## Tutorial 6: Multivariate Time Series

The following exercises should be done using R. The package MARSS (Multivariate Autoregressive State-Space Modelling) is useful as is the package vars.

```
> install.packages("MARSS")
> install.packages("vars")
```

Check the syntax of `acf`; it also works for multivariate time series.

## Multivariate Time Series in R

A VARMA(p,q) model (vector auto regressive moving average, lags $p$ and $q$ for the auto-regressive and moving average parts respectively) is a model:

$$\underline{X}(t) = \underline{\mu}_0 + t\underline{\mu}_1 + \sum_{j=1}^{p} A_j \underline{X}(t-j) + \sum_{k=1}^{q} B_k \underline{\epsilon}_{t+1-q}$$

where $\underline{\epsilon}_t \sim N(\underline{0}, \Sigma)$ are i.i.d. (the distribution is not necessarily normal, but the normality assumption, if true, leads to sharper estimation).

The MA part often leads to instability for estimation; we therefore only consider VAR(p) processes;

$$\underline{X}(t) = \underline{\mu}_0 + t\underline{\mu}_1 + \sum_{j=1}^{p} A_j \underline{X}(t-j) + \underline{\epsilon}_t$$

The package **vars** fits a vector auto regressive model:

```
> install.packages("vars")
> library(vars)
```

Within **vars**, there is a test data-set `Canada`, which contains 4 macroeconomic indicators; `prod` (labour productivity), `e` (employment), `U` (unemployment rate) and `rw` (real wages). A VAR(2) model is fitted quite simply with the command:

```
> data(Canada)
> can = VAR(Canada,p=2)
> summary(can)
```

This gives

- Parameter estimation and the estimated equations for a VAR(2) model,
- Confidence intervals for the parameter estimates,
- $R^2$ statistics, which indicate the proportion of the sum of squares belongs to the model,
- estimates of the covariance and correlation matrices for the noise.

The default model is:

$$X_t = \mu_0 + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + Z_t.$$

The default value for this (which estimates $\underline{\mu}_0$ and sets $\underline{\mu}_1 = 0$) is `const`. To set $\underline{\mu}_0 = 0$ and $\underline{\mu}_1 = 0$, type:

```
> VAR(Canada,p=2,type="none")
```

To set $\underline{\mu}_0 = 0$ while estimating an unknown trend $\underline{\mu}_1$, type:

```
> VAR(Canada,p=2,type="trend")
```

To estimate both an intercept $\underline{\mu}_0$ and a trend $\underline{\mu}_1$, type:

```
> VAR(Canada,p=2,type="both")
```

The `stability` function verifies the covariance stationarity of a VAR process, using cumulative sums of residuals. This may be carried out by:

```
> var.2c=VAR(Canada,p=2,type="const")
> stab=stability(var.2c,type="OLS-CUSUM")
> plot(stab)
```

WARNING: you may get an error for the plot, stating that there is something wrong with the margins. This is fixable (look at the R script which accompanies the lecture).

There are several tests for normality which come under `normality.test`.

```
> normality.test(var.2c)
```

The output indicates that the null hypothesis of normality cannot be rejected.
The function `serial.test` carries out the Portmanteau (i.e. Ljung-Box) test

```
> serial.test(var.2c,lags.pt=16,type="PT.adjusted")
```

Again, there is no reason to reject the null hypothesis that the residuals are white noise.

1. Firstly, work through the example above, usining the `Canada` data set from the **vars** package.

   (a) Try fitting a VAR(2) model and a VAR(3) model. Does the VAR(3) model represent an improvement? (Use standard model building criteria, such as AIC and BIC).

   (b) Does the fitted VAR(2) model exhibit Granger causailty? How about the fitted VAR(3) model?

2. Consider the monthly log stock returns, in percentages, and including dividends, of Merck & Company, Johnson & Johnson, General Electric, General Motors, Ford Motor Company and the value-weighted index from January 1960 to December 2009, found in `m-mrk2vw.txt`.

   (a) Compute the sample mean, covariance matrix and correlation matrix of the data.

   (b) Test the hypothesis: $H_0 : R_1 = \ldots = R_6 = 0$ where $R_j$ is the lag $j$ cross correlation matrix of the data. Draw conclusions based on the 5% significance level.

   (c) Is there any lead-lag relationship among the six return series?

   (d) Try fitting a VAR model. Are there any indicators that some of the variables have a causal effect on others? Try fitting a VMA model; try fitting a VARMA model. Which model fits the data best? Does it represent an improvement over White Noise?

3. The federal reserve bank of St. Louis publishes selected interest rates and U.S. financial data on its web site:
   `http://research.stlouisfed.org/fred2/`
   and the following data set is taken from there. Consider the monthly 1-year and 10-year Tresury constant maturity rates from April 1953 to October 2009 for 679 observations stored in the file `m-gsln10.txt`. The rates are in percentages.

   (a) Let $\underline{Y}_t = \underline{X}_t - \underline{X}_{t-1}$ where $\underline{X}$ represents the bivariate series. Construct an appropriate VAR(p) (vector auto regressive) model for the series $\underline{Y}$.

   (b) Build a bivarate VMA(q) model for the series $\underline{Y}$. Compare it with the bivariate VAR(p) model from the previous part.

   (c) Now consider the series $(\log X^{(1)}, \log X^{(2)})$. Build a bivariate VARMA(p,q) model for the series.

4. The monthly log returns of IBM stock and the S&P 500 index from January 1926 - December 2008, with 996 observations can be found in `m-ibmsp2608.txt`.

   (a) Plot the two time series, one above the other. Is there evidence of similarity between the two series?

   (b) Make scatterplots of IBM versus S& P 500 at lags of 0 and $\pm 1$. Make a scatter plot of S& P 500 versus S & P 500 at lag 1. Compute the acf of the bivariate time series. What are your conclusions?

   (c) Try fitting a VAR model to the data. Which model seems most appropriate?

   (d) Consider the coefficients for the VAR(5) model. Is it reasonable to take $\Phi_4 = 0$? Is there evidence that the current return of IBM stock is influence by past behaviour of the market index? Does the current return of IBM stock influence future behaviour of the market index?

   (e) Construct a VMA(5) model for the IBM, S&P bivariate series.

5. (a) Using the `VAR` function in the `vars` package, fit a multivariate VAR model to the four economic variables in the Canadian data (which can be loaded from within the `vars` package with the command `data(Canada)`).

   (b) Using the fitted VAR model, make predictions for the next year. Add these predictions to a time series plot of each variable.

6. The figures for the monthly supply of electricity (millions of kWh), beer (Ml) and chocolate-based production (tonnes) in Australia, over the period January 1958 - December 1990 is found in the file `cbe.dat` in the course data directory.

   (a) Fit a suitable SARIMA model (hint: the SARIMA$(1,1,0)(1,1,1)_{12}$ may work well) to the logarithm of the electricity production series. Verify that the residuals are approximately white noise.

   (b) Fit the same model as in (a) to the logarithm of the chocolate production series. Again, verify that the residuals are approximately white noise.

   (c) Plot the cross-correlogram of the residuals of the two fitted ARIMA models, and verify that the lag 0 correlation is significantly different from zero. Give a possible reason why this may happen.

   (d) Forecast values for the next month for each series, and add a simulated bivariate white noise term to each forecast. This gives one possible realisation. Repeat the process ten times to give ten possible future scenarios for the next month's production for each series.