Module 2: Introduction to Forecasting and Time Series

References:

- *Chapter 2 in Tsay*
- *Chapter 5 in Brooks*
- *Chapter 4 in McNeil Frey and Embrechts*

Section 1: Forecasting

(1.1) Introduction to Forecasting in Finance

- What is forecasting?
- Why do we do it?
- Why do financial institutions need to forecast?
- How do financial institutions traditionally forecast?

- Investment decisions rely on forecasts of what will happen during the period of investment.
- (Black-Scholes) Option prices are expectations of future pay-offs under a Gaussian distribution for returns.
- Asset prices themselves are meant to represent the expectation of future (e.g. earnings) performance of a company.
- Modern financial risk management specifically uses forecasts of future risk to ensure companies stay viable following extreme adverse events
- There are many ways to forecast in general.
- These can be broken up into categories in a number of ways, including quantitative vs qualitative. There are other classifications (e.g. ...?)
- For quantitative forecasting, methods can be **naive, adhoc, informal** or **formal**.

- In this unit we focus mainly on formal quantitative forecasting.
- Quantitative forecasts typically use data and a rule to forecast into the future.
- However they may also simply use a mathematical model, e.g. Black-Scholes option pricing.
- *Naive* forecasts simply use the rule: the forecast is the most recent data point.
- This would work **best** if data followed a random walk, so that $E(y_{t+1}|y_t) = y_t$ (if our data is y_t).
- On the other end, *formal* forecasting means that a statistical model has been estimated from the data and the pattern in that model extended into the future, using probability rules.
- The most common way to forecast formally is to use conditional moments, e.g. $E(y_{t+1}|y_1, \ldots, y_t)$.
- This is the conditional expected value of the data series one-step-ahead, given all the data observed up to that time.

- There are many "shades of grey" between naive and formal forecasting that we will explore.
- An example of an *adhoc* method is simply estimating $E(y_{t+1}|y_1, \ldots, y_t)$ by $\bar{y}$, without using any model to suggest that choice.
- We need some definitions to make sense of forecasting in general here:
  1. Forecast **horizon**: the number of periods ahead that wish to be forecast.
  2. Forecast **origin**: the point up to which you have information and beyond which you will forecast.
- A one-step-ahead forecast has horizon 1. A k-step-ahead forecast has horizon k.
- In any forecasting approach, only data or information available up to and at the time of origin can be used to make a forecast.
- Sounds obvious, but it is important when testing our different forecast models and methods!

- Some notation I will employ:

  1. All the information available up to and including time t will be denoted by the symbol $\mathcal{F}_t$. t is the implied origin here.
  2. A horizon 1 (one-step ahead) forecast, i.e. of y_{t+1} from origin t, will be denoted

     $$\hat{y}_{t+1|t}$$

     .
  3. A formal statistical approach uses a model to set

     $$\hat{y}_{t+1|t} = \hat{E}(y_{t+1}|\mathcal{F}_t)$$

     .

- Under the Basel Capital Accord, financial institutions should forecast risk measures at horizons 1 and 10 periods ahead.

- Forecasting texts usually agree on the steps to making a forecast. Roughly, these are:

  1. Formulate the problem in detail and the purpose/goals
  2. Gather information, both background and relevant data.
  3. Manipulate, explore and clean data
  4. Decide on forecast method, model, etc
  5. Model building and estimation
  6. Prepare forecasts
  7. Evaluate and monitor forecasts

- For example, say we have a portfolio of assets. A related problem is choosing the weights so as to minimise volatility of portfolio return, subject to a minimum expected portfolio return requirement.

- These are perhaps better chosen to be forecasted volatility and expected return levels.

- Thus our problem is formulated as being able to accurately forecast volatility and expected returns from a portfolio over the period we will invest in.
- We can gather return historical data on the list of assets in our portfolio and build a model to forecast volatility and returns.
- Naturally we should check, as we proceed with our investing strategy, that it is working well enough (by some measure).

(1.2) Assessing forecast accuracy

- It is important to know and measure whether forecasts are accurate.

- How can we measure this accuracy?

- If we knew the *true* data points we were forecasting (i.e. y_{t+1}, y_{t+2}, etc.), we could then compare these to our forecasts.

- The comparison could be done using standard distance measures, such as:

  1. *Root Mean Square Error*

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{y}_{t+i|t} - y_{t+i})^2}$$

2. *Mean Absolute Deviation*

$$\text{MAD} = \frac{1}{m}\sum_{i=1}^{m} |\hat{y}_{t+i|t} - y_{t+i}|$$

3. *Mean Absolute Percentage error*

$$\text{MAPE} = 100 \times \frac{1}{m}\sum_{i=1}^{m} \frac{|\hat{y}_{t+i|t} - y_{t+i}|}{y_{t+i}}$$

- where m is the number of forecasts made.

- RMSE and MAD are in the same units as the data points y_t.

- RMSE is more sensitive to outliers than MAD. *why?*

- MAPE is a percentage, between 0 and 100%. Non-quant managers often prefer this type of measure. But it is problematic for returns. *why?*

- Other measures are possible and make sense. For example, we may wish to assess the returns and risk levels obtained from using our forecasts to make investment decisions.

- When we make our forecasts, we cannot assess their accuracy immediately. We must wait for the actual data we are forecasting to arrive before using these measures.

- However, it is standard practice for forecasters to do the following:

  1. Split the data into two sub-parts, by time.
  2. The first sub-sample, the oldest data, is called the "in-sample" or estimation sample or "learning" period.
  3. The 2nd sub-sample, which leads up to the period to be forecast, is called the "forecast sample" or "validation sample".
  4. Use the in-sample period to estimate the model.

5. Use the estimated model to generate forecasts of the forecast sample data (without using the forecast sample data in any way).
6. Compare the forecasts generated to the actual forecast sample data.
7. If performance or accuracy is acceptable, re-estimate the model using ALL the data.
8. Then generate forecasts beyond the end of the forecast sample of data (i.e. from now).

- Forecasts can be conducted at *fixed origin* with *increasing horizon*.

- i.e. generating $\hat{y}_{t+i|t}$ for $i = 1, \ldots, m$, which are $1, 2, \ldots, m$ step-ahead forecasts from origin t.

- This approach is often favoured in business, e.g. forecasting multiple horizon sales in the next 12 months.

- OR forecasts can be generated with *moving origin* at *fixed horizon*.

- e.g. generating $\hat{y}_{t+i+h|t+i}$ for $i = 0, \ldots, m-1$, which generates m, horizon h, forecasts.

- Forecasting for investment usually makes more sense with a fixed horizon, moving origin. Basel recommends $h = 1$ and/or $h = 10$ and $m = 250$ (for daily data).

- An issue here is that e.g. $\hat{y}_{t+1|t}$ employs different data to $\hat{y}_{t+2|t+1}$, etc.

- For a formal model approach there is a decision to be made about whether a new data point warrants re-estimation of the model.

- If not, then how often should models be re-estimated?

- A second issue is should an *expanding data window*, i.e. $y_1, \ldots, y_t$ for $\hat{y}_{t+1|t}$, then

$y_1, \ldots, y_{t+1}$ for $\hat{y}_{t+2|t+1}$ be used?

- Or a *fixed size data window* be used, i.e. $y_1, \ldots, y_t$ for $\hat{y}_{t+1|t}$, then $y_2, \ldots, y_{t+1}$ for $\hat{y}_{t+2|t+1}$?

(1.3) Example

- Consider the CBA daily returns on the ASX.

- Let's compare the naive forecast approach and two *adhoc* methods for generating forecasts of stock returns:

  1. Naive: $\hat{y}_{t+1|t} = y_t$
  2. Adhoc 1 (25 day average): $\hat{y}_{t+1|t} = \frac{1}{25}(y_{t-24} + \ldots + y_t)$
  3. Adhoc 2 (5 day average): $\hat{y}_{t+1|t} = \frac{1}{5}(y_{t-4} + \ldots + y_t)$

- Let's simply forecast the last 250 days of returns, at a daily forecast horizon (i.e. fixed horizon of $h = 1$, moving origin).

- That is, we'll generate 250 one-day-ahead forecasts of CBA's stock price, using the three methods above.

- Figure 1 shows the last 25 days of CBA returns in the in-sample period, together with the actual 1st return in the forecast period, and the three forecasts of it.

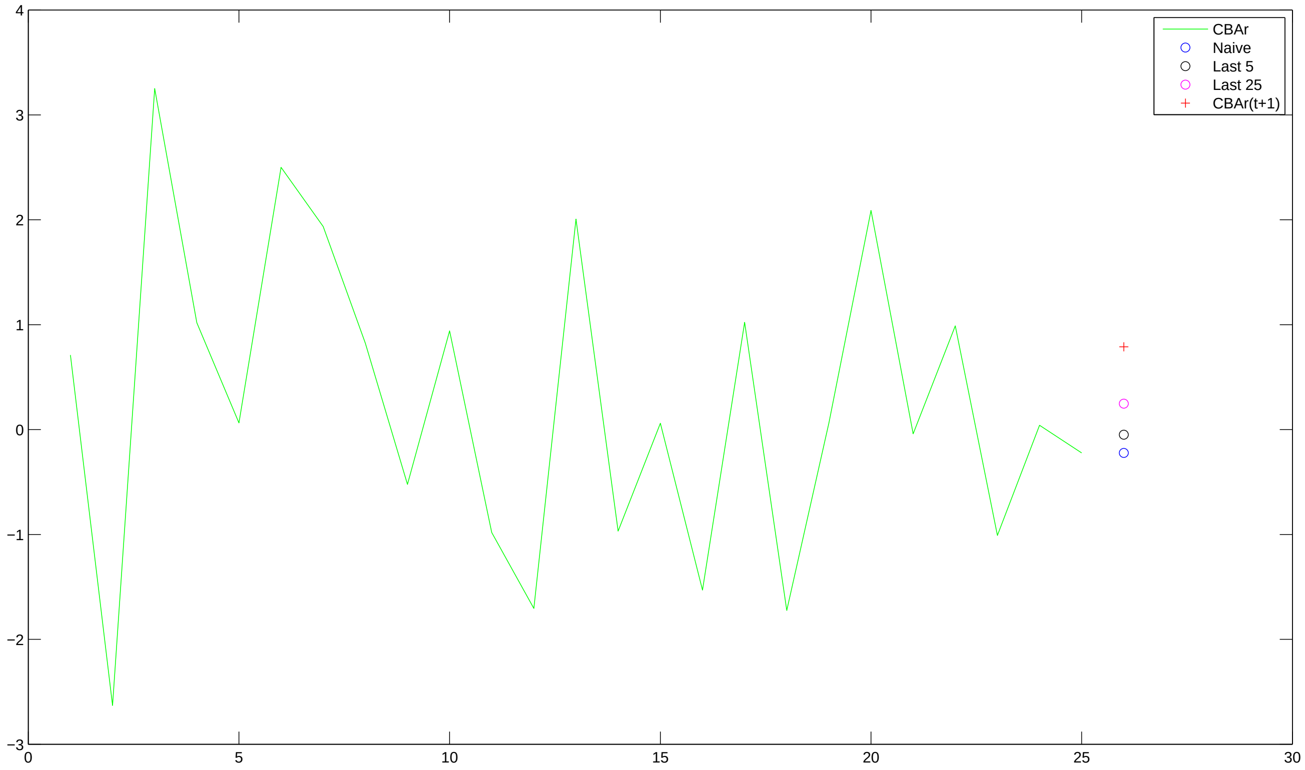

Figure 1: Log returns for CBA for last 25 days from Jan 21, 2013.

- We repeat this exercise for each of the last 250 days of returns.

- Figure 2 shows the last 25 days of CBA returns in the in-sample period, together with the actual 250 returns in the forecast period, and the three sets of forecasts of these.

- These are all horizon 1 forecasts, i.e. 1 step-ahead.

- Table 1 shows the forecast accuracy measures for these 250 days.

Table 1: Forecast accuracy measures for 100 days of CBA returns

	Naive	Adhoc 1	Adhoc 2
Measure	Last day	5 days	25 days
RMSE	1.23	1.01	0.92
MAD	0.86	0.71	0.64

- Clearly, the 25 day average is the most accurate return predictor.

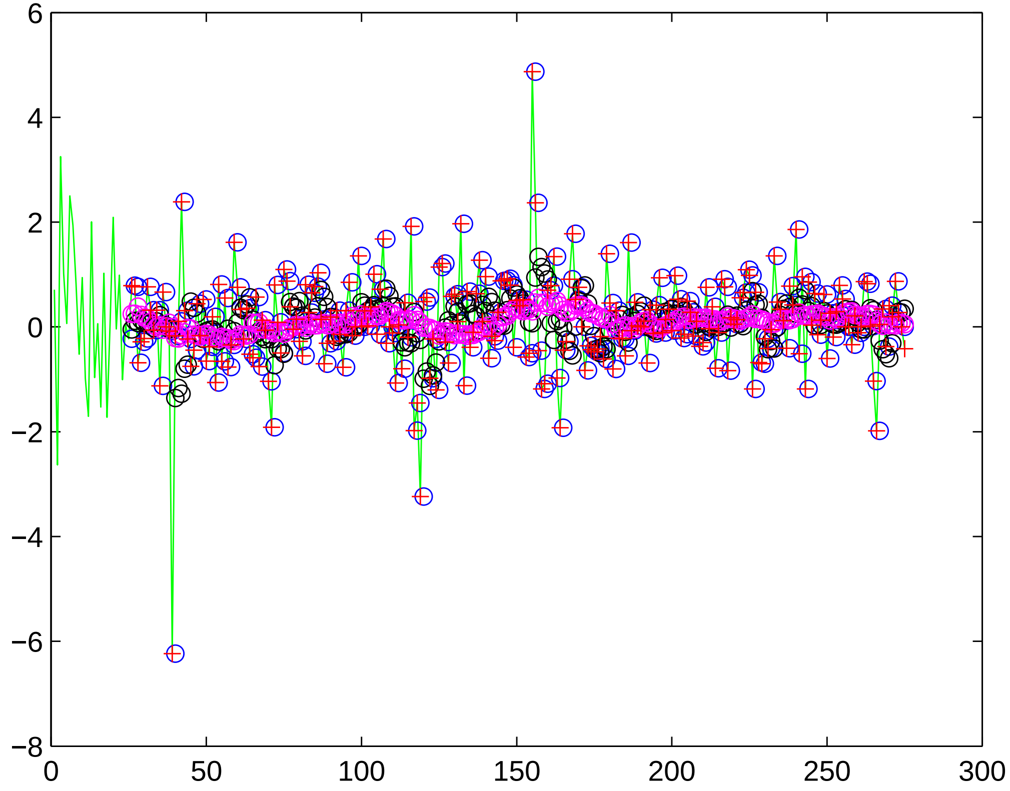

Figure 2: Log returns for CBA for last 250 days from Jan 21, 2013, plus three sets of 1 step-ahead forecasts

- Are you surprised?

- Why is MAPE not included here?

2. Introduction to time series analysis and financial price/return modelling

- Time series data is simply data that is recorded over time.

- *Stationarity* is one of the fundamental principles of time series modelling.

- A time series model is *strictly* stationary iff the joint distribution of $(y_t, y_{t+1}, \ldots, y_{t+k})$ is *independent* of the choice of t.

- Figure 3 shows daily prices (top) for CBA stock from January, 2000 to January, 2013.

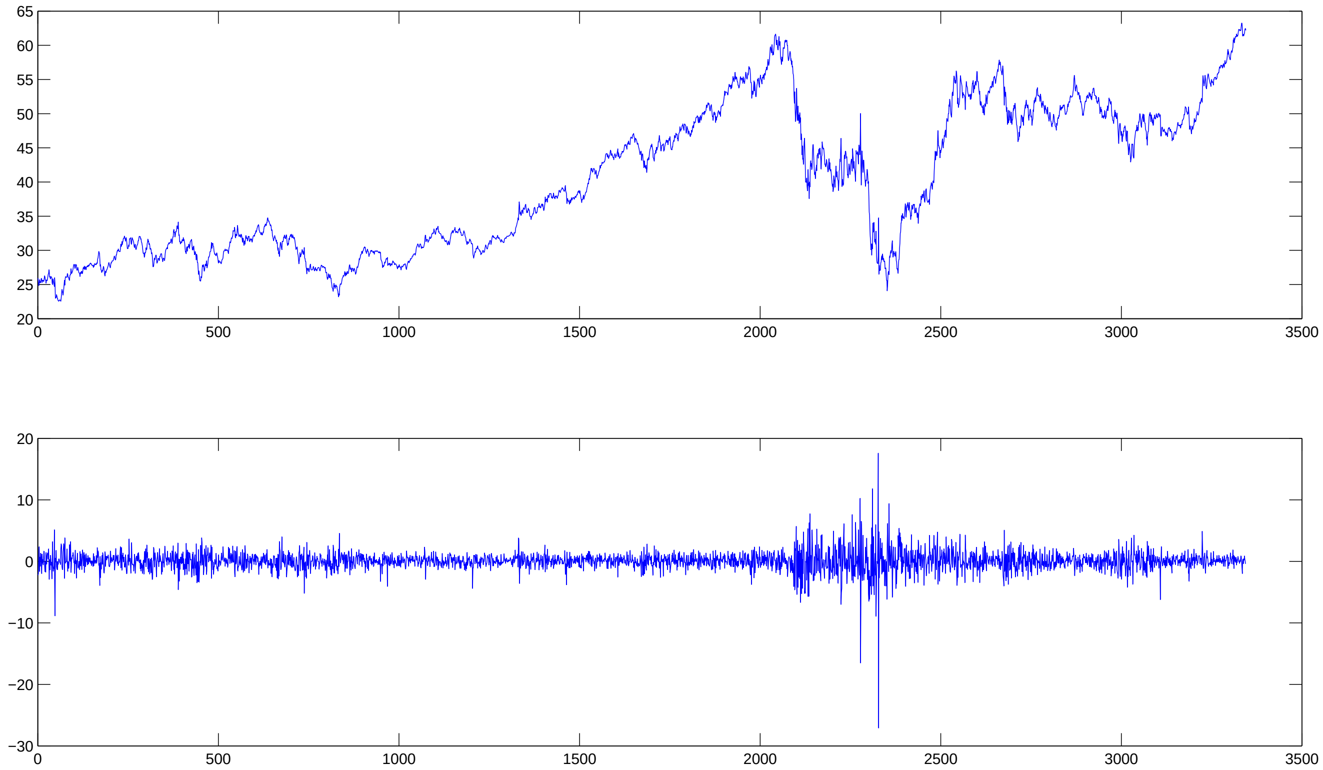

Figure 3: Prices and Log returns for CBA from Jan, 2000 to Jan, 2013.

- The bottom plot shows percentage log-returns for these prices.

- Are either of these series strictly stationary?
- A time series model is *weakly* stationary if the 1st and 2nd moment properties are constant and finite over time.
- i.e. if $E(y_t) = \mu$ and $Var(y_t) = \sigma^2$ are constant AND if

$$Cov(y_t, y_{t-k}) = \gamma_k \equiv Corr(y_t, y_{t-k}) = \rho_k$$

is constant over t, and all these terms are finite.

- Are either the price or log-return series for CBA weakly stationary??
- (Mean) stationarity for individual time series can be assessed via an autocorrelation (ACF) plot.
- The ACF plots the correlations between observations separated by a lag of k time

periods, with k on the x-axis., i.e.

$$\rho_k = corr(y_t, y_{t-k}) = \frac{Cov(y_t, y_{t-k})}{\sqrt{Var(y_t)Var(y_{t-k})}}$$

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

- A variance stabilising transform should be applied to the data first. *why?*

- A stationary in mean series has an ACF that *dies down* reasonably quickly.

- A series that is **not** mean stationary has an ACF that dies down extremely slowly.

- Figure 4 shows the CBA prices and their ACF plot.

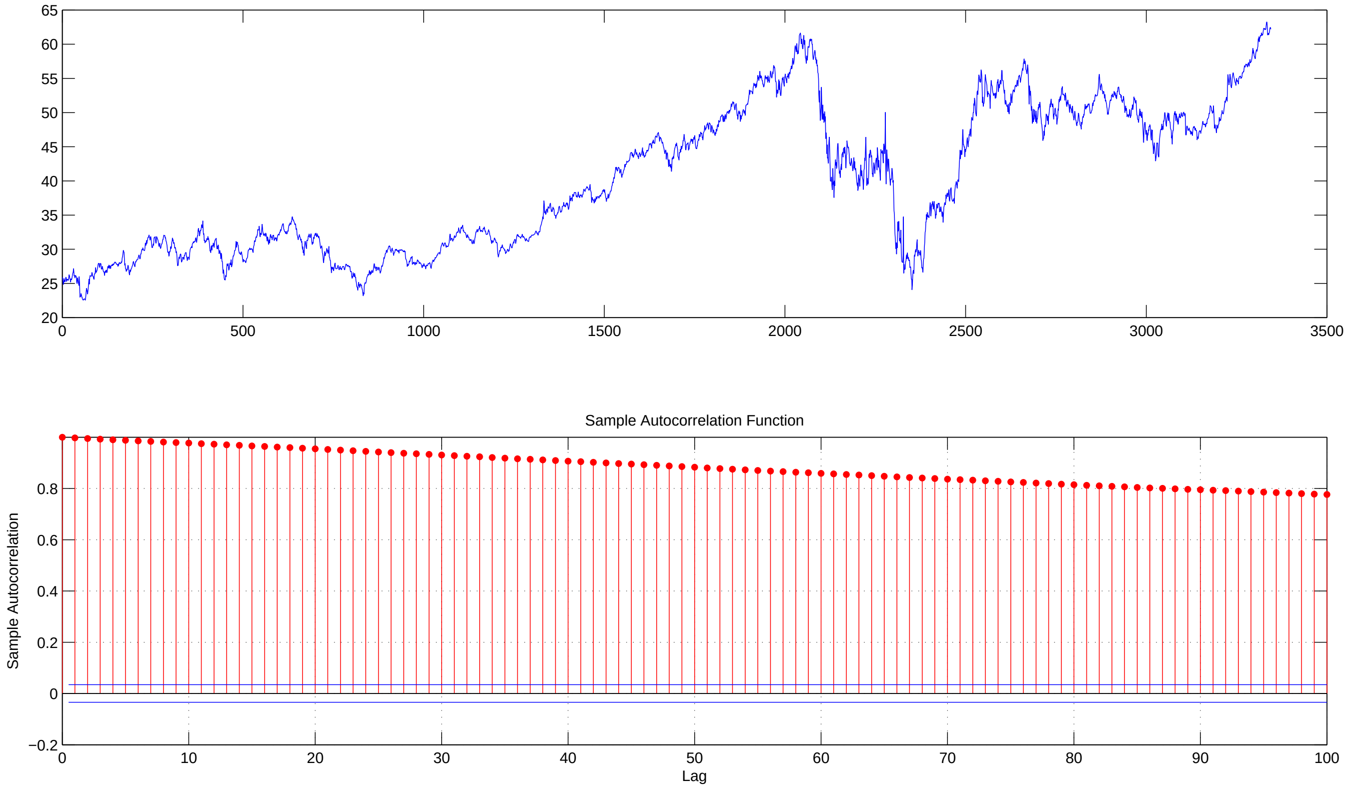

Figure 4: Prices and their ACF for CBA from Jan, 2000 to Jan, 2013.

- Figure 5 shows the CBA log-returns and their ACF plot.

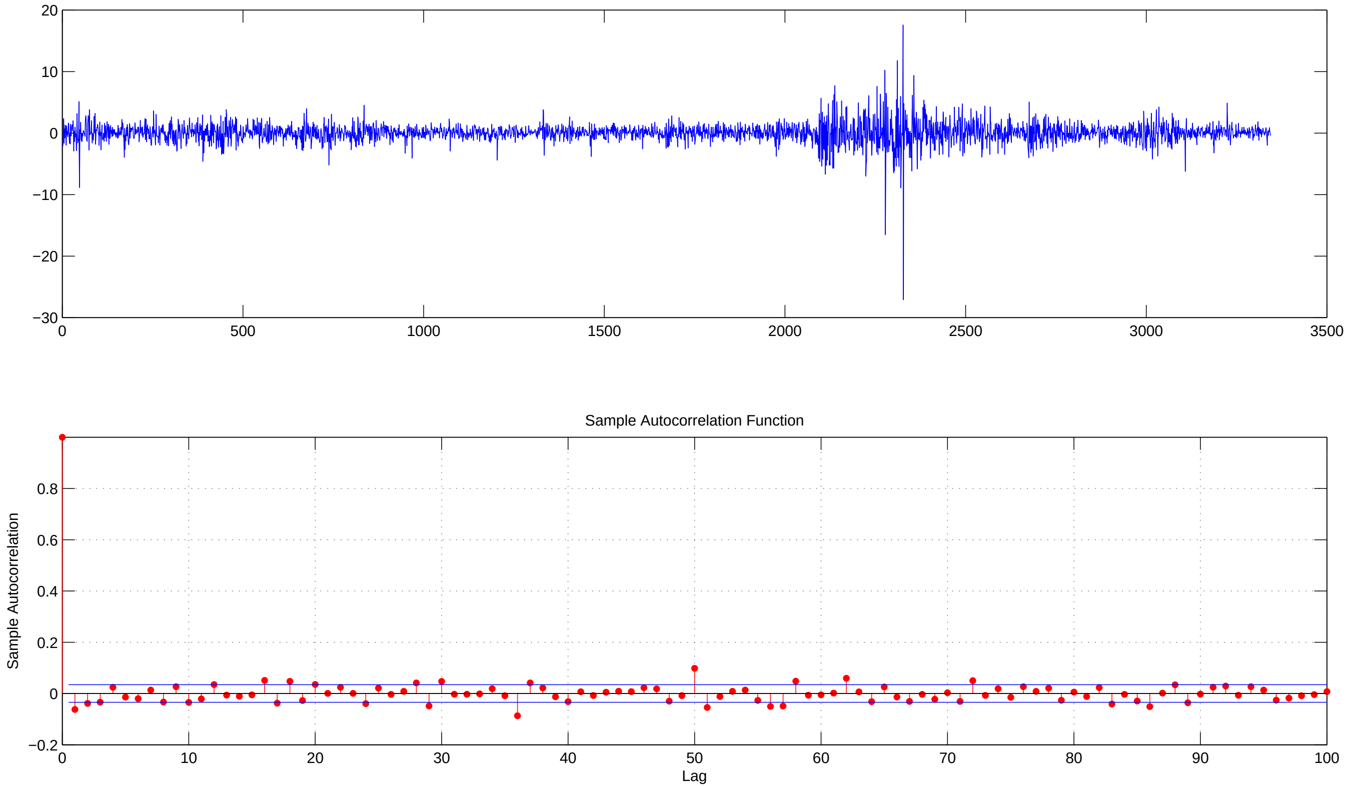

Figure 5: Log-returns and their ACF for CBA from Jan, 2000 to Jan, 2013.

- Comments?

- A *white noise* process consists of a series of i.i.d. observations with mean 0 and fixed variance.

- What would the ACF for a white noise process look like?

- The standard errors for sample autocorrelations are:

$$
\begin{aligned}
s_{\hat{\rho}_k} &= \frac{1}{\sqrt{n}} \ , \ k = 1 \\
&= \frac{1}{\sqrt{n}} \sqrt{1 + 2\sum_{i=1}^{k-1} \hat{\rho}_k^2} \ , \ k > 1.
\end{aligned}
$$

- However, Matlab assumes that all lower lag correlations are 0 and approximates with:

$$s_{\hat{\rho}_k} = \frac{1}{\sqrt{n}}$$

for all lags k.

- In large samples we can form a t-statistic, $\frac{\hat{\rho}_k}{s_{\hat{\rho}_k}}$ and test ...

(2.1) The Autoregressive model

- Recall the multiple linear regression model

$$y_t = \beta_0 + \beta_1 x_{1,t} + \ldots + \beta_p x_{p,t} + e_t \;;\; e_t \sim \text{i.i.d.}(0, \sigma^2).$$

- If $x_{i,t}$ is set as y_{t-i} then we have a time series model.

$$y_t = \phi_0 + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + e_t$$

which is called an autoregressive (AR) model of order p.

- We first focus on the AR(1) ($p = 1$) model:

$$y_t = \phi_0 + \phi_1 y_{t-1} + e_t$$

- If $\mu_t = \phi_0 + \phi_1 y_{t-1}$ then we can write:

$$E(y_t|y_{t-1}) = \phi_0 + \phi_1 y_{t-1}\,; \;\; Var(y_t|y_{t-1}) = \sigma^2$$

and

$$\begin{aligned} Cov(y_t, y_{t-1}) &= Cov(\phi_0 + \phi_1 y_{t-1} + e_t, y_{t-1}) \\ &= \phi_1 Var(y_t) \end{aligned}$$

Proof and assumptions?

- For this model we thus have: $\rho_1 = \phi_1$

- In fact: $\rho_k = \phi_1^k$, $k \geq 1$

- What happens to the ACF for an AR(1) process as k gets large?

- Figure 6 shows simulated AR(1) data and associated ACF plots for $\phi_1 = 0.7$

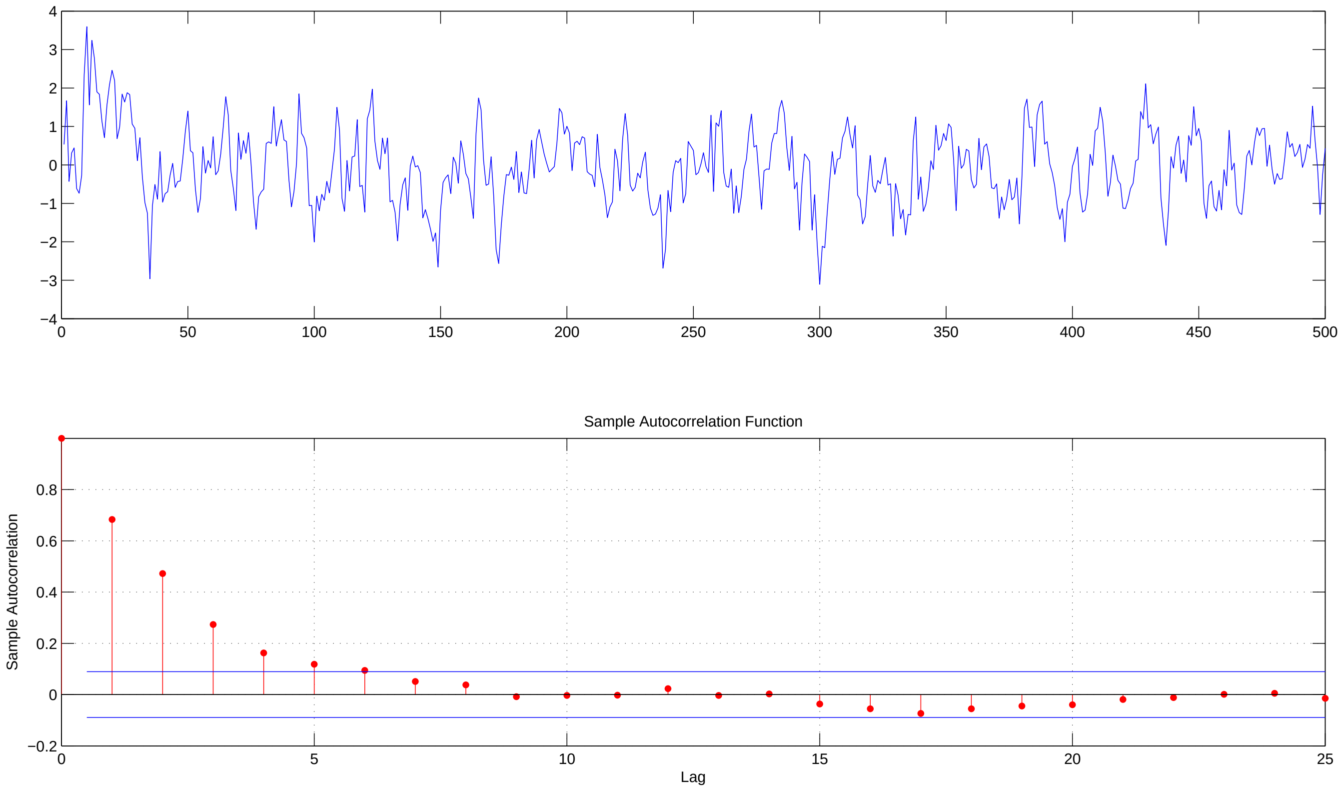

Figure 6: Simulated AR(1) data with $\phi_1 = 0.7$

- Figure 7 shows some simulated AR(1) data and associated ACF plots for $\phi_1 = 0.95$.

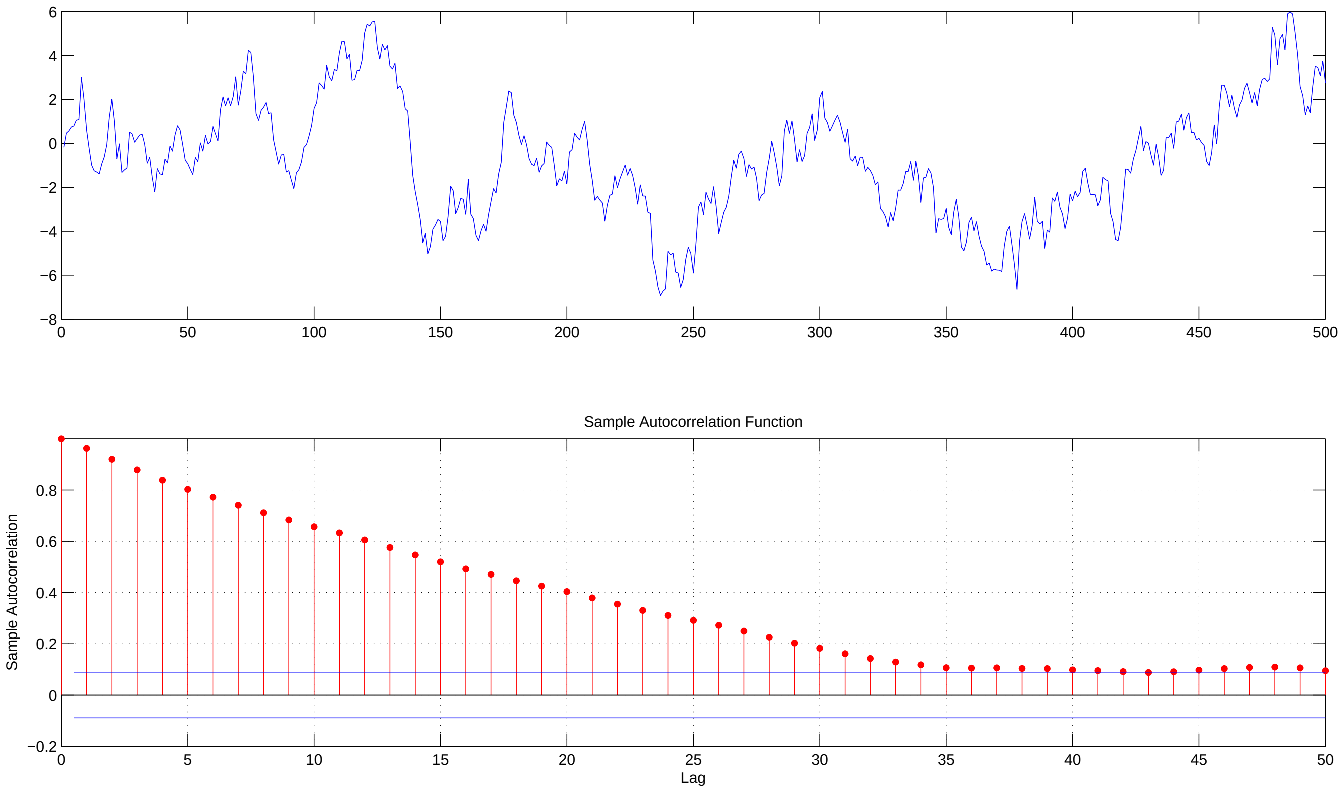

Figure 7: Simulated AR(1) data with $\phi_1 = 0.95$

- Figure 8 shows a data set with negative $\rho_1 = \phi_1 = -0.8$ for comparison.

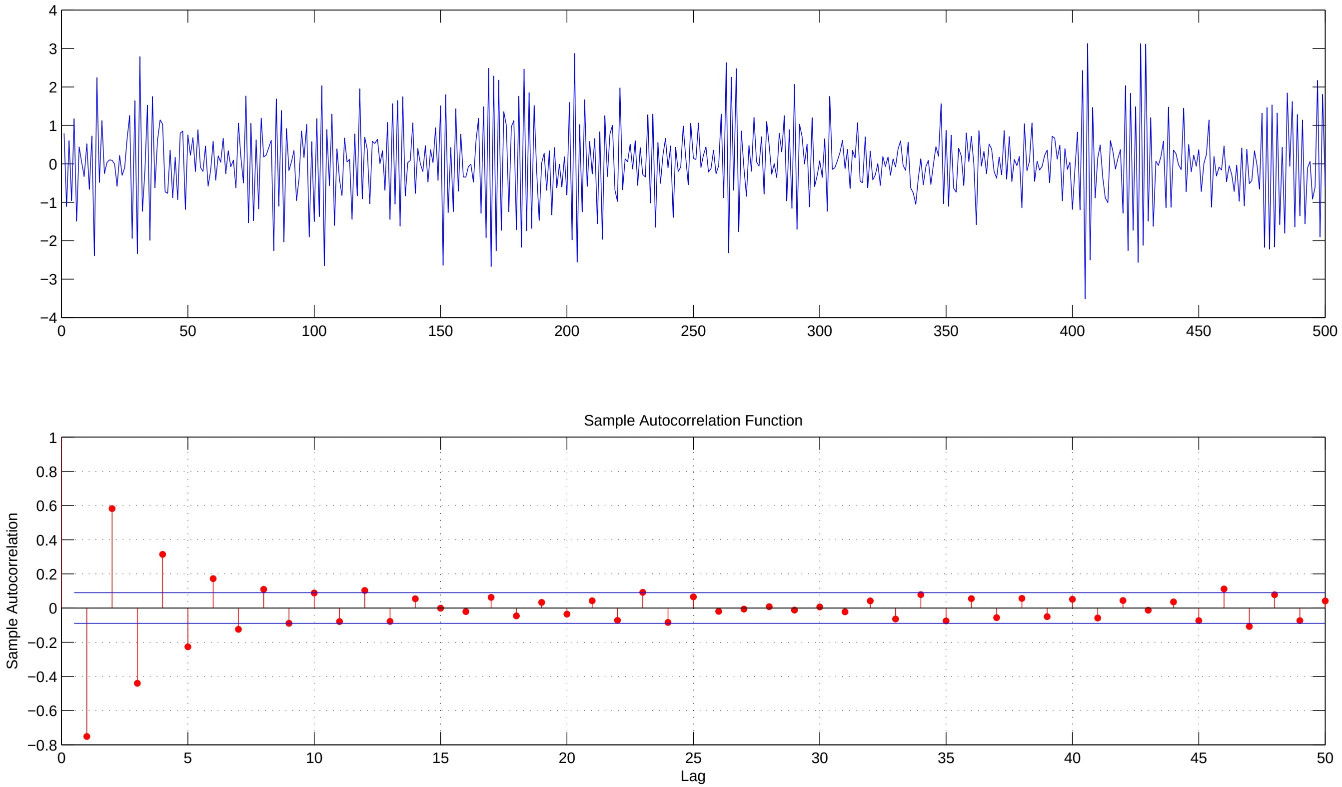

Figure 8: Simulated AR(1) data with $\phi_1 = -0.8$

(2.2) Estimation for the AR(1) model

- Some more AR(1) properties:

$$\begin{aligned} E(y_t) &= \phi_0 + \phi_1 E(y_{t-1}) \\ &= \frac{\phi_0}{1-\phi_1} \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}(y_t) &= \phi_1^2 Var(y_{t-1}) + Var(e_t) \\ &= \frac{\sigma^2}{1-\phi_1^2} \end{aligned}$$

Proof and assumptions

- Is an AR(1) model stationary? When?

- Is an AR(1) model weakly stationary? When?

- Usually, the condition $|\phi_1| < 1$ is enforced during estimation.

- **Least squares** estimation: The OLS estimator minimises:

$$\begin{aligned} \text{SSE} &= \sum_{t=1}^{n} e_t^2 \\ &= \sum_{t=2}^{n} (y_t - \phi_0 - \phi_1 y_{t-1})^2 + \tilde{e}_1^2 \end{aligned}$$

- subject to $|\phi_1| < 1$.

- What is $\tilde{e}_1$ here?

- If we ignore $t = 1$ this is called *conditional* least squares.

- We then have:

$$\begin{aligned}
\hat{\phi}_1 &= \frac{\sum_{t=2}^{n}(y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=2}^{n}(y_{t-1} - \bar{y})^2} \\
\hat{\phi}_0 &= \bar{y}(1 - \hat{\phi}_1) \\
\hat{\sigma}^2 &= \frac{1}{n-1}\sum_{t=2}^{n}(y_t - \hat{\phi}_0 - \hat{\phi}_1 y_{t-1})^2
\end{aligned}$$

- Is $-1 < \hat{\phi}_1 < 1$ here??

- If we include $t = 1$ then there is no closed form solution. The LS estimates can be obtained by numerically minimising the SSE.

- Again the restriction $-1 < \hat{\phi}_1 < 1$ is enforced.

- Under the 3 LS assumptions, there is a CLT for the LS estimate, *issues with Assumption 2?*

- which is also consistent

- However, as the true value of ϕ_1 moves closer to 1, there is some bias, so that $E(\hat{\phi}_1) < \phi_1$.

- The bias diminishes with sample size.

- Assumption 2 changes to the observations being a random sample from a mean stationary process

- Or the model errors being iid.

2.3 Forecasting for the AR(1) model

- $$\begin{aligned}\hat{y}_{t+1|t} &= E(y_{t+1}|\mathcal{F}_t) = E(\phi_0 + \phi_1 y_t + e_{t+1}|\mathcal{F}_t) \\ &= \phi_0 + \phi_1 y_t\end{aligned}$$

- This is a horizon 1 forecast.

- Under the 1st least squares assumption, we have $E(e_{t+1}|\mathcal{F}_t) = 0$

- For horizon $h = 2$:

$$\begin{aligned}\hat{y}_{t+2|t} &= E(y_{t+2}|\mathcal{F}_t) = \phi_0 + \phi_1 E(y_{t+1}|\mathcal{F}_t) + E(e_{t+2}|\mathcal{F}_t) \\ &= \phi_0 + \phi_1(\phi_0 + \phi_1 y_t) + E(e_{t+2}|\mathcal{F}_t) \\ &= \phi_0(1 + \phi_1) + \phi_1^2 y_t\end{aligned}$$

- We could show that the horizon k, or k-step-ahead forecast is:

$$\begin{aligned}\hat{y}_{t+k|t} &= E(y_{t+k}|\mathcal{F}_t)\\ &= \phi_0(1+\phi_1+\ldots+\phi_1^{k-1})+\phi_1^k y_t\end{aligned}$$

- As $k \to \infty$, $\hat{y}_{t+k|t} \to \frac{\phi_0}{1-\phi_1} = E(y_{t+k})$.

- Figures 9, 10 and 11 highlight the forecast behaviour for AR(1) models.

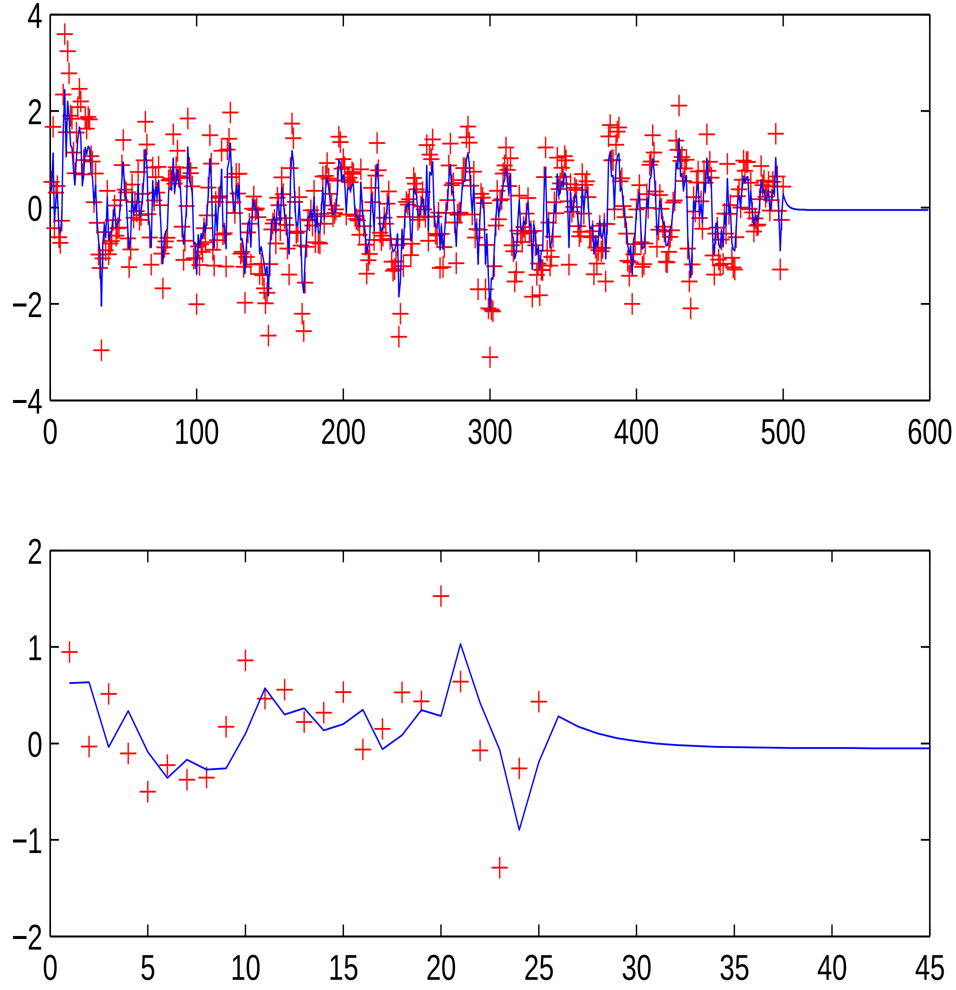

Figure 9: Simulated AR(1) data with $\phi_1 = 0.7$ plus forecasts

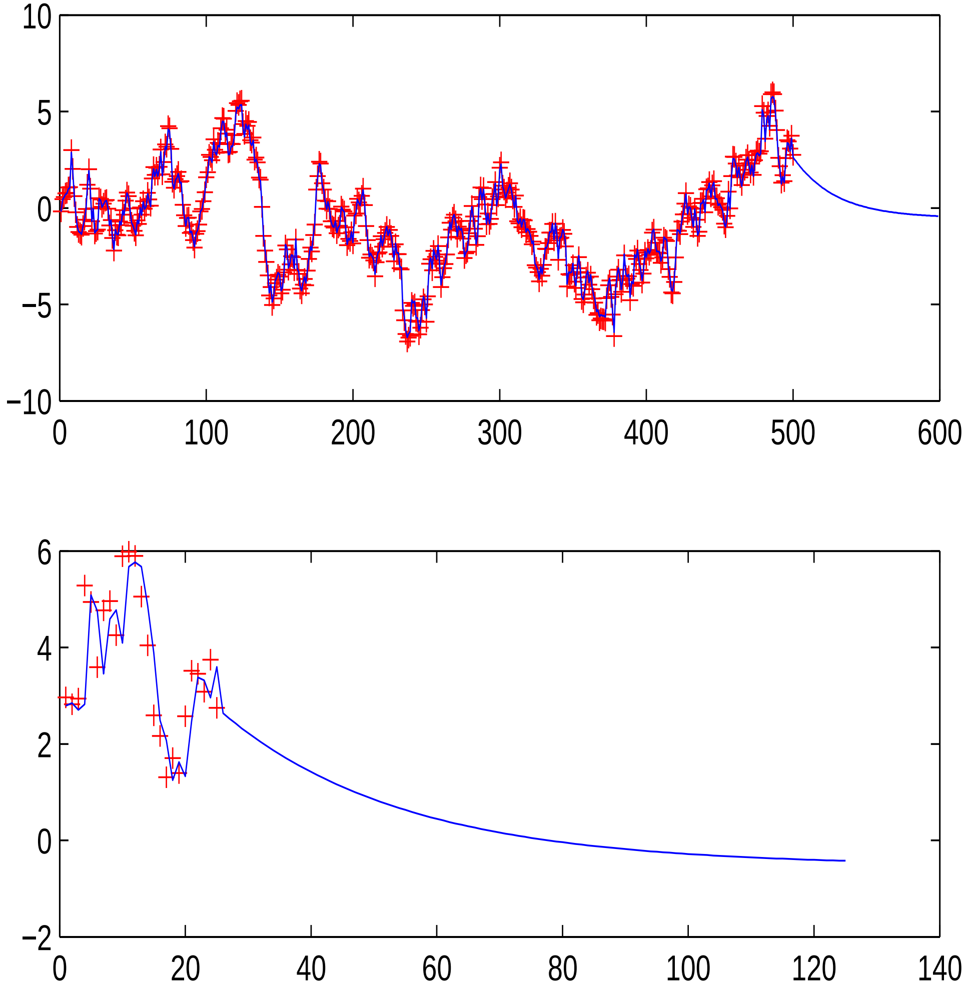

Figure 10: Simulated AR(1) data with $\phi_1 = 0.95$ plus forecasts

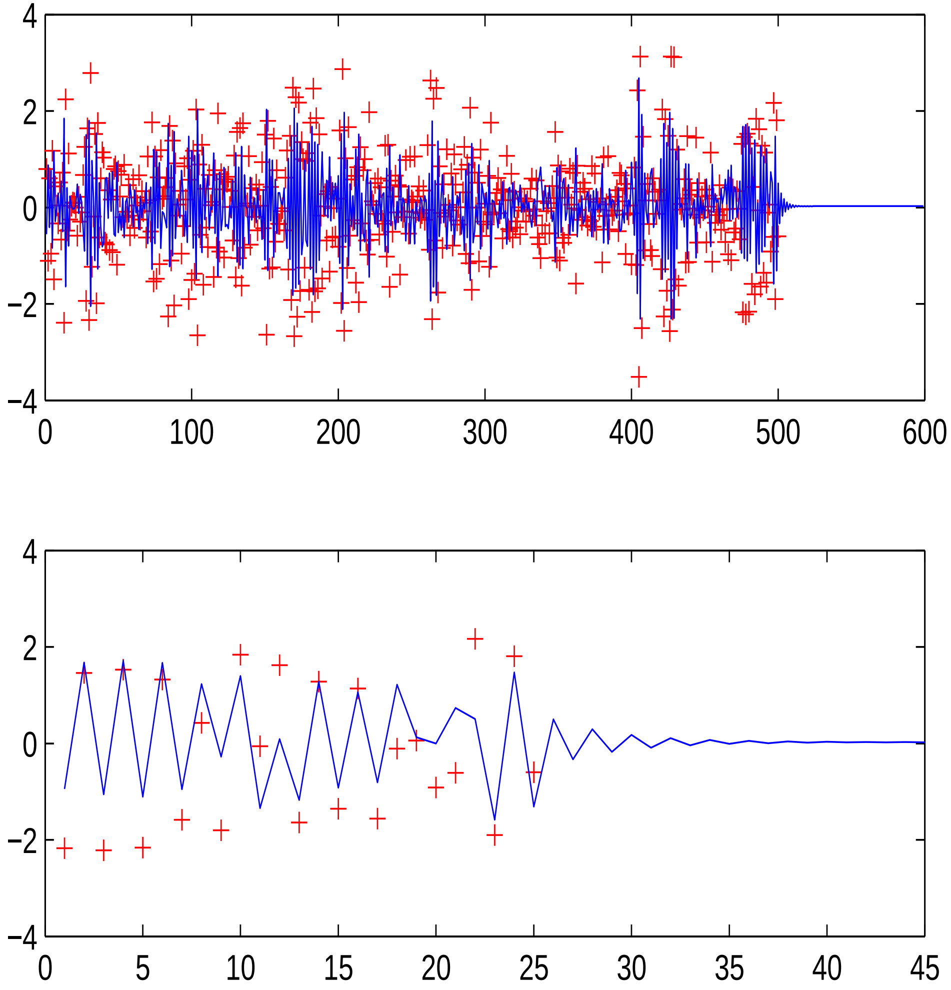

Figure 11: Simulated AR(1) data with $\phi_1 = -0.8$ plus forecasts

(2.4) Autoregressive models in general

- An AR(p) model has p lagged terms in the mean equation:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + e_t$$

- We can write:

$$\begin{aligned} E(y_t|y_{t-1}, \ldots, y_{t-p}) &= \phi_0 + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} \\ Var(y_t|y_{t-1}, \ldots, y_{t-p}) &= \sigma^2 \end{aligned}$$

- However, covariances are now harder to derive and complicated in general.

- Some more AR(p) properties:

$$\begin{aligned} E(y_t) &= \phi_0 + \phi_1 E(y_{t-1}) + \ldots + \phi_p E(y_{t-p}) \\ &= \frac{\phi_0}{1 - \phi_1 - \ldots - \phi_p} \end{aligned}$$

  but Var(y_t) is a very complicated expression.

- Is an AR(p) model stationary? When?

- Is an AR(p) model weakly stationary? When?

- Usually, the conditions $|\phi_1 + \ldots + \phi_p| < 1$ and $|\phi_p| < 1$, among others, are enforced during estimation.

- **Least squares** estimation: The conditional OLS estimator minimises:

$$\begin{aligned} \text{SSE} &= \sum_{t=p+1}^{n} e_t^2 \\ &= \sum_{t=p+1}^{n} (y_t - \phi_0 - \phi_1 y_{t-1} - \ldots - \phi_p y_{t-p})^2 \end{aligned}$$

- We then have:

$$\begin{aligned}\hat{\phi}_0 &= \bar{y}(1 - \hat{\phi}_1 - \ldots - \hat{\phi}_p) \\ \hat{\sigma}^2 &= \frac{1}{n-p} \sum_{t=p+1}^{n} (y_t - \hat{\phi}_0 - \hat{\phi}_1 y_{t-1} - \ldots - \hat{\phi}_p y_{t-p}^2)^2\end{aligned}$$

- Expressions for $\hat{\phi}_1, \ldots, \hat{\phi}_p$ are complicated.

- Under the 3 LS assumptions (2nd again modified to iid errors), there is a CLT for the LS estimates,

- which is also consistent

- However, as the true value of $\phi_1 + \ldots + \phi_p$ moves closer to 1, there is some bias, so that $E(\hat{\phi}_i) < \phi_i$.

- The bias again diminishes with sample size.

2.5 Forecasting for the AR(p) model

- $$\begin{aligned}\hat{y}_{t+1|t} &= E(y_{t+1}|\mathcal{F}_t) \\ &= E(\phi_0 + \phi_1 y_t + \ldots + \phi_p y_{t-p+1} + e_{t+1}|\mathcal{F}_t) \\ &= \phi_0 + \phi_1 y_t + \ldots + \phi_p y_{t-p+1}\end{aligned}$$

- This is a horizon 1 forecast.

- For horizon h=2:

$$\begin{aligned}\hat{y}_{t+2|t} &= \phi_0 + \phi_1 E(y_{t+1}|\mathcal{F}_t) + \ldots + \phi_p E(y_{t+2-p}|\mathcal{F}_t) + E(e_{t+2}|\mathcal{F}_t) \\ &= \phi_0 + \phi_1 \hat{y}_{t+1|t} + \phi_2 y_t + \ldots + \phi_p y_{t+2-p}\end{aligned}$$

- We could show that the horizon k, or k-step-ahead forecast, for $k > 2$, is:

$$\hat{y}_{t+k|t} \ = \ \phi_0 + \Sigma_{i=1}^{p}\phi_i E(y_{t+k-i}|\mathcal{F}_t)$$

where

$$E(y_{t+k-i}|\mathcal{F}_t) = \begin{cases} \hat{y}_{t+k-i|t} \ , \ k > i \\ y_{t+k-i} \ , \ k \leq i \end{cases}$$

- As $k \to \infty$, $\hat{y}_{t+k|t} \to \frac{\phi_0}{1-\phi_1-\ldots-\phi_p} = E(y_{t+k})$.

- Figures 13 and 15 highlight the forecast behavior for AR(3) models.

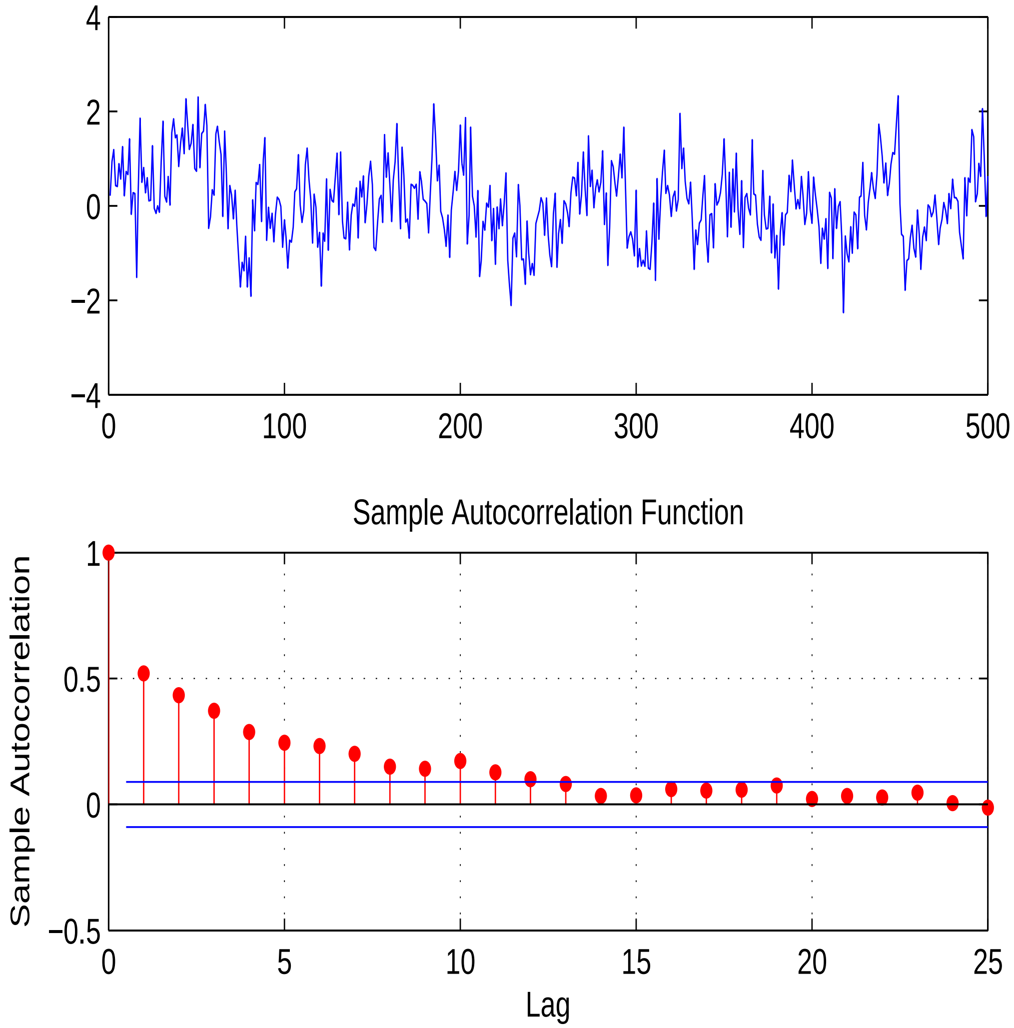

Figure 12: Simulated AR(3) data with $\Sigma\phi_i = 0.7$

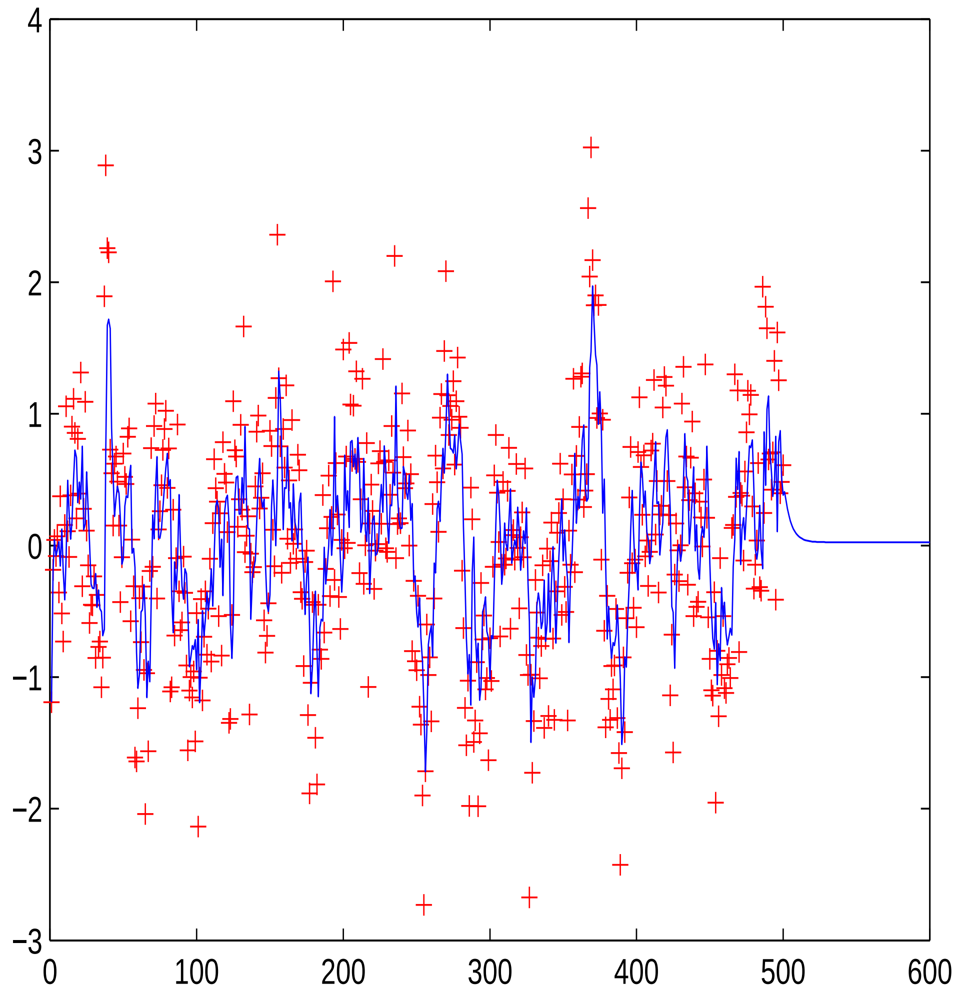

Figure 13: Simulated AR(3) data with $\phi_1 + \phi_2 + \phi_3 = 0.7$ plus forecasts

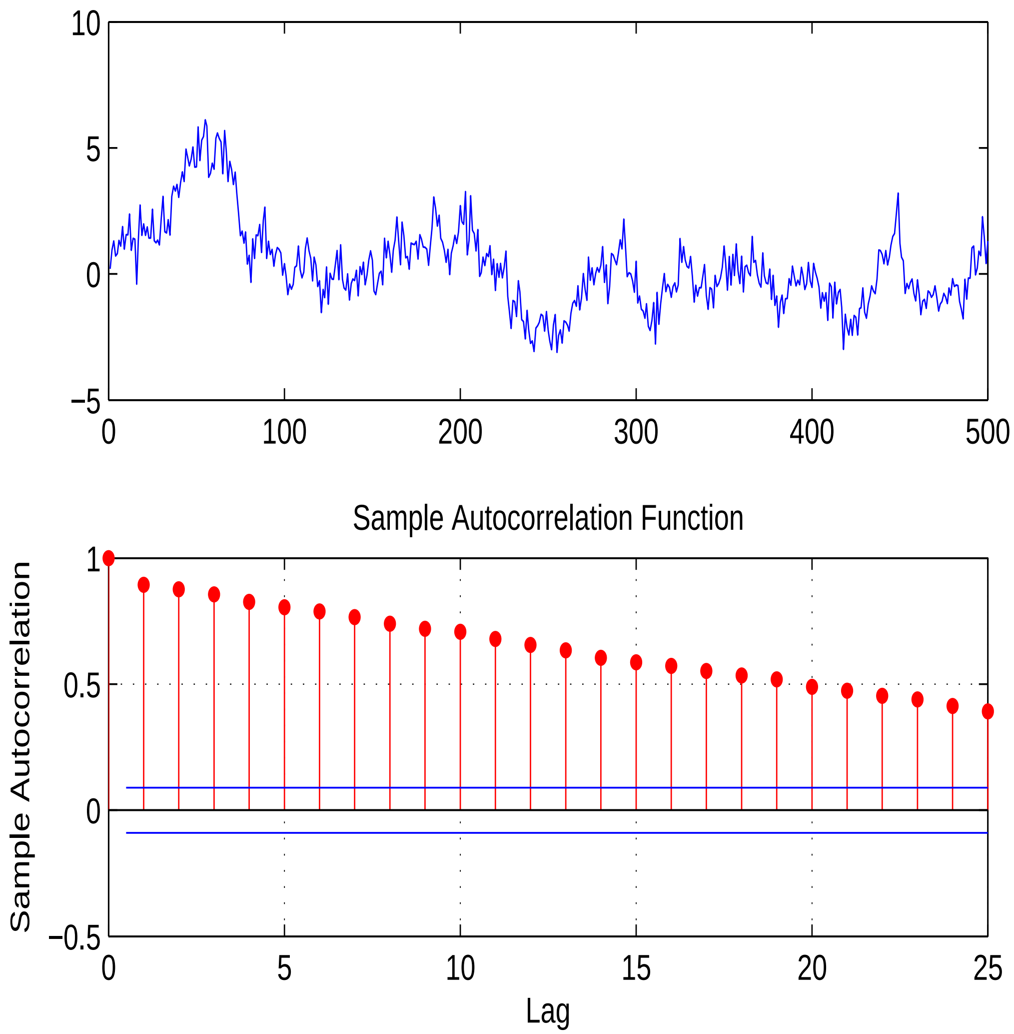

Figure 14: Simulated AR(3) data with $\Sigma\phi_i = 0.95$

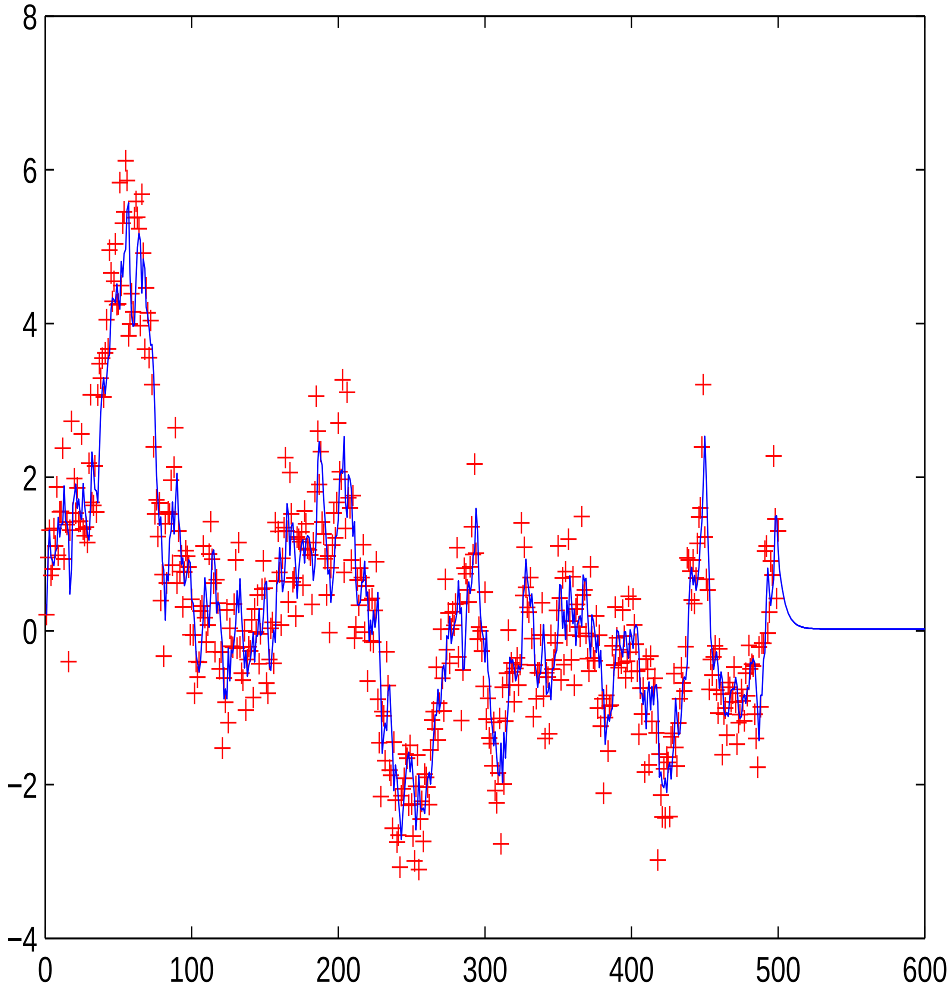

Figure 15: Simulated AR(3) data with $\phi_1 + \phi_2 + \phi_3 = 0.95$ plus forecasts

(2.4) Autoregressive Moving-Average models

- An AR(1) model has 1 lagged term in the mean equation.

- Often, it is also useful to add a lagged error term into the equation:

$$y_t = \phi_0 + \phi_1 y_{t-1} + e_t + \theta_1 e_{t-1}$$

- The lagged error terms are called moving average terms.

- The model simply indicates that previous errors may impact on future observations.

- This can be useful whenever a series is determined by human behavior, where humans will likely be watching the process.

- We can write:

$$\begin{aligned} E(y_t|y_{t-1},\ldots,y_1) &= \phi_0+\phi_1 y_{t-1}+\theta_1 e_{t-1} \\ Var(y_t|y_{t-1},\ldots,y_1) &= \sigma^2 \end{aligned}$$

- However, covariances are again harder to derive and complicated in general.

- Some more ARMA(1,1) properties:

$$\begin{aligned} E(y_t) &= \phi_0+\phi_1 E(y_{t-1})+\theta_1 E(e_{t-1})+E(e_t) \\ &= \frac{\phi_0}{1-\phi_1} \end{aligned}$$

  same as an AR(1), but Var(y_t) is again a very complicated expression.

- Is an ARMA(1,1) model stationary? When?

- Is an ARMA(1,1) model weakly stationary? When?

- Usually, the conditions $|\phi_1| < 1$ and $|\theta_1| < 1$, are enforced during estimation.

- **Least squares** estimation: The conditional OLS estimator minimises:

$$
\begin{aligned}
\text{SSE} &= \sum_{t=2}^{n} e_t^2 \\
&= \sum_{t=2}^{n} (y_t - \phi_0 - \phi_1 y_{t-1} - \theta_1 e_{t-1})^2
\end{aligned}
$$

- There is no closed form expression, so estimates are calculated numerically by search.

- Under the 3 LS assumptions, there is a CLT for the LS estimates,

- which are also consistent

- However, as the true value of ϕ_1, θ_1 move closer to 1, there is some bias, so that $E(\hat{\phi}_1) < \phi_1, E(\hat{\theta}_1) < \theta_1$.

- The bias again diminishes with sample size.

2.7 Forecasting for the ARMA(1,1) model

- $$\begin{aligned}\hat{y}_{t+1|t} &= \phi_0 + \phi_1 y_t + E(e_{t+1}|\mathcal{F}_t) + \theta_1 E(e_t|\mathcal{F}_t) \\ &= \phi_0 + \phi_1 y_t + \theta_1 \hat{e}_t\end{aligned}$$

- This is a horizon 1 forecast.

- For horizon 2:

$$\begin{aligned}\hat{y}_{t+2|t} &= \phi_0 + \phi_1 \hat{y}_{t+1|t} + \theta_1 E(e_{t+1}|\mathcal{F}_t) + E(e_{t+2}|\mathcal{F}_t) \\ &= \phi_0 + \phi_1(\phi_0 + \phi_1 y_t + \theta_1 \hat{e}_t) \\ &= \phi_0(1 + \phi_1) + \phi_1^2 y_t + \phi_1 \theta_1 \hat{e}_t\end{aligned}$$

- We could show that the horizon k, or k-step-ahead forecast, for $k > 2$, is:

$$\hat{y}_{t+k|t} \;=\; \phi_0(1 + \phi_1 + \ldots + \phi_1^{k-1}) + \phi_1^k y_t + \phi_1^{k-1}\theta_1 \hat{e}_t$$

- As $k \to \infty$, $\hat{y}_{t+k|t} \to \frac{\phi_0}{1-\phi_1} = E(y_{t+k})$.

- Figures 16, 17 and 18 highlight the forecast behaviour for ARMA(1,1) models.

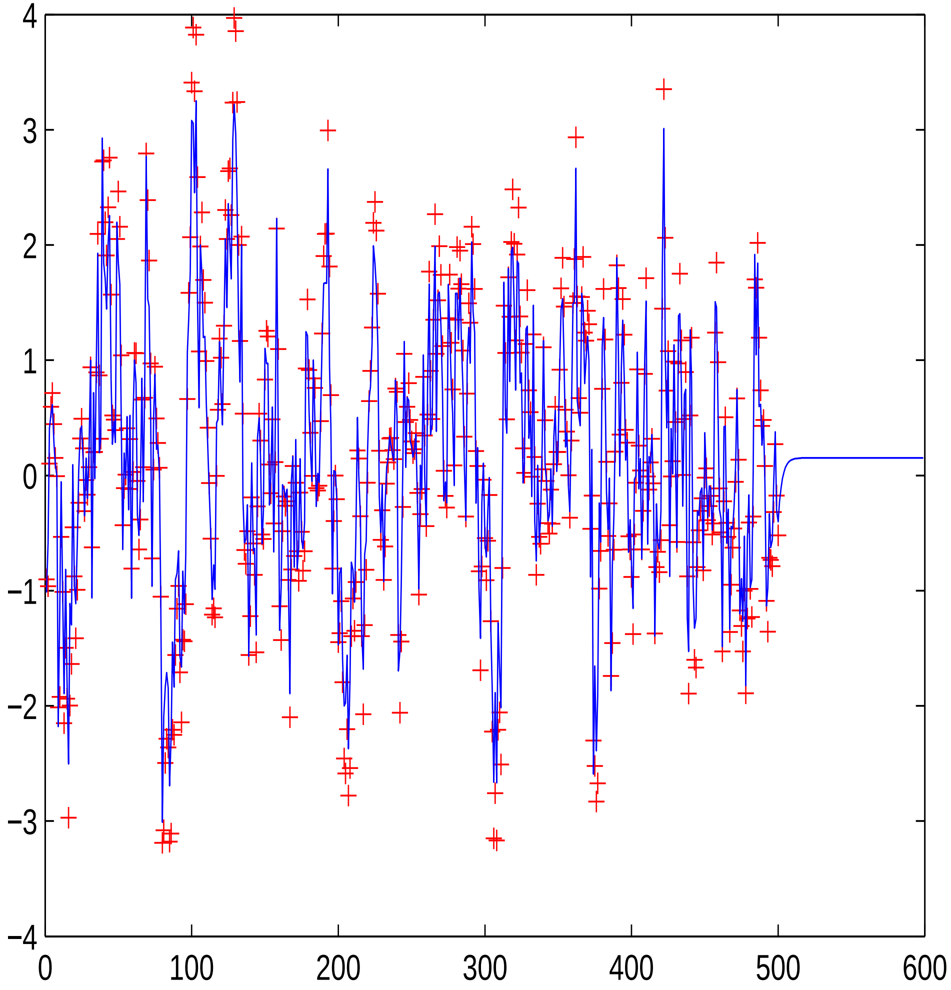

Figure 16: Simulated ARMA(1,1) data with $\phi_1 = 0.7, \theta_1 = 0.4$ plus forecasts

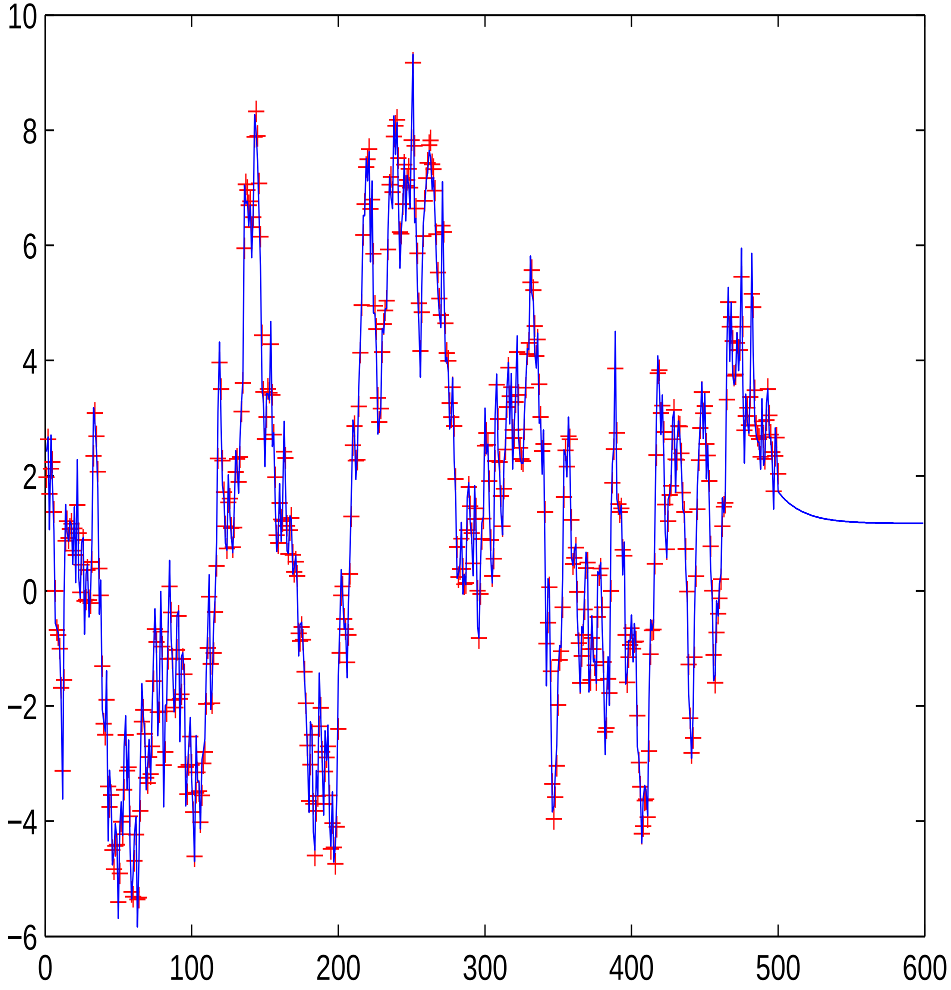

Figure 17: Simulated ARMA(1,1) data with $\phi_1 = 0.95, \theta_1 = 0.6$ plus forecasts

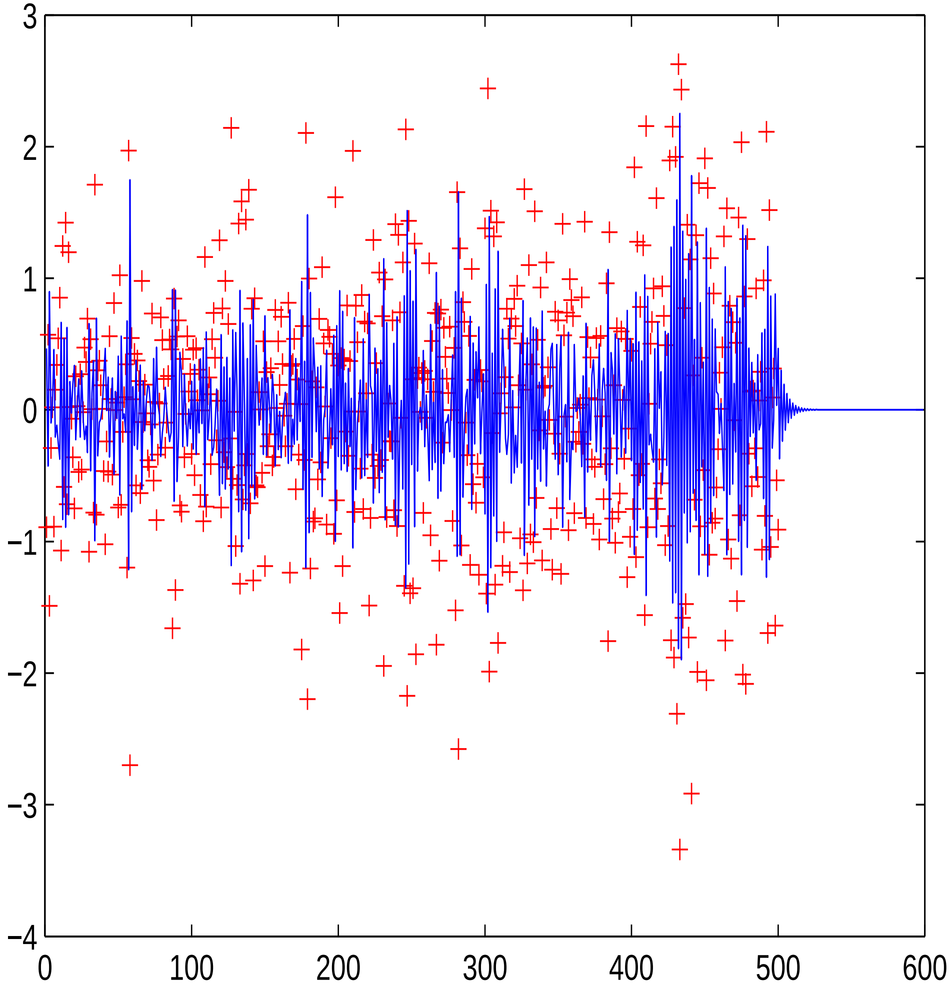

Figure 18: Simulated AR(1) data with $\phi_1 = -0.8, \theta_1 = 0.3$

(2.8) Example

- Consider again the CBA daily prices on the ASX.

- Let's compare the AR(1), AR(3) and ARMA(1,1) models for generating forecasts of stock returns:

- Let's again forecast the last 250 days of returns, at a daily forecast horizon.

- That is, we'll generate 250 one-day-ahead forecasts of CBA's stock price, using the three models above.

- I'll use a rolling window approach, where parameters are re-estimated each day, using the same sample size.

- Figure 19 shows the last 25 days of CBA returns in the in-sample period, together with the actual 1st return in the forecast period, and the three forecasts of it.

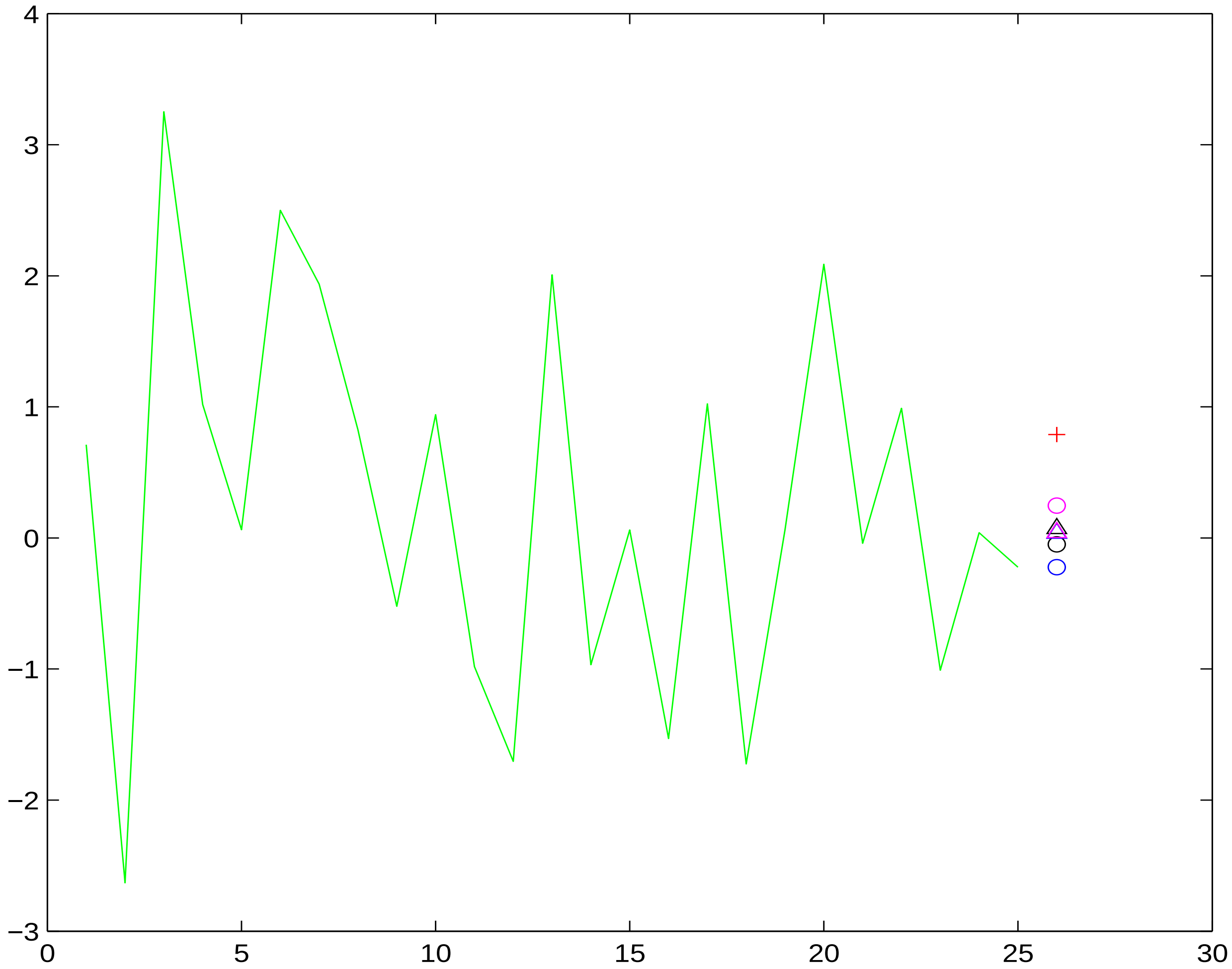

Figure 19: Log returns for CBA for last 25 days in Jan 21, 2013.

- The three estimated models are:

  1. AR(1): $y_t = 0.022 - 0.066y_{t-1} + e_t$
  2. AR(3): $y_t = 0.024 - 0.070y_{t-1} - 0.046y_{t-2} - 0.035y_{t-3} + e_t$
  3. ARMA(1,1): $y_t = 0.0096 + 0.546y_{t-1} - 0.617e_{t-1} + e_t$

- We repeat this exercise for each of the last 250 days of returns.

- Figure 20 shows the last 25 days of CBA returns in the in-sample period, together with the actual 250 returns in the forecast period, and the six sets of forecasts of these.

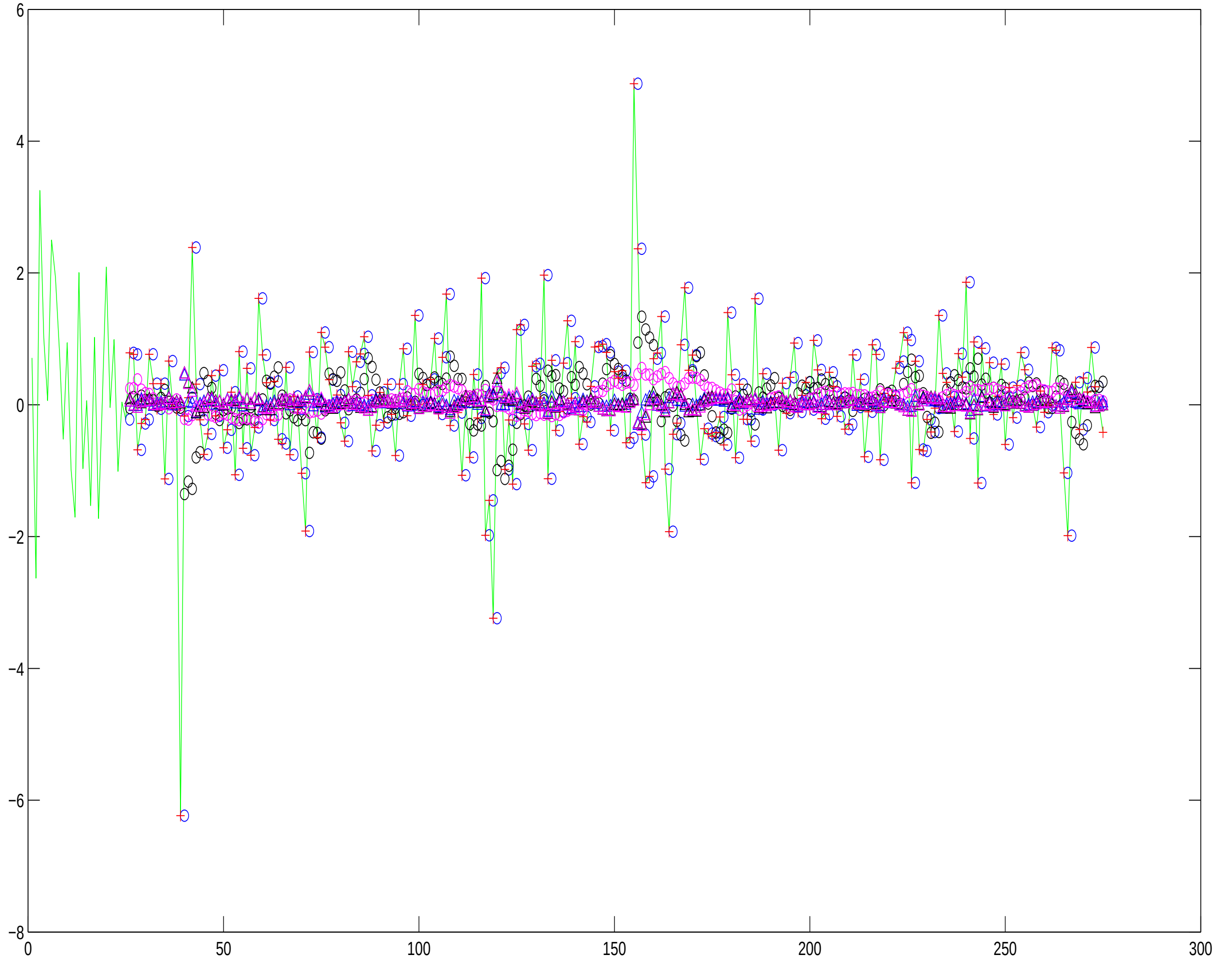

Figure 20: Log returns for CBA for last 250 days from Jan 21, 2013, plus six sets of 1 step-ahead forecasts

- These are all horizon 1 forecasts, 1 step-ahead.

- Table 2 shows the forecast accuracy measures for these 250 days.

Table 2: Forecast accuracy measures for 100 days of CBA returns

Measure	AR(1)	AR(3)	ARMA(1,1)	Naive Last day	Adhoc 1 5 days	Adhoc 2 25 days
RMSE	0.9183	0.9120	0.9138	1.228	1.014	0.922
MAD	0.6250	0.6218	0.6220	0.865	0.711	0.636

- Clearly, the ARMA models are the most accurate return predictors, though only marginally, and all pretty similar to each other, except the naive model.

- Are you surprised?

(3) Further Assessing forecast accuracy

- It is important to know and measure whether forecasts are accurate.

- If we knew the *true* data points we were forecasting (i.e. y_{t+1}, y_{t+2}, etc.), we could then compare these to our forecasts.

- The standard comparisons can be done using standard distance measures, such as RMSE, MAD, MAPE, as above.

- These allow direct comparison of several competing models.

- But they do not really measure the strength of the forecast models, nor how well they fit the forecast data

- To assess strength of fit, we could fit a regression of the actual forecast data using the forecasts themselves as the explanatory variable. i.e.:

$$y_t = \alpha + \beta \hat{y}_{t|t-1}$$

- Estimating this regression would give us two things:
  1. We could formally test whether $\alpha = 0$ and $\beta = 1$.
  2. We could calculate an R^2 for this regression.

- The first test above examines how well our forecasts track the data being forecast.

- The 2nd measure above assesses the strength of fit of the forecasts to the forecast data.

- This procedure is called a Mincer-Zarnowitz regression.

- From the example of CBA above, the table can be extended as in Table 3

Table 3: Forecast accuracy measures for 100 days of CBA returns

Measure	AR(1)	AR(3)	ARMA(1,1)	Naive Last day	Adhoc 1 5 days	Adhoc 2 25 days
RMSE	0.9183	$\boxed{0.9120}$	0.9138	1.228	1.014	0.922
MAD	0.6250	$\boxed{0.6218}$	0.6220	0.865	0.711	0.636
R^2	$\boxed{0.0077}$	0.0015	0.0003	0.0072	0.0073	0.000
$\hat{\alpha}$	0.115	0.089	0.092	$\boxed{0.086}$	0.112	0.093
$\hat{\beta}$	-1.296	0.085	$\boxed{0.187}$	0.085	-0.203	0.0082
F-stat (2,248)	7.817	$\boxed{2.768}$	3.444	209	64	8
p-val	0.0005	$\boxed{0.065}$	0.034	0.000	0.000	0.0004

- We see that all the models are pretty poor at forecasting returns!

- The p-values are from the test that $\alpha = 0, \beta = 1$. Here the formal models do better, but the estimates are very far from (0,1).

- Only the AR(3) model is not rejected at the 5% significance level.