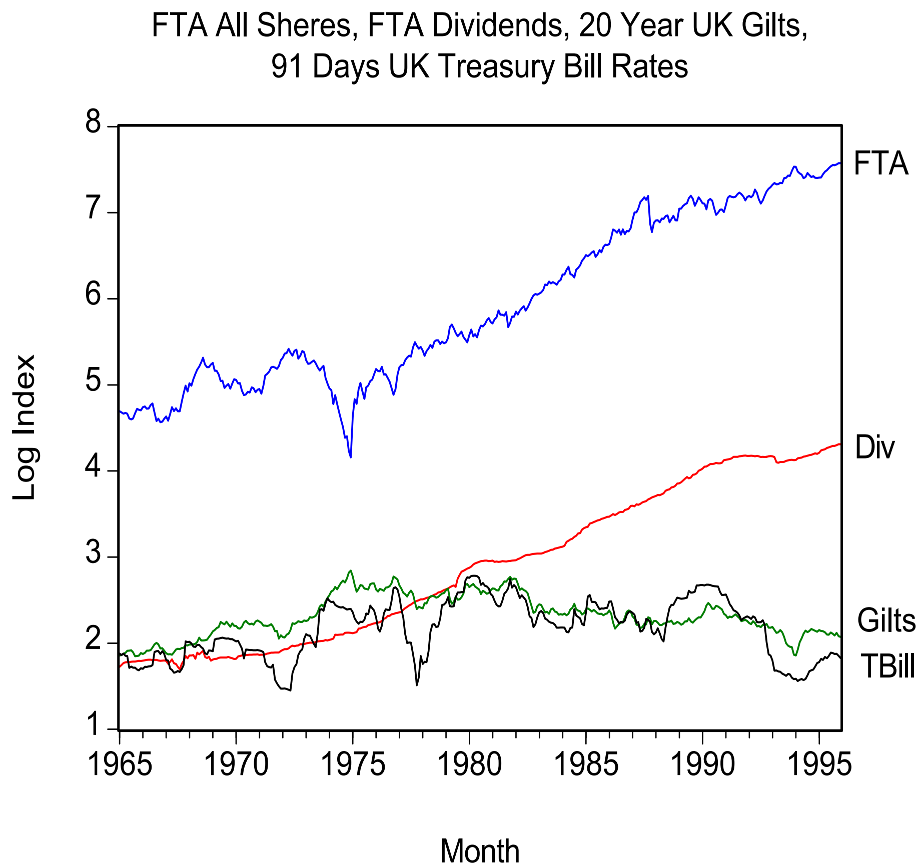
2. Multivariate Time Series

2.1 *Background*

Example. Consider the following monthly observations on FTA All Share index, the associated dividend index and the series of 20 year UK gilts and 91 day Treasury bills from January 1965 to December 1995 (372 months)



Potentially interesting questions:

1. Do some markets have a tendency to lead others?

2. Are there feedbacks between the markets?

3. How about contemporaneous movements?

4. How do impulses (shocks, innovations) transfer from one market to another?

5. How about common factors (disturbances, trend, yield component, risk)?

Most of these questions can be empirically investigated using tools developed in multivariate time series analysis.

Time series models

AR(p)-model

$$\begin{aligned} y_t &= \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t \\ \phi(L) y_t &= \mu + \epsilon_t, \end{aligned}$$

where $\epsilon_t \sim \text{WN}(0, \sigma_\epsilon^2)$ (White Noise), i.e.

$$\begin{aligned} E(\epsilon_t) &= 0, \\ E(\epsilon_t \epsilon_s) &= \begin{cases} \sigma_\epsilon^2 & \text{if } t = s \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

and $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ is the lag polynomial of order p with

$$L^k y_t = y_{t-k}$$

being the *L*ag operator ($L^0 y_t = y_t$).

The so called (weak) stationarity condition requires that the roots of the (characteristic) polynomial

$$\phi(L) = 0$$

should lie outside the unit circle, or equivalently the roots of

$$z^p - \phi_1 z^{p-1} - \cdots - \phi_{p-1} z - \phi_p = 0$$

are less than one in absolute value.

Note. Usually the series are centralized such that $\mu = 0$.

MA(q)-model

$$\begin{aligned} y_t &= \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \ldots - \theta_q \epsilon_{t-q} \\ &= \mu + \theta(L)\epsilon_t, \end{aligned}$$

where $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q$ is again a polynomial in L, this time, of order q, and $\epsilon_t \sim \text{WN}(0, \sigma_\epsilon^2)$.

Note. An MA-process is always stationary. But the so called invertibility condition requires that the roots of the characteristic polynomial $\theta(L) = 0$ lie outside the unit circle.

ARMA(p, q)-process

Compiling the two above together yields an ARMA$(p,\ q)$-process

$$\phi(L)y_t = \mu + \theta(L)\epsilon_t.$$

ARIMA(p, d, q)-process

A series is called integrated of order d, denoted as $y_t \sim I(d)$, if it becomes stationary after differencing d times. Furthermore, if

$$(1 - L)^d y_t \sim \text{ ARMA}(p, q),$$

we say that $y_t \sim$ ARIMA(p, d, q), where p denotes the order of the AR-lags, q the order of MA-lags, and d the order of differencing.

Example. Univariate time series models for the above (log) series look as follows. All the series prove to be $I(1)$.

Sample: 1965:01 1995:12
Included observations: 372

<table>
<tr><td colspan="7">FTA</td></tr>
<tr><th>Autocorrelation</th><th>Partial Correlation</th><th></th><th>AC</th><th>PAC</th><th>Q-Stat</th><th>Prob</th></tr>
<tr><td>.|********</td><td>.|********</td><td>1</td><td>0.992</td><td>0.992</td><td>368.98</td><td>0.000</td></tr>
<tr><td>.|********</td><td>.|. |</td><td>2</td><td>0.983</td><td>-0.041</td><td>732.51</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>3</td><td>0.975</td><td>0.021</td><td>1090.9</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>4</td><td>0.966</td><td>-0.025</td><td>1444.0</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>5</td><td>0.957</td><td>-0.024</td><td>1791.5</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>6</td><td>0.949</td><td>0.008</td><td>2133.6</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>7</td><td>0.940</td><td>0.005</td><td>2470.5</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>8</td><td>0.931</td><td>-0.007</td><td>2802.1</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>9</td><td>0.923</td><td>0.023</td><td>3128.8</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>10</td><td>0.915</td><td>-0.011</td><td>3450.6</td><td>0.000</td></tr>
<tr><td colspan="7">Dividends</td></tr>
<tr><th>Autocorrelation</th><th>Partial Correlation</th><th></th><th>AC</th><th>PAC</th><th>Q-Stat</th><th>Prob</th></tr>
<tr><td>.|********</td><td>.|********</td><td>1</td><td>0.994</td><td>0.994</td><td>370.59</td><td>0.000</td></tr>
<tr><td>.|********</td><td>.|. |</td><td>2</td><td>0.988</td><td>-0.003</td><td>737.78</td><td>0.000</td></tr>
<tr><td>.|********</td><td>.|. |</td><td>3</td><td>0.982</td><td>0.002</td><td>1101.6</td><td>0.000</td></tr>
<tr><td>.|********</td><td>.|. |</td><td>4</td><td>0.976</td><td>-0.004</td><td>1462.1</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>5</td><td>0.971</td><td>-0.008</td><td>1819.2</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>6</td><td>0.965</td><td>-0.006</td><td>2172.9</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>7</td><td>0.959</td><td>-0.007</td><td>2523.1</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>8</td><td>0.953</td><td>-0.004</td><td>2869.9</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>9</td><td>0.947</td><td>-0.006</td><td>3213.3</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>10</td><td>0.940</td><td>-0.006</td><td>3553.2</td><td>0.000</td></tr>
<tr><td colspan="7">T-Bill</td></tr>
<tr><th>Autocorrelation</th><th>Partial Correlation</th><th></th><th>AC</th><th>PAC</th><th>Q-Stat</th><th>Prob</th></tr>
<tr><td>.|********</td><td>.|********</td><td>1</td><td>0.980</td><td>0.980</td><td>360.26</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>**|. |</td><td>2</td><td>0.949</td><td>-0.301</td><td>698.79</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>3</td><td>0.916</td><td>0.020</td><td>1014.9</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>4</td><td>0.883</td><td>-0.005</td><td>1309.5</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>5</td><td>0.849</td><td>-0.041</td><td>1583.0</td><td>0.000</td></tr>
<tr><td>.|****** |</td><td>*|. |</td><td>6</td><td>0.811</td><td>-0.141</td><td>1833.1</td><td>0.000</td></tr>
<tr><td>.|****** |</td><td>.|. |</td><td>7</td><td>0.770</td><td>-0.018</td><td>2059.2</td><td>0.000</td></tr>
<tr><td>.|****** |</td><td>.|. |</td><td>8</td><td>0.730</td><td>0.019</td><td>2263.1</td><td>0.000</td></tr>
<tr><td>.|***** |</td><td>.|. |</td><td>9</td><td>0.694</td><td>0.058</td><td>2447.6</td><td>0.000</td></tr>
<tr><td>.|***** |</td><td>.|. |</td><td>10</td><td>0.660</td><td>-0.013</td><td>2615.0</td><td>0.000</td></tr>
<tr><td colspan="7">Gilts</td></tr>
<tr><th>Autocorrelation</th><th>Partial Correlation</th><th></th><th>AC</th><th>PAC</th><th>Q-Stat</th><th>Prob</th></tr>
<tr><td>.|********</td><td>.|********</td><td>1</td><td>0.984</td><td>0.984</td><td>362.91</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>*|. |</td><td>2</td><td>0.962</td><td>-0.182</td><td>710.80</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>3</td><td>0.941</td><td>0.050</td><td>1044.6</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>4</td><td>0.921</td><td>0.015</td><td>1365.5</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>5</td><td>0.903</td><td>0.031</td><td>1674.8</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>6</td><td>0.885</td><td>-0.038</td><td>1972.4</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>7</td><td>0.866</td><td>-0.001</td><td>2258.4</td><td>0.000</td></tr>
<tr><td>.|*******|</td><td>.|. |</td><td>8</td><td>0.848</td><td>0.019</td><td>2533.6</td><td>0.000</td></tr>
<tr><td>.|****** |</td><td>.|. |</td><td>9</td><td>0.832</td><td>0.005</td><td>2798.6</td><td>0.000</td></tr>
<tr><td>.|****** |</td><td>.|. |</td><td>10</td><td>0.815</td><td>-0.014</td><td>3053.6</td><td>0.000</td></tr>
</table>

Formally, as is seen below, the Dickey-Fuller (DF) unit root tests indicate that the series indeed all are $I(1)$. The test is based on the augmented DF-regression

$$\Delta y_t = \rho y_{t-1} + \alpha + \delta t + \sum_{i=1}^{4} \phi_i \Delta y_{t-i} + \epsilon_t,$$

and the hypothesis to be tested is

$$H_0 : \rho = 0 \text{ vs } H_1 : \rho < 0.$$

Test results:

Series	$\widehat{\rho}$	t-Stat
FTA	-0.030	-2.583
DIV	-0.013	-2.602
R20	-0.013	-1.750
T-BILL	-0.023	-2.403
ΔFTA	-0.938	-8.773
ΔDIV	-0.732	-7.300
ΔR20	-0.786	-8.129
ΔT-BILL	-0.622	-7.095
	ADF critical values	
Level	No trend	Trend
1%	-3.4502	-3.9869
5%	-2.8696	-3.4237
10%	-2.5711	-3.1345

Provided that the series are not *cointegrated* an appropriate modeling approach is VAR for the differences.

2.2 *Vector Autoregression (VAR)*

Suppose we have m time series y_{it}, $i = 1, \ldots, m$, and $t = 1, \ldots, T$ (common length of the time series). Then a *vector autoregression* model is defined as

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{mt} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{pmatrix} + \begin{pmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \cdots & \phi_{1m}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \cdots & \phi_{2m}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m1}^{(1)} & \phi_{m2}^{(1)} & \cdots & \phi_{mm}^{(1)} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \vdots \\ y_{m,t-1} \end{pmatrix} +$$

$$\cdots + \begin{pmatrix} \phi_{11}^{(p)} & \phi_{12}^{(p)} & \cdots & \phi_{1m}^{(p)} \\ \phi_{21}^{(p)} & \phi_{22}^{(p)} & \cdots & \phi_{2m}^{(p)} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m1}^{(p)} & \phi_{m2}^{(1)} & \cdots & \phi_{mm}^{(p)} \end{pmatrix} \begin{pmatrix} y_{1,t-p} \\ y_{2,t-p} \\ \vdots \\ y_{m,t-p} \end{pmatrix}$$

$$+ \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{mt} \end{pmatrix}.$$

In matrix notations

$$\mathbf{y}_t = \mu + \Phi_1 \mathbf{y}_{t-1} + \cdots + \Phi_p \mathbf{y}_{t-p} + \epsilon_t,$$

which can be further simplified by adopting the matric form of a lag polynomial

$$\Phi(L) = \mathbf{I} - \Phi_1 L - \ldots - \Phi_p L^p.$$

Thus finally we get

$$\Phi(L)\mathbf{y}_t = \epsilon_t.$$

Note that in the above model each y_{it} depends not only its own history but also on other series' history (cross dependencies). This gives us several additional tools for analyzing causal as well as feedback effects as we shall see after a while.

A basic assumption in the above model is that the residual vector follow a multivariate white noise, i.e.

$$\begin{aligned} \mathrm{E}(\epsilon_t) &= \mathbf{0} \\ \mathrm{E}(\epsilon_t \epsilon_s') &= \begin{cases} \Sigma_\epsilon & \text{if } t = s \\ \mathbf{0} & \text{if } t \neq s \end{cases} \end{aligned}$$

The coefficient matrices must satisfy certain constraints in order that the VAR-model is stationary. They are just analogies with the univariate case, but in matrix terms. It is required that roots of

$$|\mathbf{I} - \Phi_1 z - \Phi_2 z^2 - \cdots - \Phi_p z^p| = 0$$

lie outside the unit circle. Estimation can be carried out by single equation least squares.

Example. Let us estimate a VAR(1) model for the equity-bond data. First, however, test whether the series are cointegrated. As is seen below, there is no empirical evidence of cointegration (EViews results)

```
Sample(adjusted): 1965:06 1995:12
Included observations: 367 after adjusting end points
Trend assumption: Linear deterministic trend
Series: LFTA LDIV LR20 LTBILL
Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test
===================================================================
Hypothesized                Trace      5 Percent       1 Percent
No. of CE(s)  Eigenvalue  Statistic  Critical Value  Critical Value
===================================================================
None           0.047131    46.02621       47.21           54.46
At most 1      0.042280    28.30808       29.68           35.65
At most 2      0.032521    12.45356       15.41           20.04
At most 3      0.000872    0.320012        3.76            6.65
===================================================================
 *(**) denotes rejection of the hypothesis at the 5%(1%) level
 Trace test indicates no cointegration at both 5% and 1% levels
```

VAR(1) Estimates:

```
 Sample(adjusted): 1965:04 1995:12
 Included observations: 369 after adjusting end points
 Standard errors in ( ) & t-statistics in [ ]
============================================================
                       DFTA        DDIV        DR20      DTBILL
============================================================
DFTA(-1)           0.102018 -0.005389   -0.140021   -0.085696
                  (0.05407) (0.01280)   (0.02838)   (0.05338)
                  [1.88670] [-0.42107] [-4.93432]  [-1.60541]
DFTA(-2)          -0.170209  0.012231    0.014714    0.057226
                  (0.05564) (0.01317)   (0.02920)   (0.05493)
                 [-3.05895] [0.92869]   [0.50389]   [1.04180]
DDIV(-1)          -0.113741  0.035924    0.197934    0.280619
                  (0.22212) (0.05257)   (0.11657)   (0.21927)
                 [-0.51208] [0.68333]   [1.69804]   [1.27978]
DDIV(-2)           0.065178  0.103395    0.057329    0.165089
                  (0.22282) (0.05274)   (0.11693)   (0.21996)
                  [0.29252] [1.96055]   [0.49026]   [0.75053]
DR20(-1)          -0.359070 -0.003130    0.282760    0.373164
                  (0.11469) (0.02714)   (0.06019)   (0.11322)
                 [-3.13084] [-0.11530]  [4.69797]   [3.29596]
DR20(-2)           0.051323 -0.012058   -0.131182   -0.071333
                  (0.11295) (0.02673)   (0.05928)   (0.11151)
                  [0.45437] [-0.45102] [-2.21300]  [-0.63972]
DTBILL(-1)         0.068239  0.005752   -0.033665    0.232456
                  (0.06014) (0.01423)   (0.03156)   (0.05937)
                  [1.13472] [0.40412]  [-1.06672]   [3.91561]
DTBILL(-2)        -0.050220  0.023590    0.034734   -0.015863
                  (0.05902) (0.01397)   (0.03098)   (0.05827)
                 [-0.85082] [1.68858]   [1.12132]  [-0.27224]
C                  0.892389  0.587148   -0.033749  -0.317976
                  (0.38128) (0.09024)   (0.20010)   (0.37640)
                  [2.34049] [6.50626]  [-0.16867]  [-0.84479]
============================================================
```

Continues . . .

	DFTA	DDIV	DR20	DTBILL
R-squared	0.057426	0.028885	0.156741	0.153126
Adj. R-squared	0.036480	0.007305	0.138002	0.134306
Sum sq. resids	13032.44	730.0689	3589.278	12700.62
S.E. equation	6.016746	1.424068	3.157565	5.939655
F-statistic	2.741619	1.338486	8.364390	8.136583
Log likelihood	-1181.220	-649.4805	-943.3092	-1176.462
Akaike AIC	6.451058	3.569000	5.161567	6.425267
Schwarz SC	6.546443	3.664385	5.256953	6.520652
Mean dependent	0.788687	0.688433	0.052983	-0.013968
S.D. dependent	6.129588	1.429298	3.400942	6.383798

Determinant Residual Covariance	18711.41
Log Likelihood (d.f. adjusted)	-3909.259
Akaike Information Criteria	21.38352
Schwarz Criteria	21.76506

As is seen the number of estimated parameters grows rapidly very large.

Defining the order of a VAR-model

In the first step it is assumed that all the series in the VAR model have equal lag lengths. To determine the number of lags that should be included, criterion functions can be utilized the same manner as in the univariate case.

The underlying assumption with BIC and related criteria is that the residuals follow a multivariate normal distribution, i.e.

$$\epsilon \sim N_m(\mathbf{0}, \Sigma_\epsilon).$$

In the original forms AIC and BIC are defined as

$$\text{AIC} = -2 \log L + 2s$$

$$\text{BIC} = -2 \log L + s \log T$$

where L stands for the *Likelihood function*, and s denotes the number of estimated parameters.

The best fitting model is the one that minimizes the criterion function.

For example in a VAR(j) model with m equations there are $s = m(1 + jm) + m(m + 1)/2$ estimated parameters.

Under the normality assumption BIC can be simplified to

$$\text{BIC}(j) = \log \left| \hat{\Sigma}_{\epsilon,j} \right| + \frac{1}{T} m^2 j \log T, \quad j = 0, \ldots, p$$

and AIC to

$$AIC(j) = \log \left| \hat{\Sigma}_{\epsilon,j} \right| + 2\frac{j}{T}, \quad j = 0, \ldots, p,$$

where

$$\hat{\Sigma}_{\epsilon,j} = \frac{1}{T} \sum_{t=j+1}^{T} \hat{\epsilon}_{t,j} \hat{\epsilon}'_{t,j} = \frac{1}{T} \hat{\mathbf{E}}_j \hat{\mathbf{E}}'_j$$

with $\hat{\epsilon}_{t,j}$ the OLS residual vector of the VAR(j) model (i.e. VAR model estimated with j lags), and

$$\hat{\mathbf{E}}_j = \left[\hat{\epsilon}_{j+1,j}, \hat{\epsilon}_{j+2,j}, \cdots, \hat{\epsilon}_{T,j} \right]$$

an $m \times (T - j)$ matrix.

The likelihood ratio (LR) test can be also used in determining the order of a VAR. The test is generally of the form

$$\text{LR} = T(\log|\hat{\Sigma}_k| - \log|\hat{\Sigma}_p|),$$

where $\hat{\Sigma}_k$ denotes the maximum likelihood estimate of the residual covariance matrix of VAR(k) and $\hat{\Sigma}_p$ the estimate of VAR(p) ($p > k$) residual covariance matrix. If VAR(k) (the shorter model) is the true one, then

$$\text{LR} \sim \chi^2_{df},$$

where the degrees of freedom, df, equals the difference of in the number of estimated parameters between the two models.

In an m variate VAR(k)-model each series has $p - k$ lags less than those in VAR(p). Thus the difference in each equation is $m(p - k)$, so that in total $df = m^2(p - k)$.

Note that often, when T is small, a modified LR

$$\text{LR}^* = (T - mp)(\log|\hat{\Sigma}_k| - \log|\hat{\Sigma}_p|)$$

is used to correct possible small sample bias.

Example Let $p = 12$ then in the equity-bond data different VAR models yield the following results. Below are EViews results.

```
VAR Lag Order Selection Criteria
Endogenous variables: DFTA DDIV DR20 DTBILL
Exogenous variables: C
Sample: 1965:01 1995:12
Included observations: 359
```

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-3860.59	NA	26324.1	21.530	21.573	21.547
1	-3810.15	99.473	21728.6*	21.338*	21.554*	21.424*
2	-3796.62	26.385	22030.2	21.352	21.741	21.506
3	-3786.22	20.052	22729.9	21.383	21.945	21.606
4	-3783.57	5.0395	24489.4	21.467	22.193	21.750
5	-3775.66	14.887	25625.4	21.502	22.411	21.864
6	-3762.32	24.831	26016.8	21.517	22.598	21.947
7	-3753.94	15.400	27159.4	21.560	22.814	22.059
8	-3739.07	27.018*	27348.2	21.566	22.994	22.134
9	-3731.30	13.933	28656.4	21.612	23.213	22.248
10	-3722.40	15.774	29843.7	21.651	23.425	22.357
11	-3715.54	12.004	31443.1	21.702	23.649	22.476
12	-3707.28	14.257	32880.6	21.745	23.865	22.588

```
 * indicates lag order selected by the criterion
 LR: sequential modified LR test statistic
     (each test at 5% level)
 FPE: Final prediction error
 AIC: Akaike information criterion
 SC: Schwarz information criterion
 HQ: Hannan-Quinn information criterion
```

Criterion function minima are all at VAR(1) (SC or BIC just borderline). LR-tests suggest VAR(8). Let us look next at the residual autocorrelations.

To investigate whether the VAR residuals are White Noise, the hypothesis to be tested is

$$H_0 : \Upsilon_1 = \cdots = \Upsilon_h = 0$$

where $\Upsilon_k = (\rho_{ij}(k))$ is the autocorrelation matrix (see later in the notes) of the residual series with $\rho_{ij}(k)$ the cross autocorrelation of order k of the residuals series i and j. A general purpose (portmanteau) test is the Q-statistic[††]

$$Q_h = T \sum_{k=1}^{h} \text{tr}(\hat{\Upsilon}'_k \hat{\Upsilon}_0^{-1} \hat{\Upsilon}_k \hat{\Upsilon}_0^{-1}),$$

where $\hat{\Upsilon}_k = (\hat{\rho}_{ij}(k))$ are the estimated (residual) autocorrelations, and $\hat{\Upsilon}_0$ the contemporaneous correlations of the residuals. Alternatively (especially in small samples) a modified statistic is used

$$Q_h^* = T^2 \sum_{k=1}^{h} (T-k)^{-1} \text{tr}(\hat{\Upsilon}'_k \hat{\Upsilon}_0^{-1} \hat{\Upsilon}_k \hat{\Upsilon}_0^{-1}).$$

[††]See e.g. Lütkepohl, Helmut (1993). *Introduction to Multiple Time Series*, 2nd Ed., Ch. 4.4

The asymptotic distribution is χ^2 with $df = p^2(h-k)$. Note that in computer printouts h is running from $1, 2, \ldots h^*$ with h^* specified by the user.

```
VAR(1) Residual Portmanteau Tests for Autocorrelations
H0: no residual autocorrelations up to lag h
Sample: 1966:02 1995:12
Included observations: 359
================================================
Lags    Q-Stat     Prob.    Adj Q-Stat  Prob.    df
------------------------------------------------
 1     1.847020   NA*       1.852179   NA*      NA*
 2     27.66930   0.0346    27.81912   0.0332   16
 3     44.05285   0.0761    44.34073   0.0721   32
 4     53.46222   0.2725    53.85613   0.2603   48
 5     72.35623   0.2215    73.01700   0.2059   64
 6     96.87555   0.0964    97.95308   0.0843   80
 7     110.2442   0.1518    111.5876   0.1320   96
 8     137.0931   0.0538    139.0485   0.0424  112
 9     152.9130   0.0659    155.2751   0.0507  128
10     168.4887   0.0797    171.2972   0.0599  144
11     179.3347   0.1407    182.4860   0.1076  160
12     189.0256   0.2379    192.5120   0.1869  176
================================================
*The test is valid only for lags larger than the
VAR lag order. df is degrees of freedom for
(approximate) chi-square distribution
```

There is still left some autocorrelation into the VAR(1) residuals. Let us next check the residuals of the VAR(2) model

```
VAR Residual Portmanteau Tests for Autocorrelations
H0: no residual autocorrelations up to lag h
Sample: 1965:01 1995:12
Included observations: 369
==============================================
Lags   Q-Stat    Prob.   Adj Q-Stat Prob.  df
----------------------------------------------
 1    0.438464    NA*    0.439655   NA*   NA*
 2    1.623778    NA*    1.631428   NA*   NA*
 3    17.13353   0.3770  17.26832  0.3684  16
 4    27.07272   0.7143  27.31642  0.7027  32
 5    44.01332   0.6369  44.48973  0.6175  48
 6    66.24485   0.3994  67.08872  0.3717  64
 7    80.51861   0.4627  81.63849  0.4281  80
 8    104.3903   0.2622  106.0392  0.2271  96
 9    121.8202   0.2476  123.9049  0.2081 112
10    136.8909   0.2794  139.3953  0.2316 128
11    147.3028   0.4081  150.1271  0.3463 144
12    157.4354   0.5425  160.6003  0.4718 160
==============================================
*The test is valid only for lags larger than
the VAR lag order.
df is degrees of freedom for (approximate)
chi-square distribution
```

Now the residuals pass the white noise test. On the basis of these residual analyses we can select VAR(2) as the specification for further analysis. Mills (1999) finds VAR(6) as the most appropriate one. Note that there ordinary differences (opposed to log-differences) are analyzed. Here, however, log transformations are preferred.

Vector ARMA (VARMA)

Similarly as is done in the univariate case one can extend the VAR model to the vector ARMA model

$$\mathbf{y}_t = \mu + \sum_{i=1}^{p} \Phi_i \mathbf{y}_{t-i} + \epsilon_t - \sum_{j=1}^{q} \Theta_j \epsilon_{t-j}$$

or

$$\Phi(L)\mathbf{y}_t = \mu + \Theta(L)\epsilon_t,$$

where $\mathbf{y}_t$, μ, and ϵ_t are $m \times 1$ vectors, and Φ_i's and Θ_j's are $m \times m$ matrices, and

$$\begin{aligned} \Phi(L) &= \mathbf{I} - \Phi_1 L - \ldots - \Phi_p L^p \\ \Theta(L) &= \mathbf{I} - \Theta_1 L - \ldots - \Theta_q L^q. \end{aligned}$$

Provided that $\Theta(L)$ is invertible, we always can write the VARMA(p,q)-model as a VAR(∞) model with $\Pi(L) = \Theta^{-1}(L)\Phi(L)$. The presence of a vector MA component, however, complicates the analysis somewhat.

Autocorrelation and Autocovariance Matrices

The kth *cross autocorrelation* of the ith and jth time series, y_{it} and y_{jt} is defined as

$$\gamma_{ij}(k) = E(y_{it-k} - \mu_i)(y_{jt} - \mu_j).$$

Although for the usual autocovariance $\gamma_k = \gamma_{-k}$, the same is not true for the cross autocovariance, but $\gamma_{ij}(k) \neq \gamma_{ij}(-k)$. The corresponding cross autocorrelations are

$$\rho_{i,j}(k) = \frac{\gamma_{ij}(k)}{\sqrt{\gamma_i(0)\gamma_j(0)}}.$$

The kth *autocorrelation matrix* is then

$$\Upsilon_k = \begin{pmatrix} \rho_1(k) & \rho_{1,2}(k) & \dots & \rho_{1,m}(k) \\ \rho_{2,1}(k) & \rho_2(k) & \dots & \rho_{2,m}(k) \\ \vdots & \ddots & & \vdots \\ \rho_{m,1}(k) & \rho_{m,2}(k) & \dots & \rho_m(k) \end{pmatrix}.$$

Note: $\Upsilon_k = \Upsilon'_{-k}$.

The diagonal elements, $\rho_j(k)$, $j = 1, \ldots, m$ of Υ are the usual autocorrelations.

Example. Cross autocorrelations of the equity-bond data.

$$\hat{\Upsilon}_1 = \begin{array}{l} \\ \text{Div}_{t-1} \\ \text{Fta}_{t-1} \\ \text{R20}_{t-1} \\ \text{Tbl}_{t-1} \end{array} \begin{array}{cccc} \text{Div}_t & \text{Fta}_t & \text{R20}_t & \text{Tbl}_t \end{array} \\ \left(\begin{array}{rrrr} 0.0483 & -0.0099 & 0.0566 & 0.0779 \\ -0.0160 & 0.1225^* & -0.2968^* & -0.1620^* \\ 0.0056 & -0.1403^* & 0.2889^* & 0.3113^* \\ 0.0536 & -0.0266 & 0.1056^* & 0.3275^* \end{array} \right)$$

	Div_t	Fta_t	R20_t	Tbl_t
Div_{t-1}	0.0483	−0.0099	0.0566	0.0779
Fta_{t-1}	−0.0160	0.1225*	−0.2968*	−0.1620*
R20_{t-1}	0.0056	−0.1403*	0.2889*	0.3113*
Tbl_{t-1}	0.0536	−0.0266	0.1056*	0.3275*

Note: Correlations with absolute value exceeding $2 \times$ std err $=$ $2 \times /\sqrt{371} \approx 0.1038$ are statistically significant at the 5% level (indicated by $*$).

Example. Consider individual security returns and portfolio returns. For example French and Roll[‡‡] have found that daily returns of individual securities are slightly negatively correlated. The tables below, however suggest that daily returns of portfolios tend to be positively correlated!

‡‡French, K. and R. Ross (1986). Stock return variances: The arrival of information and reaction of traders. *Journal of Financial Economics*, **17**, 5–26.

One explanation might be that the cross autocorrelations are positive, which can be partially proved as follows: Let

$$r_{pt} = \frac{1}{m}\sum_{i=1}^{m} r_{it} = \frac{1}{m}\iota'\mathbf{r}_t$$

denote the return of an equal weighted index, where $\iota = (1,\ldots,1)'$ is a vector of ones, and $\mathbf{r}_t = (r_{1t},\ldots,r_{mt})'$ is the vector of returns of the securities. Then

$$\mathrm{cov}(r_{pt-1}, r_{pt}) = \mathrm{cov}\left[\frac{\iota'\mathbf{r}_{t-1}}{m}, \frac{\iota'\mathbf{r}_t}{m}\right] = \frac{\iota'\Gamma_1\iota}{m^2},$$

where Γ_1 is the first order autocovariance matrix.

Therefore

$$m^2\mathrm{cov}(r_{pt-1}, r_{pt}) = \iota'\Gamma_1\iota = \left(\iota'\Gamma_1\iota - \mathrm{tr}(\Gamma_1)\right) + \mathrm{tr}(\Gamma_1),$$

where $\mathrm{tr}(\cdot)$ is the trace operator which sums the diagonal elements of a square matrix. Consequently, because the right hand side tends to be positive and $\mathrm{tr}(\Gamma_1)$ tends to be negative, $\iota'\Gamma_1\iota - \mathrm{tr}(\Gamma_1)$, which contains only cross autocovariances, must be positive, and larger than the absolute value of $\mathrm{tr}(\Gamma_1)$, the autocovariances of individual stock returns.

2.3 *Exogeneity and Causality*

Consider the following extension of the VAR-model (*multivariate dynamic regression model*)

$$\mathbf{y}_t = \mathbf{C} + \sum_{i=1}^{p} \mathbf{A}_i' \mathbf{y}_{t-i} + \sum_{i=0}^{p} \mathbf{B}_i' \mathbf{x}_{t-i} + \epsilon_t,$$

where $p + 1 \leq t \leq T$, $\mathbf{y}_t' = (y_{1t}, \ldots, y_{mt})$, $\mathbf{C}$ is an $m \times 1$ vector of constants, $\mathbf{A}_1, \ldots, \mathbf{A}_p$ are $m \times m$ matrices of lag coefficients, $\mathbf{x}_t' = (x_{1t}, \ldots, x_{kt})$ is a $k \times 1$ vector of regressors, $\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_p$ are $k \times m$ coefficient matrices, and ϵ_t is an $m \times 1$ vector of errors having the properties

$$E(\epsilon_t) = E\{E(\epsilon_t | \mathbf{Y}_{t-1}, \mathbf{x}_t)\} = 0$$

and

$$E(\epsilon_t \epsilon_s') = E\{E(\epsilon_t \epsilon_s' | \mathbf{Y}_{t-1}, \mathbf{x}_t)\} = \begin{cases} \Sigma_\epsilon & t = s \\ \mathbf{0} & t \neq s, \end{cases}$$

where

$$\mathbf{Y}_{t-1} = (\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{y}_1).$$

We can compile this in matrix form

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{U},$$

where

$$\begin{aligned} \mathbf{Y} &= (\mathbf{y}_{p+1}, \ldots, \mathbf{y}_T)' \\ \mathbf{X} &= (\mathbf{X}_{p+1} \ldots, \mathbf{X}_T)' \\ \mathbf{X}_t &= (\mathbf{1}, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-p}, \mathbf{x}_1, \ldots, \mathbf{x}_{t-p}) \\ \mathbf{U} &= (\epsilon_{p+1}, \ldots, \epsilon_T)', \end{aligned}$$

and

$$\mathbf{B} = (\mathbf{C}', \mathbf{A}_1', \ldots, \mathbf{A}_p', \mathbf{B}_0, \ldots, \mathbf{B}_p').$$

The estimation theory for this model is basically the same s for the univariate linear regression. For example the LS and (approximate) ML estimator of $\mathbf{B}$ is

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$$

and the ML estimator of $\mathbf{\Sigma}_\epsilon$ is

$$\hat{\mathbf{\Sigma}}_\epsilon = \frac{1}{T}\hat{\mathbf{U}}'\hat{\mathbf{U}}, \qquad \hat{\mathbf{U}} = \mathbf{Y} - \mathbf{X}'\hat{\mathbf{B}}.$$

We say that $\mathbf{x}_t$ is *weakly exogenous** if the stochastic structure of $\mathbf{x}$ contains no information that is relevant for estimation of the parameters of interest, $\mathbf{B}$, and $\mathbf{\Sigma}_\epsilon$

If the conditional distribution of $\mathbf{x}_t$ given the past is independent of the history of $\mathbf{y}_t$ then $\mathbf{x}_t$ is said to be *strongly exogenous*.

*For a through discussion of *Exogeneity* see Engle, R.F., D.F. Hendry and J.F. Richard (1985). Exogeneity. *Econometrica*, **51**, 277–304.

Granger-causality and measures of feedback

One of the key questions that can be addressed to VAR-models is how useful some variables are for forecasting others.

If the history of x does not help to predict the future values of y, we say that x does not Granger-cause y.* Usually the prediction ability is measured in terms of the MSE (Mean Square Error). Hence, x fails to Granger-cause y, if for all $s > 0$

$$\begin{aligned} &\text{MSE}(\hat{y}_{t+s}|y_t, y_{t-1}, \ldots) \\ &\quad = \text{MSE}(\hat{y}_{t+s}|y_t, y_{t-1}, \ldots, x_t, x_{t-1}, \ldots), \end{aligned}$$

where (e.g.)

$$\begin{aligned} &\text{MSE}(\hat{y}_{t+s}|y_t, y_{t-1}, \ldots) \\ &\quad = E\left((y_{t+s} - \hat{y}_{t+s})^2 | y_t, y_{t-1}, \ldots\right). \end{aligned}$$

*Granger, C.W. (1969). *Econometrica* **37**, 424–438.
Sims, C.A. (1972). *American Economic Review*, **62**, 540–552.

Note. This is essentially the same that y is strongly exogenous to x.

In terms of VAR models this can be expressed as follows:
Consider the $g = m + k$ dimensional vector $\mathbf{z}'_t = (\mathbf{y}'_t, \mathbf{x}'_t)$, which is assumed to follow a VAR(p) model

$$\mathbf{z}_t = \sum_{i=1}^{p} \Pi_i \mathbf{z}_{t-i} + \nu_t,$$

where

$$\begin{aligned} E(\nu_t) &= \mathbf{0} \\ E(\nu_t \nu'_s) &= \begin{cases} \Sigma_\nu, & t = s \\ \mathbf{0}, & t \neq s. \end{cases} \end{aligned}$$

Partition the VAR of $\mathbf{z}$ as

$$\begin{aligned}
\mathbf{y}_t &= \textstyle\sum_{i=1}^{p} \mathbf{C}_{2i}\mathbf{x}_{t-i} + \sum_{i=1}^{p} \mathbf{D}_{2i}\mathbf{y}_{t-i} + \nu_{1t} \\
\mathbf{x}_t &= \textstyle\sum_{i=1}^{p} \mathbf{E}_{2i}\mathbf{x}_{t-i} + \sum_{i=1}^{p} \mathbf{F}_{2i}\mathbf{y}_{t-i} + \nu_{2t}
\end{aligned}$$

where $\nu_t' = (\nu_{1t}', \nu_{2t}')$ and Σ_ν are correspondingly partitioned as

$$\Sigma_\nu = \begin{pmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

with $E(\nu_{it}\nu_{jt}') = \Sigma_{ij}$, $i, j = 1, 2$.

Now $\mathbf{x}$ does not Granger-cause $\mathbf{y}$ if and only if $\mathbf{C}_{2i} \equiv \mathbf{0}$, or equivalently, if and only if $|\Sigma_{11}| = |\Sigma_1|$, where $\Sigma_1 = E(\eta_{1t}\eta_{1t}')$ with η_{1t} from the regression

$$\mathbf{y}_t = \sum_{i=1}^{p} \mathbf{C}_{1i}\mathbf{y}_{t-i} + \eta_{1t}.$$

Changing the roles of the variables we get the necessary and sufficient condition of $\mathbf{y}$ *not* Granger-causing $\mathbf{x}$.

It is also said that $\mathbf{x}$ is *block-exogenous* with respect to $\mathbf{y}$.

Testing for the Granger-causality of $\mathbf{x}$ on $\mathbf{y}$ reduces to testing for the hypothesis

$$H_0 \ : \ \mathbf{C}_{2i} = \mathbf{0}.$$

This can be done with the likelihood ratio test by estimating with OLS the restricted* and non-restricted† regressions, and calculating the respective residual covariance matrices:

Unrestricted:

$$\hat{\Sigma}_{11} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\nu}_{1t} \hat{\nu}'_{1t}.$$

Restricted:

$$\hat{\Sigma}_{1} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\eta}_{1t} \hat{\eta}'_{1t}.$$

*Perform OLS regressions of each of the elements in $\mathbf{y}$ on a constant, p lags of the elements of $\mathbf{x}$ and p lags of the elements of $\mathbf{y}$.

†Perform OLS regressions of each of the elements in $\mathbf{y}$ on a constant and p lags of the elements of $\mathbf{y}$.

The LR test is then

$$\text{LR} = (T-p)\left(\ln|\hat{\Sigma}_1| - \ln|\hat{\Sigma}_{11}|\right) \sim \chi^2_{mkp},$$

if H_0 is true.

Example. Granger causality between pairwise equity-bond market series

Pairwise Granger Causality Tests
Sample: 1965:01 1995:12
Lags: 12

Null Hypothesis:	Obs	F-Statistic	Probability
DFTA does not Granger Cause DDIV	365	0.71820	0.63517
DDIV does not Granger Cause DFTA		1.43909	0.19870
DR20 does not Granger Cause DDIV	365	0.60655	0.72511
DDIV does not Granger Cause DR20		0.55961	0.76240
DTBILL does not Granger Cause DDIV	365	0.83829	0.54094
DDIV does not Granger Cause DTBILL		0.74939	0.61025
DR20 does not Granger Cause DFTA	365	1.79163	0.09986
DFTA does not Granger Cause DR20		3.85932	0.00096
DTBILL does not Granger Cause DFTA	365	0.20955	0.97370
DFTA does not Granger Cause DTBILL		1.25578	0.27728
DTBILL does not Granger Cause DR20	365	0.33469	0.91843
DR20 does not Granger Cause DTBILL		2.46704	0.02377

The p-values indicate that FTA index returns Granger cause 20 year Gilts, and Gilts lead Treasury bill.

Let us next examine the block exogeneity between the bond and equity markets (two lags). Test results are in the table below.

Direction	LoglU	LoglR	2(LU-LR)	df	p-value
(Tbill, R20) --> (FTA, Div)	-1837.01	-1840.22	6.412	8	0.601
(FTA,Div) --> (Tbill, R20)	-2085.96	-2096.01	20.108	8	0.010

The test results indicate that the equity markets are Granger-causing bond markets. That is to some extend previous changes in stock markets can be used to predict bond markets.

2.4 *Geweke's[*] measures of Linear Dependence*

Above we tested Granger-causality, but there are several other interesting relations that are worth investigating.

Geweke has suggested a measure for *linear feedback from* $\mathbf{x}$ *to* $\mathbf{y}$ based on the matrices Σ_1 and Σ_{11} as

$$F_{\mathbf{x}\to\mathbf{y}} = \ln(|\Sigma_1|/|\Sigma_{11}|),$$

so that the statement that "$\mathbf{x}$ does not (Granger) cause $\mathbf{y}$" is equivalent to $F_{\mathbf{x}\to\mathbf{y}} = 0$. Similarly the *measure of linear feedback from* $\mathbf{y}$ *to* $\mathbf{x}$ is defined by

$$F_{\mathbf{y}\to\mathbf{x}} = \ln(|\Sigma_2|/|\Sigma_{22}|).$$

[*] Geweke (1982) *Journal of the American Statistical Association*, **79**, 304–324.

It may also be interesting to investigate the *instantaneous causality* between the variables. For the purpose, premultiplying the earlier VAR system of $\mathbf{y}$ and $\mathbf{x}$ by

$$\begin{pmatrix} \mathbf{I}_m & -\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{12}'\Sigma_{11}^{-1} & \mathbf{I}_k \end{pmatrix}$$

gives a new system of equations, where the first m equations become

$$\mathbf{y}_t = \sum_{i=0}^{p} \mathbf{C}_{3i}\mathbf{x}_{t-i} + \sum_{i=1}^{p} \mathbf{D}_{3i}\mathbf{y}_{t-i} + \omega_{1t},$$

with the error $\omega_{1t} = \nu_{1t} - \Sigma_{12}\Sigma_{22}^{-1}\nu_{2t}$ that is uncorrelated with $\mathbf{v}_{2t}$* and consequently with $\mathbf{x}_t$ (important!).

*$\text{cov}(\omega_{1t}, \nu_{2t}) = \text{cov}(\nu_{1t} - \Sigma_{12}\Sigma_{22}^{-1}\nu_{2t}, \nu_{2t}) = \text{cov}(\nu_{1t}, \nu_{2t}) - \Sigma_{12}\Sigma_{22}^{-1}\text{cov}(\nu_{2t}, \nu_{2t}) = \Sigma_{12} - \Sigma_{12} = \mathbf{0}$. Note further that $\text{cov}(\omega_{1t}) = \text{cov}(\nu_{1t} - \Sigma_{12}\Sigma_{22}^{-1}\nu_{2t}) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} =: \Sigma_{11.2}$. Similarly $\text{cov}(\omega_{2t}) = \Sigma_{22.1}$.

Similarly, the last k equations can be written as

$$\mathbf{x}_t = \sum_{i=1}^{p} \mathbf{E}_{3i}\mathbf{x}_{t-i} + \sum_{i=0}^{p} \mathbf{F}_{3i}\mathbf{y}_{t-i} + \omega_{2t}.$$

Denoting $\Sigma_{\omega i} = E(\omega_{it}\omega_{it}')$, $i = 1, 2$, there is *instantaneous causality* between $\mathbf{y}$ and $\mathbf{x}$ if and only if $\mathbf{C}_{30} \neq \mathbf{0}$ and $\mathbf{F}_{30} \neq \mathbf{0}$ or, equivalently, $|\Sigma_{11}| > |\Sigma_{\omega 1}|$ and $|\Sigma_{22}| > |\Sigma_{\omega 2}|$. Analogously to the linear feedback we can define *instantaneous linear feedback*

$$F_{\mathbf{x}\cdot\mathbf{y}} = \ln(|\Sigma_1|/|\Sigma_{\omega 1}|) = \ln(|\Sigma_2|/|\Sigma_{\omega 2}|).$$

A concept closely related to the idea of linear feedback is that of *linear dependence*, a measure of which is given by

$$F_{\mathbf{x},\mathbf{y}} = F_{\mathbf{x}\to\mathbf{y}} + F_{\mathbf{y}\to\mathbf{x}} + F_{\mathbf{x}\cdot\mathbf{y}}.$$

Consequently the linear dependence can be decomposed additively into three forms of feedback. Absence of a particular causal ordering is then equivalent to one of these feedback measures being zero.

Using the method of least squares we get estimates for the various matrices above as

$$\widehat{\Sigma}_i = (T-p)^{-1} \sum_{t=p+1}^{T} \widehat{\eta}_{it}\widehat{\eta}_{it}',$$

$$\widehat{\Sigma}_{ii} = (T-p)^{-1} \sum_{t=p+1}^{T} \widehat{\nu}_{it}\widehat{\nu}_{it}',$$

$$\widehat{\Sigma}_{\omega i} = (T-p)^{-1} \sum_{t=p+1}^{T} \widehat{\omega}_{it}\widehat{\omega}_{it}',$$

for $i = 1, 2$. For example

$$\widehat{F}_{\mathbf{x}\to\mathbf{y}} = \ln(|\widehat{\Sigma}_1|/|\widehat{\Sigma}_{11}|).$$

With these estimates one can test the particular dependencies,

No Granger-causality: $\mathbf{x} \to \mathbf{y}$, $H_{01} : F_{\mathbf{x}\to\mathbf{y}} = 0$

$$(T-p)\hat{F}_{\mathbf{x}\to\mathbf{y}} \sim \chi^2_{mkp}.$$

No Granger-causality: $\mathbf{y} \to \mathbf{x}$, $H_{02} : F_{\mathbf{y}\to\mathbf{x}} = 0$

$$(T-p)\hat{F}_{\mathbf{y}\to\mathbf{x}} \sim \chi^2_{mkp}.$$

No instantaneous feedback: $H_{03} : F_{\mathbf{x}\cdot\mathbf{y}} = 0$

$$(T-p)\hat{F}_{\mathbf{x}\cdot\mathbf{y}} \sim \chi^2_{mk}.$$

No linear dependence: $H_{04} : F_{\mathbf{x},\mathbf{y}} = 0$

$$(T-p)\hat{F}_{\mathbf{x},\mathbf{y}} \sim \chi^2_{mk(2p+1)}.$$

This last is due to the asymptotic independence of the measures $F_{\mathbf{x}\to\mathbf{y}}$, $F_{\mathbf{y}\to\mathbf{x}}$ and $F_{\mathbf{x}\cdot\mathbf{y}}$.

There are also so called Wald and Lagrange Multiplier (LM) tests for these hypotheses that are asymptotically equivalent to the LR test.

Note that in each case $(T-p)\hat{F}$ is the LR-statistic.

Example. The LR-statistics of the above measures and the associated χ^2 values for the equity-bond data are reported in the following table with $p = 2$.

[$\mathbf{y} = (\Delta \log \text{FTA}_t, \Delta \log \text{DIV}_t)$ and $\mathbf{x}' = (\Delta \log \text{Tbill}_t, \Delta \log \text{r20}_t)$]

	LR	DF	P-VALUE
x-->y	6.41	8	0.60118
y-->x	20.11	8	0.00994
x.y	23.31	4	0.00011
x,y	149.83	20	0.00023

Note. The results lead to the same inference as in Mills (1999), p. 251, although numerical values are different [in Mills VAR(6) is analyzed and here VAR(2)].

2.5 *Variance decomposition and innovation accounting*

Consider the VAR(p) model

$$\Phi(L)\mathbf{y}_t = \epsilon_t,$$

where

$$\Phi(L) = \mathbf{I} - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_p L^p$$

is the (matric) lag polynomial.

Provided that the stationary conditions hold we have analogously to the univariate case the vector MA representation of $\mathbf{y}_t$ as

$$\mathbf{y}_t = \Phi^{-1}(L)\epsilon_t = \epsilon_t + \sum_{i=1}^{\infty} \Psi_i \epsilon_{t-i},$$

where Ψ_i is an $m \times m$ coefficient matrix. Now ϵ_t's represent shocks in the system. Suppose we have a unit change in ϵ_t then its effect in $\mathbf{y}$ s periods ahead is

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \epsilon_t} = \Psi_s.$$

Accordingly the interpretation of the Ψ matrices is that they represent marginal effects, or the model's response to a unit shock (or innovation) at time point t in *each* of the variables. Economists call such parameters are as *dynamic multipliers*.

For example if we were told that the first element in ϵ_t changes by δ_1 at the same time that the second element changed by $\delta_2, \ldots,$ and the mth element by δ_m, then the combined effect of these changes on the value of the vector $\mathbf{y}_{t+s}$ would be given by

$$\Delta \mathbf{y}_{t+s} = \frac{\partial \mathbf{y}_{t+s}}{\partial \epsilon_{1t}} \delta_1 + \cdots + \frac{\partial \mathbf{y}_{t+s}}{\partial \epsilon_{mt}} \delta_m = \Psi_s \delta,$$

where $\delta' = (\delta_1, \ldots, \delta_m)$.

The response of y_i to a unit shock in y_j is given the sequence, known as the *impulse multiplier function*,

$$\psi_{ij,1}, \psi_{ij,2}, \psi_{ij,3}, \cdots,$$

where $\psi_{ij,k}$ is the ijth element of the matrix Ψ_k $(i, j = 1, \ldots, m)$. *Generally an impulse response function traces the effect of a one-time shock to one of the innovations on current and future values of the endogenous variables.*

What about exogeneity (or Granger-causality)?

Suppose we have a bivariate VAR system such that x_t does not Granger cause y. Then we can write

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \phi_{11}^{(1)} & 0 \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \cdots + \begin{pmatrix} \phi_{11}^{(p)} & 0 \\ \phi_{21}^{(p)} & \phi_{22}^{(p)} \end{pmatrix} \begin{pmatrix} y_{t-p} \\ x_{t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}.$$

Then under the stationarity condition

$$(\mathbf{I} - \Phi(L))^{-1} = \mathbf{I} + \sum_{i=1}^{\infty} \Psi_i L^i,$$

where

$$\Psi_i = \begin{pmatrix} \psi_{11}^{(i)} & 0 \\ \psi_{21}^{(i)} & \psi_{22}^{(i)} \end{pmatrix}.$$

Hence, we see that variable $\mathbf{y}$ does not react to a shock of $\mathbf{x}$.

Generally, if a variable, or a block of variables, are strictly exogenous, then the implied zero restrictions ensure that these variables do not react to a shock to any of the endogenous variables. Nevertheless it is advised to be careful when interpreting the possible (Granger) causalities in the philosophical sense of causality between variables.

Othogonalized impulse response function

Noting that $E(\epsilon_t \epsilon_t') = \Sigma_\epsilon$, we observe that the components of ϵ_t are contemporaneously correlated, meaning that they have overlapping information to some extend.

Example. For example in equity-bond data the contemporaneous VAR(2)-residual correlations are

	FTA	DIV	R20	TBILL
FTA	1			
DIV	0.123	1		
R20	-0.247	-0.013	1	
TBILL	-0.133	0.081	0.456	1

Many times, however, it is of interest to know how "new" information on y_{jt} makes us revise our forecasts on $\mathbf{y}_{t+s}$. To single out the individual effects the residuals must be first orthogonalized, such that they become contemporaneously uncorrelated (they are already serially uncorrelated).

Unfortunately orthogonalization, however, is not unique in the sense that changing the order of variables in $\mathbf{y}$ chances the results. Nevertheless there are some guidelines (based on the economic theory) how the ordering might be done in a specific situation.

Whatever the case, if we define a lower triangular matrix $\mathbf{S}$ such that $\mathbf{SS}' = \Sigma_\epsilon$, and

$$\nu_t = \mathbf{S}^{-1}\epsilon_t,$$

then $\mathbf{I} = \mathbf{S}^{-1}\Sigma_\epsilon\mathbf{S}'^{-1}$, implying
$E(\nu_t\nu_t') = \mathbf{S}^{-1}E(\epsilon_t\epsilon_t')\mathbf{S}'^{-1} = \mathbf{S}^{-1}\Sigma_\epsilon\mathbf{S}'^{-1} = \mathbf{I}$.

Consequently the new residuals are both uncorrelated over time as well as across equations. Furthermore, they have unit variances.

The new vector MA representation becomes

$$\mathbf{y}_t = \sum_{i=0}^{\infty} \psi_i^*\nu_{t-i},$$

where $\psi_i^* = \psi_i \mathbf{S}^{-1}$ ($m \times m$ matrices) so that $\psi_0^* = \mathbf{S}^{-1}$. The impulse response function of y_i to a unit shock in y_j is then given by

$$\psi_{ij,0}^*, \psi_{ij,1}^*, \psi_{ij,2}^*, \ldots$$

Variance decomposition

The uncorrelatedness of ν_ts allow the error variance of the s step-ahead forecast of y_{it} to be decomposed into components accounted for by these shocks, or innovations (this is why this technique is usually called *innovation accounting*). Because the innovations have unit variances (besides the uncorrelatedness), the components of this error variance accounted for by innovations to y_j is given by

$$\sum_{k=0}^{s} \psi_{ij,k}^{*2}.$$

Comparing this to the sum of innovation responses we get a relative measure how important variable js innovations are in the explaining the variation in variable i at different step-ahead forecasts, i.e.,

$$R^2_{ij,s} = 100 \frac{\sum_{k=0}^{s-1} \psi^{*2}_{ij,k}}{\sum_{h=1}^{m} \sum_{k=0}^{s-1} \psi^{*2}_{ih,k}}.$$

Thus, while impulse response functions traces the effects of a shock to one endogenous variable on to the other variables in the VAR, variance decomposition separates the variation in an endogenous variable into the component shocks to the VAR.

Letting s increase to infinity on gets the portion of the total variance of y_j that is due to the disturbance term ϵ_j of y_j.

On the ordering of variables

As was mentioned earlier, when there is contemporaneous correlation between the residuals, i.e., $\text{cov}(\epsilon_t) = \Sigma_\epsilon \neq \mathbf{I}$ the orthogonalized impulse response coefficients are not unique. There are no statistical methods to define the ordering. It must be done by the analyst!

It is recommended that various orderings should be tried to see whether the resulting interpretations are consistent. The principle is that the first variable should be selected such that it is the only one with potential immediate impact on all other variables. The second variable may have an immediate impact on the last $m-2$ components of $\mathbf{y}_t$, but not on y_{1t}, the first component, and so on. Of course this is usually a difficult task in practice.

Selection of the $\mathbf{S}$ matrix

Selection of the $\mathbf{S}$ matrix, where $\mathbf{SS}' = \Sigma_\epsilon$, actually defines also the ordering of variables. Selecting it as a lower triangular matrix implies that the first variable is the one affecting (potentially) the all others, the second to the $m-2$ rest (besides itself) and so on.

One generally used method in choosing $\mathbf{S}$ is to use *Cholesky decomposition* which results to a lower triangular matrix with positive main diagonal elements.

Example. Let us choose in our example two orderings. One according to the feedback analysis

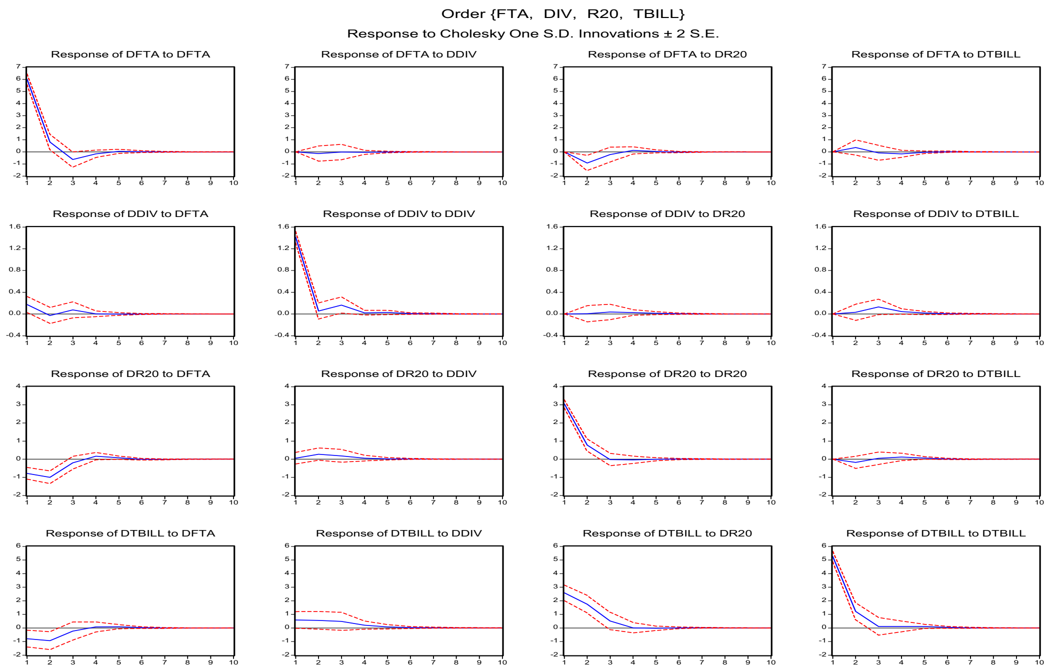
[(I: FTA, DIV, R20, TBILL)],

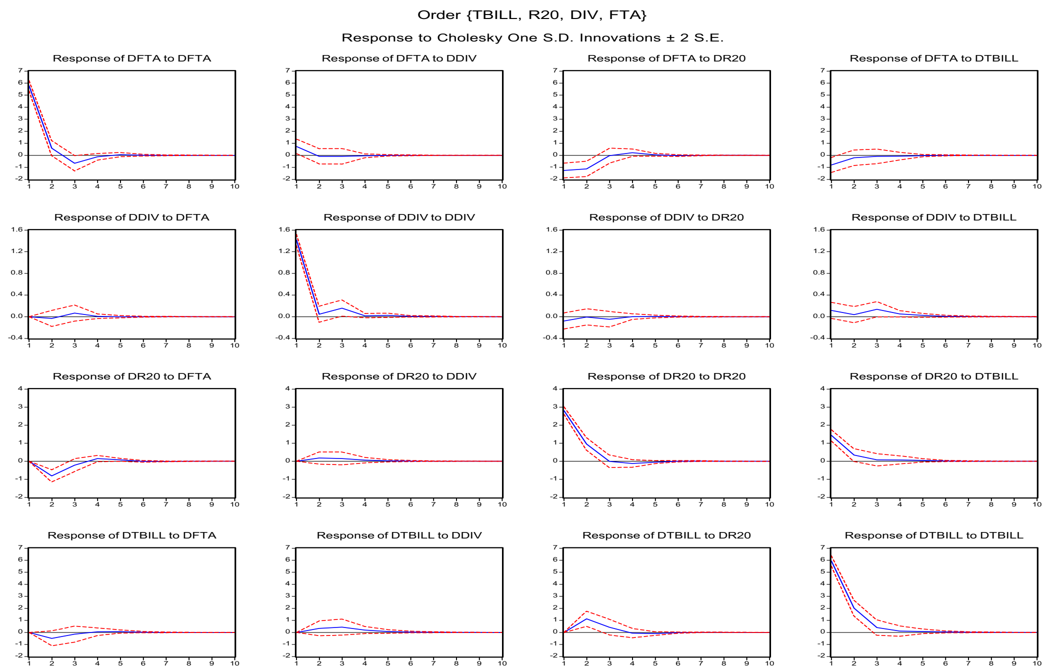
and an ordering based upon the relative timing of the trading hours of the markets

[(II: TBILL, R20, DIV, FTA)].

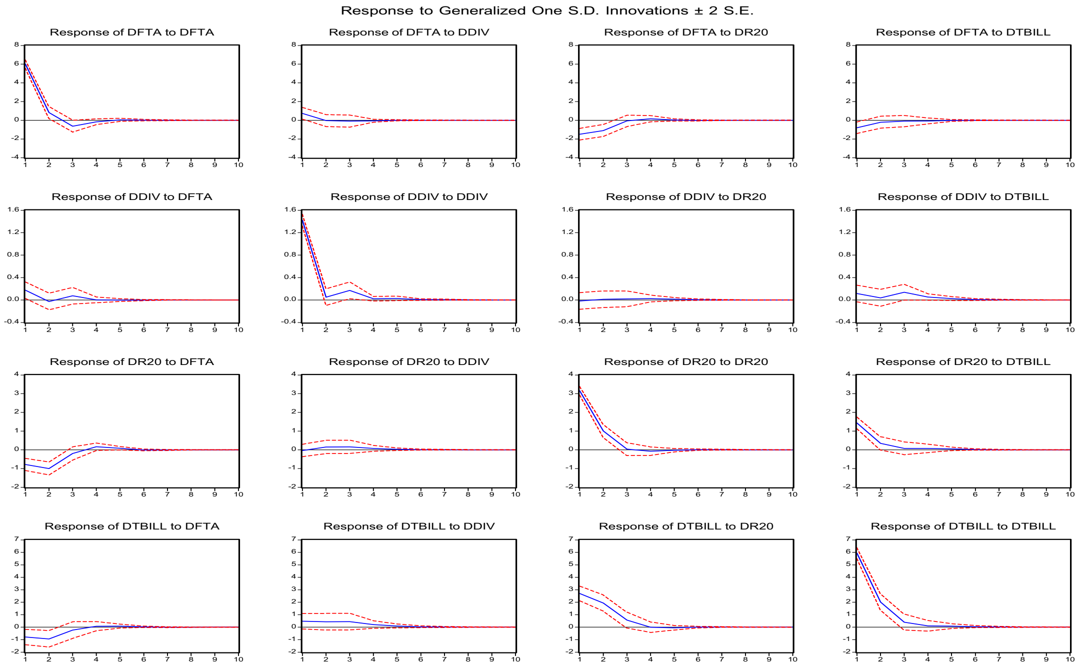
In EViews the order is simply defined in the Cholesky ordering option. Below are results in graphs with I: FTA, DIV, R20, TBILL; II: R20, TBILL DIV, FTA, and III: General impulse response function.

Impulse responses:





Impulse responses continue:



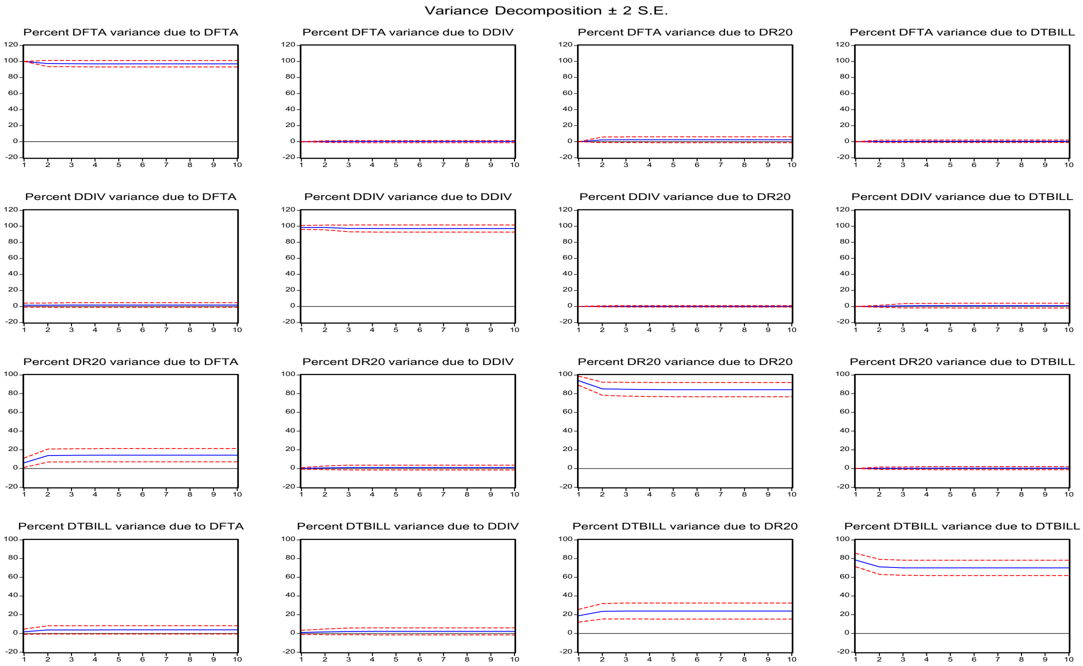
The general impulse response function are defined as[‡‡]

$$GI(j, \delta_i, \mathcal{F}_{t-1}) = \mathrm{E}[\mathbf{y}_{t+j}|\epsilon_{it} = \delta_i, \mathcal{F}_{t-1}] - \mathrm{E}[\mathbf{y}_{t+j}|\mathcal{F}_{t-1}].$$

That is difference of conditional expectation given an one time shock occurs in series j. These coincide with the orthogonalized impulse responses if the residual covariance matrix Σ is diagonal.

[‡‡]Pesaran, M. Hashem and Yongcheol Shin (1998). Impulse Response Analysis in Linear Multivariate Models, *Economics Letters*, 58, 17-29.

Variance decomposition graphs of the equity-bond data



On estimation of the impulse response coefficients

Consider the VAR(p) model

$$\mathbf{y} = \Phi_1 \mathbf{y}_{t-1} + \cdots + \Phi_p \mathbf{y}_{t-p} + \epsilon$$

or

$$\Phi(L)\mathbf{y} = \epsilon.$$

Then under stationarity the vector MA representation is

$$\mathbf{y} = \epsilon + \Psi_1 \epsilon_{t-1} + \Psi_2 \epsilon_{t-2} + \cdots$$

When we have estimates of the AR-matrices Φ_i denoted by $\hat{\Phi}_i$, $i = 1, \ldots, p$ the next problem is to construct estimates for MA matrices Ψ_j. It can be shown that

$$\Psi_j = \sum_{i=1}^{j} \Psi_{j-i} \Phi_i$$

with $\Psi_0 = \mathbf{I}$, and $\Phi_j = \mathbf{0}$ when $i > p$. Consequently, the estimates can be obtained by replacing Φ_i's by the corresponding estimates.

Next we have to obtain the orthogonalized impulse response coefficients. This can be done easily, for letting $\mathbf{S}$ be the Cholesky decomposition of Σ_ϵ such that

$$\Sigma_\epsilon = \mathbf{S}\mathbf{S}',$$

we can write

$$\begin{aligned} \mathbf{y}_t &= \sum_{i=0}^{\infty} \Psi_i \epsilon_{t-i} \\ &= \sum_{i=0}^{\infty} \Psi_i \mathbf{S}\mathbf{S}^{-1} \epsilon_{t-i} \\ &= \sum_{i=0}^{\infty} \Psi_i^* \nu_{t-i}, \end{aligned}$$

where

$$\Psi_i^* = \Psi_i \mathbf{S}$$

and $\nu_t = \mathbf{S}^{-1}\epsilon_t$. Then

$$\text{Cov}(\nu_t) = \mathbf{S}^{-1}\Sigma_\epsilon \mathbf{S}'^{-1} = \mathbf{I}.$$

The estimates for Ψ_i^* are obtained by replacing Ψ_t with its estimates and using Cholesky decomposition of $\hat{\Sigma}_\epsilon$.

2.6 *Cointegration*

(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

$$\mathbf{y}_t = \mu + \Phi_1\mathbf{y}_{t-1} + \Phi_2\mathbf{y}_{t-2} + \epsilon_t,$$

where $\epsilon \sim \text{NID}(\mathbf{0}, \Sigma)$ Which we can rewrite into the VECM (Vector Error Correction Model) form

$$\Delta\mathbf{y}_t = \mu + \Pi\mathbf{y}_{t-1} + \Gamma\Delta\mathbf{y}_{t-1} + \epsilon_t,$$

where $\Pi = -(\mathbf{I} - \Phi_1 - \Phi_2)$, and $\Gamma = -\Phi_2$. Note that the left hand side of the equation 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, $\Delta\mathbf{y}_t = \mu + \Phi\Delta\mathbf{y}_{t-1} + \mathbf{v}_t$ leads to a misspesification!

*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 $ax_t + by_t \sim I(1)$.

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

Definition of cointegration: 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.* I.e., $(x_t, y_t)' \sim CI(1, 1)$.

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!

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

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

yields nonsense results!

*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 $\mathbf{y}_t \sim CI(d, b)$.

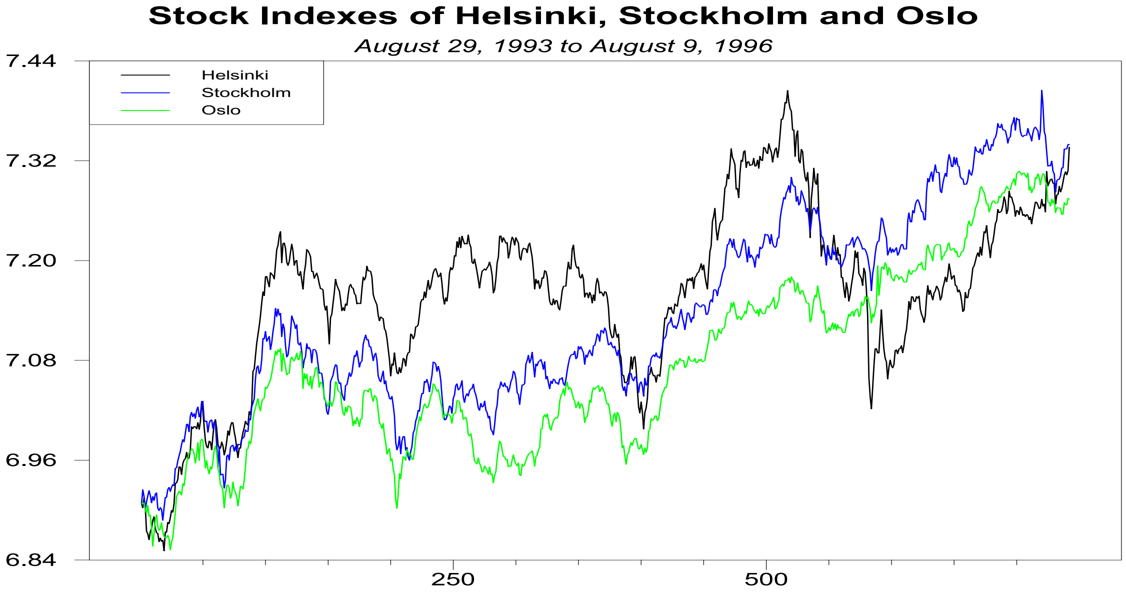
In the general case we can decompose $\Pi = \alpha\beta'$, where α and β are $m \times r$ matrices $r \leq m$ is the rank of Π. Matrix β contains the long run relationships between variables in $\mathbf{y}_t$ and α contains the short run adjusting parameters towards the long run steady state relationship $\beta'\mathbf{y}_t$.

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 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 α and β are obtained.

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



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

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.

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.

Note that several other tests also exist 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.

Example. 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. 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.

Note.

- In testing the unit root of the individual series, acceptance of the null hypothesis led the conclusion that the series are integrated.

- In testing for cointegration with ADF or PP tests, rejection of the hypothesis leads to the conclusion that the residuals are not integrated, which implies that the x_t and y_t series are cointegrated. Johansen's procedure test the number of cointegration vectors.

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

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

where

$$z_t = y_t - Ax_t \sim I(0),$$

$$\delta(L) = 1 + \delta_1 L + \cdots + \delta_p L^p.$$

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

The relation

$$y_t = \beta_0 + \beta_1 x_t$$

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 $x_t, y_t \sim \text{CI}(1,1)$ 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 $I(0)$.

(*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}, $f_{t,h}(J_t)$

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

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. 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
```

Trend and Intercept

Consider the following ECM

$\Delta \mathbf{x}_t = \alpha\beta'\mathbf{x}_{t-1} + \Gamma_1 \Delta \mathbf{x}_{t-1} + \cdots + \Gamma_p \Delta \mathbf{x}_{t-p} + \mu + \delta t + \epsilon_t.$

Then we can write

$$\begin{aligned} \delta &= \alpha\delta_1 + \alpha_\perp \delta_2 \\ \mu &= \alpha\mu_1 + \alpha_\perp \mu_2, \end{aligned}$$

where $\alpha_\perp$ is a $m \times (m-r)$ matrix such that $\alpha'\alpha_\perp = \mathbf{0}$ (orthogonal complement of the column space spanned by the columns of α),
$\delta_1 = (\alpha'\alpha)^{-1}\alpha'\delta$ is a r dimensional vector of *linear trend coefficients in the cointegrations relations*,
$\delta_2 = (\alpha'_\perp \alpha_\perp)^{-1}\alpha'_\perp \delta$ is an $(m-r)$-dimensional vector of *quadratic trend coefficients in the data*,
$\mu_1 = (\alpha'\alpha)^{-1}\alpha'\mu$ is a r-dimensional vector of *intercepts in the cointegration relations*
$\mu_2 = (\alpha'_\perp \alpha_\perp)^{-1}\alpha'_\perp \mu$ is an $(m-r)$-vector of *linear trend coefficients in the data.*

Using this decomposition, we can write

$$\Delta \mathbf{x}_t = \alpha \begin{pmatrix} \beta \\ \mu_1' \\ \delta_1' \end{pmatrix}' \tilde{\mathbf{x}}_{t-1} + \Gamma_1 \Delta \mathbf{x}_{t-1} + \cdots + \Gamma_p \Delta \mathbf{x}_{t-p} + \alpha_\perp \mu_2 + \alpha_\perp \delta_2 t + \epsilon_t.$$

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 δ and μ*: Then the model is consistent with linear trend in differences, $\Delta \mathbf{x}_t$, and quadratic tend in $\mathbf{x}_t$.

Case 2. $\delta_2 = 0$, δ_1, μ_1 and μ_2 unrestricted: When $\delta_2 = 0$ the model is restricted to exclude quadratic trends in data. $\delta_1 \neq 0$ implies that the cointegration relations may have linear trends, and $\mu_2 \neq 0$ implies that the cointegration relation has an intercept as well. $\mu_1 \neq 0$ implies that there is a linear trend in also in $\mathbf{x}_t$.

Case 3. $\delta = 0$, μ_1 and μ_2 unrestricted: This implies that there may be linear trend in $\mathbf{x}_t$ ($\mu_2 \neq 0$), but no trend in the cointegration relations. There may also be intercepts ($\mu_1 \neq 0$) in the cointegration relations.

Case 4. $\delta = 0, \mu_2 = 0$ and μ_1 unrestricted: No trend in data ($\delta_2 = 0$, $\mu_2 = 0$), and no trend in cointegration relation ($\delta_1 = 0$), but there may be intercepts in the cointegration relations ($\mu_1 \neq 0$).

Case 5. $\delta = 0$ and $\mu = 0$: 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.

Note: Probably the most common cases in practice are the cases 3 and 4.

Testing hypotheses

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

Hypotheses about β can be presented in the form different forms. In EViews the hypotheses are spcified by imposing restriction formulas between the parameters.

Example. Let $m = 4$, $r = 2$ and

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

then we can write set in EViews $B(1,2) = 0$, $B(2,1) = -B(1,1)$, $B(3,1) = 0$, $B(4,2) = 0$.

Special cases:

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

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

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

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. Consider the (log) indexes of Helsinki, Stockholm, Oslo and Copenhagen. Using CATS the Johansen's test give

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

Unrestricted Cointegration Rank Test
=================================================================
Hypothesized                Trace      5 Percent       1 Percent
No. of CE(s) Eigenvalue Statistic  Critical Value  Critical Value
-----------------------------------------------------------------
None **         0.0498      56.54         47.21            54.46
At most 1       0.0137      18.92         29.68            35.65
At most 2       0.0097       8.78         15.41            20.04
At most 3       0.0022       1.62          3.76             6.65
=================================================================
 *(**) denotes rejection of the hypothesis at the 5%(1%) level
 Trace test indicates 1 cointegrating equation(s)
 at both 5% and 1% levels
```

There is possibly one significant cointegration relation, as already found earlier.

Estimating β under this restriction yields.

```
Cointegration equation
 LFIN   LSWE     LNOR     LSWE    C
 1.00  109.02  -126.73   21.20  -35.78
```

Test next whether Helsinki is needed in the cointegration relation. In EViews this can be accomplished by setting $B(1,1) = 0$ in the estimation.

Imposing this restriction yields

```
Cointegration Restrictions:
      B(1,1)=0
Convergence achieved after 3 iterations.
Not all cointegrating vectors are identified
LR test for binding restrictions (rank = 1):
Chi-square(1)   0.027825
Probability     0.867520
========================================
Cointegrating Eq:
LFIN  LSWE     LNOR    LDEN        C
----------------------------------------
0.00  46.77  -54.01  8.872722  -13.22640
========================================
```

Hence, the data support the hypothesis that Helsinki is not needed in the cointegration relation.

Restriction on α

Restrictions on α (in EViews) can be specified in the same manner as in the β matrix.

Weak exogeneity

In $\Pi = \alpha\beta'$ the matrix α represents the speed of adjustment to the disequilibrium. If all α_{ij} in row i of α 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. If x_t above is weakly exogenous then $\alpha_{31} = \alpha_{32} = 0$. This also implies that we can model the system as

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

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

Example. Test just for illustration purposes, whether Copenhagen is weakly exogenous (under the restriction on β). The EViews results are

```
 Vector Error Correction Estimates
 Sample(adjusted): 6 742
 Included observations: 737 after adjusting endpoints
 Standard errors in ( ) & t-statistics in [ ]
=====================================================
Cointegration Restrictions:
      B(1,1)=0,A(4,1)=0
Convergence achieved after 5 iterations.
Not all cointegrating vectors are identified
LR test for binding restrictions (rank = 1):
Chi-square(2)   0.747561
Probability     0.688128
===========================================
Cointegrating Eq:
LFIN    LSWE       LNOR       LDEN       C
-------------------------------------------
0.00   46.56     -53.33      7.88     -9.69
===========================================

=========================================================
Error Correction: D(LFIN)    D(LSWE)    D(LNOR)   D(LDEN)
---------------------------------------------------------
CointEq1       -0.001286  -0.000699   0.000834   0.000000
                (0.00041)  (0.00031)   (0.00027) (0.00000)
               [-3.14380] [-2.23172]  [ 3.07083] [   NA   ]
=========================================================
```

The hypothesis can be accepted.

Conditioning the model with weakly exogenous variables sometimes improve stochastic properties of the model. This can be done by first saving the ci-relation and continuing the analysis with traditional system equation econometrics.

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. Following this the residual analysis of CATS yields

```
RESIDUAL ANALYSIS

Correlation matrix
     DLFIN     DLSWE     DLNOR     DLDEN
  1.000000
  0.478891  1.000000
  0.424876  0.550324  1.000000
  0.338403  0.442204  0.444650  1.000000

Standard deviations of residuals
  0.011727  0.009371  0.008113  0.007387

MULTIVARIATE STATISTICS

LOG(DET(SIGMA))          =     -38.65190
INFORMATION CRITERIA: SC =     -38.04059
                      HQ =     -38.30229
TRACE CORRELATION        =       0.04519

TEST FOR AUTOCORRELATION
L-B(183), CHISQ(2864)= 2770.503, p-val = 0.89
LM(1),   CHISQ(16)   =  25.555, p-val = 0.06
LM(4),   CHISQ(16)   =  22.791, p-val = 0.12
```

```
TEST FOR NORMALITY
CHISQ(8)             = 634.572, p-val = 0.00

UNIVARIATE STATISTICS

     MEAN      STD.DEV     SKEWNESS     KURTOSIS      MAXIMUM      MINIMUM
 0.000000     0.011727    -0.329214     6.375007     0.054239    -0.073806
 0.000000     0.009371     0.153852     4.673587     0.048026    -0.036842
-0.000000     0.008113    -0.049348     5.237731     0.043042    -0.035961
 0.000000     0.007387    -0.097416     3.905279     0.023646    -0.030046

     ARCH(4)   Normality   R-squared
      37.094     161.390       0.062
      22.496      59.313       0.034
      50.391      96.711       0.041
      17.333      22.013       0.050
```

As is seen normality of residuals cannot be accepted. There is obvious ARCH-effect in the residuals. All have kurtosis larger than 3 (the kurtosis of a normal distribution). Hence, the analysis results may not be fully reliable. Nevertheless, we do not try to improve the stochastic properties of the model here for now.

Estimates of the short run dynamics

The main focus in cointegration analysis is in the matrices α and β, 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:

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

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

- Condition on any (weakly) exognous variables thus obtaining a conditional VECM.

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