ECON90033 – QUANTITATIVE ANALYSIS OF FINANCE I

TUTORIAL 10

Download the t10e1 Excel data file from the subject website and save it to your computer or USB flash drive. Read this handout and complete the tutorial exercises before your tutorial class so that you can ask for help during the tutorial if necessary.

Vector Autoregression (*VAR*)

On the week 9 lectures we briefly discussed the difference between structural *VAR* and reduced form *VAR*. A structural *VAR* is a set of autoregressive structural equations, i.e., the time path of each left-hand side (endogenous) variable is determined by its own history and by current and past realizations of other left-hand side variables and by an error term.

To keep the discussion simple, we considered a bivariate *SVAR* system with a single lag:

$$y_t = b_{10} - b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \varepsilon_{yt}$$
$$z_t = b_{20} - b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{zt}$$

where $\{y_t\}$, $\{z_t\}$ are supposed to be $I(0)$ variables, and $\{\varepsilon_{yt}\}$, $\{\varepsilon_{zt}\}$ are uncorrelated white-noise errors.

Granted that $b_{12} b_{21} \neq 1$, this *SVAR*(1) system can be manipulated to obtain its reduced form, i.e., *VAR*(1):

$$y_t = a_{10} + a_{11} y_{t-1} + a_{12} z_{t-1} + u_{1t}$$
$$z_t = a_{20} + a_{21} y_{t-1} + a_{22} z_{t-1} + u_{2t}$$

where $\{u_{1t}\}$ and $\{u_{2t}\}$ might be correlated contemporaneously.

In a *VAR* system every equation has the same set of right-hand-side variables, apart from the error term. Consequently, each equation can be estimated individually by OLS. A *SVAR* model, however, cannot be estimated directly due to the (possible) contemporaneous correlation between the endogenous variables. It might be though recovered from the estimated *VAR* model under certain conditions.

Exercise 1

The *t10e1.xlsx* file contains data on the US producer price index (ppi_t) and the *M1* measure of money supply (*m1*) from 1960Q1 through 2002Q1. The goal of this exercise is to model these two variables simultaneously using a *VAR*.

a) Launch *RStudio*, create a new project and script, and name both *t10e1*. Import the data set from the *t10e1.xlsx* file to *RStudio*, save it as *t10e1.RData*, and attach it to your *R* project.

Construct the rate of inflation rate (p_t) and the growth rate of money supply (m_t) as

$$m_t = 100\left[\ln(m1_t) - \ln(m1_{t-1})\right]$$
$$p_t = 100\left[\ln(ppi_t) - \ln(ppi_{t-1})\right]$$

Plot these variables and briefly comment on the figures.

Execute the following commands:

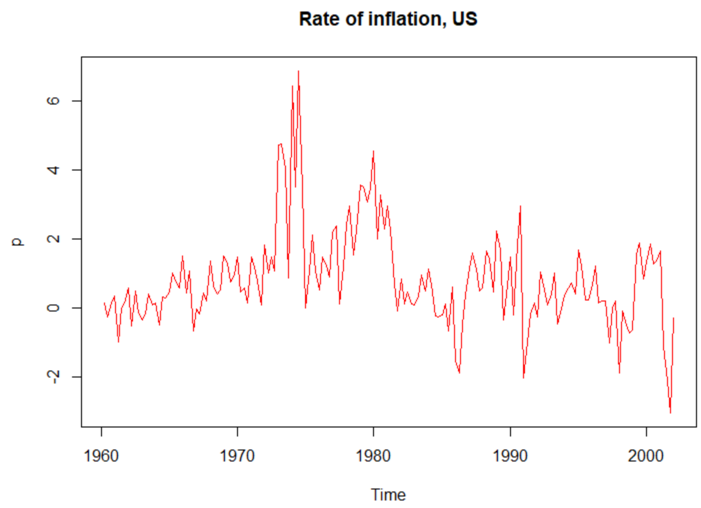
```
attach(t10e1)
ppi = ts(ppi, start = c(1960,1), end = c(2002,1), frequency = 4)
m1 = ts(m1, start = c(1960,1), end = c(2002,1), frequency = 4)

p = 100*diff(log(ppi))
plot(p, main = "Rate of inflation, US", col = "red")
m = 100*diff(log(m1))
plot(m, main = "Rate of money supply, US", col = "darkgreen")
```

You should get the figures on the next page. Apparently, both time series fluctuate around some constant mean.

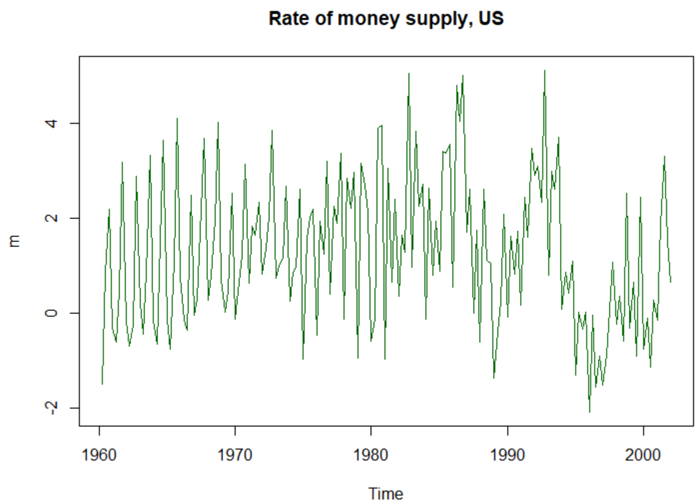
b) Perform the *ADF* and *KPSS* tests on the levels and first differences of *p* and *m*. What conclusions do you draw from these tests about the order of integration of these variables?

The time series plots suggest that for both variables the best is to use Model 2 in the tests on their levels and Model 1 (when possible) on their first differences.

You can perform these tests like in Tutorials 8 and 9,



.



Namely,

ADF tests:

library(urca)
adf_p = ur.df(p, type = "drift", selectlags = "BIC")
summary(adf_p)
adf_dp = ur.df(diff(p), type = "none", selectlags = "BIC")
summary(adf_dp)
adf_m = ur.df(m, type = "drift", selectlags = "BIC")
summary(adf_m)
adf_dm = ur.df(diff(m), type = "none", selectlags = "BIC")
summary(adf_dm)

KPSS tests:

```
kpss_p = ur.kpss(p, type = "mu")
summary(kpss_p)
kpss_dp = ur.kpss(diff(p), type = "mu")
summary(kpss_dp)
kpss_m = ur.kpss(m, type = "mu")
summary(kpss_m)
kpss_dm = ur.kpss(diff(m), type = "mu")
summary(kpss_dm)
```

The printouts are not shown here, but the results are summarised in the table below:[1]

The results are unambiguous, both tests indicate that *p* and *m* are stationary, i.e., *I*(0).

	Number of detected unit roots (α = 0.10)	
p	***ADF***	***KPSS***
Level	0	0
Differenced	0	0
	I(0)	*I*(0)
m	***ADF***	***KPSS***
Level	0	0
Differenced	0	0
	I(0)	*I*(0)

c) Consider a *VAR* model with a constant of *p* and *m*, and determine the optimal lag length with the *VARselect()* function of the *vars* package.

VARselect(y, lag.max = , type =

where y is the list of endogenous variables and type is "const" or "trend" or "both" or "none", returns information criteria and final prediction error for VAR(p) models where the lag length (p) is increased gradually from 1 to k, and k is based on the same sample size.

[1] Take this opportunity to practice and try to verify this table.

Execute,

data = cbind(p, m)
library(vars)
VARselect(data, type = "const")

You should get this printout:

```
$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
    10      5      5     10

$criteria
              1        2        3         4         5         6
AIC(n) 1.195658 1.125373 1.143558 0.6106545 0.4474812 0.4819956
HQ(n)  1.242890 1.204092 1.253765 0.7523485 0.6206628 0.6866647
SC(n)  1.311959 1.319208 1.414927 0.9595577 0.8739185 0.9859669
FPE(n) 3.305764 3.081496 3.138278 1.8420912 1.5650730 1.6205106
               7         8         9        10
AIC(n) 0.4784176 0.4552138 0.4160713 0.4125227
HQ(n)  0.7145743 0.7228581 0.7152030 0.7431420
SC(n)  1.0599230 1.1142533 1.1526448 1.2266302
FPE(n) 1.6153711 1.5791493 1.5195439 1.5154113
```

HQ and *SC* take their smallest values for 5 lags, while *AIC* and *FPE* select 10 lags. Estimate the more parsimonious *VAR*(5) model and test the residuals for first and second order autocorrelation with the Breusch-Godfrey *LM* test.

The

var5 = VAR(data, p = 5, type = "const")
serial.test(var5, lags.bg = 2, type = "BG")

commands return this printout:

```
        Breusch-Godfrey LM test

data:  Residuals of VAR object var5
Chi-squared = 7.2637, df = 8, p-value = 0.5085
```

Given the large *p*-value (0.5085), we maintain the null hypothesis that there is not first and second order residual serial correlation and accept the *VAR*(5) model.

The VAR(5) printout is on the next two pages.

```
VAR Estimation Results:
=========================
Endogenous variables: p, m
Deterministic variables: const
Sample size: 163
Log Likelihood: -474.151
Roots of the characteristic polynomial:
0.9612 0.9255 0.9243 0.9243 0.7266 0.7266 0.6083 0.6083 0.5751 0.02102
Call:
VAR(y = data, p = 5, type = "const")


Estimation results for equation p:
==================================
p = p.l1 + m.l1 + p.l2 + m.l2 + p.l3 + m.l3 + p.l4 + m.l4 + p.l
const

      Estimate Std. Error t value Pr(>|t|)
p.l1   0.42719    0.07999   5.341 3.32e-07 ***
m.l1  -0.01141    0.07446  -0.153   0.8784
p.l2   0.13541    0.08767   1.545   0.1245
m.l2   0.11578    0.05636   2.055   0.0416 *
p.l3   0.11144    0.08712   1.279   0.2028
m.l3   0.07521    0.05712   1.317   0.1899
p.l4   0.17996    0.08651   2.080   0.0392 *
m.l4  -0.08942    0.05754  -1.554   0.1222
p.l5  -0.10288    0.08250  -1.247   0.2143
m.l5   0.08774    0.07550   1.162   0.2470
const -0.01918    0.15937  -0.120   0.9043
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Residual standard error: 1.076 on 152 degrees of freedom
Multiple R-Squared: 0.4605,     Adjusted R-squared: 0.425
F-statistic: 12.97 on 10 and 152 DF,  p-value: 3.256e-16
```

```
Estimation results for equation m:
===================================
m = p.l1 + m.l1 + p.l2 + m.l2 + p.l3 + m.l3 + p.l4 + m.l4 + p.
const

      Estimate Std. Error t value Pr(>|t|)
p.l1  -0.17988    0.07952  -2.262 0.025110 *
m.l1   0.29317    0.07403   3.960 0.000115 ***
p.l2  -0.07998    0.08716  -0.918 0.360261
m.l2   0.04157    0.05603   0.742 0.459278
p.l3  -0.00891    0.08661  -0.103 0.918202
m.l3  -0.04305    0.05678  -0.758 0.449513
p.l4  -0.08691    0.08601  -1.010 0.313890
m.l4   0.72009    0.05720  12.588  < 2e-16 ***
p.l5   0.32073    0.08202   3.910 0.000139 ***
m.l5  -0.25608    0.07505  -3.412 0.000827 ***
const  0.35880    0.15844   2.265 0.024947 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 1.07 on 152 degrees of freedom
Multiple R-Squared: 0.5671,     Adjusted R-squared: 0.5386
F-statistic: 19.91 on 10 and 152 DF,  p-value: < 2.2e-16



Covariance matrix of residuals:
          p         m
p  1.158061 -0.001948
m -0.001948  1.144586

Correlation matrix of residuals:
          p         m
p  1.000000 -0.001692
m -0.001692  1.000000
```

The first part of the printout shows, among others, the lengths of the estimated characteristic roots. Since this is a bivariate system with 5 lags, there are 10 characteristic roots. The absolute values (lengths) of their point estimates are all smaller than one (i.e., inside the unit circle), indicating that this *VAR* is stable.

The second part of the printout shows the two estimated equations of *VAR*(5). Both equations are acceptable as they are strongly significant and have reasonable adjusted R^2 statistics, but there are also many individually insignificant t ratios. This is not unusual in *VAR* models, and it is not an issue because in *VAR* analyses the individual coefficients are of little importance.[2]

[2] Still, to save degrees of freedom, you might wish to eliminate the individually insignificant variables. Remember, however, that if you do so and as a result various equations have different sets of right-hand-side variables, the

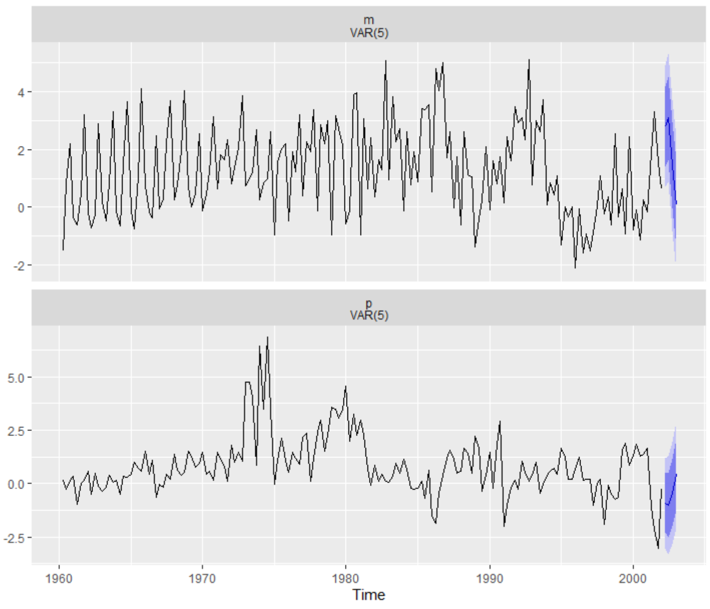
d) Use the estimated *VAR*(5) model to forecast *p* and *m* 1- 4 quarters ahead.

Execute

library(forecast)
var5_ea = forecast(var5, h = 4)
print(var5_ea)
autoplot(var5_ea)

to get

```
p
        Point Forecast     Lo 80     Hi 80     Lo 95    Hi 95
2002 Q2     -0.9192595 -2.298379 0.4598596 -3.028440 1.189921
2002 Q3     -1.0296777 -2.529458 0.4701027 -3.323394 1.264038
2002 Q4     -0.4549721 -2.025189 1.1152450 -2.856412 1.946468
2003 Q1      0.4337493 -1.199049 2.0665476 -2.063400 2.930899

m
        Point Forecast       Lo 80    Hi 80      Lo 95    Hi 95
2002 Q2     2.78926909  1.41819705 4.160341  0.6923954 4.886143
2002 Q3     3.08278112  1.63250881 4.533053  0.8647810 5.300781
2002 Q4     1.45645420 -0.03302575 2.945934 -0.8215088 3.734417
2003 Q1     0.08911688 -1.41868065 1.596914 -2.2168605 2.395094
```



system becomes a near *VAR* and it should be estimated using seemingly unrelated regression (*SUR*).

As usual, the *forecast()* function provides the point predictions and the 80% and 95% prediction bands. On the plots, the narrower dark blue areas illustrate the 80% prediction bands and the light blue areas the 95% prediction bands. Unfortunately, even the narrower 80% prediction bands appear to be quite wide, so the point predictions are of little precision.

Granger Causality

In econometrics causality is a synonym for predictability. Considering two stationary variables, Y and Z, Z is said to be Granger causal to Y (denoted as $Z \rightarrow Y$) if and only if y_{t+1} can be predicted better when the information set includes z_t, z_{t-1}, …. , and Y is said to be Granger causal to Z (denoted as $Y \rightarrow Z$) if and only if z_{t+1} can be predicted better when the information set includes y_t, y_{t-1}, …. If Z is Granger causal to Y and Y is Granger causal to Z, there is a two-way (or feedback) Granger causal relationship (denoted as $Z \leftrightarrow Y$) between the two variables. The last possibility is that there is no Granger causal relationship at all between Y and Z (denoted as $Z \nrightarrow Y$ and $Y \nrightarrow Z$).

When we set up a *VAR* model, the left-hand side variables are considered endogenous variables. Whether they are indeed endogenous, i.e., are determined within the system, as opposed to exogenous variables that are fully determined outside the system, can be checked with joint Granger causality tests. Namely, a variable is an endogenous variable in the given system if the other variables jointly Granger cause it, and it is exogenous otherwise.

Granger causality can be tested with the general F-test or the Wald chi-square test on all lags of a variable (or several variables) jointly. Under the null hypothesis all these lags have zero coefficients, while under the alternative hypothesis some lag(s) has (have) non-zero coefficient(s).

Based on vector autoregression, Granger causality can be tested with the following function of the *bruceR* package:

granger_causality(model, test =)

where model is an estimated VAR object and test is "F" or "Chisq" (the default is both), performs Granger (predictive) causality tests between multivariate time series in VAR framework.

We return to Exercise 1 to illustrate Granger causality testing.

e) Use the $VAR(5)$ model to test for Granger causality between p and m at the 5% significance level.

The

library(bruceR)
granger_causality(var5)

commands return the following printout:

```
> granger_causality(var5)

Granger Causality Test (Multivariate)

F test and Wald χ² test based on VAR(5) model:
──────────────────────────────────────────────────────
                 F df1 df2     p    Chisq df     p
──────────────────────────────────────────────────────
 --------
 p <= m      2.13   5 152  .065 .   10.64  5  .059 .
 p <= ALL    2.13   5 152  .065 .   10.64  5  .059 .
 --------
 m <= p      4.00   5 152  .002 **  20.02  5  .001 **
 m <= ALL    4.00   5 152  .002 **  20.02  5  .001 **
──────────────────────────────────────────────────────
```

The top part of the printout shows the Granger causality test results for the null hypothesis that m does not cause p, and the bottom part shows the Granger causality test results for the null hypothesis that p does not cause m. In both panels two pairs of tests are reported. The first test in each pair is for causality from one variable to the other, and the second for causality from all other variables to the one on the left. In this case, however, there are only two variables, so in the top and bottom parts alike the two tests are the same.

This time the *F* and *Chisq* tests lead to the same conclusions at the conventional significance levels. Namely, at the 5% level both tests indicate that m is not causing p (i.e., $m \nrightarrow p$) but p is causing m ($p \rightarrow m$). At the 10% significance level, however, both tests indicate that m is causing p and p is causing m, so there is a two-way (feedback) Granger causal relationship between them ($p \leftrightarrow m$).

These results imply that in this bivariate *VAR* system of the rate of inflation rate (p_t) and the growth rate of money supply (m_t), at the 10% significance level p and m both prove to be endogenous variables, while at the 5% level m is still endogenous while p appears to be exogenous.

In this example there is agreement between the *F* and *Chisq* versions of the Granger causality test. Occasionally, however, they might lead to contradicting, or at least ambiguous, conclusions and it might be important to decide which test to rely on. To make this call, remember that the *F* test assumes normality while the *Chisq* does not.

After having estimated a *VAR* model, multivariate Jarque-Bera tests and multivariate skewness and kurtosis tests for the residuals can be performed with the *normality.test* function. The Jarque-Bera test, in general, is based on the comparison of the skewness and kurtosis statistics to the skewness and kurtosis parameters of a multivariate normal distributions whose expected values and standard deviations are equal to the corresponding sample means and sample standard deviations.

This time the

normality.test(var5)

command returns the following printout:

```
$JB

        JB-Test (multivariate)

data:  Residuals of VAR object var5
Chi-squared = 133.07, df = 4, p-value < 2.2e-16


$Skewness

        Skewness only (multivariate)

data:  Residuals of VAR object var5
Chi-squared = 1.3049, df = 2, p-value = 0.5208


$Kurtosis

        Kurtosis only (multivariate)

data:  Residuals of VAR object var5
Chi-squared = 131.76, df = 2, p-value < 2.2e-16
```

The *p*-value of the multivariate *JB* test is practically zero, so the null hypothesis of normality can be safely rejected. The middle and bottom part of this printout focus on the two crucial components of the *JB* test, skewness and kurtosis. The *p*-value for skewness is 0.5208, quite large, so in terms of skewness the residuals might be normally distributed. The *p*-value of kurtosis, however, is practically zero, so in terms of kurtosis the residuals are not normally distributed, and that's why the *JB* test rejects normality.

Save your *R* code and quit *RStudio*.