3 Cointegration

3.1 Definition

(Engle and Granger (1987). Econometrica.)

Let $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{mt})'$, where $y_{it} \sim I(1)$, $i = 1, \ldots, m$, i.e., $\Delta \mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$ is stationary.

Suppose one wants to fit a VAR(1) model. Should $\mathbf{y}_t$ be differenced or not?*

To answer this question, consider

(1) $$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{y}_{t-2} + \boldsymbol{\epsilon}_t,$$

where $\epsilon \sim \text{NID}(\mathbf{0}, \boldsymbol{\Sigma})$.

*Recall that, if x_t and y_t are $I(1)$, and z_t is stationary then for any $a, b (\neq 0)$ (*i*) $ax_t + bz_t \sim I(1)$ (*ii*) usually $ax_t + by_t \sim I(1)$.

We can rewrite equation (1) into the VECM (Vector Error Correction Model) form

(2) $\Delta \mathbf{y}_t = \boldsymbol{\mu} + \Pi \mathbf{y}_{t-1} + \Gamma \Delta \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$,

where $\Pi = -(\mathbf{I} - \Phi_1 - \Phi_2)$, and $\Gamma = -\Phi_2$.

Note that the left hand side (2) is stationary! If $\Pi \neq \mathbf{0}$ and the model holds, then $\Pi \mathbf{y}_{t-1}$ must be stationary.

As a consequence, fitting purely the differenced model, $\Delta \mathbf{y}_t = \boldsymbol{\mu} + \Phi \Delta \mathbf{y}_{t-1} + \mathbf{v}_t$ leads to a misspesification!

Under what conditions is $\Pi \mathbf{y}_t$ stationary?

Definition 3.1: Let x_t and y_t be $I(1)$. Then x_t and y_t are cointegrated if there exist $a \neq 0$ such that $y_t - ax_t \sim I(0)$ (i.e. $y_t - ax_t$ is stationary).* We denote, $(x_t, y_t)' \sim CI(1,1)$.

It can be noted that in this simple case the cointegration parameter a is unique.

Furthermore, it is a very special relationship, because, generally, any linear combination of two integrated series is integrated!

*Generally: Let $\mathbf{y}_t = (y_{1t}, \ldots, \mathbf{y}_{mt})'$ with $y_{it} \sim I(d)$. Then y_{it} are cointegrated of order $b > 0$ if there exist a vector $\mathbf{a} = (a_1, \ldots, a_m)'$ such that $\mathbf{a}'\mathbf{y}_t \sim I(d-b)$. We denote $\mathbf{y}_t \sim CI(d,b)$.

It is also important to note that if x_t and y_t are integrated but not cointegrated then estimation of the regression model

$$y_t = \beta_0 + \beta_1 x_t + u_t \tag{3}$$

yields nonsense results!

In the general case we can decompose $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $m \times r$ matrices $r \leq m$ is the rank of Π.

Matrix β contains the long run relationships between variables in $\mathbf{y}_t$ and $\boldsymbol{\alpha}$ contains the short run adjusting parameters towards the long run steady state relationship $\boldsymbol{\beta}'\mathbf{y}_t$.

The case $r = m$:

If $r = m$ then Π is of full rank, and the variables in $\mathbf{y}_t$ are stationary, i.e. $I(0)$. If $r = 0$ then $\Pi = \mathbf{0}$, and the variables are not cointegrated.

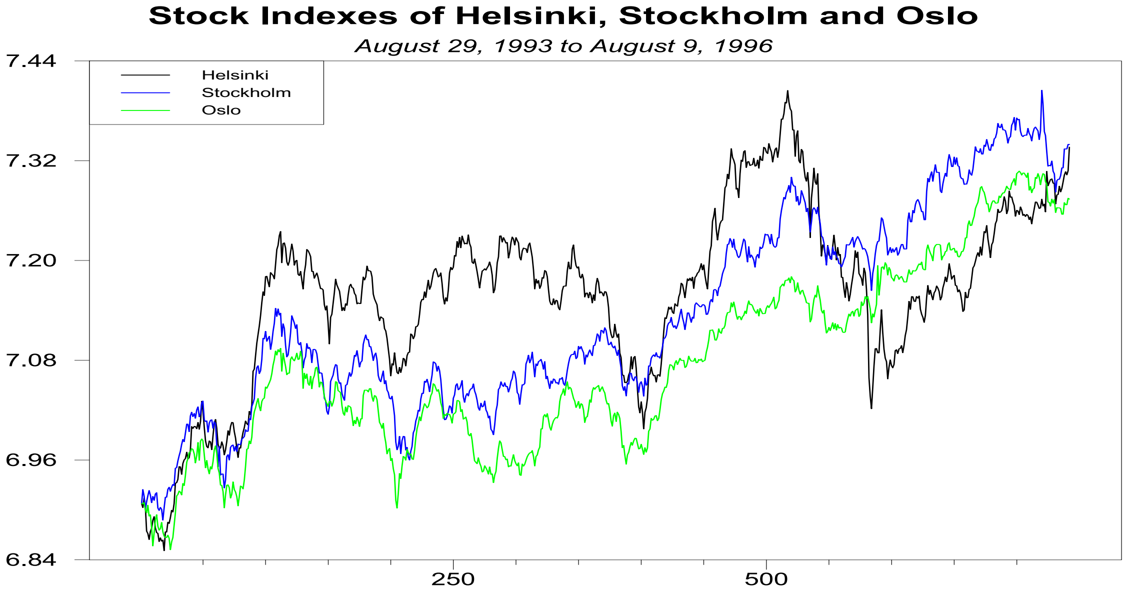
The case $r < m$:

The interesting case is when $0 < r \leq m - 1$. Then Π has a reduced rank, and there are r cointegrating vectors (equilibrium conditions).

In testing cointegration, the first step is to find the cointegration order, r.

In the general case this can be accomplished with the Johansen's (1988) procedure, where also estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are obtained.

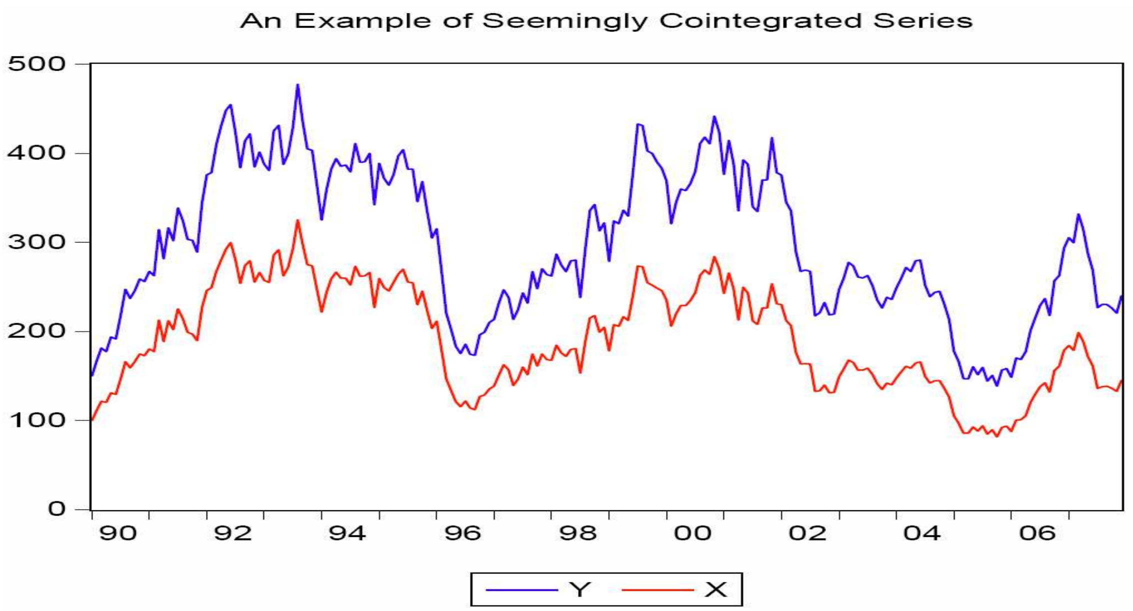
Example 3.1: Consider Helsinki, Oslo and Stockholm stock indexes (see figure).



We observe that especially Stockholm and Oslo seem to follow each other rather closely.

Remark 3.1: Graphs may be fully misleading!

Example 3.2: As will be demonstrated later, the two time series in the graph below are not cointegrated although they might look like.



Generally certain economic series are such that they should not drift too far apart from each other.

This means that the variables may depart from each other in short run, but there is certain mechanism that forces them back to a common path after some periods.

Examples of such series are prices and wages, price of a certain commodity sold in two different locations, wages and expenditure, etc.

3.2 Testing for cointegration

The testing procedure is carried out in two phases. Using, e.g. ADF, in the first step we test whether both of the series are $I(1)$. In the second step the cointegration of the series is tested.

Remark 3.2: There are several other tests for testing the unit root. E.g. Phillips-Perron test is one which is frequently used. Furthermore, modifications for the testing procedure must be carried out if there is trend in the series.

Johansen's test procedure:*

In the Johansen testing there are two test statistics;

the <u>trace statistics</u> and
the <u>maximum eigenvalue</u> statistic.

The trace statistic tests the null hypothesis: "there are at most r cointegrating relations" against the alternative of "m cointegrating relations" (i.e., the series are stationary), $r = 0, 1, \ldots, m-1$.

The maximum eigenvalue statistic test the null hypohesis: "there are r coinegrating relations" against the alternative: "there are $r+1$ cointegrating relations".

*Johansen, Søren (1991). Estimating and testing cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59, 1551–1580.

Example 3.3: Test the integration of the three stock indexes shares price series. (EViews output). ADF with four lags.

Stock Exchange	ADF Test Statistic
Finland	-2.51
Norway	-1.07
Sweden	-1.35

1%	Critical Value*	-3.4418
5%	Critical Value	-2.8658
10%	Critical Value	-2.5691

*MacKinnon critical values for rejection of hypothesis of a unit root.

None of the test statistics are statistically significant. Hence all (logarithmic) stock indexes are integrated, and a further check of differences reveals that each index is integrated of order one.

If the series are integrated of the same order. Then the next step is to test whether the series are cointegrated.

Example 3.4: Test pair wise possible cointegration of the three stock index series

```
Included observations: 737
Test assumption: Linear deterministic trend in the data

Lags interval: 1 to 4
----------------------------------------------------------
         Likelihood     Critical Values      Hypothesized
Eigenvalue  Ratio    5 Percent     1 Percent    No. of CE(s)
----------------------------------------------------------
Series: LFIN LNOR
 0.011019  8.818672   15.41          20.04         None
 0.000885  0.652729    3.76          6.65       At most 1
Series: LFIN LSWE
 0.009158  7.464918   15.41         20.04          None
 0.000928  0.684529    3.76          6.65       At most 1
Series: LNOR LSWE
 0.026201 20.76955    15.41         20.04          None **
 0.001629  1.201820    3.76          6.65       At most 1
----------------------------------------------------------
*(**) denotes rejection of the hypothesis
at 5%(1%) significance level

Normalized Cointegrating Coefficients: Norway and Sweden

LNOR          LSWE          C
 1.000000  -0.938233  -0.381887
           (0.04471)
```

The test results suggest that Sweden and Norway are cointegrated.

Example 3.5: Consider the case series in Example 3.2.

```
Unit root tests: (ADF with intecept)

==============================
Series       ADF    p-val
------------------------------
  x        -2.201151    0.2066
  y        -2.445337    0.1306
------------------------------
 diff(x)  -15.35371     0.0000
 diff(y)  -15.31091     0.0000
==============================
```

ADF-tests indicate strong evidence that both series are $I(1)$.

However, the Johansen CI-tests below indicate that there is no sign of cointegration

```
Unrestricted Cointegration Rank Test (Trace)
==============================================================
Hypothesized                Trace                       0.05
No. of CE(s)  Eigenvalue  Statistic   Critical Value   Prob.**
--------------------------------------------------------------
None            0.050303    11.54141      15.49471       0.1803
At most 1       0.003082    0.651297      3.841466       0.4196
==============================================================
 Trace test indicates no cointegration at the 0.05 level
 * denotes rejection of the hypothesis at the 0.05 level
 **MacKinnon-Haug-Michelis (1999) p-values
```

3.3 Properties of cointegrated series

Granger's representation theorem:

Let $x_t \sim I(1)$ and $y_t \sim I(1)$ then x_t and y_t are cointegrated if and only if there exist the *error correction model* (ECM) such that

$$
\begin{aligned}
\Delta x_t &= -\gamma_1 z_{t-1} + \text{lagged}(\Delta x_t, \Delta y_t) + \delta(L)\epsilon_{1t} \\
\Delta y_t &= -\gamma_2 z_{t-1} + \text{lagged}(\Delta x_t, \Delta y_t) + \delta(L)\epsilon_{2t},
\end{aligned}
\tag{4}
$$

where

$$z_t = y_t - Ax_t \sim I(0), \tag{5}$$

$$\delta(L) = 1 + \delta_1 L + \cdots + \delta_p L^p. \tag{6}$$

is the same in both equations, and $|\gamma_1| + |\gamma_2| \neq 0$ (Engle and Granger 1987).

The relation

$$y_t = \beta_0 + \beta_1 x_t \tag{7}$$

can be interpreted as the equilibrium state between x_t and y_t. The residual term z_t is called the equilibrium error.

One implication of the theorem is that if $x_t, y_t \sim \mathrm{CI}(1,1)$ then there exist so called *Granger causality* at least in one direction, meaning that the one of the series can be predicted by the history of the other.

Note: On efficient markets there should not exist cointegration between for share prices, otherwise one could predict the values of one share by the historical values of the other series.

Further properties:*

(*i*) If x_t, y_t are cointegrated, so will be x_t and $by_{t-k}+w_t$, for any k where $w_t \sim I(0)$, with a possible change in cointegrating parameter. Formally, if x_t is $I(1)$ then x_t and x_{t-k} will be cointegrated for any k, but this is not an interesting property as it is true for any $I(1)$ process and so does not suggest a special relationship, unlike cointegration of a pair of $I(1)$ series. It follows that if x_t and y_t are cointegrated but are only observed with measurement error, then the two series will also be cointegrated if all measurement errors are $I(0)$.

*C.W.J. Granger (1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics and Statistics*, **48**:3, 213–228.

(*ii*) If x_t is $I(1)$ and $f_{n,h}(J_n)$ is the optimal forecast of x_{n+h}, based on the information set J_n available at time n, then x_{t+h}, $f_{t,h}(J_t)$ are cointegrated if J_n is a proper information set, that is if it includes x_{n-j}, $j \geq 0$. If J_n is not a proper information set, x_{t+n} and its optimum forecast are only cointegrated if x_t is cointegrated with variables in J_t.

(*iii*) If x_{n+h}, y_{n+h} are cointegrated series with parameter A and are optimally forecast using the information set $J_n : x_{n-j}, y_{n-j}$, $j \geq 0$, then the h-step forecast $f^x_{n,h}, f^y_{n,h}$ will obey $f^x_{n,h} = Af^y_{n,h}$ as $h \to \infty$. Thus, long-term forecast of x_t, y_t will be tied together by the equilibrium relationships. Forecasts formed without cointegration terms such as univariate forecasts will not necessarily have this property.

(*iv*) If T_t is an $I(1)$ target variable and x_t is an $I(1)$ controllable variable, then T_t, x_t will be cointegrated if optimum control is applied.

(*v*) If x_t, y_t are $I(1)$ and cointegrated, there must be Granger causality in at least one direction, as one variable can help forecast the other. This follows directly from the condition $|\rho_1| + |\rho_2| \neq 0$, as z_t must occur in at least one equation and thus knowledge of z_t must improve forecastability of at least one of x_t, y_t. Here causality is respect information set J_t defined in (*iii*).

(*vi*) If x_t, y_t are a pair of prices from a jointly efficient, speculative market, they cannot be cointegrated. This follows directly from (*v*) as if the two prices are cointegrated, one can be used to forecast the other and this would contradict the efficient market assumption. Thus, for example, cold and silver prices, if generated by an efficient market, cannot move closely together in the long-run.

Example 3.6: Vector Error Correction Model (VECM) for Norway and Sweden (EViews)

```
 Sample(adjusted): 4 742
 Included observations: 739 after adjusting end points
 t-statistics in parentheses

Cointegrating Eq:     CointEq1
LNOR(-1)              1.000000
LSWE(-1)             -0.936439
                     (0.04354)
                    (-21.5059)
C                    -0.394794

Error Correction:    D(LNOR)     D(LSWE)
CointEq1            -0.027201    0.032955
                   (-2.15480)   (2.25424)

D(LNOR(-1))         -0.020720   -0.038613
                   (-0.47054)  (-0.75717)

D(LNOR(-2))          0.017904   -0.045137
                   (0.40970)   (-0.89189)

D(LSWE(-1))     0.099523     0.087900
                   (2.60472)    (1.98651)

D(LSWE(-2))          0.038129    0.032427
                   (0.99712)    (0.73224)

C                    0.000417    0.000546
                   (1.37480)    (1.55140)

 R-squared           0.023823    0.011391
 Adj. R-squared      0.017165    0.004648
 S.E. equation       0.008220    0.009519
```

3.4 Trend and Intercept

Consider the following ECM

$$\Delta \mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \Gamma_1 \Delta \mathbf{x}_{t-1} + \cdots + \Gamma_p \Delta \mathbf{x}_{t-p} + \boldsymbol{\mu} + \boldsymbol{\delta} t + \boldsymbol{\epsilon}_t. \quad (8)$$

Then we can write

$$\begin{aligned} \boldsymbol{\delta} &= \boldsymbol{\alpha}\boldsymbol{\delta}_1 + \boldsymbol{\alpha}_\perp \boldsymbol{\delta}_2 \\ \boldsymbol{\mu} &= \boldsymbol{\alpha}\boldsymbol{\mu}_1 + \boldsymbol{\alpha}_\perp \boldsymbol{\mu}_2, \end{aligned} \quad (9)$$

where $\boldsymbol{\alpha}_\perp$ is a $m \times (m - r)$ matrix such that $\boldsymbol{\alpha}'\boldsymbol{\alpha}_\perp = \mathbf{0}$ (orthogonal complement of the column space spanned by the columns of $\boldsymbol{\alpha}$), $\boldsymbol{\delta}_1 = (\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\boldsymbol{\delta}$ is a r dimensional vector of *linear trend coefficients in the cointegrations relations*,
$\boldsymbol{\delta}_2 = (\boldsymbol{\alpha}_\perp'\boldsymbol{\alpha}_\perp)^{-1}\boldsymbol{\alpha}_\perp'\boldsymbol{\delta}$ is an $(m-r)$-dimensional vector of *quadratic trend coefficients in the data*,
$\boldsymbol{\mu}_1 = (\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\boldsymbol{\mu}$ is a r-dimensional vector of *intercepts in the cointegration relations*
$\boldsymbol{\mu}_2 = (\boldsymbol{\alpha}_\perp'\boldsymbol{\alpha}_\perp)^{-1}\boldsymbol{\alpha}_\perp'\boldsymbol{\mu}$ is an $(m - r)$-vector of *linear trend coefficients in the data*.

Using this decomposition, we can write

$$\Delta \mathbf{x}_t = \boldsymbol{\alpha} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\mu}_1' \\ \boldsymbol{\delta}_1' \end{pmatrix}' \tilde{\mathbf{x}}_{t-1} + \Gamma_1 \Delta \mathbf{x}_{t-1} + \cdots + \Gamma_p \Delta \mathbf{x}_{t-p} + \boldsymbol{\alpha}_\perp \mu_2 + \boldsymbol{\alpha}_\perp \boldsymbol{\delta}_2 t + \boldsymbol{\epsilon}_t. \tag{10}$$

where $\tilde{\mathbf{x}}_t = (\mathbf{x}_t', 1, t)'$.

As a consequence discussion of deterministic trend and constants must be related to cointegration space as well.

Case 1. *No restrictions on* $\boldsymbol{\delta}$ *and* $\boldsymbol{\mu}$: Then the model is consistent with linear trend in differences, $\Delta \mathbf{x}_t$, and quadratic tend in $\mathbf{x}_t$.

Case 2. $\boldsymbol{\delta}_2 = 0$, $\boldsymbol{\delta}_1, \mu_1$ and $\boldsymbol{\mu}_2$ unrestricted: When $\boldsymbol{\delta}_2 = 0$ the model is restricted to exclude quadratic trends in data. $\boldsymbol{\delta}_1 \neq 0$ implies that the cointegration relations may have linear trends, and $\boldsymbol{\mu}_2 \neq 0$ implies that the cointegration relation has an intercept as well. $\boldsymbol{\mu}_1 \neq 0$ implies that there is a linear trend in also in $\mathbf{x}_t$.

Case 3. $\boldsymbol{\delta} = 0$, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ unrestricted: This implies that there may be linear trend in $\mathbf{x}_t$ ($\boldsymbol{\mu}_2 \neq 0$), but no trend in the cointegration relations. There may also be intercepts ($\boldsymbol{\mu}_1 \neq 0$) in the cointegration relations.

Case 4. $\boldsymbol{\delta} = 0, \boldsymbol{\mu}_2 = 0$ and $\boldsymbol{\mu}_1$ unrestricted: No trend in data ($\boldsymbol{\delta}_2 = 0$, $\boldsymbol{\mu}_2 = 0$), and no trend in cointegration relation ($\boldsymbol{\delta}_1 = 0$), but there may be intercepts in the cointegration relations ($\boldsymbol{\mu}_1 \neq 0$).

Case 5. $\boldsymbol{\delta} = 0$ and $\boldsymbol{\mu} = 0$: No deterministic components in the data, and all intercepts in the cointegration relations are zero. This is an unlikely case, for the intercepts in cointegration relations are usually needed to account for the units of measurements.

Remark 3.3: The most common cases in practice are the cases 3 and 4.

3.5 Testing hypotheses

An important aspect in cointegration analysis is testing hypotheses about the cointegration relations and the adjustment coefficients.

Hypotheses about $\boldsymbol{\beta}$ can be presented in the form different forms.

In EViews the hypotheses are spcified by imposing restriction formulas between the parameters. Matrix A refers to matrix $\boldsymbol{\alpha}$ and B to $\boldsymbol{\beta}'$.

Example 3.7: Let $m = 4$, $r = 2$, and

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{11} & 0 \\ -\beta_{11} & \beta_{22} \\ 0 & \beta_{32} \\ \beta_{41} & 0 \end{pmatrix}. \tag{11}$$

Eviews refers to matrix $\boldsymbol{\beta}'$, such that $B = \boldsymbol{\beta}'$.

For example $B(2,3) = \beta_{32}$.

Accordingly in Eviews the restrictions are $B(2,1) = 0$, $B(1,2) = -B(1,1)$, $B(1,3) = 0$, $B(2,4) = 0$.

Special cases:

Consider $\mathbf{z}_t = (y_{1t}, y_{2t}, x_t)'$

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \\ \Delta x_t \end{pmatrix} = \Gamma_1 \begin{pmatrix} \Delta y_{1t-1} \\ \Delta y_{2t-1} \\ \Delta x_{t-1} \end{pmatrix}$$

$$+ \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{pmatrix}. \tag{12}$$

Exclusion of variables in the long-run relations

If we assume that x_{t-1} is not needed in the stationary long-run relations, then $\beta_{31} = \beta_{32} = 0$.

Example 3.8: Consider the (log) indexes of Helsinki, Stockholm, Oslo and Copenhagen. The Johansen's test give

```
Trend assumption: Linear deterministic trend (restrited)
Series: LFIN LSWE LNOR LDEN
Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Race) [EViews 5.1]
===============================================================
Hypothesized                  Trace         0.05
No. of CE(s)    Eigenvalue  Statistic   Critical Value  Prob.**
---------------------------------------------------------------
None    *        0.049872    69.05861     63.87610       0.0172
At most 1        0.020661    31.35471     42.91525       0.4239
At most 2        0.013643    15.96774     25.87211       0.4952
At most 3        0.007897    5.843383     12.51798       0.4804
===============================================================
 Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
 * denotes rejection of the hypothesis at the 0.05 level
 **MacKinnon-Haug-Michelis (1999) p-values
```

There seems to be one significant cointegration relation (as already found earlier).

Estimating β under this restriction yields.

```
Cointegration equation:

Normalized cointegrating coefficients (t-value in brackets)
===================================================================
LOG(FIN)   LOG(SWE)  LOG(NOR)  LOG(DEN)  @TREND(8/30/93)  const
-------------------------------------------------------------------
0.016332  1.000000 -1.225587   0.234336    2.34E-05     -0.218903
[0.31852]          [-7.30173]  [1.90740]  [ 0.38327]    [0.38327]
===================================================================
```

Test next whether Helsinki and Denmark are needed in the cointegration relation.

In EViews this can be accomplished by setting $B(1,1) = 0$ [$B(1,2) = -1$ scaling] and $B(1,4) = 0$ in the estimation.

Imposing this restriction yields

```
Cointegration Restrictions:
      B(1,1)=0, B(1,2) = 1, B(1,4)=0
Convergence achieved after 5 iterations.
Restrictions identify all cointegrating vectors
LR test for binding restrictions (rank = 1):
Chi-square(2)    2.598224
Probability      0.272774

Cointegrating Eq:
====================================================================
        LOG(FIN)  LOG(SWE)   LOG(NOR)  LOG(DEN)   @TREND     const
--------------------------------------------------------------------
           0         1      -0.934692     0      -7.73E-05  -0.492051
(std)                       (0.06881)            (3.6E-05)
[t-val]                    [-13.5840]            [-2.17208]
====================================================================
```

Hence, the data support these restrictions ($\chi^2 = 2.598$ with $df = 2$ has a p-value of 0.273).

Furthermore, it seems that Norway's coefficient does not deviate significantly from -1.

Imposing this restriction yields:

```
Cointegration Restrictions:
      B(1,1)=0, B(1,2) = 1, B(1,3) = -1, B(1,4) = 0
Convergence achieved after 2 iterations.
Restrictions identify all cointegrating vectors
LR test for binding restrictions (rank = 1):
Chi-square(3)    3.190732
Probability      0.363143

===================================================================
        LOG(FIN)  LOG(SWE)   LOG(NOR)  LOG(DEN)  @TREND     const
           0         1          -1        0      -4.65E-05 -0.041441
(std)                                             (1.8E-05)
[t-val]                                          [-2.60265]
===================================================================
```

Again this restriction is accepted.

We conclude that the cointegration relations between the Nordic countries is the trending spread between Sweden and Norway.

Restriction on $\boldsymbol{\alpha}$

Restrictions on $\boldsymbol{\alpha}$ (in EViews) can be specified in the same manner as in the $\boldsymbol{\beta}$ matrix.

Weak exogeneity

In $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$ the matrix $\boldsymbol{\alpha}$ represents the speed of adjustment to the disequilibrium.

If all α_{ij} in row i of $\boldsymbol{\alpha}$ are equal to zero, then the corresponding cointegration vector does not enter the equation determining the ith element of $\Delta\mathbf{x}_t$.

In this case the variable is *weakly exogenous* to the system.

Example 3.9: If x_t is weakly exogenous in model (12), then $\alpha_{31} = \alpha_{32} = 0$. This also implies that we can model the system as

$$\Delta \mathbf{y}_t = \Gamma_0 \Delta x_t + \tilde{\Gamma}_1 \Delta \mathbf{z}_{t-1} + \tilde{\alpha}\beta \mathbf{z}_{t-1} + \tilde{\epsilon}, \qquad (13)$$

where $\Delta \mathbf{y}_t = (y_{1t}, y_{2t})'$, $\Gamma_0 = (\gamma_{01}, \gamma_{02})'$, $\tilde{\Gamma}_1$ is a (2×3) matrix, and $\tilde{\boldsymbol{\alpha}}$ is a (2×2) matrix.

Example 3.10: The α-estimate of Copenhagen does not seem to be statistically significant, while for Helsinki it is (results not shown here).

Thus, because Copenhagen does not belong to the cointegration relation and its alpha seems to be zero, we can exclude Denmark from the forthcoming analysis.

Next we test whether Sweden is weakly exogenous to the long run relationship

The EViews results are:

```
Cointegration Restrictions:
      B(1,1) = 0, B(1,2) = 1, B(1,3) = -1, A(2,1)=0
Convergence achieved after 2 iterations.
Restrictions identify all cointegrating vectors
LR test for binding restrictions (rank = 1):
Chi-square(3)     6.426996
Probability       0.092586
```

The overall hypothesis indicates acceptance, but if we look at the chi-square difference of unrestricted and restricted model, which is $6.426996 - 0.978181 = 5.449$, which with 1 degree of freedom has a p-value of 0.0196, and hence does not support that the coefficient is zero given the earlier restrictions.

After specifying the long run relation we can move to analyze the short run effects more closely.

Before that we consider some points regarding the residual analysis.

3.6 Residual analysis

Residual analysis provides the user various descriptive statistics and misspesification tests.

Johansen suggests using residuals from the unrestricted model to decide whether the model is acceptable or not.

The assumption is that the error terms are independent and

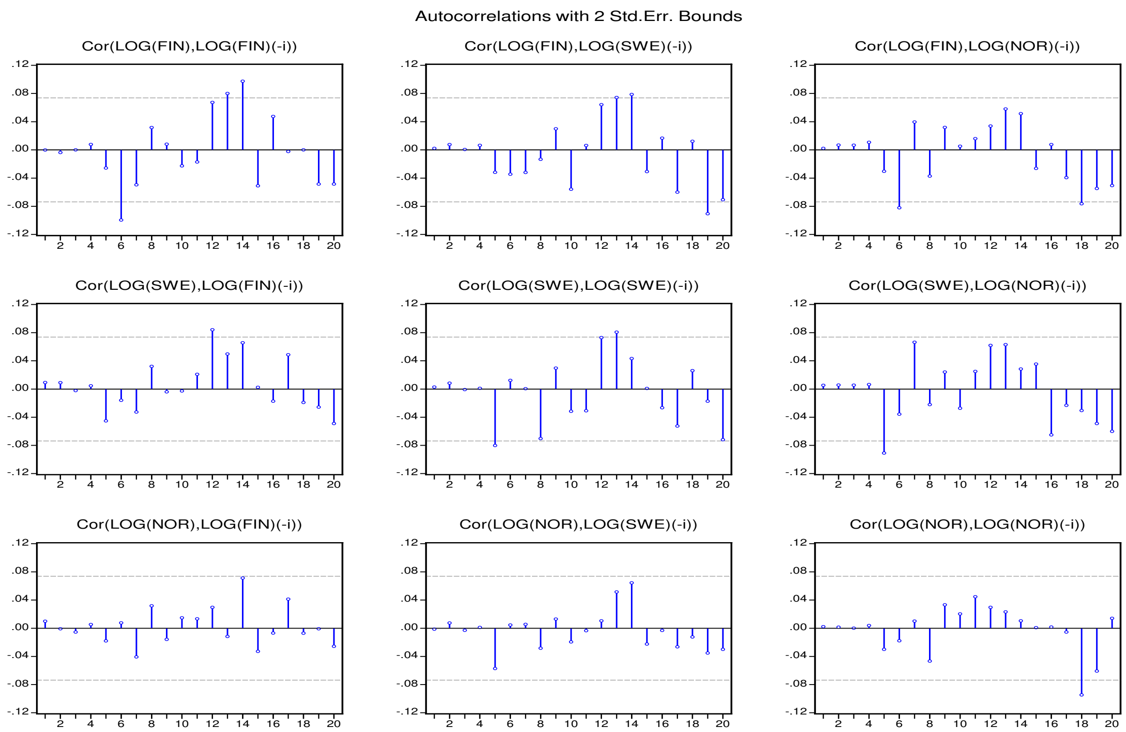
$$\epsilon_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}). \tag{14}$$

This implies that one should check: the normality, autocorrelaion of the residuals, and heteroscedasticity.

Autocorrelation

Visual inspection of the correlograms gives quick overview whether there is left significant correlation in the residuals.

In the above example the VAR(5) residual autocorrelaitons are:



Thus, there is potentially some autocorrelation left into the residuals.

Using the portmanteau test (28) of Section 2.2 yields:

VAR Residual Portmanteau Tests for Autocorrelations
H0: no residual autocorrelations up to lag h
Date: 04/10/07 Time: 00:43
Sample: 8/27/1993 8/08/1996
Included observations: 737

Lags	Q-Stat	Prob.	Adj Q-Stat	Prob.	df
1	0.188209	NA*	0.188465	NA*	NA*
2	0.472229	NA*	0.473258	NA*	NA*
3	0.565575	NA*	0.566985	NA*	NA*
4	0.682799	NA*	0.684849	NA*	NA*
5	9.412860	NA*	9.474541	NA*	NA*
6	22.38088	0.0077	22.54900	0.0073	9
7	34.21203	0.0119	34.49360	0.0109	18
8	46.69438	0.0107	47.11293	0.0096	27
9	49.30719	0.0688	49.75804	0.0633	36
10	55.81576	0.1295	56.35614	0.1194	45
11	62.33375	0.2040	62.97289	0.1886	54
12	71.83347	0.2085	72.62984	0.1904	63
13	83.91262	0.1592	84.92588	0.1415	72
14	94.74395	0.1410	95.96695	0.1226	81
15	99.84530	0.2242	101.1743	0.1977	90
16	107.9511	0.2530	109.4600	0.2219	99
17	117.1676	0.2572	118.8941	0.2228	108
18	131.9173	0.1636	134.0130	0.1345	117
19	142.8649	0.1446	145.2503	0.1156	126
20	151.8788	0.1521	154.5157	0.1200	135

*The test is valid only for lags larger than the VAR lag order.
df is degrees of freedom for (approximate) chi-square distribution

The Q-test indicate that there is obvious autocorrelation still in the residuals.

However, if we look at the residuals of the restricted ECM model, we have

```
VEC Residual Portmanteau Tests for Autocorrelations
H0: no residual autocorrelations up to lag h
Date: 04/10/07   Time: 00:51
Sample: 8/27/1993 8/08/1996
Included observations: 737
===============================================
Lags  Q-Stat     Prob.   Adj Q-Stat  Prob.   df
-----------------------------------------------
 1    0.894271    NA*    0.895486    NA*    NA*
 2    1.622145    NA*    1.625341    NA*    NA*
 3    2.474462    NA*    2.481141    NA*    NA*
 4    3.270679    NA*    3.281703    NA*    NA*
 5    11.31190   0.2549  11.37785    0.2507   9
 6    23.44896   0.1739  23.61453    0.1681  18
 7    37.94858   0.0787  38.25319    0.0739  27
 8    50.90791   0.0509  51.35473    0.0467  36
 9    54.15326   0.1647  54.64021    0.1537  45
10    60.55402   0.2513  61.12901    0.2353  54
11    68.52547   0.2954  69.22124    0.2756  63
12    78.34483   0.2846  79.20312    0.2622  72
13    90.99733   0.2098  92.08281    0.1879  81
14    101.3846   0.1937  102.6712    0.1704  90
15    107.4499   0.2639  108.8626    0.2339  99
16    115.0620   0.3031  116.6436    0.2682 108
17    123.7651   0.3165  125.5522    0.2778 117
18    137.4486   0.2290  139.5782    0.1927 126
19    147.2028   0.2231  149.5905    0.1846 135
20    156.3008   0.2284  158.9423    0.1865 144
===============================================
*The test is valid only for lags larger than the VAR lag order.
df is degrees of freedom for (approximate) chi-square distribution
```

The above tests suggest that there is in fact no material autocorrelation left behind.

This discrepancy may be due to some big jumps (kinds of outliers) in the return series that affect more in the unrestricted VAR.

Testing for Normality

The normality test relies usually on the skewness and kurtosis of the residuals.

In the previous example for the unrestricted VAR(5) test results are:

VAR Residual Normality Tests
Orthogonalization: Cholesky (Lutkepohl)
H0: residuals are multivariate normal
Sample: 8/27/1993 8/08/1996. Included observations: 737

Component	Skewness	Chi-sq	df	Prob.
1	-0.306533	11.54169	1	0.0007
2	-0.047667	0.279100	1	0.5973
3	0.117985	1.709895	1	0.1910
Joint		13.53068	3	0.0036

Component	Kurtosis	Chi-sq	df	Prob.
1	5.876937	254.1658	1	0.0000
2	5.182095	146.2189	1	0.0000
3	6.312256	336.9024	1	0.0000
Joint		737.2870	3	0.0000

Component	Jarque-Bera	df	Prob.
1	265.7074	2	0.0000
2	146.4980	2	0.0000
3	338.6123	2	0.0000
Joint	750.8177	6	0.0000

The error terms are obviously not normal.

The VECM estimation is to some extend robust for non-normality provided that the errors are symmetrically distributed.

Here even the symmetry is questionable.

Thus, caution must be exercised in interpreting the results.

Non-normality with fat tails is typical for return series.

Heteroscdasticity

In EViews there are available White heteroscedasticity test.

Choosing the White test with no cross terms yields

```
VAR Residual Heteroskedasticity Tests:
No Cross Terms (only levels and squares)
Sample: 8/27/1993 8/08/1996
Included observations: 737

   Joint test:
=========================
Chi-sq       df  Prob.
-------------------------
 369.1247   180  0.0000
=========================

   Individual components:
===============================================================
Dependent   R-squared   F(30,706)   Prob.   Chi-sq(30)  Prob.
---------------------------------------------------------------
res1*res1    0.106181    2.795633   0.0000  78.25531    0.0000
res2*res2    0.062923    1.580233   0.0261  46.37462    0.0286
res3*res3    0.090260    2.334867   0.0001  66.52170    0.0001
res2*res1    0.098325    2.566244   0.0000  72.46561    0.0000
res3*res1    0.069859    1.767484   0.0074  51.48593    0.0086
res3*res2    0.077629    1.980614   0.0015  57.21232    0.0020
===============================================================
```

Again the residual homoscedasticity is rejected.

A stylized fact is that return series have ARCH effect, which is causes here the heteroscedasticity.

However, there is some evidence that at least the cointegration rank tests are robust against moderate ARCH effects.

3.7 Estimates of the short run dynamics

The main focus in cointegration analysis is in the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, but the programs also give estimates for the short run matrices Γ_i.

Interpreting them includes the same difficulties as in the usual VAR-analysis, especially if there are many variables.

Structural equation modeling (SEM) can also be applied in cointegration frame work.

A general approach in modeling cointegrated systems may be done as follows:

a) Use Johansen approach to obtain the long-run cointegration relationships.

b) Estimate the VECM with the cointegration relationships explicitly included to obtain a parsimonious representation of the system.

c) Condition on any (weakly) exogenous variables thus obtaining a conditional VECM.

d) Model any simultaneous effects between the variables in the (conditional) model and test to ensure that the resulting restricted model parsimoniously encompasses the VECM.

Example 3.11. From the cointegration analysis we found one cointegration equation, which we identified as

(15)

$$ci_t = \log(\texttt{swe}) - \log(\texttt{nor}) - 0.04179 - 0.00004557t.$$

Next we can proceed to analyze the lag lengths in the equations [Step b)].

This can be worked out in VAR analysis (on the differences) with ci_{t-1} as an exogenous series.

From the VAR(5), the lag exclusion test gives

```
VAR Lag Exclusion Wald Tests
Sample: 8/27/1993 8/08/1996
Included observations: 736

Chi-squared test statistics for lag exclusion:
Numbers in [ ] are p-values
=====================================================
       DLOG(FIN)    DLOG(SWE)   DLOG(NOR)    Joint
-----------------------------------------------------
Lag 1     28.38679     6.702421     7.867347     31.41502
        [3.01e-06] [ 0.082012] [ 0.048834] [ 0.000251]

Lag 2     1.667570     1.941398     3.128891     8.742661
       [ 0.644168] [ 0.584658] [ 0.372177] [ 0.461360]

Lag 3     7.547036     3.497441     0.829432     13.63461
       [ 0.056362] [ 0.321094] [ 0.842415] [ 0.135929]

Lag 4     7.761285     2.391969     3.176845     15.36624
       [ 0.051212] [ 0.495131] [ 0.365155] [ 0.081355]

Lag 5     1.327498     6.814797     2.862830     9.582267
       [ 0.722614] [ 0.078041] [ 0.413261] [ 0.385348]
------------------------------------------------------
df          3             3             3            9
======================================================
```

The test indicates that only the first lag should be retained (in Swedish equation even that may be not needed).

Next we can proceed to the final models.

Because none of the series proved to be weakly exogenous, we skip step c).

The contemporaneous correlations are pretty high as is seen from the residual correlation matrix below.

Residual correlations:

	DLOG(FIN)	DLOG(SWE)	DLOG(NOR)
DLOG(FIN)	1.000000		
DLOG(SWE)	0.468284	1.000000	
DLOG(NOR)	0.427152	0.548415	1.000000

Unfortunately, we do not really have idea about the simultaneous causalities, we cannot properly perform the simultaneus equation modeling.

Instead we proceed by estimating each equation separately, in which case we can also model the ARCH-effect.

The final model estimation results e.g. for Finland are:

```
Dependent Variable: DLOG(FIN)
Method: ML - ARCH
Sample (adjusted): 8/31/1993 8/08/1996
Included observations: 740 after adjustments
Convergence achieved after 15 iterations
Bollerslev-Wooldrige robust standard errors & covariance
Variance backcast: ON
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

===========================================================
                Coefficient Std. Error  z-Statistic  Prob.
-----------------------------------------------------------
C                0.000654    0.000398    1.642192    0.1006
CI(-1)          -0.057754    0.015806   -3.654007    0.0003
DLOG(SWE(-1))    0.225844    0.048629    4.644254    0.0000
===========================================================

    Variance Equation
===========================================================
C               1.47E-05     7.47E-06    1.966886    0.0492
RESID(-1)^2     0.125340     0.058372    2.147252    0.0318
GARCH(-1)       0.770224     0.087350    8.817710    0.0000
===========================================================
R-squared             0.033462  Mean dependent var    0.000586
Adjusted R-squared    0.026878  S.D. dependent var    0.012140
S.E. of regression    0.011976  Akaike info crit.    -6.100877
Sum squared resid     0.105271  Schwarz criterion    -6.063526
Log likelihood        2263.324  F-statistic           5.082328
Durbin-Watson stat    1.939048  Prob(F-statistic)     0.000137
===========================================================
```

From the estimation result we observe the estimated adjustment coefficient $\widehat{\alpha} = -0.0578$ indicates a slow adjustment.

A 1 percent disequilibrium causes next day on average a 0.0578 percent adjustment in Finnish stock markets.

3.8 Cointegration and common trends

Stock and Watson (1988)* observation that cointegrated variables share common stochastic trends provides a useful way to understand cointegration relationships.

Consider two cointegrated I(1) variables y_t and x_t.

We can write each variable as

$$y_t = z_t + u_t \tag{16}$$

$$x_t = w_t + v_t, \tag{17}$$

where z_t and w_t are random walk precesses representing the (stochastic) trends of the variables and u_t and v_t are stationary processes.

*Stock, James and Mark Watson (1988). Testing for common trends *Journal of the American Statistical Association* 83, 1097–1107.

Because y_t and x_t are cointegrated there is a constant $a \neq 0$ such that $y_t - ax_t$ is stationary.

Assume for simplicity that $a = 1$.

Then $y_t - x_t = (z_t - w_t) + (u_t - v_t)$. Because this is stationary the random walk component must be zero, i.e., $z_t - w_t = 0$, or $z_t = w_t$.

Thus cointegration of x_t and y_t implies that they share the same common stochastic random walk componet.

Generally if we have m cointegrated I(1) series with cointegration rank $r < m$, then these series have $m - r$ common trends.

In the above example there are thus two common stochastic trends. However, it is not obvious what they are.