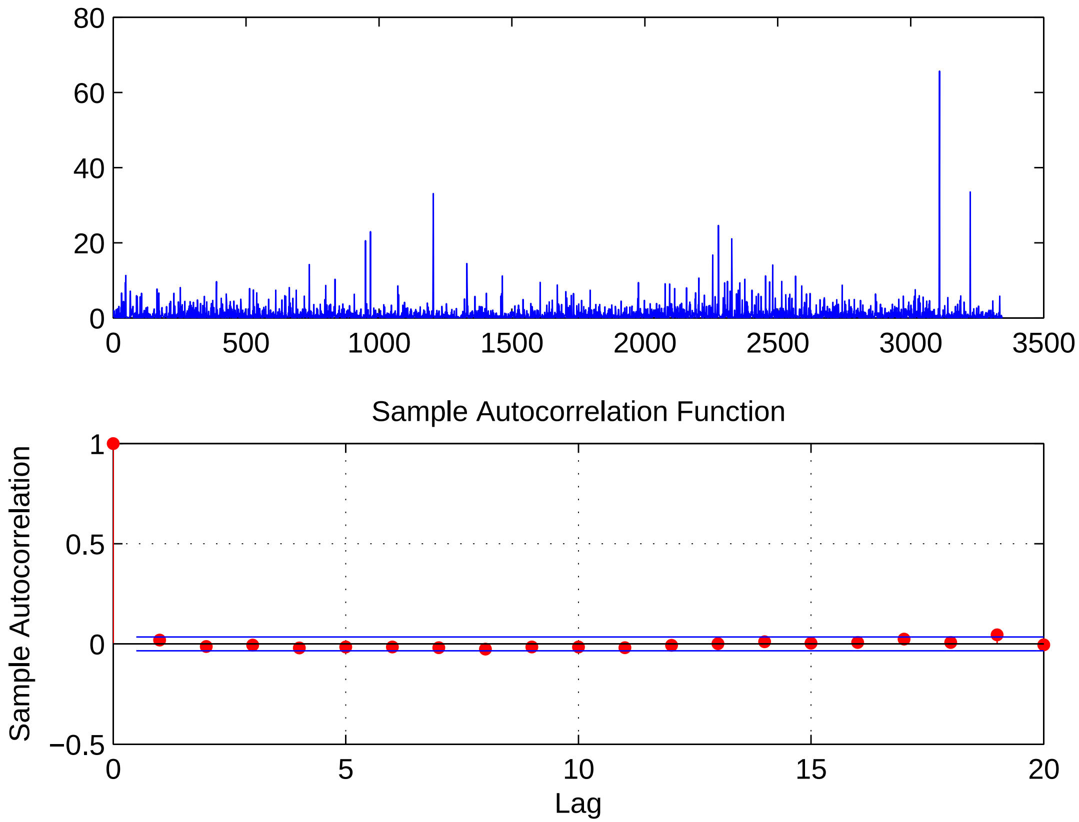

Figure 18: The squared standardised residuals for CBA from an ARCH(9) model.

- Ljung-Box statistics, using $m = 14$ and $m = 19$ (why?), show p-values of $0.004, 0.025$ for the residuals and $0.043, 0.022$ for the squared residuals.

- These suggest significant remaining significant ARCH effects in the residuals and suggest the ARCH(9) volatility equation is inadequate.

- The residuals now also display significant auto-correlation, indicating the constant mean equation is also not well-specified.

- Figure 19 shows the histogram and qq-plot for the standardised residuals.

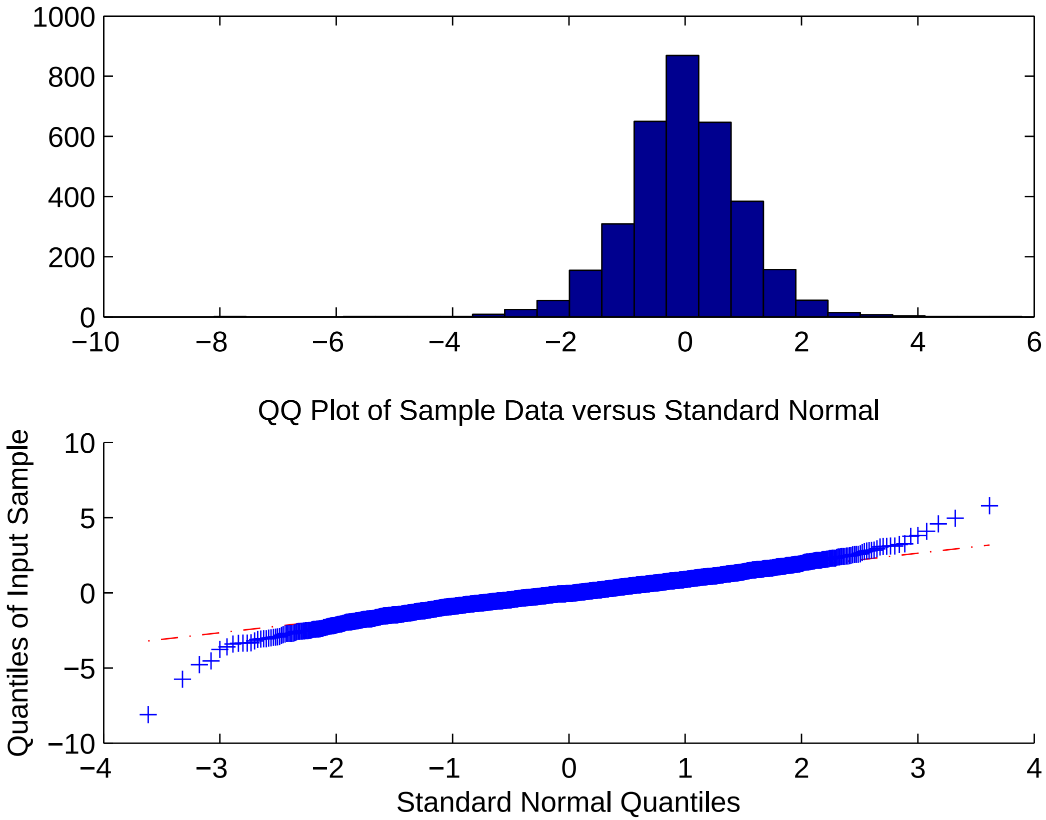

Figure 19: The standardised residuals for CBA from an ARCH(9) model: histogram and Gaussian QQ-plot.

- Clearly the standardised residuals still have fatter tails than a Gaussian.

- Their skewness is -0.20 and kurtosis is 5.82.

- The ARCH(9) has only slightly improved the fit to the CBA from the ARCH(1).