Module 2: Introduction to Forecasting and Time Series

References:

- *Chapter 2 in Tsay*
- *Chapter 5 in Brooks*

Section 2: Forecast combination and Autocorrelation testing

(2.1) Forecast combination

- When forecasting is done, often more than one set of forecasts are produced
- In these cases, usually all forecasts are assessed jointly.
- Managers often change forecasts subjectively based on their own pre-conceived ideas.

- Question: are there optimal ways to combine different forecasts? Objectively?

- Formal methods include specifically choosing weights that allow forecasts to be combined into a weighted average forecast.

- If we have K forecasts, denoted $\hat{r}_{i;t+1|t}$ (adhoc choice of horizon 1), a weighted average of these is:

$$\hat{r}_{t+1|t} = \sum_{i=1}^{K} w_i \hat{r}_{i;t+1|t}$$

- The question is, what weights should we choose?

- Some common objective choices include:
  1. Equal weights: $w_i = \frac{1}{K}$
  2. Weights chosen to minimise (forecast) variance (e.g. Markowitz, hedging)
  3. Weights chosen to maximise previous forecast accuracy (or model fit in-sample)
  4. Weights chosen to optimise some other criteria (e.g. profit, manager's beliefs, etc)

- As you can see, when we are forecasting financial returns, forecast combination is related to portfolio optimisation.

- A portfolio of K assets has a log-return, in period $t+1$, that is:

$$r_{p;t+1} = \sum_{i=1}^{K} w_i r_{t+1}$$

- Generally, it is hard to beat equal weighting.

- You will practice with combining forecasts in lab session 6. Here I give a bit of the theory behind it.

- Classic reference: Bates, J. M., and C. W. J. Granger (1969). The combination of forecasts, *Operational Research Quarterly*, 20, 451–468.
  They provide the following illustration

TABLE 1. ERRORS IN FORECASTS (ACTUAL LESS ESTIMATED) OF PASSENGER MILES FLOWN, 1953

Month	Brown's exponential smoothing forecast errors	Box–Jenkins adaptive forecasting errors	Combined forecast ($\frac{1}{2}$ Brown + $\frac{1}{2}$ Box–Jenkins) errors
Jan	1	−3	−1
Feb.	6	−10	−2
March	18	24	21
April	18	22	20
May	3	−9	−3
June	−17	−22	−19·5
July	−24	10	−7
Aug.	−16	2	−7
Sept.	−12	−11	−11·5
Oct.	−9	−10	−9·5
Nov.	−12	−12	−12
Dec.	−13	−7	−10
Variance of errors	196	188	150

- Variance reduction:

  If you have two unbiased forecasts $\hat{y}^{(1)}_{T+1|T}$ and $\hat{y}^{(2)}_{T+1|T}$ with the corresponding variances σ_1^2 and σ_2^2, then we can combine them linearly

  $$\hat{y}^{c}_{T+1|T} = w * \hat{y}^{(1)}_{T+1|T} + (1-w) * \hat{y}^{(2)}_{T+1|T}.$$

  The variance of the combined forecast will be

  $$\sigma_c^2 = w^2 * \sigma_1^2 + (1-w)^2 * \sigma_2^2 + 2 * \rho * w\sigma_1 * (1-w)\sigma_2.$$

  It will have minimum at

  $$w^* = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

  In case where $\hat{y}^{(1)}_{T+1|T}$ and $\hat{y}^{(2)}_{T+1|T}$ are uncorrelated ($\rho = 0$), then $w^* = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$.

- The variance of the combined forecast will be no greater than the smaller of the two individual forecast variances.

- Weights estimation: Adaptive weight is recommended. The simplest version is

$$\widehat{w}^{*}_{T+1} = \frac{\sum_{t=T-v}^{T}(e_{2,t})^2}{\sum_{t=T-v}^{T}(e_{1,t})^2 + \sum_{t=T-v}^{T}(e_{2,t})^2},$$

where $e_{1,t}$ and $e_{2,t}$ are known forecasting errors.

- Numerous other alternatives are available. For those interested in this topic, a few key references are:

Timmermann (2006) Forecast combination, Handbook of Economic Forecasting, vol 1, pp 135–196

Clemen, R. T. (1989). Combining forecasts: a review and annotated bibliography, *International Journal of Forecasting*, 5, 559–583

Hendry, D. F., and M. P. Clements (2004). Pooling of forecasts, *The Econometrics Journal*, 7, 1–31

Gibbs, C. G. and A. L. Vasnev (2017). Conditionally Optimal Weights and Forward-Looking Approaches to Combining Forecasts, ssrn.com/abstract=2919117

- Applications:

Stock and Watson (2004) Combination Forecasts of Output Growth in a Seven-Country Data Set, J. of Forecasting, 23, 405–430.

Vasnev A, Skirtun M and Pauwels L (2013) Forecasting Monetary Policy Decisions in Australia: A Forecast Combination Approach, Journal of Forecasting, 32, 151–156.

Matsypura, D., R. Thompson and A. L. Vasnev (2017). Optimal Selection of Expert Forecasts with Integer Programming, ssrn.com/abstract=2894083

(2.2) Assessing and testing for autocorrelation

- (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.

- 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?

- 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 ...

- To assess when we should fit an AR or MA model, we can do tests on these auto-correlations.

- A joint test considers the hypotheses:

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

  H_1: $\rho_i \neq 0$ for some $i \in \{1, 2, \ldots, m\}$

- Box and Pierce (1970, JASA) introduced the Box-Pierce statistic for such a test.

- They noted that asymptotically $\hat{\rho}_l \sim N(0, \frac{1}{T})$ and introduced the Box-Pierce statistic as

$$BP(m) = T \sum_{l=1}^{m} \hat{\rho}_l^2$$

- If $\{y_t\}$ is an iid sequence weakly stationary process, $BP(m)$ is asymptotically distributed as χ^2_m, i.e. chi-squared with m degrees of freedom.

- Ljung and Box (1978, Biometrika) modified the $BP(m)$ statistic to have more *power* in finite samples.

- The power of a statistical test measures the probability of rejecting a false null hypothesis.

- The more powerful test of Ljung and Box is thus better at detecting non-zero autocorrelations in a time-series process.

- 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.

- It is the preferred choice among econometricians and time series analysts to test for (absence of) autocorrelation.

- In practice, the choice of m can significantly affect the performance of the statistic.

- Quant. analysts often use $m \approx \ln(T)$ as a rule of thumb; or simply $m = 5, 10, 15$.

- 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 AR(p) model to data and using the Q(m) statistic above on the residuals, the correct degrees of freedom would be $m - p$.

- Significant auto-correlations in $\{y_t\}$ indicate that an AR, MA or ARMA model may be useful.

(2.3) Example

- Consider again the CBA daily prices on the ASX.
- First, are there autocorrelation effects that may need to be modelled?
- Figure 1 shows the returns and their ACF.

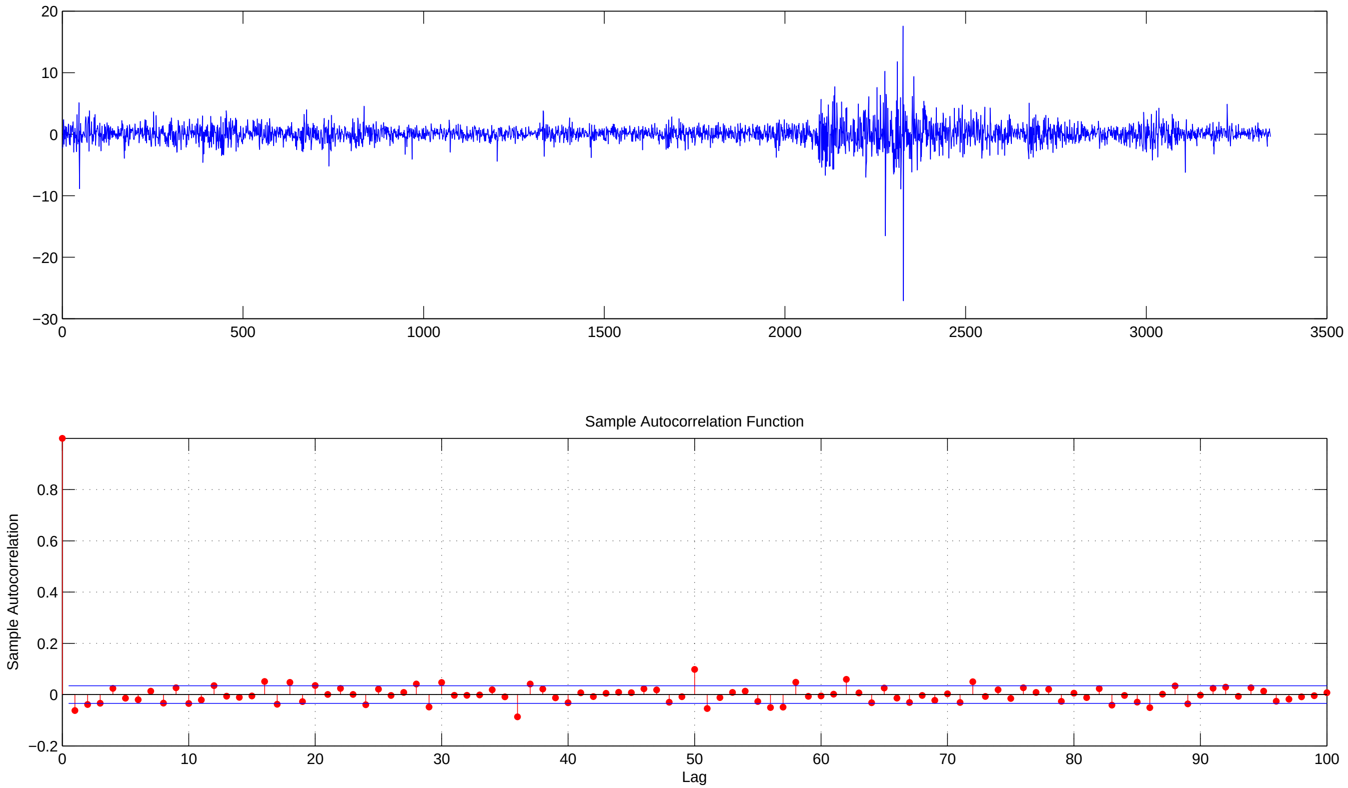

Figure 1: Log-returns and their ACF for CBA from Jan, 2000 to Feb, 2011.

- The Ljung-Box test was carried out on the first $m = 8$ autocorrelations.

- The null hypothesis is H_0: $\rho_1 = \rho_2 = \ldots = \rho_8 = 0$.

- The p-value from the $Q(8)$ statistic is 0.0002.

- As such we reject the null and conclude that at least one of the first 8 sample auto-correlations in the CBA return data are significantly different to 0.

- We consider the AR(1), AR(2) and ARMA(1,1) models to capture this autocorrelation.

- The three estimated models are:

  1. AR(1): $y_t = 0.022 - 0.066 y_{t-1} + e_{1t}$
  2. AR(2): $y_t = 0.023 - 0.069 y_{t-1} - 0.044 y_{t-2} + e_{2t}$
  3. ARMA(1,1): $y_t = 0.010 + 0.55 y_{t-1} - 0.61 e_{3,t-1} + e_{3t}$

- If the models adequately capture the correlation or time series pattern in the mean of the data, then their residuals should be iid and have no remaining autocorrelation in them.

- Ljung-Box tests were applied to the residual series e_1, e_2, e_3, using the same degrees of freedom in each case of 5, 8 and 10.

- Figure 2 shows the time series plots for each of e_1, e_2, e_3. Under the assumptions of the models fit, we expect these to show no pattern over time.

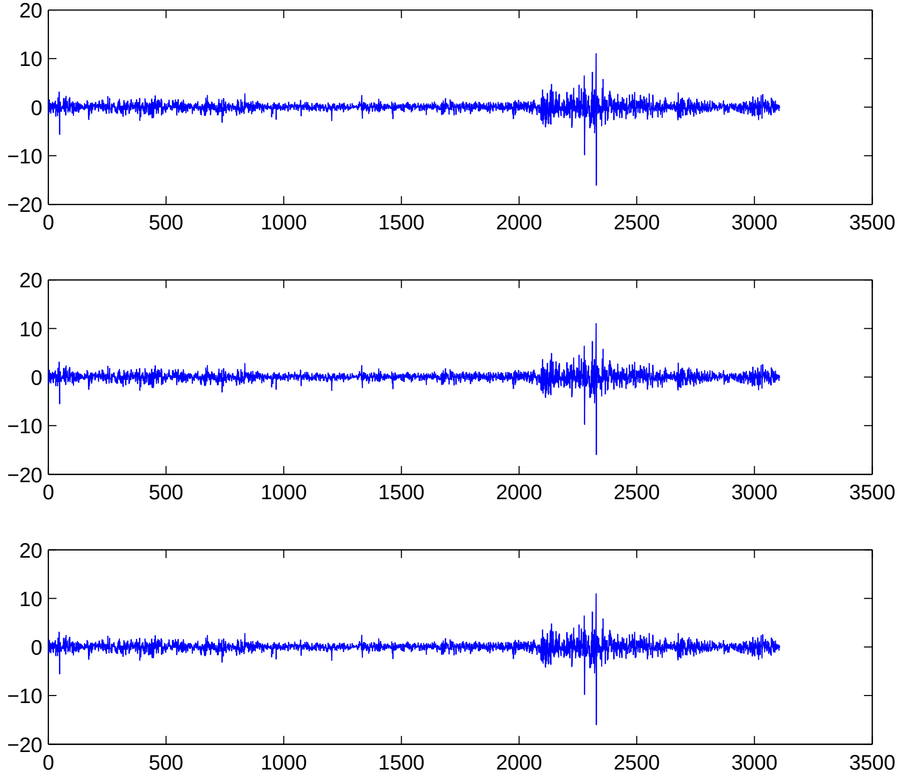

Figure 2: Residuals from AR(1), AR(2) and ARMA(1,1) models applied to CBA returns

- Figure 3 shows the ACF plots for each of e_1, e_2, e_3.

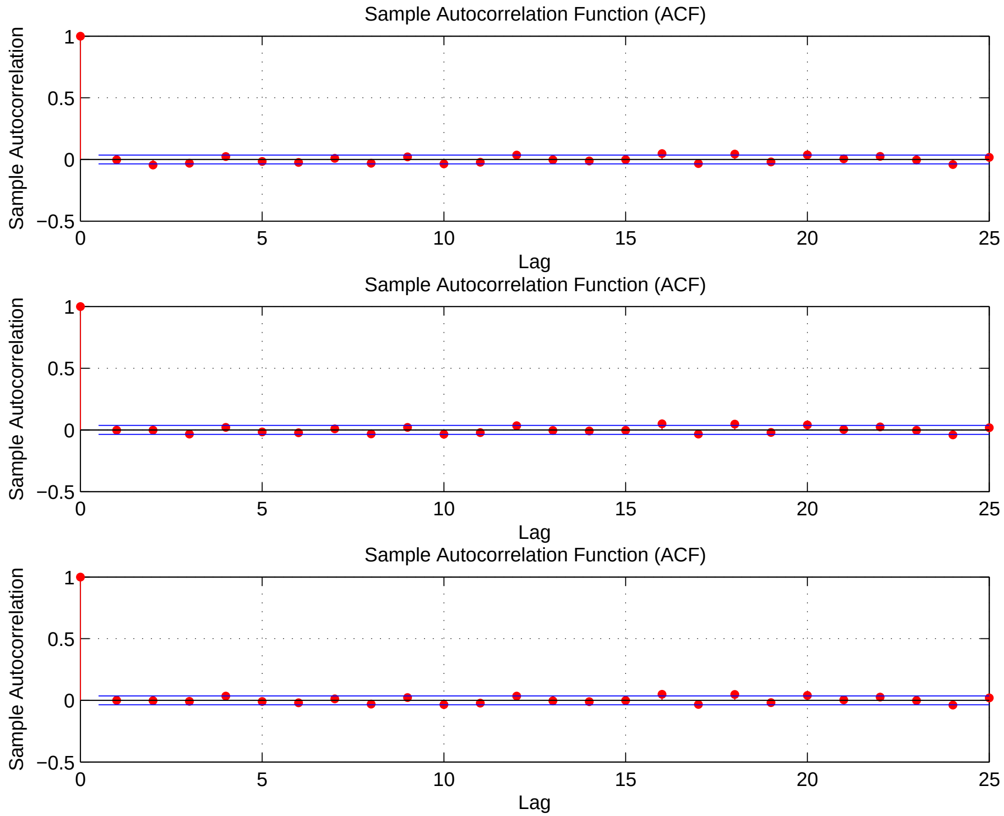

Figure 3: ACF for residuals from AR(1), AR(2) and ARMA(1,1) models applied to CBA returns

- Table 1 shows the results of these tests.

Table 1: Ljung-Box tests applied to the residual series e_1, e_2, e_3

Model	lags	df	p-val
AR(1)	6	5	0.02
	9	8	0.02
	11	10	0.008
AR(2)	7	5	0.17
	10	8	0.037
	12	10	0.018
ARMA(1,1)	7	5	0.33
	10	8	0.075
	12	10	0.034

- The ARMA(1,1) model appears to fit the data the best, while the AR(1) model fits the worst in terms of capturing the autocorrelation pattern in the data.

Time Series Regression

- Regression terms can be added to ARMA models OR ARMA terms can be added to regression models

- In the context of forecasting, it would be most usual to include lagged explanatory variables in the model. *why??*

- Figure 4 shows the daily index and log return values for the AORD and S&P500 indices, from January, 2000 until Feb, 2013.

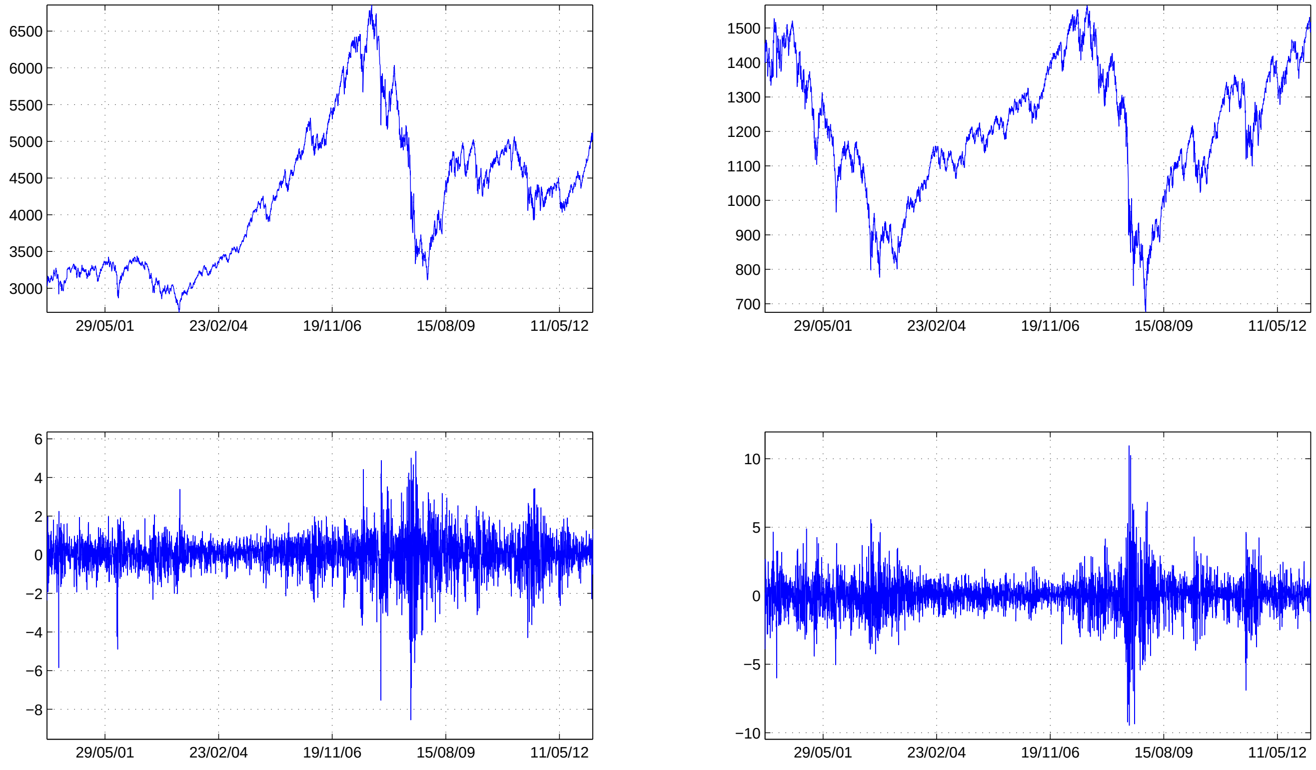

Figure 4: Index values and daily log returns for AORD and SP500 indices

- The sample from Jan, 2008 - Feb, 2013 will be used as as forecast sample; the learning sample from January, 2000 to January, 2008

- Question: is there a relationship between AORD and S&P500 daily returns that we can use to forecast?

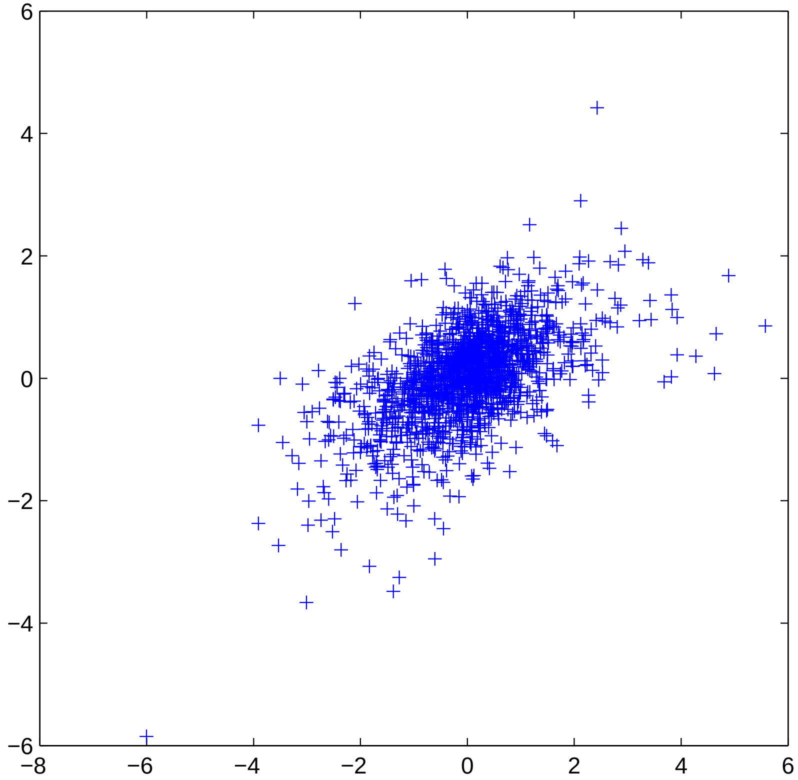

Figure 5: Scatterplot of daily log returns for AORD vs lagged SP500 indices

- The estimated regression relationship is:

$$\text{AORD}_t = 0.031 + 0.396 \times \text{S\&P500}_{t-1} + \hat{\epsilon}_t$$

OR

$$\hat{\text{AORD}}_t = 0.031 + 0.396 \times \text{S\&P500}_{t-1}$$

- $R^2 = 40\%$ indicating that 40% of the variation in AORD returns is captured by the straight line relationship with S&P500 returns.
- In other words, 40% of the daily movements in the AORD index are explained by the previous day's movements in the S&P500 index.
- Should we add ARMA terms to this model?

- Figure 6 shows residuals over time and then their ACF from this regression model.

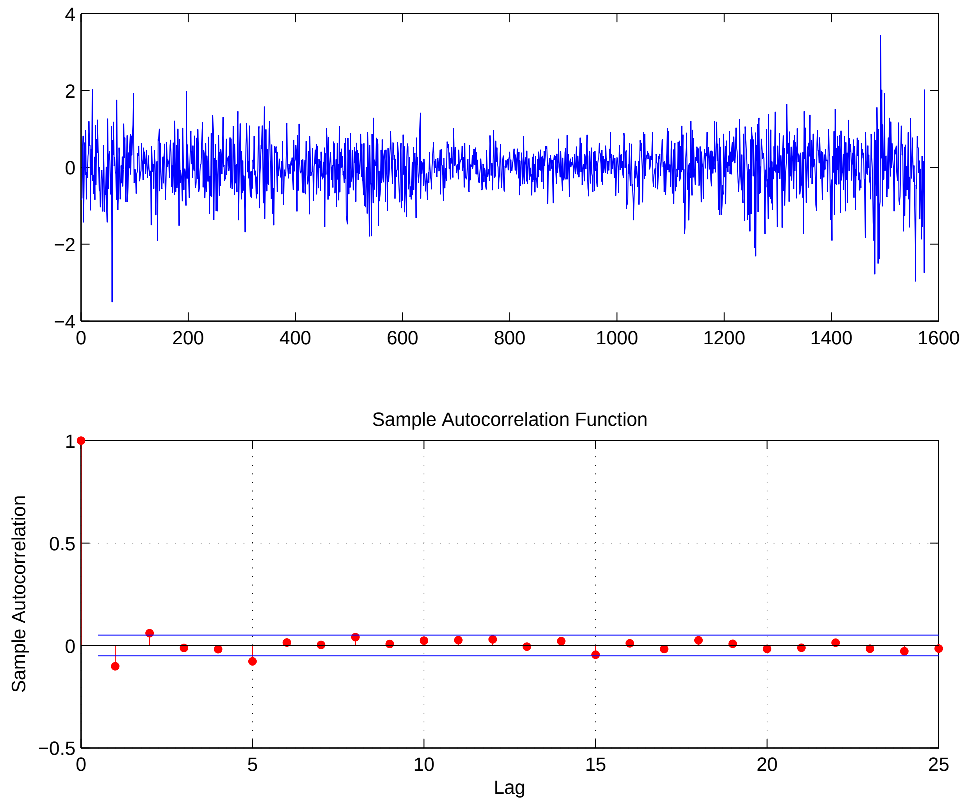

Figure 6: Residuals from regression of AORD(t) on SP500(t-1) for Jan, 2000 to Feb, 2013.

- Significant autocorrelations seem to exist at lags 1, 2 and 5.

- LB statistics using 5 and 10 degrees of freedom (using 6 and 11 lags) reveal statistically significant correlations, with p-values both close to 0.

- Based on this I will try three more time series regression models:

$$\begin{aligned}
\text{AORD}_t &= \alpha + \beta \times \text{S\&P500}_{t-1} + \phi_1 \times \text{AORD}_{t-1} + \epsilon_t \\
\text{AORD}_t &= \alpha + \beta \times \text{S\&P500}_{t-1} + \phi_1 \times \text{AORD}_{t-1} \\
&+ \phi_2 \times \text{AORD}_{t-2} + \epsilon_t \\
\text{AORD}_t &= \alpha + \beta \times \text{S\&P500}_{t-1} + \phi_1 \times \text{AORD}_{t-1} \\
&+ \phi_2 \times \text{AORD}_{t-2} + \phi_3 \times \text{AORD}_{t-3} \\
&+ \phi_4 \times \text{AORD}_{t-4} + \phi_5 \times \text{AORD}_{t-5} + \epsilon_t
\end{aligned}$$

- These are time series regression models: I label them Reg-AR(1), Reg-AR(2) and, Reg-AR(5).

- It is also possible to add extra lags of the X variable: e.g. here we could add $\text{S\&P500}_{t-2}, \text{S\&P500}_{t-3}$, etc.

- I leave that to you.

- I will now forecast the last 1000 days: from March, 2009 - Jan, 2013 of AORD returns using these models.

- I will use a forecast horizon of 1 day, a fixed size moving window of $T = 1574$ and I will re-estimate all parameters with each new forecast.

- I also include the naive model and the 25 day sample mean forecast model.

- I use Gaussian maximum likelihood to estimate all the models, but the estimates will be only negligibly away from their LS counterparts.

- The models are initially estimated as:

$$
\begin{aligned}
\text{AORD}_t &= 0.032 + 0.396 \times \text{S\&P500}_{t-1} - 0.044 \times \text{AORD}_{t-1} \\
&+ \epsilon_t \\
\text{AORD}_t &= 0.030 + 0.396 \times \text{S\&P500}_{t-1} - 0.042 \times \text{AORD}_{t-1} \\
&+ 0.062 \times \text{AORD}_{t-2} + \epsilon_t \\
\text{AORD}_t &= 0.032 + 0.394 \times \text{S\&P500}_{t-1} - 0.045 \times \text{AORD}_{t-1} \\
&+ 0.064 \times \text{AORD}_{t-2} + 0.012 \times \text{AORD}_{t-3} \\
&- 0.030 \times \text{AORD}_{t-4} - 0.062 \times \text{AORD}_{t-5} + \epsilon_t
\end{aligned}
$$

- Figure 7 shows the data being forecast and the forecast from the five time series regression models.

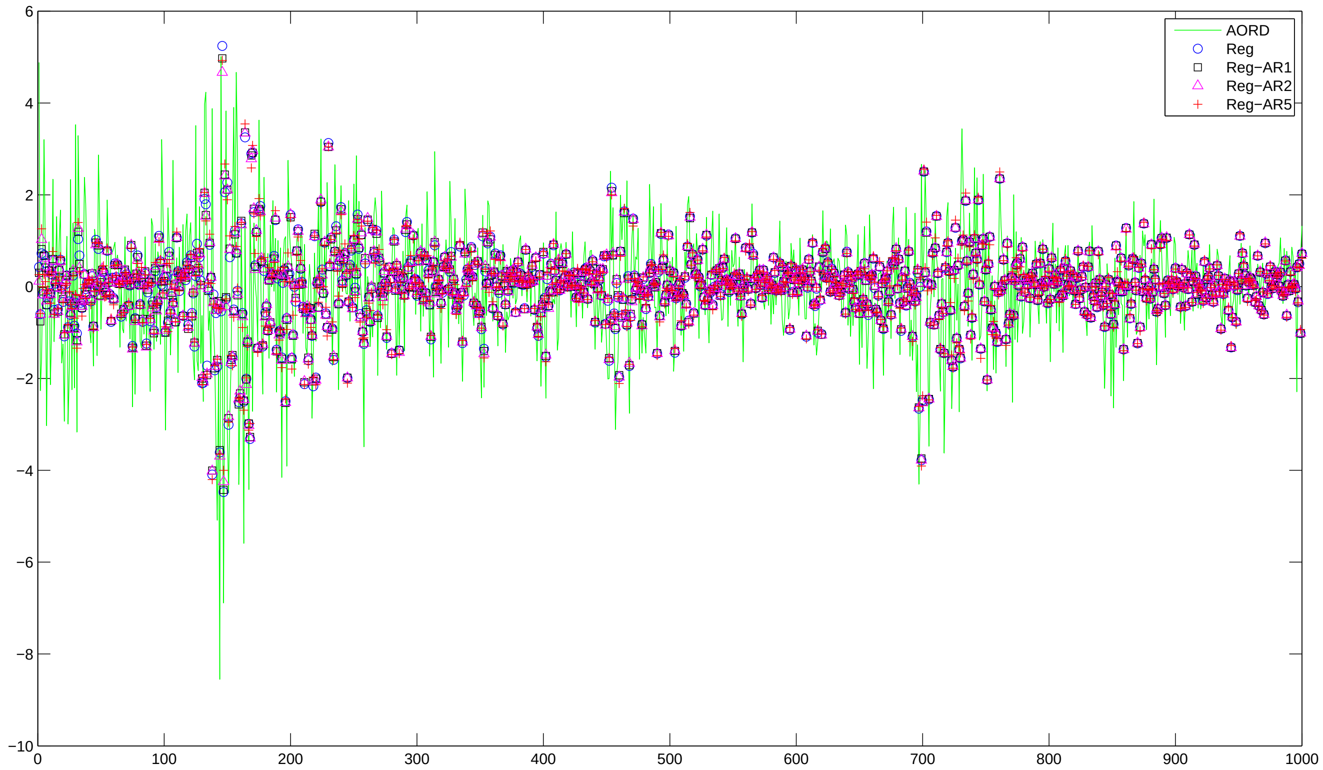

Figure 7: Forecasts from time series regression models of AORD for March, 2009 to Jan, 2013.

- Comments?

- Forecast accuracy is measured by RMSE, MAD and by Mincer-Zarnawitz regressions, without outputs in Table 2:

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

Measure	Reg	Reg-AR(1)	Reg-AR(2)	Reg-AR(5) Last day	Naive Last day	Adhoc 25 days
RMSE	1.201	$\boxed{1.198}$	1.200	1.207	**1.846**	1.331
MAD	0.839	$\boxed{0.838}$	0.839	0.844	**1.348**	0.962
R^2	0.187	$\boxed{0.189}$	0.186	0.182	**0.0001**	0.0003
$\hat{\alpha}$	0.0072	$\boxed{0.0072}$	0.0074	0.0066	**0.0142**	0.0142
$\hat{\beta}$	0.722	$\boxed{0.730}$	0.726	0.707	**0.0087**	0.0823
F-stat (2,998)	34.177	$\boxed{31.78}$	32.58	38.07	**984.50**	33.87
p-val	0.000	0.000	0.000	0.000	**0**	0.000

- The AR terms only marginally improve on the original regression in terms of forecast accuracy.

- R^2 during the forecast period is a maximum, of 19%, for the Reg-AR(1) model.

- All models can be strongly rejected as accurate forecasters in the sense of $\alpha = 0; \beta = 1$ in a Mincer-Zarnawitz regression.

- Clearly all the βs are significantly below 1.

- The naive and adhoc models do the worst under all measures.

- Regarding forecast accuracy, this is about the maximum accuracy that can be achieved for forecasting returns.

- For most assets and portfolios, the forecast accuracy of the best models will be much less.

- You'll get a taste of this in the lab session.