

## 06 Multivariate models: Granger causality, VAR and VECM models

Andrius Buteikis, [andrius.buteikis@mif.vu.lt](mailto:andrius.buteikis@mif.vu.lt)  
<http://web.vu.lt/mif/a.buteikis/>

# Introduction

In the present chapter, we discuss methods which involve more than one equation.

To motivate why multiple equation methods are important, we begin by discussing **Granger causality**.

Later, we move to discussing the most popular class of multiple-equation models: so-called Vector Autoregressive (VAR) models. VARs can be used to investigate Granger causality, but are also useful for many other things in finance.

Time series models for integrated series are usually based on applying VAR to first differences. However, differencing eliminates valuable information about the relationship among integrated series - this is where **Vector Error Correction model (VECM)** is applicable.

## Granger Causality

In our discussion of regression, we were on a little firmer ground, since we attempted to use common sense in labeling one variable the dependent variable and the others the explanatory variables. In many cases, because the latter 'explained' the former it was reasonable to talk about  $X$  'causing'  $Y$ . For instance, the price of the house can be said to be 'caused' by the characteristics of the house (e.g., number of bedrooms, number of bathrooms, etc.).

However, one can run a regression of  $Y =$  stock prices in Country B on  $X =$  stock prices in Country A. It is possible that stock price movements in Country A cause stock markets to change in Country B (i.e.,  $X$  causes  $Y$ ). For instance, if Country A is a big country with an important role in the world economy (e.g., the USA), then a stock market crash in Country A could also cause panic in Country B. However, if Country A and Country B were neighboring countries (e.g., Thailand and Malaysia) then an event which caused panic in either country could affect both countries.

In other words, the causality could run in either direction - or both! Hence, when using the word 'cause' with regression or correlation results a great deal of caution has to be taken and common sense has to be used.

However, with time series data we can make slightly stronger statements about causality simply by exploiting the fact that time does not run backward! That is, if event A happens before event B, then it is possible that A is causing B. However, it is not possible that B is causing A. In other words, events in the past can cause events to happen today. Future events cannot.

These intuitive ideas can be investigated through regression models incorporating the notion of **Granger** or regressive **causality**. The basic idea is that a variable **X Granger causes Y** if past values of X can help explain Y.

Of course, if Granger causality holds this does not guarantee that X causes Y. This is why we say 'Granger causality' rather than just 'causality'. Nevertheless, if past values of X have explanatory power for current values of Y, it at least suggests that X might be causing Y.

Granger causality is only relevant with time series variables.

To illustrate the basic concepts we will consider Granger causality between two variables ( $X$  and  $Y$ ) which are both stationary. A nonstationary case, where  $X$  and  $Y$  have unit roots but are cointegrated, will be mentioned below. Since  $X$  and  $Y$  are both stationary, an ADL model is appropriate. Suppose that the following simple ADL (only lags on the right hand side!) model holds:

$$Y_t = \alpha + \phi Y_{t-1} + \beta_1 X_{t-1} + \epsilon_t$$

This model implies that last period's value of  $X$  has explanatory power for the current value of  $Y$ . The coefficient  $\beta_1$  is a measure of the influence of  $X_{t-1}$  on  $Y_t$ . If  $\beta_1 = 0$ , then past values of  $X$  have no effect on  $Y$  and there is **no way** that  $X$  could Granger cause  $Y$ .

In other words, if  $\beta_1 = 0$  then  $X$  does not Granger cause  $Y$ .

Since we know how to estimate the ADL and carry out hypothesis tests, it is simple to test Granger causality or, in other words, to test  $H_0 : \beta_1 = 0$ : if  $\hat{\beta}_1$  is statistically **significant** (i.e. its  $p$ -value  $< 0.05$ ), then we conclude that  $X$  Granger causes  $Y$ .

Note that the null hypothesis being tested here is hypothesis that Granger causality **does not occur**. We will refer to this procedure as a **Granger causality test**.

In general, we could assume that the  $(X, Y)$  interaction is described by an  $ADL(p, q)$  model of the form (only lags on the right hand side!) [this is also called the **unrestricted model**]:

$$Y_t = \alpha + \delta t + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \beta_1 X_{t-1} + \dots + \beta_q X_{t-q} + \epsilon_t$$

we say that **X does not Granger cause Y** if all  $\beta_i = 0$ . Using the previously described lag selection technique for an ADL model, we can test the joint significance of  $\hat{\beta}_i$ : we conclude that X Granger causes Y if any (or all) of  $\hat{\beta}_1, \dots, \hat{\beta}_q$  are statistically significant. If X at any time in the past has explanatory power for the current value of Y, then we say that X Granger causes Y. Since we are assuming X and Y do not contain unit roots, OLS regression analysis can be used to estimate this model [also called the **restricted model**]:

$$Y_t = \alpha + \delta t + \phi Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t$$

We do not reject the null hypothesis  $H_0 : \beta_1 = 0, \dots, \beta_q = 0$  if the models are 'more or less the same', i.e. if  $SSR_{UR} \approx SSR_R$ .

The most popular here is the  $F$  test: if test statistic

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(T - Q - (p + 2))}$$

is greater than the 0.95 quantile of the  $F$  distribution with  $(q, T - q - (p + 2))$ , we say that  $X$  Granger causes  $Y$ .

In many cases, it is not obvious which way causality should run. For instance, should stock markets in country A affect markets in country B or should the reverse hold?

In such cases, when causality may be in either direction, it is important that you check for it. If  $Y$  and  $X$  are the two variables under study, in addition to running a regression of  $Y$  on lags of itself and lags of  $X$  (as above), you should also run a regression of  $X$  on lags of itself and lags of  $Y$ .

In other words, you should work with two separate equations: one with  $Y$  being the dependent variable and one with  $X$  being the dependent variable.

This brief discussion of Granger causality has focused on two variables, X and Y . However, there is no reason why these basic techniques cannot be extended to the case of many variables.

For instance, if we had three variables, X , Y and Z , and were interested in investigating whether X or Z Granger cause Y, we would simply regress Y on lags of Y , lags of X and lags of Z .

If, say, the lags of Z were found to be significant and the lags of X not, then we could say that Z Granger causes Y, but X does not.



Testing for Granger causality among **cointegrated variables** is very similar to the method outlined above. Remember that, if variables are found to be cointegrated (something which should be investigated using unit root and cointegration tests), then you should work with an error correction model (ECM) involving these variables. In the case where you have two variables, this is given by:

$$\Delta Y_t = \phi + \delta t + \lambda e_{t-1} + \gamma_1 \Delta Y_{t-1} + \dots + \gamma_p \Delta Y_{t-p} + \omega_1 \Delta X_{t-1} + \dots + \omega_q \Delta X_{t-q} + \epsilon_t$$

This is essentially an ADL model except for the presence of the term  $\lambda e_{t-1}$ , where  $e_{t-1} = Y_{t-1} - \alpha - \beta X_{t-1}$ .

X Granger causes Y if past values of X have explanatory power for current values of Y. Applying this intuition to the ECM, we can see that past values of X appear in the terms  $\Delta X_{t-1}, \dots, \Delta X_{t-q}$  and  $e_{t-1}$ . This implies that X does not Granger cause Y if  $\omega_1 = \dots = \omega_q = \lambda = 0$ .

t-statistics and p-values can be used to test for Granger causality in the same way as the stationary case. Also, the F - tests can be used to carry out a formal test of  $H_0 : \omega_1 = \dots = \omega_q = \lambda = 0$

**The bottom line - if X Granger-causes Y , this does not mean that X causes Y, it only means that X improves Y 's predictability (i.e., reduces residuals of the model).**

## VAR: Estimation and forecasting

Our discussion of Granger causality naturally leads us to an interest in models with several equations and the topic of **Vector Autoregressions or VARs**.

Initially, we will assume that all variables are stationary. If the original variables have unit roots, then we assume that differences have been taken such that the model includes the changes in the original variables (which do not have unit roots). The end of this section will consider the extension of this case to that of cointegration.

When we investigated Granger causality between X and Y, we began with an  $ADL(p, q)$  model for Y as the dependent variable. We used it to investigate if X Granger caused Y. We then went on to consider causality in the other direction, which involved switching the roles of X and Y in the ADL. In particular, X became the dependent variable. We can write the two equations as follows:

$$Y_t = \alpha_1 + \delta_1 t + \phi_{11} Y_{t-1} + \dots + \phi_{1p} Y_{t-p} + \beta_{11} X_{t-1} + \dots + \beta_{1q} X_{t-q} + \epsilon_{1t}$$
$$X_t = \alpha_2 + \delta_2 t + \phi_{21} Y_{t-1} + \dots + \phi_{2p} Y_{t-p} + \beta_{21} X_{t-1} + \dots + \beta_{2q} X_{t-q} + \epsilon_{2t}$$

The first of these equations tests whether X Granger causes Y; the second, whether Y Granger causes X. Note that now the coefficients have subscripts indicating which equation they are in. The errors now have subscripts to denote the fact that they will be different in the two equations.

These two equations comprise a VAR. A VAR is the extension of the autoregressive (AR) model to the case in which there is more than one variable under study. A VAR has more than one dependent variable (e.g., Y and X) and, thus, has more than one equation. Each equation uses as its explanatory variables lags of all the variables under study (and possibly a deterministic trend).

The term 'VAR' becomes more transparent if we use a matrix notation. A **first order VAR in two variables** would be given by:

$$\begin{aligned} Y_t &= \alpha_1 + \phi_{11} Y_{t-1} + \phi_{12} X_{t-1} + \epsilon_{1t} \\ X_t &= \alpha_2 + \phi_{21} Y_{t-1} + \phi_{22} X_{t-1} + \epsilon_{2t} \end{aligned}$$

where  $\epsilon_{1t}$  and  $\epsilon_{2t}$  are two white noise processes (independent of the history of X and Y) that **may be correlated**.

If, for example  $\phi_{12} \neq 0$ , this means that the history of X helps explaining Y, that is, X is a Granger cause of Y. The system can be written as:

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

or, with relevant definitions, as:

$$\vec{Y}_t = \vec{\alpha} + \Theta_1 \vec{Y}_{t-1} + \vec{\epsilon}_t$$

This extends the first order autoregressive model AR(1) to the higher dimensional case.

In general, a VAR( $p$ ) model for a  $d$ -dimensional vector  $\vec{Y}_t$  is given by:

$$\vec{Y}_t = \vec{\alpha} + \vec{\delta}t + \Theta_1 \vec{Y}_{t-1} + \dots + \Theta_p \vec{Y}_{t-p} + \vec{\epsilon}_t$$

where  $\Theta_j$  is a  $d \times d$  matrix and  $\vec{\epsilon}_t$  is a  $d$  - dimensional vector of white noise terms with a covariance matrix  $\Sigma$ .

Similarly to one-dimensional case, a VAR( $p$ ) is **stationary** if all the roots of the equation  $\det(I_k - \Theta_1 z - \Theta_2 z^2 - \dots - \Theta_p z^p) = 0$  are **outside** a unit complex circle. The VAR is said to have a single unit root if the above equation has exactly **one** root  $z = +1$ , i.e.

$$z = +1 \iff \det(I_k - \Theta_1 - \Theta_2 - \dots - \Theta_p) = 0$$

This will hold, if **at least one** of the variables in the VAR contains a unit root.

Why we would want to work with such models? One reason has to be Granger causality testing. That is, VARs provide a framework for testing for Granger causality between each set of variables.

Determining the lag length  $p$  in an empirical application is not always easy and univariate autocorrelation or partial autocorrelation functions will not help. A reasonable strategy is to estimate a VAR model for different values of  $p$  and then select on the basis of the Akaike or Schwarz information criteria.

Once the order  $p$  has been established, we have to estimate the coefficients. It appears that to get BLUE estimators, we can apply OLS to every equation individually (this is what, in R, VAR of the the package `vars` or `lineVar` of the package `tsDyn`, do).

Below we shall examine forecasting a VAR(2) model, but at first we present a brief introduction to some of the practical issues and intuitive ideas relating to forecasting. All our discussion will relate to forecasting with VARs but it is worth noting that the ideas also relate to forecasting with univariate time series models. After all, an AR model is just a VAR with only one equation.

## Forecasting

Forecasting is usually done using time series variables. The idea is that you use your observed data to predict what you expect to happen in the future. In more technical terms, you use data for periods  $t = 1, \dots, T$  to forecast periods  $T + 1, T + 2, \dots$ . To provide some intuition for how forecasting is done, consider a VAR(1) involving two variables,  $Y$  and  $X$ :

$$\begin{aligned}Y_t &= \alpha_1 + \delta_1 t + \phi_{11} Y_{t-1} + \phi_{12} X_{t-1} + \epsilon_{1t} \\X_t &= \alpha_2 + \delta_2 t + \phi_{21} Y_{t-1} + \phi_{22} X_{t-1} + \epsilon_{2t}\end{aligned}$$

You cannot observe  $Y_{T+1}$  but you want to make a guess of what it is likely to be. Using the first equation of the VAR and setting  $t = T + 1$ , we obtain an expression for  $Y_{T+1}$ :

$$Y_{T+1} = \alpha_1 + \delta_1(T + 1) + \phi_{11} Y_T + \phi_{12} X_T + \epsilon_{1,T+1}$$

This equation cannot be directly used to obtain  $Y_{T+1}$  since we don't know what unpredictable shock  $\epsilon_{1,T+1}$  or surprise will hit the economy next period. Furthermore, we do not know what the coefficients are.

However, if we ignore the error term (which cannot be forecast since it is unpredictable) and replace the coefficients by their estimates we obtain a forecast which we denote as  $\hat{Y}_{T+1}$ :

$$\hat{Y}_{T+1} = \hat{\alpha}_1 + \hat{\delta}_1(T + 1) + \hat{\phi}_{11}Y_T + \hat{\phi}_{12}X_T$$

We can use the same strategy for two periods, provided that we make one extension - since our data only runs until period  $T$ , we do not know what  $Y_{T+1}$  and  $X_{T+1}$  are. Consequently, we replace  $Y_{T+1}$  and  $X_{T+1}$  by  $\hat{Y}_{T+1}$  and  $\hat{X}_{T+1}$  (this called a dynamic forecast). The same can be done to calculate  $\hat{X}_{T+2}$ :

$$\begin{aligned}\hat{Y}_{T+2} &= \hat{\alpha}_1 + \hat{\delta}_1(T + 2) + \hat{\phi}_{11}\hat{Y}_{T+1} + \hat{\phi}_{12}\hat{X}_{T+1} \\ \hat{X}_{T+2} &= \hat{\alpha}_2 + \hat{\delta}_2(T + 2) + \hat{\phi}_{21}\hat{Y}_{T+1} + \hat{\phi}_{22}\hat{X}_{T+1}\end{aligned}$$

We can use the general strategy of ignoring the error, replacing coefficients by their estimates and replacing lagged values of variables that are unobserved by forecasts, to obtain forecasts for any number of periods in the future for any VAR(p).



# VAR: Summary

Building a VAR model involves three steps:

- (a) use some information criterion to identify the order,
- (b) estimate the specified model by using the least squares method and, if necessary, re-estimate the model by removing statistically insignificant parameters
- (c) use the Portmanteau test statistic of the residuals to check the adequacy of a fitted model (this is a multivariate analogue of the Ljung-Box Q-stat in an ARIMA model and is to test for autocorrelation and cross-correlation in residuals). If the fitted model is adequate, then it can be used to obtain forecasts.

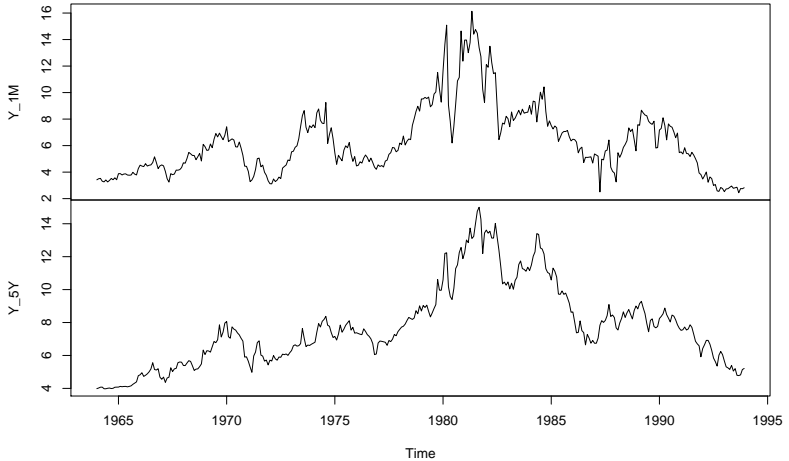
## Example

The data file `us-tbill.txt` contains monthly, 1964:01 through 1993:12, interest rates of US treasure bills for maturities of one month `Y_1M` and five years `Y_5Y`. Both series are integrated, therefore we fit a VAR(p) model to the **first differences**.

```
suppressPackageStartupMessages({
  suppressWarnings({
    library(forecast)
    library(urca)
    library(vars)
  })
})
txt1 <- "http://uosis.mif.vu.lt/~rlapinskas/(data%20R&GRETTL/"
txt2 <- "us-tbill.txt"
rate <- read.table(paste0(txt1, txt2), header = TRUE)
rate <- ts(rate, start = 1964, frequency = 12)
rate <- rate[, c(2,4)]
```

```
plot.ts(rate, main = "Interest rates of 1M & 5Y US T-bills")
```

Interest rates of 1M & 5Y US T-bills



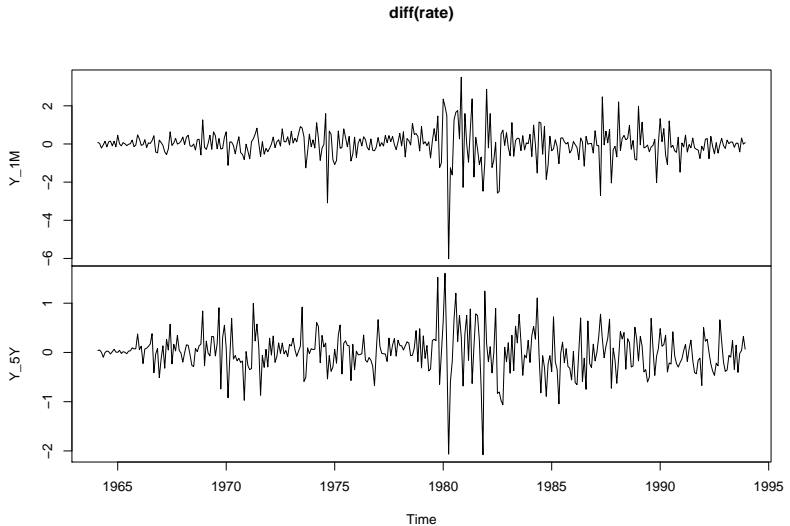
To select the order of the VAR process, we use the VARselect function and Schwarz's SC statistic:

```
d.rate = diff(rate)
VARselect(d.rate, lag.max = 4, type="const")
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      4      1      1      4
##
## $criteria
##              1              2              3              4
## AIC(n) -2.34112246 -2.34704986 -2.34649802 -2.3542181
## HQ(n)  -2.31508701 -2.30365745 -2.28574864 -2.2761117
## SC(n)  -2.27567822 -2.23797612 -2.19379478 -2.1578853
## FPE(n)  0.09621965  0.09565129  0.09570471  0.0949698
```

The \$selection shows the selected VAR order for different statistics. Since  $SC(n) = -2.3022769$  and it is the smallest  $SC(n)$  value, when  $n = 1$ , we are going with a VAR(1) model.

```
plot.ts(d.rate, main = "diff(rate)")
```



Because d.rate does not have a drift, we will create a VAR model without one:

```
var.diff = VAR(d.rate, p = 1, type = "none")
coefficients(var.diff)
```

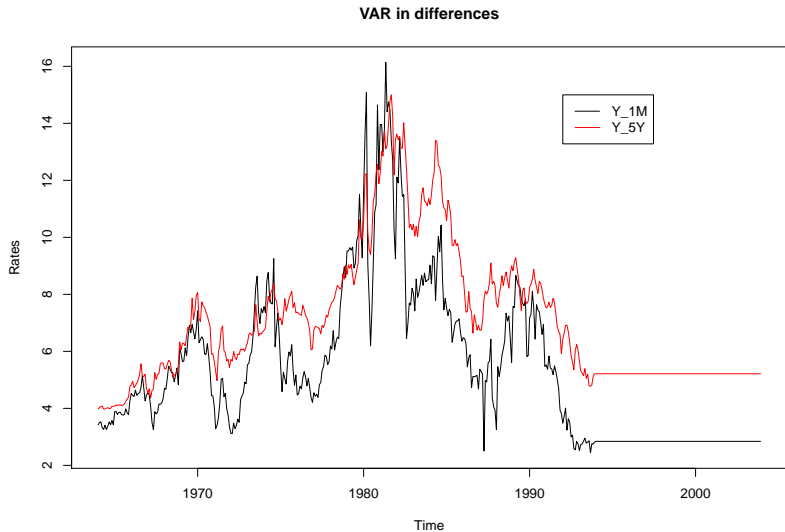
```
## $Y_1M
##           Estimate Std. Error   t value   Pr(>|t|)
## Y_1M.l1 -0.2341646 0.05524366 -4.238760 2.867202e-05
## Y_5Y.l1  0.6343748 0.10732304  5.910891 7.970641e-09
##
## $Y_5Y
##           Estimate Std. Error   t value   Pr(>|t|)
## Y_1M.l1 -0.007842584 0.02987634 -0.2625015 0.7930866
## Y_5Y.l1  0.074123953 0.05804140  1.2770876 0.2024036
```

We note that the estimation results show that **Y\_5Y is the Granger cause** of Y\_1M (in the 1st model the p-value of Y\_5Y.l1 is  $< 0.05$ ) but **not** vice versa (in the 2nd model the p-value of Y\_5Y.l1 is  $> 0.05$ ) (the R-squared is also larger in the 1st equation).

The model `var.diff` is for differences. Since  $Y_{T+h} = \Delta Y_{T+h} + \dots + \Delta Y_{T+1} + Y_T$ , the forecast  $Y_{T+h,T}$  equals the **cumulative sum** of the forecasts for differences:

```
N <- 120 #forecast for 120 periods
NN <- nrow(rate)
var.pred = predict(var.diff, n.ahead = N, ci = 0.95)
R1.d = numeric(NN+N); R2.d = numeric(NN+N)
# insert historical data
R1.d[1:NN] = rate[,1]
R2.d[1:NN] = rate[,2]
# predict levels from differences
R1.d[(NN+1):(NN+N)] = R1.d[NN] + cumsum(var.pred$fcst[["Y_1M"]][,1])
R2.d[(NN+1):(NN+N)] = R2.d[NN] + cumsum(var.pred$fcst[["Y_5Y"]][,1])
#Transform to a time series
Ra1.d=ts(R1.d,start=1964,freq=12)
Ra2.d=ts(R2.d,start=1964,freq=12)
```

```
plot(Ra1.d,ylab="Rates",main="VAR in differences")  
lines(Ra2.d,col=2)  
legend(1992,15,c("Y_1M","Y_5Y"),lty=1,col=1:2)
```





Note that VAR model should always include variables with **the same order of integration** - this allows us to create a model in *levels* (instead of differences).

```
VARselect(rate, lag.max = 4, type="const")
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      2      2      2      2
##
## $criteria
##              1              2              3              4
## AIC(n) -2.33873185 -2.41654626 -2.41234643 -2.40861749
## HQ(n)  -2.31275339 -2.37324883 -2.35173002 -2.33068212
## SC(n)  -2.27342403 -2.30769989 -2.25996151 -2.21269403
## FPE(n)  0.09644995  0.08922959  0.08960571  0.08994149
```

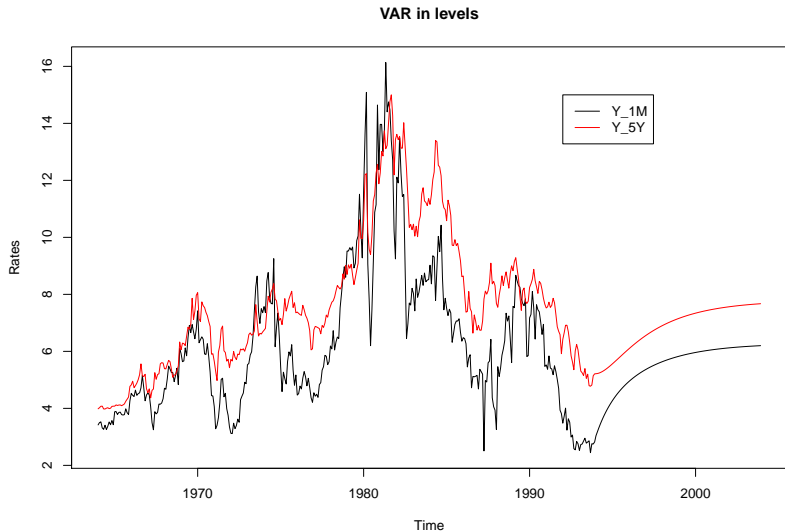
For the levels, a VAR(2) model is more appropriate.

```
var.lev=VAR(rate, p = 2, type = "const")
coefficients(var.lev)
```

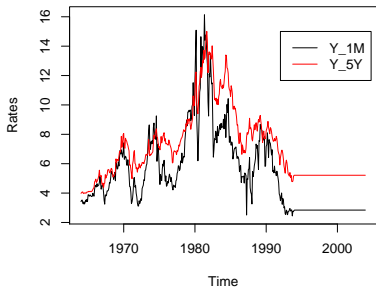
```
## $Y_1M
##           Estimate Std. Error   t value   Pr(>|t|)
## Y_1M.l1  0.7103028 0.05664413 12.5397410 4.364259e-30
## Y_5Y.l1  0.7025051 0.10873206  6.4608826 3.455831e-10
## Y_1M.l2  0.1796765 0.05690023  3.1577469 1.726953e-03
## Y_5Y.l2 -0.6292672 0.10615386 -5.9278785 7.305456e-09
## const    0.1209171 0.13943867  0.8671706 3.864376e-01
##
## $Y_5Y
##           Estimate Std. Error   t value   Pr(>|t|)
## Y_1M.l1  0.01033561 0.03085654  0.3349568 7.378566e-01
## Y_5Y.l1  1.03517258 0.05923111 17.4768387 1.016894e-49
## Y_1M.l2  0.02965198 0.03099604  0.9566375 3.394049e-01
## Y_5Y.l2 -0.09011028 0.05782665 -1.5582829 1.200623e-01
## const    0.17660319 0.07595834  2.3250005 2.063839e-02
```

```
var.pred.lev=predict(var.lev, n.ahead = N, ci = 0.95)
R1.lev=numeric(NN+N)
R2.lev=numeric(NN+N)
R1.lev[1:NN]=rate[,1]
R2.lev[1:NN]=rate[,2]
R1.lev[(NN+1):(NN+N)]=var.pred.lev$fcst[["Y_1M"]][,1]
R2.lev[(NN+1):(NN+N)]=var.pred.lev$fcst[["Y_5Y"]][,1]
Ra1.lev=ts(R1.lev,start=1964,freq=12)
Ra2.lev=ts(R2.lev,start=1964,freq=12)
```

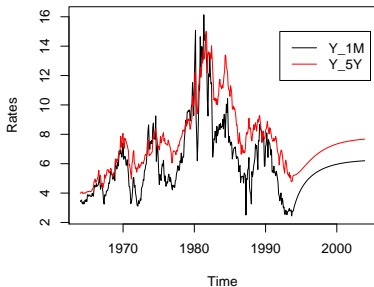
```
plot(Ra1.lev,ylab="Rates",main="VAR in levels")  
lines(Ra2.lev,col=2)  
legend(1992,15,c("Y_1M","Y_5Y"),lty=1,col=1:2)
```



VAR in differences



VAR in levels



Both levels revert to (the values close to) their means (i.e. a stationary behavior) which contradicts the unit root (i.e. non-stationary) behavior of each series.

The explanation lies in the fact that we estimated an unrestricted VAR model while actually the coefficients should reflect **cointegration** and obey some constraints (this will be later discussed in the VECM section).

## VAR: Impulse-Response Function

The **impulse-response function** is yet another device that helps us to learn about the dynamic properties of vector autoregressions of interest to forecasters. The question of interest is simple and direct: How does a unit innovation to a series affect it, now and in the future?

To clarify the issue, let us start with one-dimensional case. Let  $Y_1 = \dots = Y_{T-1} = 0$ ,  $\epsilon_1 = \dots = \epsilon_{T-1} = 0$  and at moment  $t = T$  a **unit shock** comes:  $\epsilon_T = \sigma$ ,  $\epsilon_{T+1} = \dots = 0$ .

- ▶ If  $Y_t = \epsilon_t$ , i.e.  $Y_t$  is a WN, then  $Y_T = \sigma$ ,  $Y_{T+1} = \epsilon_{T+1} = 0$  and  $Y_{T+h} = \epsilon_{T+h} = 0$  - WN has no memory, no dynamics;
- ▶ If  $Y_t = \phi Y_{t-1} + \epsilon_t$ ,  $|\phi| < 1$ , i.e.  $Y_t$  is an AR(1), then  $Y_T = \phi \cdot 0 + \sigma = \sigma$ ,  $Y_{T+1} = \phi\sigma + 0 = \phi\sigma$ ,  $\dots$ ,  $Y_{T+h} = \phi^h\sigma$  as  $h \rightarrow \infty$  - the impulse response is dying down.

Now consider again the two-variable, first-order system:

$$\begin{aligned} Y_t &= \phi_{11} Y_{t-1} + \phi_{12} X_{t-1} + \epsilon_{1t} \\ X_t &= \phi_{21} Y_{t-1} + \phi_{22} X_{t-1} + \epsilon_{2t} \end{aligned}$$

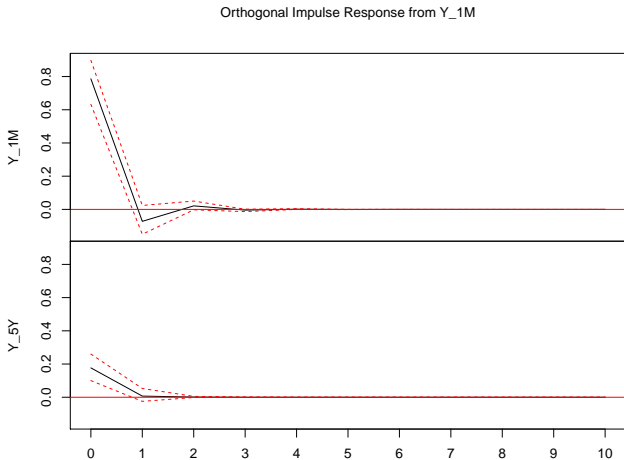
- ▶ A perturbation in  $\epsilon_{1t}$  has an immediate one-for-one effect on  $Y_t$ , but no effect on  $X_t$ .
- ▶ In period  $t + 1$ , that perturbation in  $Y_t$  affects  $Y_{t+1}$  through the 1st equation, and also affects  $X_{t+1}$ , through the 2nd equation.
- ▶ These effects work through to period  $t + 2$  etc.

Thus a perturbation in one innovation in the VAR sets up a chain reaction over time in all variables in the VAR. Impulse response functions calculate these chain reactions.

## Example (continued):

The IRF (impulse-response function) can be calculated in R. The reaction to the unit Y\_1M impulse:

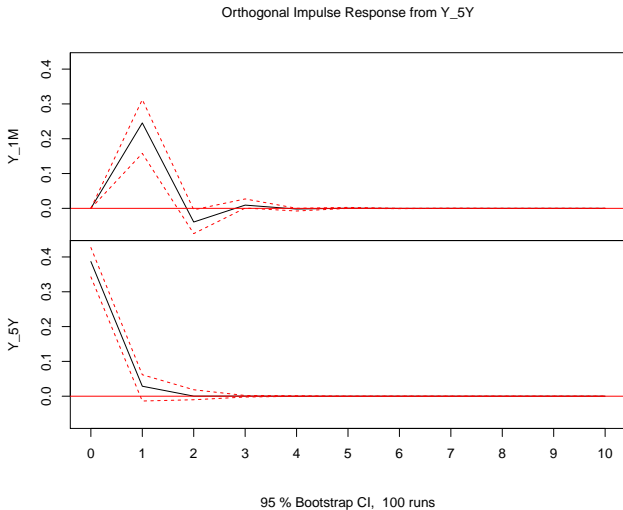
```
plot(irf(var.diff, impulse = "Y_1M"))
```



95 % Bootstrap CI, 100 runs

The reaction to the unit Y\_5Y impulse:

```
plot(irf(var.diff, impulse = "Y_5Y"))
```





## Example 2

Suppose our VAR(1) model is:

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} = \begin{pmatrix} 0.4 & 0.1 \\ 0.2 & 0.5 \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}, \quad \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 16 & 14 \\ 14 & 25 \end{pmatrix} \right)$$

We can show that this model is stationary:

$$\begin{aligned} \det \left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.5 \end{bmatrix} z \right) &= \det \left( \begin{bmatrix} 1 - 0.4z & -0.1z \\ -0.2z & 1 - 0.5z \end{bmatrix} \right) \\ &= 1 - 0.5z - 0.4z + 0.2z^2 - 0.02z^2 = 0.18z^2 - 0.9z + 1 \end{aligned}$$

```
# Find roots of equation: 1 - 0.9 * z + 0.18 * z^2 = 0  
roots <- polyroot(c(1, -0.9, 0.18))  
paste0("Absolute: |z_i| = ", round(abs(roots), 4))
```

```
## [1] "Absolute: |z_i| = 1.6667" "Absolute: |z_i| = 3.3333"
```

Because the roots are greater than one in absolute value, the VAR(1) process is stationary.

If we set:

$$(Y_0, X_0) = (0, 0), \vec{\epsilon}_1 = (4, 0)', \vec{\epsilon}_j = 0, j > 1$$

i.e. initialize  $\vec{Y}_t$  values at zero and set a one-standard-deviation innovation in the first equation and a zero innovation in the second equation in period one and assume no further shocks from innovations.

The first few  $\vec{Y}_t$  values are:

$$\begin{pmatrix} Y_1 \\ X_1 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \end{pmatrix}, \begin{pmatrix} Y_2 \\ X_2 \end{pmatrix} = \Theta_1 \begin{pmatrix} Y_1 \\ X_1 \end{pmatrix} + \vec{\epsilon}_2 = \begin{pmatrix} 1.6 \\ 0.8 \end{pmatrix}, \begin{pmatrix} Y_3 \\ X_3 \end{pmatrix} = \begin{pmatrix} 0.72 \\ 0.72 \end{pmatrix}$$

If we set:

$$(Y_0, X_0) = (0, 0), \vec{\epsilon}_1 = (0, 5)', \vec{\epsilon}_j = 0, j > 1$$

Then:

$$\begin{pmatrix} Y_1 \\ X_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \end{pmatrix}, \begin{pmatrix} Y_2 \\ X_2 \end{pmatrix} = \Theta_1 \begin{pmatrix} Y_1 \\ X_1 \end{pmatrix} + \vec{\epsilon}_2 = \begin{pmatrix} 0.5 \\ 2.5 \end{pmatrix}, \begin{pmatrix} Y_3 \\ X_3 \end{pmatrix} = \begin{pmatrix} 0.45 \\ 1.35 \end{pmatrix}$$

An objection to the procedure just illustrated for the computation `irf` is that the innovations in the VAR are, in general, not contemporaneously independent of one another, i.e. the shock covariance matrix  $\Sigma$  is **not diagonal**. So cases when one innovation receives a perturbation and the other does not is implausible. **Ignoring this gives incorrect irf results!**

A widely used solution is to transform  $\vec{\epsilon}_t$  to produce a new set of uncorrelated **unit variance** innovations  $\vec{u}_t$  (i.e. diagonalization of  $\vec{\epsilon}_t$ ).

Let:  $\mathbb{E}(u_{1,t}, u_{2,t}) = (0, 0)$  and  $\Sigma_u = I$ . We want to find a matrix **P** such that:

$$\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} = \mathbf{P} \vec{u}_t = \begin{pmatrix} p_{11} & 0 \\ p_{21} & p_{22} \end{pmatrix} \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

$\Downarrow$

$$\epsilon_{1,t} = p_{11} u_{1,t}, \quad \epsilon_{2,t} = p_{21} u_{1,t} + p_{22} u_{2,t}$$

And the covariance matrix of  $\mathbf{P} \vec{u}_t$  is the same as the (sample) covariance matrix, i.e.  $\Sigma$  (or  $\hat{\Sigma}$ ). So, we need to solve:

$$\text{Var}(\epsilon_{1,t}) = p_{11}^2 = \sigma_1^2$$

$$\text{Var}(\epsilon_{2,t}) = p_{21}^2 + p_{22}^2 = \sigma_2^2$$

$$\text{Cov}(\epsilon_{1,t}, \epsilon_{2,t}) = \mathbb{E}[p_{11} u_{1,t} \cdot (p_{21} u_{1,t} + p_{22} u_{2,t})] = p_{11} p_{21} = \sigma_{12}$$

its solution is:  $p_{11} = \sigma_1$ ,  $p_{21} = \sigma_{12}/\sigma_1$ ,  $p_{22} = \sqrt{\sigma_2^2 - (\sigma_{12}/\sigma_1)^2}$ .

In higher-dimensional VAR's, the equation that is 1st in the ordering has only one uncorrelated innovation,  $u_{1,t}$ . The equation that is 2nd has only  $u_{1,t}$  and  $u_{2,t}$ , the equation that is 3rd has only  $u_{1,t}$ ,  $u_{2,t}$  and  $u_{3,t}$ , etc.

## Error Covariance Matrix Cholesky Decomposition

What is really important here are the equations relating  $\vec{\epsilon}$  and  $\vec{u}$ :

$$\vec{u}_t = \mathbf{P}^{-1}\vec{\epsilon}_t \iff \vec{\epsilon}_t = \mathbf{P}\vec{u}_t$$

These relations imply the **Cholesky decomposition (factorization)** of the (sample) covariance matrix:

$$\Sigma = \mathbf{P}\mathbf{P}'$$

where  $\mathbf{P}$  - a lower triangular matrix and  $\mathbf{P}'$  - an upper triangular matrix. This imposes an ordering of the variables in the VAR and attributes all of the effect of any common component to the variable that comes first in the VAR system.

This method is called the **error orthogonalization** (or error covariance matrix *diagonalization*) procedure.

Note that responses can change dramatically if you change the ordering of the variables.

## Example 2 (continued)

We can decompose our covariance matrix via Cholesky decomposition using R:

```
Sigma = matrix(c(16, 14, 14, 25), ncol = 2, byrow = TRUE)
print(Sigma)
```

```
##      [,1] [,2]
## [1,]  16  14
## [2,]  14  25
```

```
round(P_mat <- t(chol(Sigma)), 4)
```

```
##      [,1] [,2]
## [1,]  4.0 0.0000
## [2,]  3.5 3.5707
```

(We can also calculate this manually from  $\Sigma = \mathbf{P}\mathbf{P}'$ )

So:

$$\mathbf{P} = \begin{pmatrix} 4 & 0 \\ 3.5 & 3.5707 \end{pmatrix}$$

Suppose, that we postulate  $\vec{u}_1 = (1, 0)'$  and  $\vec{u}_j = (0, 0)'$ ,  $j > 1$ . This vector gives a one standard deviation perturbation in the first component. This results in:

$$\vec{\epsilon}_1 = \mathbf{P}\vec{u}_1 = \begin{pmatrix} 4 & 0 \\ 3.5 & 3.5707 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 3.5 \end{pmatrix}$$

So, the second element in  $\vec{\epsilon}_1$  is now non-zero compared to the case, where we ignored the non-diagonal nature of the error covariance matrix.

We can calculate the values  $\vec{Y}_1, \vec{Y}_2, \dots$  as we did before.

```
Theta_1 = matrix(c(0.4, 0.1, 0.2, 0.5), ncol = 2, byrow = TRUE)
#Y_1:
print(Y_1 <- P_mat %*% c(1, 0))
```

```
##      [,1]
## [1,] 4.0
## [2,] 3.5
```

```
#Y_2:
print(Y_2 <- Theta_1 %*% Y_1)
```

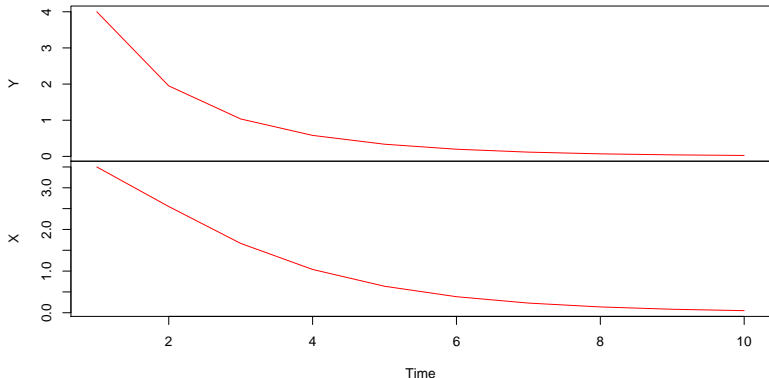
```
##      [,1]
## [1,] 1.95
## [2,] 2.55
```

```
#Y_3:
print(Y_3 <- Theta_1 %*% Y_2)
```

```
##      [,1]
## [1,] 1.035
## [2,] 1.665
```

```
Y_j <- matrix(0, nrow = 10, ncol = 2)
Y_j[1, ] <- Y_1
for(j in 2:10){
  Y_j[j, ] <- Theta_1 %*% Y_j[j - 1, ]
}
colnames(Y_j) <- c("Y", "X")
plot.ts(Y_j, col = "red",
        main = "Orthogonal Impulse Response from (1, 0)")
```

Orthogonal Impulse Response from (1, 0)



Compared with the earlier assumption of a one standard deviation in just  $\epsilon_{11}$ , there is now an important impact on  $X$  in the first period, followed by noticeably greater impacts in subsequent periods.

If a perturbation of 1 standard deviation in the second innovation of  $\vec{\epsilon}_2$ , then:

$$\vec{\epsilon}_1 = \mathbf{P}\vec{u}_1 = \begin{pmatrix} 4 & 0 \\ 3.5 & 3.5707 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 3.5707 \end{pmatrix}$$

```
print(t(Y_1 <- P_mat %*% c(0, 1)))
```

```
##          [,1]      [,2]
## [1,]      0 3.570714
```

```
print(t(Y_2 <- Theta_1 %*% Y_1))
```

```
##          [,1]      [,2]
## [1,] 0.3570714 1.785357
```

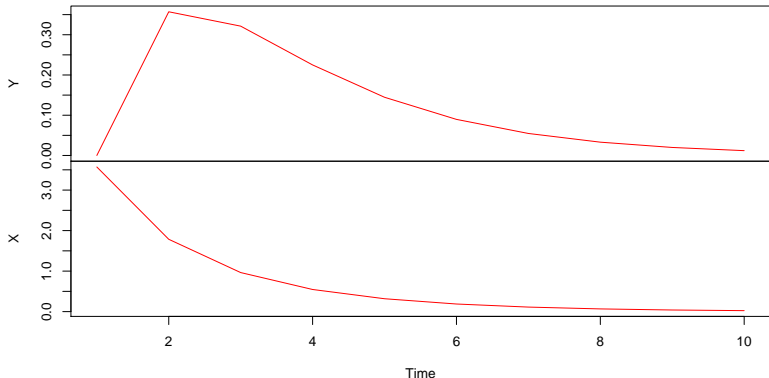
```
print(t(Y_3 <- Theta_1 %*% Y_2))
```

```
##          [,1]      [,2]
## [1,] 0.3213643 0.9640928
```



```
Y_j <- matrix(0, nrow = 10, ncol = 2)
Y_j[1, ] <- P_mat %*% c(0, 1)
for(j in 2:10){
  Y_j[j, ] <- Theta_1 %*% Y_j[j - 1, ]
}
colnames(Y_j) <- c("Y", "X")
plot.ts(Y_j, col = "red",
        main = "Orthogonal Impulse Response from (0, 1)")
```

Orthogonal Impulse Response from (0, 1)



If we switch the variable order, so we have  $(X_t, Y_t)$ , instead of  $(Y_t, X_t)$ , then, because  $\mathbf{P}$  is a lower triangular matrix, the effects also change:

```
Sigma = matrix(c(25, 14, 14, 16), ncol = 2, byrow = TRUE)
round(P_mat <- t(chol(Sigma)), 4)
```

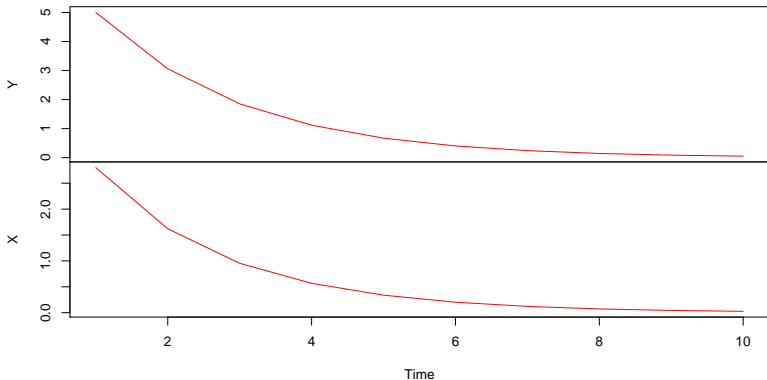
```
##      [,1]  [,2]
## [1,]  5.0  0.0000
## [2,]  2.8  2.8566
```

```
Theta_1 = matrix(c(0.5, 0.2, 0.1, 0.4), ncol = 2, byrow = TRUE)
Y_1 <- P_mat %*% c(1, 0); Y_2 <- Theta_1 %*% Y_1
Y_3 <- Theta_1 %*% Y_2
print(cbind(Y_1, Y_2, Y_3))
```

```
##      [,1] [,2] [,3]
## [1,]  5.0  3.06  1.854
## [2,]  2.8  1.62  0.954
```

```
Y_j <- matrix(0, nrow = 10, ncol = 2)
Y_j[1, ] <- P_mat %*% c(1, 0)
for(j in 2:10){
  Y_j[j, ] <- Theta_1 %*% Y_j[j - 1, ]
}
colnames(Y_j) <- c("Y", "X")
plot.ts(Y_j, col = "red",
        main = "Orthogonal Impulse Response from (1, 0)")
```

Orthogonal Impulse Response from (1, 0)



Note that the difference in variable ordering in this case results in a different magnitude

Uncorrelated innovations were developed to deal with the problem of non-zero correlations between the original innovations.

*However, the solution of one problem creates another.* The new problem is that the **order** in which the  $\bar{u}$  variables are **entered** can have dramatic effects on the numerical results.

The interpretation of impulse response functions is thus a somewhat hazardous operation, and there has been intense debate on their possible economic significance.

One way to deal with it (for the case when  $\mathbf{P}$  is the lower triangular matrix) is to order with a decreasing order of exogeneity. That is:

- ▶ the first variable (placed on top) is the most exogeneous - it is affected only by a shock to itself;
- ▶ the last variable is affected by all the previous variable shocks (i.e. it is the least exogeneous);

The exogeneity property can also be formulated in terms of economic theory (i.e. which of the variables is the most and the least exogenous).

# Error Correction Model

## Refresh: Univariate case (ECM)

Recall that the AR(p) process:

$$Y_t = \mu + \delta t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t$$

can be rearranged into

$$\Delta Y_t = \mu + \delta t + \rho Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \dots + \gamma_{p-1} \Delta Y_{t-p+1} + \epsilon_t$$

the latter form is more convenient for the unit root testing:  $H_0 : \rho = 0$  implies that  $Y_t$  has unit root.

## New: Multivariate case (VECM)

Similarly, the  $d$  - dimensional VAR process:

$$\vec{Y}_t = \vec{\mu}_t + \Theta_1 \vec{Y}_{t-1} + \dots + \Theta_p \vec{Y}_{t-p} + \vec{\epsilon}_t$$

where  $\vec{\mu}_t = \vec{\mu}$  or  $\vec{\mu}_t = \vec{\mu} + \vec{\delta}t$  or  $\vec{\mu}_t = \vec{\mu} + \vec{\delta}t + \vec{\gamma}t^2$ , can be rewritten in VEC form:

$$\Delta \vec{Y}_t = \vec{\mu}_t + \Pi \vec{Y}_{t-1} + \Gamma_1 \Delta \vec{Y}_{t-1} + \dots + \Gamma_{p-1} \Delta \vec{Y}_{t-p+1} + \vec{\epsilon}_t$$

where the long-run matrix  $\Pi = \sum_{i=1}^p \Theta_i - I$  and  $\Gamma_i = -\sum_{j=i+1}^p \Theta_j$ .

Note:

- ▶ if  $\text{rank}(\Pi) = r$  (i.e.  $r$  linearly independent cointegrating relations), it can be written as  $\Pi_{d \times d} = \alpha_{d \times r} \beta_{r \times d}^T$ .
- ▶ the rows of the matrix  $\beta^T$  form a basis for the  $r$  cointegrating vectors and the elements of  $\alpha$  distribute the impact of the cointegrating vectors to the evolution of  $\Delta Y_t$  (they are usually interpreted as speed of adjustment to equilibrium coefficients).

## Multivariate unit root test

For the equation:

$$\Delta \vec{Y}_t = \vec{\mu}_t + \Pi \vec{Y}_{t-1} + \Gamma_1 \Delta \vec{Y}_{t-1} + \dots + \Gamma_{p-1} \Delta \vec{Y}_{t-p+1} + \vec{\epsilon}_t$$

We shall use the **Johansen test** to test whether  $\Pi$  equals 0 in some sense: in  $\det(\Pi) = 0$  or in  $\text{rank}(\Pi) = r < d$  sense. The test is aimed to test the number  $r$  of cointegrating relationships. thus, the Johansen approach can be interpreted as a **multivariate unit root test**.

- ▶ In the previous discussion of VARs, we assumed that **all** variables were **stationary**.
- ▶ If **all** of the original variables have unit roots and are **not cointegrated**, then they should be **differenced** and the resulting stationary variables should be used in the VAR.
- ▶ This covers every case except one where the variables have **unit roots and are cointegrated**. Recall that in this case in the discussion of **Granger causality**, we recommended that you work with an ECM. The same strategy can be employed here. In particular, instead of working with a vector autoregression (VAR), you should work with a **vector error correction model (VECM)**.

To outline the strategy of dealing with multivariate time series, it is better to start with two-dimensional case:  $\vec{Y}_t = (Y_t, X_t)'$  and recall the **Engle-Granger** (EG) procedure:

1. Test whether each series,  $Y_t$  and  $X_t$ , is integrated of the **same** order.
2. If both series are  $I(0)$ , estimate VAR model in levels (no need for VECM).
3. If both series are  $I(1)$ , estimate the **cointegration regression**  $Y_t = \gamma_0 + \gamma_1 X_t + Z_t$ , then test whether the residuals  $\hat{Z}_t$  are stationary (this is called the **Engle-Granger** (EG) test, it is close to the ADF test).
4. If  $Z_t$  is  $I(1)$ , estimate a VAR model in differences,  $\Delta Y_t$  and  $\Delta X_t$ .
5. If  $Z_t$  is stationary,  $Y_t$  and  $X_t$  are **cointegrated** - in this case, estimate the VEC model:

$$\begin{cases} \Delta Y_t = \alpha_2 \hat{Z}_{t-1} + \mu_2 + \gamma_{21,1} \Delta X_{t-1} + \gamma_{22,1} \Delta Y_{t-1} + \dots + \gamma_{21,p} \Delta X_{t-p} + \gamma_{22,p} \Delta Y_{t-p} + \epsilon_{Y,t} \\ \Delta X_t = \alpha_1 \hat{Z}_{t-1} + \mu_1 + \gamma_{11,1} \Delta X_{t-1} + \gamma_{12,1} \Delta Y_{t-1} + \dots + \gamma_{11,p} \Delta X_{t-p} + \gamma_{12,p} \Delta Y_{t-p} + \epsilon_{X,t} \end{cases}$$

The order  $p$  of this VEC model is chosen such that VAR( $p+1$ ) model fitted to the **levels** has a minimum AIC or SC. If  $p=0$ , i.e. a level VAR(1) has minimum, this may indicate that the original series is **stationary**.



6. If necessary, use the model obtained to forecast  $Y_t$  and  $X_t$  (this can be done by rewriting the VEC model as a VAR model). For example, the model:

$$\Delta \vec{Y}_t = \alpha \beta^T \vec{Y}_{t-1} + \Gamma_1 \Delta \vec{Y}_{t-1} + \vec{\epsilon}_t$$

can be expressed as a VAR(2):

$$\vec{Y}_t = (I + \Gamma_1 + \alpha \beta^T) \vec{Y}_{t-1} - \Gamma_1 \vec{Y}_{t-2} + \vec{\epsilon}_t$$

- ▶ Thus, in the case where our data consists of **two**  $I(1)$  components, we use the EG test for cointegration.
- ▶ In a **multidimensional** case we use another cointegration test called the **Johansen test**.

The first thing to note is that it is possible for **more than one** cointegrating relationship to exist if you are working with several time series variables (all of which you have tested and found to have unit roots).

To be precise, if you are working with  $d$  variables, then it is possible to have up to  $d - 1$  cointegrating relationships (and, thus, up to  $d - 1$  cointegrating residuals included in the VECM).

To begin with, let us return to equation:

$$\Delta \vec{Y}_t = \vec{\mu}_t + \alpha \beta^T \vec{Y}_{t-1} + \Gamma_1 \Delta \vec{Y}_{t-1} + \dots + \Gamma_{t-p} \Delta \vec{Y}_{t-p+1} + \vec{\epsilon}_t$$

- ▶ If  $\text{rank}(\beta) = 0$ , then only  $\beta^T \vec{Y}_{t-1} = 0 \cdot Y_{1,t-1} + \dots + 0 \cdot Y_{d,t-1} = 0$  is stationary. In other words,  $\vec{Y}_t$  is not cointegrated and VECM reduces to  $\text{VAR}(p-1)$  in differences.
- ▶ If  $0 < \text{rank}(\beta) = r < d$ ,  $\vec{Y}_t$  is  $I(1)$  with  $r$  linearly independent cointegrating vectors and  $\beta^t \vec{Y}_{t-1} \sim I(0)$ .
- ▶ If  $\text{rank}(\beta) = d$ , then  $\Pi$  has full rank and is invertible, therefore  $\vec{Y}_{t-1}$  will be a linear combination of stationary differences, therefore, stationary itself.

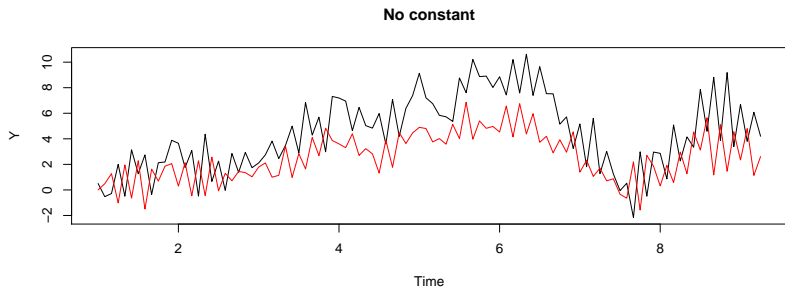
Thus, it is often of interest to **test**, not simply for whether cointegration is present or not, but also for the **number of cointegrating relationships**. Recall that any hypothesis is rejected if the test statistics **exceeds** the critical value. However, these critical values depend on the **deterministic** components of VECM such as constants and linear trends.

In a similar situation when testing for a unit root, i.e., the hypothesis  $H_0 : \rho = 0$ , we used two different critical values for the  $t$  – statistics of  $\hat{\rho}$  depending on the presence, or absence, of a deterministic trend. Now we have **five** different variants for calculating  $p$  – values. The five cases are:

1.  $\vec{\mu}_t = \vec{0}$  (**no constant**) - all the series in  $\vec{Y}_t$  are  $I(1)$  without a drift and the cointegrating relations:  $\beta^T \vec{Y}_t = \beta_1 Y_{1,t} + \dots + \beta_M Y_{M,t}$  have zero mean:

$$\Delta \vec{Y}_t = \alpha \beta^T \vec{Y}_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta \vec{Y}_{t-i} + \vec{\epsilon}_t$$

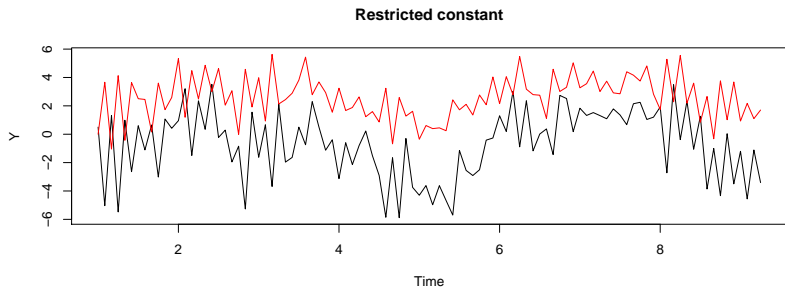
(no components have a drift)



2.  $\vec{\mu}_t = \vec{\mu}_0 = \alpha \vec{\rho}_0$  (**restricted constant**) - the series series in  $\vec{Y}_t$  are  $I(1)$  without a drift and the cointegrating relations  $\beta^T \vec{Y}_t$  have a non-zero mean:

$$\Delta \vec{Y}_t = \alpha (\beta^T \vec{Y}_{t-1} + \vec{\rho}_0) + \sum_{i=1}^{p-1} \Gamma_i \Delta \vec{Y}_{t-i} + \vec{\epsilon}_t$$

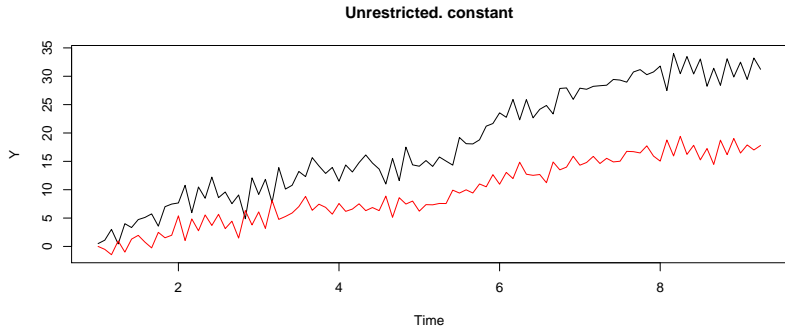
(no components have a drift)



3.  $\vec{\mu}_t = \vec{\mu}_0$  (**unrestricted constant**) - the series series in  $\vec{Y}_t$  are  $I(1)$  with a drift vector  $\vec{\mu}_0$  and the cointegrating relations  $\beta^T \vec{Y}_t$  have a non-zero mean:

$$\Delta \vec{Y}_t = \vec{\mu}_0 + \alpha \beta^T \vec{Y}_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta \vec{Y}_{t-i} + \vec{\epsilon}_t$$

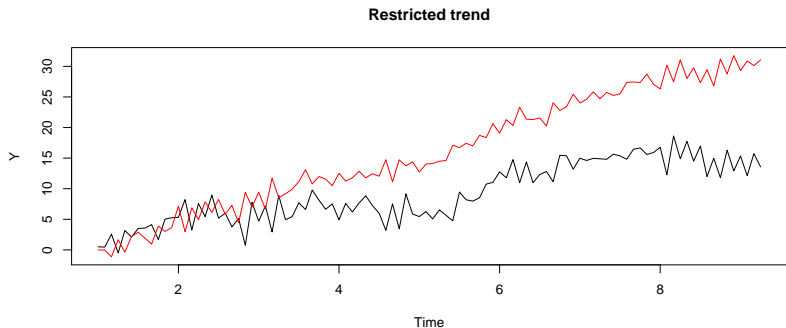
(at least one component drifts)



4.  $\vec{\mu}_t = \vec{\mu}_0 + \alpha \vec{\rho}_1 t$  (**restricted trend**) - the series series in  $\vec{Y}_t$  are  $I(1)$  with a drift vector  $\vec{\mu}_0$  and the cointegrating relations  $\beta^T \vec{Y}_t$  have a linear trend  $\vec{\rho}t$ :

$$\Delta \vec{Y}_t = \vec{\mu}_0 + \alpha(\beta^T \vec{Y}_{t-1} + \vec{\rho}_1 t) + \sum_{i=1}^{p-1} \Gamma_i \Delta \vec{Y}_{t-i} + \vec{\epsilon}_t$$

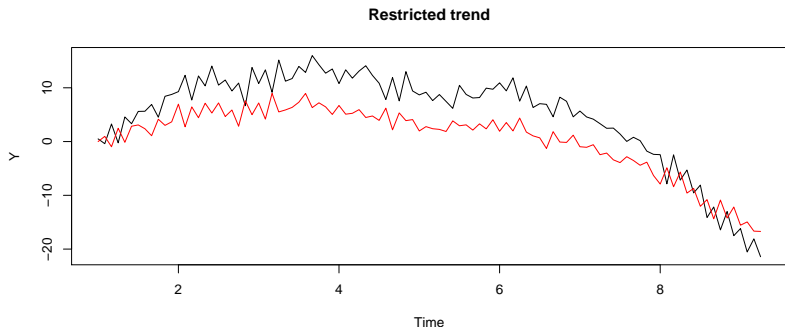
(at least one component drifts)



5.  $\vec{\mu}_t = \vec{\mu}_0 + \vec{\mu}_1 t$  (**unrestricted constant and trend**) - the series series in  $\vec{Y}_t$  are  $I(1)$  with a linear trend in VECM (and a **quadratic** trend in **levels**) and the cointegration relations  $\beta^T \vec{Y}_t$  have a linear trend:

$$\Delta \vec{Y}_t = \vec{\mu}_0 + \vec{\mu}_1 t + \alpha \beta^T \vec{Y}_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta \vec{Y}_{t-i} + \vec{\epsilon}_t$$

(at least one component has a quadratic trend)



Note:

- ▶ if no components of  $\vec{Y}_t$  drift - we use Cases 1 or 2.
- ▶ if at least one component of  $\vec{Y}_t$  drifts - we use Case 3 or 4.
- ▶ if at least one component of  $\vec{Y}_t$  has a quadratic trend - we use Case 5.

Generally:

- ▶ Case 1 is not really relevant for empirical work.
- ▶ The restricted constant Case 2 is appropriate for non-trending  $I(1)$  data like interest rates and exchange rates.
- ▶ The unrestricted constant Case 3 is appropriate for trending  $I(1)$  data like asset prices, macroeconomic aggregates (real GDP, consumption, employment etc).
- ▶ The restricted trend Case 4 is also appropriate for trending  $I(1)$  as in Case 3. However, notice the deterministic trend in the cointegrating residual in Case 4 as opposed to the stationary residuals in Case 3.
- ▶ The unrestricted trend Case 5 is appropriate for  $I(1)$  data with a quadratic trend. An example might be nominal price data during times of extreme inflation.



The above-given figures and considerations are important in choosing the right variant to define critical values of the Johansen test. The basic steps in Johansen's methodology are (we assume that all the series in  $\vec{Y}_t$  are  $I(1)$ ):

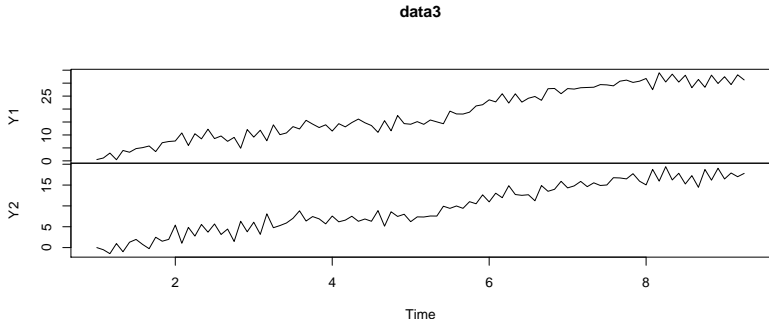
1. Choose the right order  $p$  for a  $VAR(p)$  model for **levels** (i.e. for  $\vec{Y}_t$  and not  $\Delta\vec{Y}_t$ ).
2. Choose the right case out of five ones (use graphs of  $\vec{Y}_t$ ).
3. Apply Johansen's test and find the number of cointegrating relations.
4. Create a VECM.
5. Use it to forecast  $\vec{Y}_t$ .

## Example

Following the above comments, we will estimate the VECM of the following data:/colcodeR

```
txt1 <- "http://uosis.mif.vu.lt/~rlapinskas/(data%20R&GRETTL/"  
txt2 <- "VECM3.txt"  
data3 <- data.frame(read.table(paste0(txt1, txt2), header = T))  
data3 <- ts(data3, frequency = 12)
```

```
plot.ts(data3)
```



## 0. Test for a unit root

```
mdl1 <- dynlm::dynlm(d(Y1) ~ L(Y1) + L(d(Y1), 1) + time(Y1),  
                    data = data3)  
round(summary(mdl1)$coeff, 4)
```

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	0.2282	0.4782	0.4772	0.6343
## L(Y1)	-0.2598	0.0962	-2.6996	0.0082
## L(d(Y1), 1)	-0.6426	0.0797	-8.0619	0.0000
## time(Y1)	0.9663	0.3828	2.5239	0.0133

t-statistic = -2.6996 > 3.45 so we **do not reject** the null hypothesis  $H_0$  : *the process has a unit root* .

```
mdl2 <- dynlm::dynlm(d(Y2) ~ L(Y2) + L(d(Y2), 1) + time(Y2),  
                    data = data3)  
round(summary(mdl2)$coeff, 4)
```

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	-0.6287	0.4360	-1.4418	0.1527
## L(Y2)	-0.4141	0.1205	-3.4374	0.0009
## L(d(Y2), 1)	-0.5792	0.0852	-6.7950	0.0000
## time(Y2)	0.9290	0.2823	3.2906	0.0014

t-statistic = -3.4374 > 3.45 so we **do not reject** the null hypothesis  $H_0$  : *the process has a unit root* .

## 1. Select the VAR lag order

```
VARselect(data3, lag.max = 5, type="both")
```

```
## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)
##      1      1      1      1
##
## $criteria
##              1          2          3          4          5
## AIC(n) 0.3492889 0.4014374 0.4698795 0.5101213 0.5599392
## HQ(n)  0.4361906 0.5317899 0.6436829 0.7273756 0.8206444
## SC(n)  0.5643522 0.7240323 0.9000061 1.0477796 1.2051292
## FPE(n) 1.4182001 1.4944737 1.6010811 1.6681027 1.7553224
```

The recommended VAR order is  $p = 1$ .

## 2. Choosing the right Case out of 5

As both  $Y_1$  and  $Y_2$  are drifting, we should try both Case 3 and Case 4 and choose the better one.

### 3. Find the number of cointegrating relationships

To find the number of cointegrating relationships (the rank of the matrix  $\Pi$ ) (in our two dimensional case, it can be zero or one), we shall apply Johansen's test.

If there are a couple of competing models, select the one with the smallest AIC value, i.e. estimate `VECM(..., estim="ML", LRinclude = c("none", "const", "trend", "both"))` for different competing Cases and compare the AIC:

- ▶ Case 1 `include = 'none'`.
- ▶ Case 2 `LRinclude = 'const'`.
- ▶ Case 3 `include = 'const'`, thus no need for parameters.
- ▶ Case 4 `LRinclude = 'trend'`.
- ▶ Case 5 `include = 'both'`.

For Case 3:

```
suppressWarnings({  
  library(tsDyn)  
  C3 <- VECM(data3, lag = 1, estim = "ML", include = 'const')  
  summary(rank.test(C3))  
})
```

```
##      r      trace trace_pval trace_pval_T      eigen eigen_pval  
## 1 0 62.222227      <0.001      <0.001 60.802033      <0.001  
## 2 1  1.420193      0.2334      0.2398  1.420193      0.2334
```

The output of the test allows us to determine the rank  $r$  of  $\Pi$ .

The estimated *eigenvalues*,  $\lambda_i$ , are sorted from largest to smallest and we apply a sequential procedure: both `trace_pval_T` and `eigen_pval` rejects the null hypothesis  $H_0 : r = 0$ , but accepts the next null hypothesis  $H_0 : r = 1$ .

Thus, if we decide to describe our data with Case 3, we should take  $r = 1$ :  
C3

	ECT	Intercept	Y1 -1	Y2 -1
Equation Y1	-0.7073693	1.852102	-0.03455440	0.11606697
Equation Y2	0.6451750	-1.217922	-0.07083753	0.06764216

For Case 4:

```
suppressWarnings({  
  library(tsDyn)  
  C4 <- VECM(data3, lag = 1, estim = "ML", LRinclude = 'trend')  
  summary(rank.test(C4))  
})
```

```
##      r      trace trace_pval trace_pval_T      eigen eigen_pval  
## 1 0 70.586968      <0.001      <0.001 63.063750      <0.001  
## 2 1  7.523218      0.302      0.3027  7.523218      0.3024
```

we derive the same conclusion, i.e.  $r = 1$ .

C4

```
                ECT Intercept          Y1 -1          Y2 -1  
Equation Y1 -0.6509582  1.767273 -0.06022577  0.1081974  
Equation Y2  0.6502533 -1.276647 -0.05833625  0.1501910
```

```
data.frame(AIC = AIC(C3), BIC = BIC(C3))
```

```
##          AIC          BIC  
## 1 40.21628 63.48098
```

```
data.frame(AIC = AIC(C4), BIC = BIC(C4))
```

```
##          AIC          BIC  
## 1 37.95456 61.21927
```

We note that the *BIC* of the Restricted Trend case is lower than the *BIC* of Case 3. So, the model from Case 4 is our final VEC model.

Alternatively, we can use the `ca.jo` test:

```
H1 <- ca.jo(data3, ecdet = "trend", K = 2)
txt_out <- capture.output(summary(H1))
cat(paste0(txt_out[c(2:15)], collapse = "\n"))
```

```
## #####
## # Johansen-Procedure #
## #####
##
## Test type: maximal eigenvalue statistic (lambda max) , with linear trend in
##
## Eigenvalues (lambda):
## [1] 4.745539e-01 7.389488e-02 2.003298e-17
##
## Values of teststatistic and critical values of test:
##
##          test 10pct  5pct  1pct
## r <= 1 |  7.52 10.49 12.25 16.26
## r = 0  | 63.06 16.85 18.96 23.65
```

- ▶ At  $r = 0$  we have  $test = 63.06 > 18.96$ , so we reject  $H_0 : r = 0$  at the 5% significance level;
- ▶ At  $r \leq 1$  we have  $test = 7.52 < 12.25$ , so we do not reject  $H_0 : r \leq 1$  at the 5% significance level. In other words, given the previous result at  $r = 0$ , we do not reject  $H_0 : r = 1$ .



While `ca.jo` in package `urca` and `rank.test` both implement Johansen tests, there are a few differences:

- ▶ `rank.test` gives p-values, while `ca.jo` gives only critical values.
- ▶ `rank.test` allows for five different specifications of deterministic terms, `ca.jo` for only three.
- ▶ `ca.jo` allows for seasonal and exogenous regressors, which is not available in `rank.test`.
- ▶ The lag is specified differently:  $K$  from `ca.jo` corresponds to  $\text{lag} + 1$  in `rank.test`.

The last point is the reason why we set  $K = 2$  in `ca.jo`, because it corresponds to  $\text{lag} = 1$ .

## 4. Create a VECM

```
C4$coefficients
```

```
##                ECT Intercept          Y1 -1          Y2 -1
## Equation Y1 -0.6509582  1.767273 -0.06022577  0.1081974
## Equation Y2  0.6502533 -1.276647 -0.05833625  0.1501910
```

```
t(C4$model.specific$beta)
```

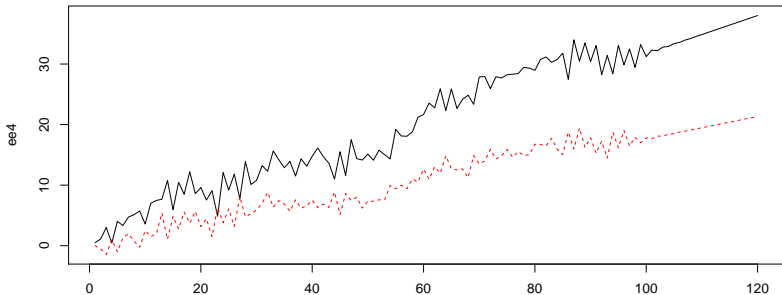
```
##      Y1          Y2      trend
## r1  1 -1.883276  0.03712441
```

$$\begin{cases} \Delta Y_{1,t} &= 1.767 - 0.65 \cdot (Y_{1,t-1} - 1.88 Y_{2,t-1} + 0.0371 \cdot (t-1)) \\ \Delta Y_{2,t} &= -1.277 + 0.65 \cdot (Y_{2,t-1} - 1.88 Y_{1,t-1} + 0.0371 \cdot (t-1)) \end{cases}$$

Coefficient ( $\alpha =$ ) ECT = 0.65 is called an adjustment coefficient. It indicates that  $Y_2$  will return to equilibrium in  $1/0.65 \sim 2$  steps, *ceteris paribus*.

## 5. Forecast

```
pred_C4 <- predict(C4, n.ahead = 20)
ee4 = rbind(data3, pred_C4)
matplot(ee4, type = "l")
```



## Remarks

- ▶ If we have **two variables, both Engle-Granger and Johansen tests are appropriate** for cointegration.
- ▶ In a multidimensional case, with  $d > 2$ , the EG suffers from omitted variable bias, so in those cases **when  $d > 2$  it is better to use Johansen test.**
- ▶ The most restricted model (Case 1) is unlikely to find general use, because at least a constant will usually be included in the cointegration equation.
- ▶ The least restricted model (Case 5) allows for quadratic trends in the data which occurs quite rarely.
- ▶ The choice between Case 2 and Case 3 rests upon whether there is a need to allow for the possibility of linear trends in the data, a preliminary graphing of the data is often helpful in this respect.

- ▶ If Case 3 is preferred to Case 2, only then does Case 4 need to be considered since the data has to have a linear trend if we are to consider allowing a trend in the cointegration equation.
- ▶ As a rough guide, use Case 2 if none of the series appear to have a trend.
- ▶ For trending series, use Case 3 if you believe all trends are stochastic.
- ▶ If you believe some of the series are trend stationary, use Case 4.

However, the simplest way that also allows one to “automate” selection is to choose the case according to the minimum of AIC or BIC of the model.

Multivariate VAR is a complicated model. Below we present some useful facts without proofs:

- ▶ To estimate the coefficients of  $\vec{Y}_t = \vec{\alpha} + \vec{\delta}t + \Theta_1 \vec{Y}_{t-1} + \dots + \Theta_p \vec{Y}_{t-p} + \vec{\epsilon}_t$  use the (conditional) maximum likelihood estimation method, assuming that the innovations  $\vec{\epsilon}_t$  have a multivariate normal distribution. This is equivalent to OLS method applied to each equation separately.
- ▶ Maximum likelihood estimates are consistent even if the true innovations are non-Gaussian.
- ▶ Standard OLS  $t$  and  $F$  statistics applied to the coefficients of any single equation of the VAR are asymptotically valid.
- ▶ The goal of unit root tests is to find a parsimonious representation of the data that gives a reasonable approximation of the true process, as opposed to determining whether or not the true process is literally  $I(1)$ .
- ▶ If  $\vec{Y}_t$  is cointegrated, a VAR estimated in levels is not misspecified but involves a loss of efficiency.

- ▶ Let  $\vec{Y}_t$  have a unit root, but no cointegration. A VAR in levels is not subject to the spurious regression problem discussed above for single equation regressions.
- ▶ Even if there is no cointegration among the variables in  $\vec{Y}_t$ , equation-by-equation OLS estimation of VAR in levels delivers consistent estimates of the VAR parameters. Unlike a univariate regression, differencing is not required to obtain consistent estimates. Nevertheless, the small sample properties of the estimator may be improved by estimating the VAR in differences.
- ▶ Suppose that some of the  $M$  variables are stationary while the other variables are each individually  $I(1)$  and also cointegrated by, say, a single cointegration relation. One can find an explanation of how to construct a VEC model in this case. For example, if  $M = 4$ ,  $Y_t^{(1)} = Y_{1,t}$  is stationary and  $\vec{Y}_t^{(2)} = (Y_{2,t}, Y_{3,t}, Y_{4,t})$  are cointegrated  $I(1)$ , then the VEC representation will be:

$$\begin{bmatrix} Y_t^{(1)} \\ \Delta \vec{Y}_t^{(2)} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vec{\mu}_2 \end{bmatrix} + \begin{bmatrix} \vec{\pi}_1 \\ \vec{\pi}_2 \end{bmatrix} \vec{Y}_{t-1}^{(2)} + \begin{bmatrix} \gamma_{11}^{(1)} & \vec{\gamma}_{12}^{(1)} \\ \gamma_{21}^{(1)} & \vec{\gamma}_{22}^{(1)} \end{bmatrix} \begin{bmatrix} Y_{t-1}^{(1)} \\ \Delta \vec{Y}_{t-1}^{(2)} \end{bmatrix} + \cdots + \begin{bmatrix} \epsilon_t^{(1)} \\ \vec{\epsilon}_t^{(2)} \end{bmatrix}$$

where the  $4 \times 3$  matrix  $(\vec{\pi}_1 \ \vec{\pi}_2)'$  is restricted to be of the form  $\vec{b}\vec{a}'$ , where  $\vec{b}$  is  $4 \times 1$  and  $\vec{a}'$  is  $1 \times 3$ .