The vector error correction model using R Software

Presentation · July 2019

DOI: 10.13140/RG.2.2.19399.65446

1 author:

Sami Mestiri
Faculté des Sciences Économiques et de Gestion de Mahdia

The vector error correction model
with R Software

Sami Mestiri

EAS-Mahdia Research Unit
Faculty of Science Mahdia Management and Economics, University of Monastir, Tunisia.

27 juillet 2019

The cointegration

The cointegration analysis consists in analyzing the errors between the non-stationary components of several series. It makes it possible to clearly identify the true relationship between two variables by searching for the existence of a cointegrating vector and eliminating its effect. Two series x_t and y_t are cointegrated if the two conditions are satisfied :

- They are affected by a stochastic trend of the same order of integration d.
- A linear combination of these series can be reduced to a series

Let :

$$x_t \rightarrow I(d)$$

$$y_t \rightarrow I(b)$$

Such $Z_t = \alpha_1 x_t + \alpha_2 y_t \rightarrow I(d-b)$ with $d \geq b \geq 0$.
We notice $x_t, y_t \rightarrow CI(d,b)$
with $\alpha = [\alpha_1, \alpha_2]$ is the vector cointegration.

Cointegration test between two variables

Step 1 : test the order of integration of the variables :
A necessary condition of cointegration is that the series must be integrated of the same order. If the series are not integrated of the same order, they can not be cointegrated. If the statistical series studied are not integrated of the same order, the procedure is stopped, there is no risk of cointegration.
Let : $X_t \to I(d)$ and $Y_t \to I(d)$

Cointegration test between two variables

Step 2 : Estimation of the long-term relation : If the necessary condition is verified, the OLS estimates the long-term relationship between the variables :

$$Y_t = a_1 X_t + a_0 + e_t$$

For the cointegration relation to be accepted, the e_t residue from this regression must be stationary :
$e_t = Y_t - \hat{a} X_t - \hat{a}_0$
The stationarity of the residue is tested using DF or DFA tests. In this case, we can no longer use the Dickey and Fuller tables.
If the residue is stationary then we can estimate the error correction model.

Estimation of the two-step error correction model (ECM) :

Let the series Y_t and $X_t \rightarrow I(d)$ the OLS estimation of the long-term relationship indicates a stationarity of the residue. The seriesy_t and x_t are therefore noted $CI(1, 1)$.
The ECM can be used to simultaneously model short-term dynamics (represented by the first-difference variables) over the long term (represented by the variables in level).

Estimation of the two-step error correction model (ECM) :

Step 1 : OLS estimation of the long-term relation :

$$y_t = \hat{\alpha} + \hat{\beta} x_t + e_t$$

Step 2 : estimation by the MCO of the relation of the dynamic model (short term) :

$$\Delta y_t = \alpha_1 \Delta x_t + \alpha_2 e_{t-1} + \mu_t ; \alpha_2 < 0$$

The coefficient α_2 must be significantly of type otherwise, it is necessary to reject a meaning of type *ECM*.

Application

Let there be two statistical series y_t and x_t.

We ask to estimate the relation between these two variables ($y_t = a_0 + a_1 x_t + e_t$) by testing a possible cointegration (in this case estimate the error correction model).

The cointegration test is performed from the model estimation residual :

Soit : $y_t = \underset{(6,3)}{0.55} x_t + \underset{(41,46)}{10,38} + e_t$

$n = 30; R^2 = 0,58; (.) = \text{t of Student}.$

We can verify that the residue is stationary, so there is a risk of cointegration between the two variables.

```
tab2=read.table("modeleECM.txt",h=T)
attach(tab2)
library("vars")
adf1 <- summary(ur.df(X, type = "trend", lags = 1))
adf3 <- summary(ur.df(Y, type = "trend", lags = 2))
mod3=lm(Y~X)
summary(mod3)
resd=as.vector(mod3$residuals)
adf5 <- summary(ur.df(resd, type = "trend", lags = 2))
adf6 <- summary(ur.df(resd, type = "drift", lags = 2))
resd1=resd[-1]
Y1=diff(Y)
X1=diff(X)
mod4=lm(Y1~X1+resd1-1)
summary(mod4)
```

Application

We calculate, first, the residual (from the previous model) shifted by one period, ie : $e_{t-1} = y_{t-1} - 0,55x_{t-1} - 10,38$
Then we estimate (by the MCO) the model :

$$\Delta y_t = \alpha_1 \Delta x_t + \alpha_2 e_{t-1} + \mu_t$$

$$\Delta y_t = \underset{(3,09)}{0,610}\Delta x_t - \underset{(5,22)}{1,02}e_{t-1} + u_t$$

$n = 29; R2 = 0,60; (.) =$ t Student.

The coefficient (callback) of e_{t-1} is significantly negative, the error-correcting representation is validated.

The vector error correction model VECM (2)

Let us first examine the case of a VAR (2) process with k variables in matrix form :

$$Y_t = A_0 + A_1 Y_{t-1} + A_2 Y_{t-2} + \epsilon$$

with :
Y_t : dimension vector (k, 1) consisting of k variables $(y_{1t}, y_{2t}.., Y_{kt})$,
A_0 : dimension vector (k, 1),
A_p : dimension matrix (k, k).
The VECM representation is valid either

$$\Delta Y_t = A_0 + A_2 \Delta Y_{t-1} + (A_1 + A_2 - I) Y_{t-1} + \epsilon$$

$\Delta Y_t = A_0 + B_1 \Delta Y_{t-1} + \pi Y_{t-1} + \epsilon$ with : $A_2 = -B_1$ et
$\pi = (A_1 + A_2 - I)$.

The vector error correction model VECM(p)

This result can be generalized to a representation VAR (p) with k variables in matrix form :

$$Y_t = A_0 + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + \epsilon$$

This model can be written in first differences in two ways :
$\Delta Y_t = A_0 + (A_1 - I)\Delta Y_{t-1} + (A_1 + A_2 - I)\Delta Y_{t-2} + ... + (A_{p-1} + ... + A_2 + A_1 - I)\Delta Y_{t-p+1} + \pi Y_{t-p} + \epsilon$
or according to Y_{t-1} :

$$\Delta Y_t = A_0 + B_1 \Delta Y_{t-1} + B_2 \Delta Y_{t-2} + ... + B_{p-1} \Delta Y_{t-p+1} + \pi Y_{t-1} + \epsilon$$

The matrix B_i being functions of the matrix A_i and $\pi = \sum_{i=1}^{p} (A_i - I)$
If the rank of the pi matrix (noted r) is $(1 < r < k - 1)$, then there are r cointegration relationships and the ECM representation is valid :

$$\Delta Y_t = A_0 + B_1 \Delta Y_{t-1} + B_2 \Delta Y_{t-2} + ... + B_{p-1} \Delta Y_{t-p+1} + \alpha e_{t-1} + \epsilon$$

With $e_t = \beta' Y_t$

Cointegration relationship tests

Step 1 : calculating two residues u_t and v_t
We perform two regressions :
First regression :
$\Delta Y_t = A_0 + A_1 \Delta Y_{t-1} + A_2 \Delta Y_{t-2} + ... + A_p \Delta Y_{t-p} + u_t$
Second regression :
$Y_t = A_0 + A_1 \Delta Y_{t-1} + A_2 \Delta Y_{t-2} + ... + A_p \Delta Y_{t-p} + v_t$
with $Y_t = [y_{1,t} y_{2,t}, ..., y_{k,t}]$
We have the same explanatory variables, only the specification of the block of the variable to be explained is modified.
u_t and v_t are the matrices of the residuals of dimension (k, n) with k = number of variables, n = number of observations.

Cointegration relationship tests

Step 2 : calculation of the matrix allowing the calculation of the eigenvalues

We compute four matrices of the variances-covariances of dimension (k, k) from the residuals u_t and v_t.

$$\Sigma_{uu} = \frac{1}{n}\sum_{t=1}^{n} u_t u_t^{'}; \Sigma_{vv} = \frac{1}{n}\sum_{t=1}^{n} v_t v_t^{'};$$

$$\Sigma_{uv} = \frac{1}{n}\sum_{t=1}^{n} u_t v_t^{'}; \Sigma_{vu} = \frac{1}{n}\sum_{t=1}^{n} v_t u_t^{'}$$

Then we extract the k eigenvalues of the matrix M of dimension k, k calculated as follows :

$$M = \Sigma_{vv}^{-1}\Sigma_{uv}^{-1}\Sigma_{uu}^{-1}$$

Trace tests

From these eigenvalues, we compute a statistic

$$\lambda_{trace} = -n \sum_{i=r+1}^{k} Ln(1 - \lambda_i)$$

with n = number of observations, λ_i = i th eigenvalue of the matrix M, k = number of variables, r = rank of the matrix.

This Johansen test works by excluding alternative hypotheses :

Rank of the matrix π equal 0 (r = 0), $H_0 : r = 0$ against H_1 : r> 0 ; if H_0 is refused, we proceed to the next test (if $\lambda_{trace} <$ to the critical value read in the table, we reject H_0) ;

Rank of the matrix π equal 1 (r = 1), that is $H_0 : r = 1$ against H_1 : r> 1 ; if H_0 is refused, we proceed to the next test ;

Rank of the matrix π equal 2 (r = 2), $H_0 : r = 2$ against H_1 r> 2 ; if H_0 is refused, proceed to the next test.

If, after refusing the different H_0 assumptions at the end of the procedure,

Application

Let three variables $y_{1,t}, y_{2,t}$ and $y_{3,t}$ be observed over 30 periods.
We ask to test a possible cointegration and to estimate a VAR model or an error-correcting vector model if applicable.
- **First step** : determination of the number of delays of the representation VAR in level
The calculation of the AIC and SC information criteria for delays ranging from 1 to 3 - we do not go further because of the small number of observations - does not pose a problem.

```
y = cbind (Y1, Y2, Y3)
VARselect (y, lag.max = 3)
```

Application

Existence of a constant in the long term relation and not in the data (no deterministic trend),

```
library(urca)
sjf.vecm1 <- ca.jo(y, type = "trace", ecdet ="const")
summary(sjf.vecm1)
```

The three eigenvalues of the matrix π, estimated by the maximum of likelihood, are equal to
$\lambda_1 = 0,605; \lambda_2 = 0,257; \lambda_3 = 0.139.$

Application

Second step Rank of the matrix π equals 0 (r = 0), H0 : r = 0 against H1 : r> 0.

Values of test statistic and critical values :

	test	10pct	5pct	1pct
r <= 2	4.19	7.52	9.24	12.97
r <= 1	12.51	17.85	19.96	24.60
r = 0	38.51	32.00	34.91	41.07

Let's calculate the Johansen statistic :

$$\lambda_{trace} = -n \sum_{i=r+1}^{k} Ln(1-\lambda_i) pour \quad r = 0$$

$$\lambda_{trace} = -n.(Ln(1-\lambda_1) + Ln(1-\lambda_2) + Ln(1-\lambda_3))$$

$$\lambda_{trace} = -28.0.928 + 0.297 + 0.150 = 38.50$$

Application

Critical values are 34.91 for a threshold of 5% and 41.07 for a threshold of 1% ; we therefore reject the hypothesis H0, The rank of the matrix is not 0 (the series are not stationary).
Rank of the matrix π equal 1 (r = 1), ie H0 : r = 1 against H1 : r> 1
$\lambda_{trace} = -28.0.297 + 0.150 = 12.51$ Critical values are 19.96 for a threshold of 5% and 24.60 for a threshold of 1% ; we can not reject the hypothesis H0 ni at 5% nor at 1%, we consider that the rank of the matrix π is equal to 1. We therefore accept the hypothesis of a cointegrating relationship.

Application

Third step : estimation of the vector model with error correction
The two preceding specifications are estimated with or without a constant in the data and therefore, in both cases, with a single cointegration relation between $y_{1,t}$, $y_{2,t}$ and $y_{3,t}$.
The first specification is rejected because the three constants of the three equations are not significantly different from 0.

Application

```
vecm.r1 <- cajorls(sjf.vecm1, r = 1)
vecm.r1
lm(formula = substitute(form1), data = data.mat)
Coefficients :
Y1.d Y2.d Y3.d
ect1 -0.85 0.32 0.0089
Y1.dl1 -0.08 0.43 0.54
Y2.dl1 0.02 -0.62 -0.46
Y3.dl1 0.20 -0.25 -0.25
$beta
ect1
Y1.l2 1.00
Y2.l2 -0.87
Y3.l2 -0.54
const 12.92
```

Application

The final estimate of the VECM on 28 observations is therefore :

$\Delta y_{1,t} = -0.08\Delta y_{1,t-1} + 0.32\Delta y_{2,t-1} + 0.54\Delta Y_{3,t-1} - 0.85.(y_{1,t-1} - 0.87y_{2,t-1} - 0.54y_{3,t-1} - 12.92)$

$\Delta y_{2,t} = 0.2\Delta y_{1,t-1} - 0.62\Delta y_{2,t-1} - 0.64\Delta Y_{3,t-1} + 0.32.(y_{1,t-1} - 0.87y_{2,t-1} - 0.54y_{3,t-1} - 12.92)$

$\Delta y_{3,t} = 0.2\Delta y_{1,t-1} - 0.25\Delta y_{2,t-1} - 0.25\Delta Y_{3,t-1} + 0.0089(y_{1,t-1} - 0.87y_{2,t-1} - 0.54y_{3,t-1} - 12.92)$

THANK YOU FOR YOUR ATTENTION