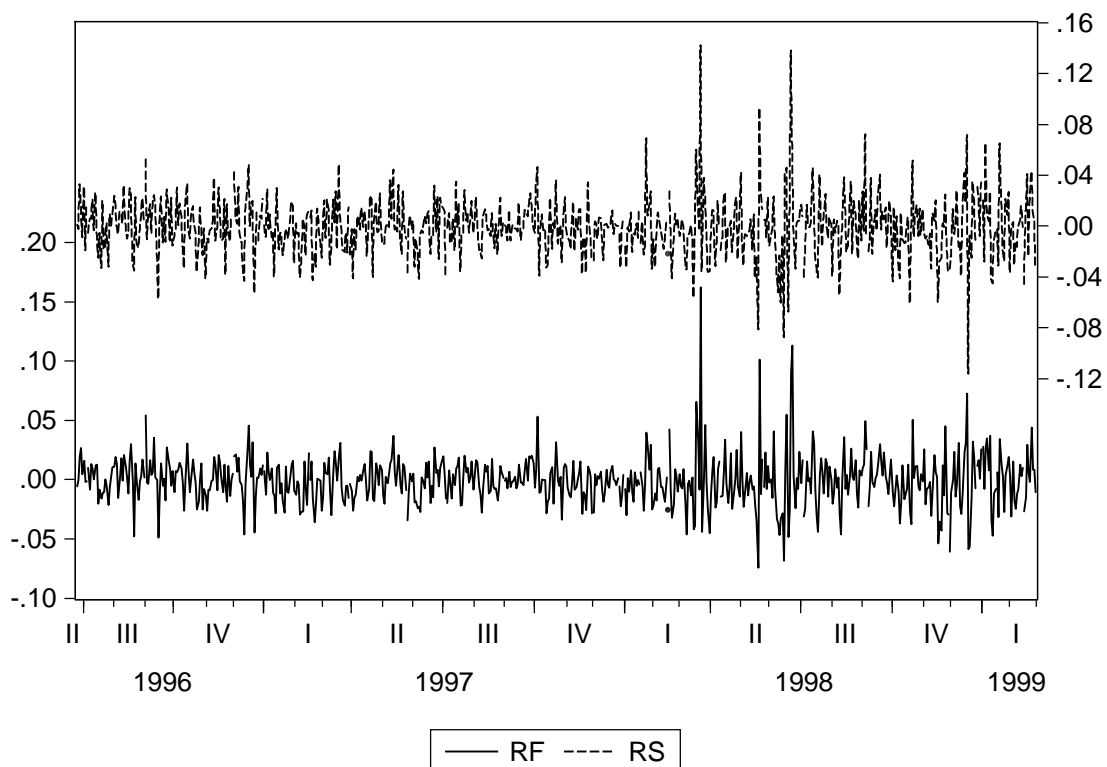


## 4. Multivariate Time Series Models

Consider the crude oil spot and near futures prices from 24 June 1996 to 26 February 1999 below.

Crude Oil Spot and Futures Returns



If we wish to forecast a stationary series not only based upon its own past realizations, but additionally taking realizations of other stationary series into account, then we can model the series as a vector autoregressive process (VAR, for short), provided the corresponding price series are not cointegrated. We shall define cointegration later.

## 4.1 Vector Autoregressions

Vector autoregressions extend simple autoregressive processes by adding the history of other series to the series' own history. For example, a vector autoregression of order 1 (VAR(1)) on a bivariate system is:

$$(1) \quad y_{1t} = \phi_{01} + \phi_{11}y_{1,t-1} + \phi_{12}y_{2,t-1} + \epsilon_{1t},$$

$$(2) \quad y_{2t} = \phi_{02} + \phi_{21}y_{1,t-1} + \phi_{22}y_{2,t-1} + \epsilon_{2t}.$$

The error terms  $\epsilon_{it}$  are assumed to be white noise processes, which may be contemporaneously correlated, but are uncorrelated with any past or future disturbances.

This may be compiled in matrix form as

(3)

$$\mathbf{y}_{t(2 \times 1)} = \Phi_{0(2 \times 1)} + \Phi_{1(2 \times 2)}\mathbf{y}_{t-1(2 \times 1)} + \epsilon_{t(2 \times 1)}$$

and is easily extended to contain more than 2 time series and more than one lag as shown on the following slide.

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

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{mt} \end{pmatrix} = \begin{pmatrix} \phi_1^{(0)} \\ \phi_2^{(0)} \\ \vdots \\ \phi_m^{(0)} \end{pmatrix} + \begin{pmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \cdots & \phi_{1m}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \cdots & \phi_{2m}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m1}^{(1)} & \phi_{m2}^{(1)} & \cdots & \phi_{mm}^{(1)} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \vdots \\ y_{m,t-1} \end{pmatrix} + \dots + \begin{pmatrix} \phi_{11}^{(p)} & \phi_{12}^{(p)} & \cdots & \phi_{1m}^{(p)} \\ \phi_{21}^{(p)} & \phi_{22}^{(p)} & \cdots & \phi_{2m}^{(p)} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m1}^{(p)} & \phi_{m2}^{(p)} & \cdots & \phi_{mm}^{(p)} \end{pmatrix} \begin{pmatrix} y_{1,t-p} \\ y_{2,t-p} \\ \vdots \\ y_{m,t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{mt} \end{pmatrix}.$$

In matrix notations

$$(4) \quad \mathbf{y}_t = \Phi_0 + \Phi_1 \mathbf{y}_{t-1} + \cdots + \Phi_p \mathbf{y}_{t-p} + \epsilon_t,$$

where  $\mathbf{y}_t$ ,  $\Phi_0$ , and  $\epsilon_t$  are  $(m \times 1)$  vectors and  $\Phi_1, \dots, \Phi_p$  are  $(m \times m)$  coefficient matrices introducing cross-dependencies between the series. It is common practice to work with centralized series, such that  $\Phi_0 = 0$ .

The representation (4) can be further simplified by adopting the matrix form of a lag polynomial ( $\mathbf{I}$  denoting the identity matrix)

$$(5) \quad \Phi(L) = \mathbf{I} - \Phi_1 L - \dots - \Phi_p L^p.$$

Thus finally we get for centralized series in analogy to the univariate case

$$(6) \quad \Phi(L)\mathbf{y}_t = \epsilon_t,$$

which is now a matrix equation containing cross-dependencies between the series.

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

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

which allows for estimation by OLS, because each individual residual series is assumed to be serially uncorrelated with constant variance. Note that  $\Sigma_\epsilon$  is not required to be diagonal, that is, while shocks must be serially uncorrelated, simultaneous correlation of the shocks between different series is allowed.

## Example.

Fitting a VAR(1) model to the spot and future returns from the introductory section in EViews (Quick/Estimate VAR) yields

Vector Autoregression Estimates

Vector Autoregression Estimates		
Date: 01/31/13 Time: 13:35		
Sample: 6/24/1996 2/26/1999		
Included observations: 671		
Standard errors in ( ) & t-statistics in [ ]		
	RS	RF
RS(-1)	0.197536 (0.06586) [ 2.99917]	0.716341 (0.05159) [ 13.8849]
RF(-1)	-0.253450 (0.07409) [-3.42068]	-0.536303 (0.05804) [-9.24057]
C	-0.000769 (0.00091) [-0.84181]	-0.000584 (0.00072) [-0.81610]
R-squared	0.017418	0.234962
Adj. R-squared	0.014476	0.232671
Sum sq. resids	0.373856	0.229385
S.E. equation	0.023657	0.018531
F-statistic	5.920711	102.5796
Log likelihood	1561.677	1725.558
Akaike AIC	-4.645834	-5.134302
Schwarz SC	-4.625676	-5.114143
Mean dependent	-0.000725	-0.000708
S.D. dependent	0.023830	0.021155
Determinant resid covariance (dof adj.)		4.23E-08
Determinant resid covariance		4.19E-08
Log likelihood		3795.402
Akaike information criterion		-11.29479
Schwarz criterion		-11.25447

implying

(7)

$$\hat{\Phi}_0 = \begin{pmatrix} -0.000796 \\ -0.000584 \end{pmatrix}, \quad \hat{\Phi}_1 = \begin{pmatrix} 0.197536 & -0.253450 \\ 0.716341 & -0.536303 \end{pmatrix}.$$

The same coefficient estimates may be obtained by estimating the scalar equations (1) and (2) separately for both the spot and the futures return series with OLS.

The fact that all entries in the coefficient matrix  $\Phi_1$  are significant implies that past spot and future returns have an impact upon both current spot and future returns. We say that the series Granger cause each other. A more precise definition of Granger causality will be given later.

If the off-diagonal elements in  $\Phi_1$  had been insignificant it would have implied, that the return series are only influenced by their own history, but not the history of the other series, implying that there is no Granger causality.

We may also find that one off-diagonal element is significant while the other one is not. In that case there is Granger causality from the series corresponding to the column of the significant entry to the series corresponding to the row of the significant entry, but not the other way round.

Analogous to the univariate case, in order for the VAR model to be stationary the coefficient matrices must satisfy the following constraint. It is required that the roots of the determinant equation

$$(8) \quad |\mathbf{I} - \Phi_1 z - \Phi_2 z^2 - \dots - \Phi_p z^p| = 0$$

lie outside the unit circle.

Example: (continued.)

$$\begin{aligned} |\mathbf{I} - \Phi_1 z| &= \begin{vmatrix} 1 - 0.197536z & 0.253450z \\ -0.716341z & 1 + 0.536303z \end{vmatrix} \\ &= 1 + 0.338767z + 0.0756175z^2 = 0 \end{aligned}$$

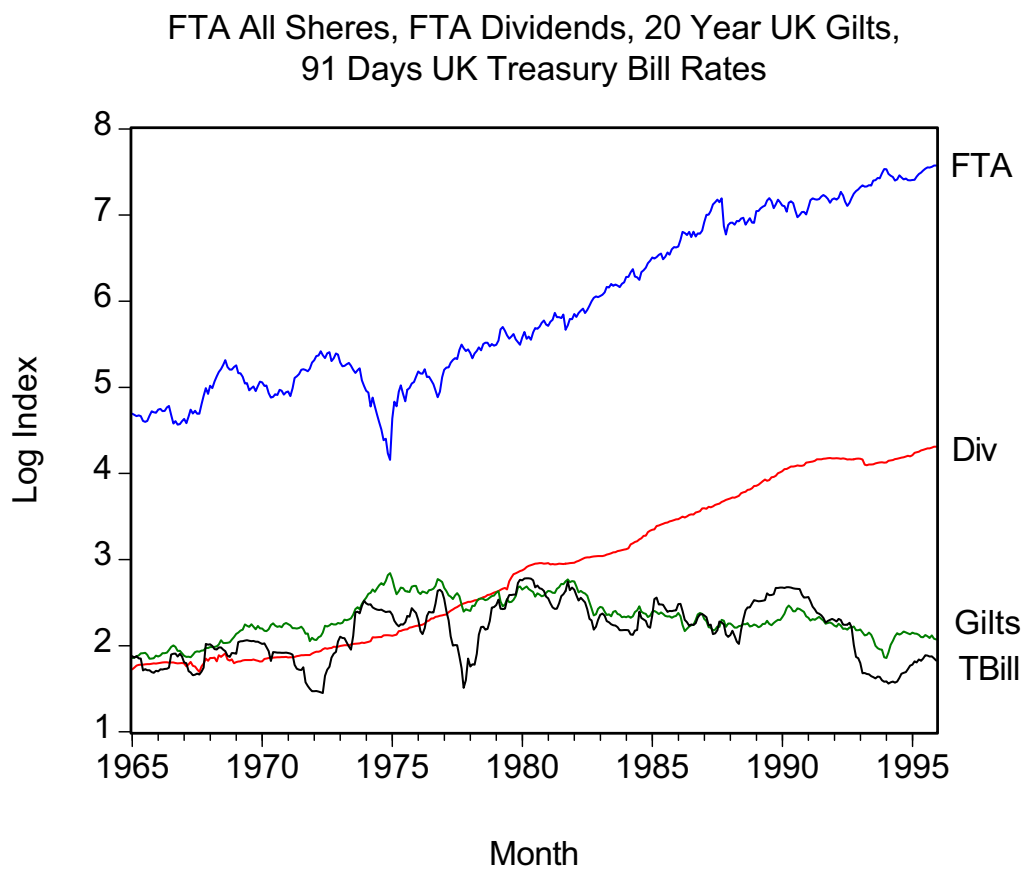
is solved by  $z = -2.24 \pm 2.86i$

with modulus  $|z| = 3.64 > 1$ .

Hence, if the VAR(1) model is correctly specified, then possible differences in sample correlations over different subperiods should be attributed to sampling error alone, because the model is covariance stationary.

## 4.2 Defining the order of a VAR-model

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)





The correlograms for the above (log) series look as follows. All the series prove to be  $I(1)$ .

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

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
*****	*****	1	0.992	0.992	368.98	0.000
*****	. .	2	0.983	-0.041	732.51	0.000
*****	. .	3	0.975	0.021	1090.9	0.000
*****	. .	4	0.966	-0.025	1444.0	0.000
*****	. .	5	0.957	-0.024	1791.5	0.000
*****	. .	6	0.949	0.008	2133.6	0.000
*****	. .	7	0.940	0.005	2470.5	0.000
*****	. .	8	0.931	-0.007	2802.1	0.000
*****	. .	9	0.923	0.023	3128.8	0.000
*****	. .	10	0.915	-0.011	3450.6	0.000

Dividends

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
*****	*****	1	0.994	0.994	370.59	0.000
*****	. .	2	0.988	-0.003	737.78	0.000
*****	. .	3	0.982	0.002	1101.6	0.000
*****	. .	4	0.976	-0.004	1462.1	0.000
*****	. .	5	0.971	-0.008	1819.2	0.000
*****	. .	6	0.965	-0.006	2172.9	0.000
*****	. .	7	0.959	-0.007	2523.1	0.000
*****	. .	8	0.953	-0.004	2869.9	0.000
*****	. .	9	0.947	-0.006	3213.3	0.000
*****	. .	10	0.940	-0.006	3553.2	0.000

T-Bill

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
*****	*****	1	0.980	0.980	360.26	0.000
*****	** .	2	0.949	-0.301	698.79	0.000
*****	. .	3	0.916	0.020	1014.9	0.000
*****	. .	4	0.883	-0.005	1309.5	0.000
*****	. .	5	0.849	-0.041	1583.0	0.000
*****	*. .	6	0.811	-0.141	1833.1	0.000
*****	. .	7	0.770	-0.018	2059.2	0.000
*****	. .	8	0.730	0.019	2263.1	0.000
*****	. .	9	0.694	0.058	2447.6	0.000
*****	. .	10	0.660	-0.013	2615.0	0.000

Gilts

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
*****	*****	1	0.984	0.984	362.91	0.000
*****	*. .	2	0.962	-0.182	710.80	0.000
*****	. .	3	0.941	0.050	1044.6	0.000
*****	. .	4	0.921	0.015	1365.5	0.000
*****	. .	5	0.903	0.031	1674.8	0.000
*****	. .	6	0.885	-0.038	1972.4	0.000
*****	. .	7	0.866	-0.001	2258.4	0.000
*****	. .	8	0.848	0.019	2533.6	0.000
*****	. .	9	0.832	0.005	2798.6	0.000
*****	. .	10	0.815	-0.014	3053.6	0.000

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	$\hat{\rho}$	$t$ -Stat	$p$ -value
FTA	-0.029	-2.474	0.341
DIV	-0.059	-2.555	0.302
R20	-0.013	-1.779	0.713
T-BILL	-0.024	-2.399	0.380
$\Delta$ FTA	-1.001	-9.013	0.000
$\Delta$ DIV	-2.600	-12.18	0.000
$\Delta$ R20	-0.786	-8.147	0.000
$\Delta$ T-BILL	-0.620	-7.028	0.000
ADF critical values			
Level	No trend	Trend	
1%	-3.448	-3.983	
5%	-2.869	-3.422	
10%	-2.571	-3.134	

# VAR(2) estimation output on percentual logreturns:

## Vector Autoregression Estimates

Vector Autoregression Estimates				
Date: 02/07/13 Time: 13:14				
Sample (adjusted): 1965M04 1995M12				
Included observations: 369 after adjustments				
Standard errors in ( ) & t-statistics in [ ]				
	DFTA	DDIV	DR20	DTBILL
DFTA(-1)	0.020769 (0.13045) [ 0.15921]	-0.676248 (0.12977) [-5.21126]	-0.057670 (0.06832) [-0.84416]	-0.040784 (0.12673) [-0.32181]
DFTA(-2)	0.024780 (0.13211) [ 0.18757]	-0.128730 (0.13141) [-0.97958]	-0.026210 (0.06918) [-0.37885]	0.038316 (0.12834) [ 0.29855]
DDIV(-1)	0.054781 (0.12618) [ 0.43415]	-0.140806 (0.12552) [-1.12178]	-0.039136 (0.06608) [-0.59225]	-0.001885 (0.12259) [-0.01538]
DDIV(-2)	-0.129923 (0.05558) [-2.33744]	-0.168589 (0.05529) [-3.04905]	-0.001959 (0.02911) [-0.06732]	0.013840 (0.05400) [ 0.25629]
DR20(-1)	-0.425101 (0.11868) [-3.58202]	-0.469488 (0.11806) [-3.97685]	0.291443 (0.06215) [ 4.68935]	0.378387 (0.11530) [ 3.28190]
DR20(-2)	0.204949 (0.12205) [ 1.67927]	0.163345 (0.12141) [ 1.34543]	-0.138976 (0.06391) [-2.17439]	-0.076893 (0.11857) [-0.64850]
DTBILL(-1)	0.065007 (0.06114) [ 1.06321]	0.106466 (0.06082) [ 1.75046]	-0.024714 (0.03202) [-0.77185]	0.244177 (0.05940) [ 4.11071]
DTBILL(-2)	-0.057654 (0.06029) [-0.95627]	-0.037304 (0.05997) [-0.62200]	0.037882 (0.03157) [ 1.19981]	-0.012430 (0.05857) [-0.21222]
C	0.797370 (0.33567) [ 2.37547]	1.522351 (0.33391) [ 4.55917]	0.137442 (0.17579) [ 0.78187]	-0.036121 (0.32610) [-0.11077]
R-squared	0.072595	0.432941	0.123435	0.143822
Adj. R-squared	0.051986	0.420340	0.103956	0.124796
Sum sq. resids	13604.34	13462.22	3731.042	12840.15
S.E. equation	6.147343	6.115151	3.219318	5.972192
F-statistic	3.522471	34.35688	6.336766	7.559173
Log likelihood	-1189.144	-1187.206	-950.4561	-1178.478
Akaike AIC	6.494005	6.483504	5.200304	6.436193
Schwarz SC	6.589390	6.578889	5.295689	6.531578
Mean dependent	0.770651	0.681399	0.052983	-0.013968
S.D. dependent	6.313643	8.031942	3.400942	6.383798
Determinant resid covariance (dof adj.)	62164.98			
Determinant resid covariance	56318.40			
Log likelihood	-4112.558			
Akaike information criterion	22.48541			
Schwarz criterion	22.86695			

By default, EViews always suggests a VAR model of order  $p=2$ . In order to determine the most appropriate order of the model, multivariate extensions of criterion functions like SC and AIC can be utilized in the same manner as in the univariate case. These are given in the lower part of the estimation output and calculated as

$$(9) \quad AIC = -2l/T + 2n/T,$$

$$(10) \quad SC = -2l/T + n \log T/T,$$

where

$$(11) \quad l = -\frac{T}{2} \left[ m(1 + \log 2\pi) + \log |\hat{\Sigma}_\epsilon| \right]$$

is the log likelihood value of a multivariate normal distribution,  $T$  is the number of time points,  $m$  is the number of equations,  $|\hat{\Sigma}_\epsilon|$  is the determinant of the estimated variance matrix of residuals, and  $n = m(1 + pm)$  is the total number of parameters to be estimated.

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

$$(12) \quad LR = T(\log |\hat{\Sigma}_p| - \log |\hat{\Sigma}_q|)$$

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

$$LR \sim \chi_{df}^2,$$

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

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

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

$$(13) \quad LR^* = (T - mq)(\log |\hat{\Sigma}_p| - \log |\hat{\Sigma}_q|)$$

is used to correct for small sample bias.

## Example: (continued.)

Both the LR-test and the information criteria are obtained from EViews under 'View/Lag Structure/Lag Length Criteria' after estimating a VAR of arbitrary order. For our data:

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-4415.547	NA	441799.9	24.35012	24.39303	24.36718
1	-4066.698	688.0881	70596.77	22.51624	22.73081*	22.60153*
2	-4050.003	32.56117*	70329.52*	22.51241*	22.89864	22.66594
3	-4040.595	18.14318	72936.57	22.54873	23.10661	22.77048
4	-4032.265	15.87856	76096.65	22.59099	23.32052	22.88098
5	-4023.542	16.43737	79228.74	22.63109	23.53227	22.98930
6	-4019.160	8.161322	84495.96	22.69509	23.76793	23.12154
7	-4006.563	23.18114	86137.49	22.71384	23.95834	23.20852
8	-3996.916	17.54005	89263.79	22.74885	24.16499	23.31176

\* 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

The Schwarz criterion selects VAR(1), whereas AIC and the LR-test suggest VAR(2).

These are only preliminary suggestions for the order of the model. For the chosen model it is crucial that the residuals fulfill the assumption of multivariate white noise.

### 4.3 Model Diagnostics

To investigate whether the VAR residuals are white noise, the hypothesis to be tested is

$$(14) \quad H_0 : \Upsilon_1 = \cdots = \Upsilon_h = 0,$$

where  $\Upsilon_k = (\gamma_{ij}(k))$  denotes the matrix of the  $k$ 'th cross autocovariances of the residuals series  $\epsilon_i$  and  $\epsilon_j$ :

$$(15) \quad \gamma_{ij}(k) = E(\epsilon_{i,t-k} \cdot \epsilon_{j,t}),$$

whose diagonal elements reduce to the usual autocovariances  $\gamma_k$ . Note, however, that cross-autocovariances, unlike univariate autocovariances, are not symmetric in  $k$ , that is  $\gamma_{i,j}(k) \neq \gamma_{i,j}(-k)$ , because the covariance between residual series  $i$  and residual series  $j$   $k$  steps ahead is in general not the same as the covariance between residual series  $i$  and the residual series  $j$   $k$  steps before. Stationarity ensures, however, that  $\Upsilon_k = \Upsilon'_{-k}$  (exercise).

In order to test  $H_0 : \Upsilon_1 = \dots = \Upsilon_h = 0$ , we may use the (Portmanteau)  $Q$ -statistics<sup>††</sup>

$$(16) \quad Q_h = T \sum_{k=1}^h \text{tr}(\hat{\Upsilon}'_k \hat{\Upsilon}_0^{-1} \hat{\Upsilon}_k \hat{\Upsilon}_0^{-1})$$

where

$$\hat{\Upsilon}_k = (\hat{\gamma}_{ij}(k)) \text{ with } \hat{\gamma}_{ij}(k) = \frac{1}{T-k} \sum_{t=k}^T \hat{\epsilon}_{t-k,i} \hat{\epsilon}_{t,j}$$

are the estimated (residual)cross autocovariances and  $\hat{\Upsilon}_0$  the contemporaneous covariances of the residuals. Alternatively (especially in small samples) a modified statistic is used

$$(17) \quad Q_h^* = T^2 \sum_{k=1}^h (T-k)^{-1} \text{tr}(\hat{\Upsilon}'_k \hat{\Upsilon}_0^{-1} \hat{\Upsilon}_k \hat{\Upsilon}_0^{-1}).$$

The statistics are asymptotically  $\chi^2$  distributed with  $m^2(h-p)$  degrees of freedom. Note that in computer printouts  $h$  is running from  $1, 2, \dots, h^*$  with  $h^*$  specified by the user.

<sup>††</sup>See e.g. Lütkepohl, Helmut (1993). *Introduction to Multiple Time Series*, 2nd Ed., Ch. 4.4



The  $Q$ -statistics of the VAR(1) residuals available from EViews under 'View/Residual Tests/Portmanteau Autocorrelation Test' imply that the residuals don't pass the white noise test:\*

Lags	Q-Stat	Prob.	Adj Q-Stat	Prob.	df
1	13.88141	NA*	13.91903	NA*	NA*
2	34.35804	0.0049	34.50695	0.0046	16
3	43.98856	0.0771	44.21618	0.0738	32

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

The residuals of the VAR(2) model, however, may be regarded as multivariate white noise:

Lags	Q-Stat	Prob.	Adj Q-Stat	Prob.	df
1	0.561311	NA*	0.562836	NA*	NA*
2	11.82149	NA*	11.88438	NA*	NA*
3	18.76786	0.2809	18.88769	0.2745	16

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

We therefore adopt the VAR(2) model.

\*Some versions of EViews use the wrong degrees of freedom! Check that  $df = m^2(h - p)$ !

## 4.4 Vector ARMA (VARMA)

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

$$(18) \quad \mathbf{y}_t = \Phi_0 + \sum_{i=1}^p \Phi_i \mathbf{y}_{t-i} + \epsilon_t + \sum_{j=1}^q \Theta_j \epsilon_{t-j}$$

or

$$(19) \quad \Phi(L)\mathbf{y}_t = \Phi_0 + \Theta(L)\epsilon_t,$$

where  $\mathbf{y}_t$ ,  $\Phi_0$ , and  $\epsilon_t$  are  $m \times 1$  vectors, and  $\Phi_i$ 's and  $\Theta_j$ 's are  $m \times m$  matrices, and

$$(20) \quad \begin{aligned} \Phi(L) &= \mathbf{I} - \Phi_1 L - \dots - \Phi_p L^p \\ \Theta(L) &= \mathbf{I} + \Theta_1 L + \dots + \Theta_q L^q. \end{aligned}$$

Provided that  $\Theta(L)$  is invertible, we always can write the VARMA( $p, q$ )-model as a VAR( $\infty$ ) model with  $\Pi(L) = \Theta^{-1}(L)\Phi(L)$ . The presence of a vector MA component, however, implies that we can no longer find parameter estimates by ordinary least squares. We do not pursue our analysis to this direction.

## 4.5 Exogeneity and Causality

Suppose you got two time series  $x_t$  and  $y_t$  as in the introductory crude oil spot and futures returns example. We say  $x$  Granger causes  $y$  if

(21)

$$E(y_t|y_{t-1}, y_{t-2}, \dots) \neq E(y_t|y_{t-1}, y_{t-2}, \dots, x_{t-1}, x_{t-2}, \dots),$$

that is, if we can improve the forecast for  $y_t$  based upon its own history by additionally considering the history of  $x_t$ . In the other case

(22)

$$E(y_t|y_{t-1}, y_{t-2}, \dots) = E(y_t|y_{t-1}, y_{t-2}, \dots, x_{t-1}, x_{t-2}, \dots),$$

where adding history of  $x_t$  does not improve the forecast for  $y_t$ , we say that  $x$  does not Granger cause  $y$ , or  $x$  is exogeneous to  $y$ .

Note that Granger causality is not the same as causality in the philosophical sense. Granger causality does not claim that  $x$  is the reason for  $y$  in the sense like, for example,  $y$  moves because  $x$  moves. It just says that  $x$  is helpful in forecasting  $y$ , which might happen for other reasons than direct causality. There might be, for example, a third series  $z$  which has a fast causal impact upon  $x$  and a slower causal impact upon  $y$ . Then we can use the reaction of  $x$  in order to forecast the reaction in  $y$ , such that  $x$  Granger causes  $y$ .

## Testing for Exogeneity: The bivariate case

Consider a bivariate VAR( $p$ ) model written out in scalar form as

$$(23) \quad x_t = \phi_1 + \sum_{i=1}^p \phi_{11}^{(i)} x_{i,t-i} + \sum_{i=1}^p \phi_{12}^{(i)} y_{i,t-i} + \epsilon_{1t},$$

$$(24) \quad y_t = \phi_2 + \sum_{i=1}^p \phi_{21}^{(i)} x_{i,t-i} + \sum_{i=1}^p \phi_{22}^{(i)} y_{i,t-i} + \epsilon_{2t}.$$

Then the test for Granger causality from  $x$  to  $y$  is an  $F$ -test for the joint significance of  $\phi_{21}^{(1)}, \dots, \phi_{21}^{(p)}$  in the OLS regression (24).

Similarly, the test for Granger causality from  $y$  to  $x$  is an  $F$ -test for the joint significance of  $\phi_{12}^{(1)}, \dots, \phi_{12}^{(p)}$  in the OLS regression (23).

Recall from STAT1010 and Econometrics 1 (see also the section about the  $F$ -test for general linear restrictions in chapter 1) that the  $F$ -test for testing

(25)  $H_0: \beta_{k-q+1} = \beta_{k-q+2} = \dots = \beta_k = 0$   
against

(26)  $H_1: \text{some } \beta_{k-i} \neq 0, \quad i = 0, 1, \dots, q-1$   
in the model

(27)  $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$   
is

$$(28) \quad F = \frac{(\text{SSE}_r - \text{SSE}_{ur})/q}{\text{SSE}_{ur}/(n - k - 1)},$$

where  $\text{SSE}_r$  is the residual sum of squares from the restricted model under  $H_0$  and  $\text{SSE}_u$  is the residual sum of squares for the unrestricted model (27).

Under the null hypothesis the test statistic (28) is  $F$ -distributed with  $q = df_r - df_{ur}$  and  $n - k - 1$  degrees of freedom, where  $df_r$  is the degrees of freedom of  $\text{SSE}_r$  and  $df_{ur}$  is the degrees of freedom of  $\text{SSE}_{ur}$ .

In our case of considering Granger causality in a bivariate VAR( $p$ ) model we are considering to drop  $q = p$  variables in a model with  $n = T$  observations and  $k = 2p$  variables beyond the constant. Hence,

$$(29) \quad F = \frac{(SSE_r - SSE_{ur})/p}{SSE_{ur}/(T - 2p - 1)} \sim F(p, T - 2p - 1)$$

under  $H_0: y$  is exogeneous to  $x$  in (23), and under  $H_0: x$  is exogeneous to  $y$  in (24).

Example:

(Oil Spot and Futures Returns continued.)

Consider fitting a VAR(2) model to the crude oil spot and futures returns discussed earlier. The EViews output can be found on the next slide.

Vector Autoregression Estimates

Vector Autoregression Estimates		
Date: 01/31/13 Time: 12:55		
Sample: 6/24/1996 2/26/1999		
Included observations: 671		
Standard errors in ( ) & t-statistics in [ ]		
	RS	RF
RS(-1)	0.274978 (0.08238) [ 3.33796]	0.959797 (0.06259) [ 15.3356]
RS(-2)	0.203046 (0.08964) [ 2.26519]	0.479233 (0.06810) [ 7.03710]
RF(-1)	-0.381912 (0.10492) [-3.63993]	-0.912483 (0.07971) [-11.4470]
RF(-2)	-0.222814 (0.08333) [-2.67394]	-0.381616 (0.06331) [-6.02798]
C	-0.000831 (0.00091) [-0.91219]	-0.000627 (0.00069) [-0.90549]
R-squared	0.027856	0.287949
Adj. R-squared	0.022017	0.283673
Sum sq. resids	0.369885	0.213498
S.E. equation	0.023567	0.017904
F-statistic	4.770877	67.33160
Log likelihood	1565.260	1749.639
Akaike AIC	-4.650553	-5.200117
Schwarz SC	-4.616955	-5.166519
Mean dependent	-0.000725	-0.000708
S.D. dependent	0.023830	0.021155
Determinant resid covariance (dof adj.)		3.51E-08
Determinant resid covariance		3.46E-08
Log likelihood		3859.801
Akaike information criterion		-11.47482
Schwarz criterion		-11.40762

Denoting the spot series with  $x$  and the futures returns with  $y$ ,  $SSR_{ur}$  is 0.3699 in (23) and 0.2135 in (24). The same sum of squared residuals are obtained by running the corresponding univariate regressions.

The restricted sum of squared residuals dropping the futures returns in (23) and dropping the spot returns in (24) are 0.3780 and 0.2900, respectively. The  $F$ -statistics are

$$(30) \quad F = \frac{0.3780 - 0.3699}{0.3699} \cdot \frac{671 - 4 - 1}{2} = 7.29$$

for Granger causality from futures to spot returns and

$$(31) \quad F = \frac{0.2900 - 0.2135}{0.2135} \cdot \frac{671 - 4 - 1}{2} = 119.3$$

for Granger causality from spot to futures.

Past spot returns are thus decisively more helpful in explaining current futures returns than past futures returns are in explaining current spot returns, but there is Granger causality in both directions, as both  $F$ -statistics exceed the 0.1% critical value of 6.98, which can be obtained e.g. from Excel with the command `FINV(0.001;2;666)`. EViews does this test under 'View/ Coefficient Diagnostics/ Wald-Test - Coefficient Restrictions'.



Recall from Econometrics 1 that the  $F$ -tests above are strictly valid only in the case of strictly exogenous regressors and normally distributed error terms. In our case, however, we have included lagged dependent variables into the regression, such that the regressors are only contemporaneously exogenous and the  $F$ -tests are only asymptotically valid for large sample sizes.

Since the  $F$ -test are only asymptotically valid, we may just as well use the Wald, Lagrange Multiplier, or Likelihood Ratio tests which we got acquainted to in section 8 of chapter 1. These are in our case:

$$(32) \quad W = \frac{SSE_R - SSE_U}{SSE_U / (T - 2p - 1)},$$

$$(33) \quad LM = \frac{SSE_R - SSE_U}{SSE_R / (T - p - 1)},$$

$$(34) \quad LR = T(\log SSE_R - \log SSE_U),$$

all of which are asymptotically  $\chi^2$ -distributed with  $df = p$  under the null hypothesis that the other time series is exogenous.

Example: (Oil spot and futures continued.)  
Consider now as an illustration only Granger causality from spot to futures returns.

The Wald statistic (32) is

$$W = (671 - 5) \cdot \frac{0.2900 - 0.2135}{0.2135} = 238.63.$$

The Lagrange Multiplier statistic (33) is

$$LM = (671 - 3) \cdot \frac{0.2900 - 0.2135}{0.2135} = 239.35.$$

The Likelihood Ratio test statistic (34) is

$$LR = 671 \cdot (\log 0.2900 - \log 0.2135) = 205.51.$$

All of these exceed by far the 0.1% critical value for a  $\chi^2(2)$  distribution of 13.8 and are therefore highly significant, implying that spot returns Granger cause futures returns.

EViews displays the Wald test (32) under 'View/ Lag Structure/ Granger Causality/ Block Exogeneity Tests'.

## 4.6 Testing for Exogeneity: The general case

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

$$(35) \quad \mathbf{z}_t = \sum_{i=1}^p \Pi_i \mathbf{z}_{t-i} + \nu_t$$

where

$$(36) \quad \begin{aligned} E(\nu_t) &= \mathbf{0} \\ E(\nu_t \nu'_s) &= \begin{cases} \Sigma_\nu, & t = s \\ \mathbf{0}, & t \neq s. \end{cases} \end{aligned}$$

We wish to investigate Granger causality between the  $(m \times 1)$  vector  $\mathbf{y}_t$  and the  $(k \times 1)$  vector  $\mathbf{x}_t$ , that is, whether the time series contained in  $\mathbf{x}_t$  improve the forecasts of the time series contained in  $\mathbf{y}_t$  beyond using  $\mathbf{y}$ 's own history, and vice versa.

If  $\mathbf{x}$  does not Granger cause  $\mathbf{y}$ , then we say in this context that  $\mathbf{x}$  is block-exogeneous to  $\mathbf{y}$  in order to stress that the vectors contain more than a single time series.

For that purpose, partition the VAR of  $\mathbf{z}$  as  
(37)

$$\mathbf{y}_t = \sum_{i=1}^p \mathbf{C}_{2i} \mathbf{x}_{t-i} + \sum_{i=1}^p \mathbf{D}_{2i} \mathbf{y}_{t-i} + \nu_{1t}$$

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

where  $\nu'_t = (\nu'_{1t}, \nu'_{2t})$  and  $\Sigma_\nu$  are correspondingly partitioned as

$$(38) \quad \Sigma_\nu = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

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  $\eta_{1t}$  from the regression

$$(39) \quad \mathbf{y}_t = \sum_{i=1}^p \mathbf{C}_{1i} \mathbf{y}_{t-i} + \eta_{1t}.$$

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

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

$$H_0: \mathbf{C}_{2i} = \mathbf{0} \quad \text{against} \quad H_1: \mathbf{C}_{2i} \neq \mathbf{0}.$$

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

Unrestricted:

$$(40) \quad \hat{\Sigma}_{11} = \frac{1}{T-p} \sum_{t=p+1}^T \hat{v}_{1t} \hat{v}'_{1t}$$

Restricted:

$$(41) \quad \hat{\Sigma}_1 = \frac{1}{T-p} \sum_{t=p+1}^T \hat{\eta}_{1t} \hat{\eta}'_{1t}$$

The LR test is then

$$(42) \quad LR = (T-p) \left( \ln |\hat{\Sigma}_1| - \ln |\hat{\Sigma}_{11}| \right) \sim \chi_{m_k p}^2,$$

if  $H_0$  is true.

<sup>\*</sup>Perform OLS regressions of each of the elements in  $y$  on a constant,  $p$  lags of the elements of  $x$  and  $p$  lags of the elements of  $y$ .

<sup>†</sup>Perform OLS regressions of each of the elements in  $y$  on a constant and  $p$  lags of the elements of  $y$ .

## Example:

(UK stock and bond time series continued.)

Let us investigate whether the stock indices Granger cause the interest rate series.

Vector Autoregression Estimates

Vector Autoregression Estimates		
Date: 02/13/13 Time: 13:47		
Sample: 1965M04 1995M12		
Included observations: 369		
Standard errors in ( ) & t-statistics in [ ]		
	DR20	DTBILL
DR20(-1)	0.338822 (0.05885) [ 5.75756]	0.384765 (0.10772) [ 3.57201]
DR20(-2)	-0.121348 (0.05981) [-2.02889]	-0.090376 (0.10948) [-0.82552]
DTBILL(-1)	-0.022618 (0.03228) [-0.70069]	0.246859 (0.05909) [ 4.17797]
DTBILL(-2)	0.030179 (0.03164) [ 0.95368]	-0.016474 (0.05792) [-0.28441]
C	0.041421 (0.16938) [ 0.24455]	-0.029794 (0.31003) [-0.09610]
R-squared	0.095495	0.139897
Adj. R-squared	0.085556	0.130445
Sum sq. resids	3849.967	12899.01
S.E. equation	3.252204	5.952886
F-statistic	9.607556	14.80129
Log likelihood	-956.2451	-1179.322
Akaike AIC	5.210000	6.419087
Schwarz SC	5.262992	6.472079
Mean dependent	0.052983	-0.013968
S.D. dependent	3.400942	6.383798
Determinant resid covariance (dof adj.)	294.3394	
Determinant resid covariance	286.4167	
Log likelihood	-2090.976	
Akaike information criterion	11.38740	
Schwarz criterion	11.49339	

So  $|\hat{\Sigma}_1| = 286.4167$  in the restricted model.

Vector Autoregression Estimates

Vector Autoregression Estimates		
Date: 02/13/13 Time: 10:33		
Sample (adjusted): 1965M04 1995M12		
Included observations: 369 after adjustments		
Standard errors in ( ) & t-statistics in [ ]		
	DR20	DTBILL
DR20(-1)	0.291443 (0.06215) [ 4.68935]	0.378387 (0.11530) [ 3.28190]
DR20(-2)	-0.138976 (0.06391) [-2.17439]	-0.076893 (0.11857) [-0.64850]
DTBILL(-1)	-0.024714 (0.03202) [-0.77185]	0.244177 (0.05940) [ 4.11071]
DTBILL(-2)	0.037882 (0.03157) [ 1.19981]	-0.012430 (0.05857) [-0.21222]
C	0.137442 (0.17579) [ 0.78187]	-0.036121 (0.32610) [-0.11077]
DFTA(-1)	-0.057670 (0.06832) [-0.84416]	-0.040784 (0.12673) [-0.32181]
DFTA(-2)	-0.026210 (0.06918) [-0.37885]	0.038316 (0.12834) [ 0.29855]
DDIV(-1)	-0.039136 (0.06608) [-0.59225]	-0.001885 (0.12259) [-0.01538]
DDIV(-2)	-0.001959 (0.02911) [-0.06732]	0.013840 (0.05400) [ 0.25629]
R-squared	0.123435	0.143822
Adj. R-squared	0.103956	0.124796
Sum sq. resids	3731.042	12840.15
S.E. equation	3.219318	5.972192
F-statistic	6.336766	7.559173
Log likelihood	-950.4561	-1178.478
Akaike AIC	5.200304	6.436193
Schwarz SC	5.295689	6.531578
Mean dependent	0.052983	-0.013968
S.D. dependent	3.400942	6.383798
Determinant resid covariance (dof adj.)	290.0625	
Determinant resid covariance	276.0856	
Log likelihood	-2084.198	
Akaike information criterion	11.39403	
Schwarz criterion	11.58480	

From the previous EViews output we find that  $|\hat{\Sigma}_{11}| = 276.0856$  in the unrestricted VAR model of order  $p = 2$ .

The LR test statistics (42) is therefore

$$LR = (369 - 2)(\ln 286.4167 - \ln 276.0856) = 13.48.$$

The LR test statistics without small sample correction (dropping -2) is 13.556, which may be also obtained directly from the EViews Log likelihood output as

$$LR = -2(-2090.976 - (-2084.198)) = 13.556.$$

The 10% and 5% critical values for the  $\chi^2(8)$  distribution are 13.36 and 15.51, respectively.

Hence, while there is some indication that current stock returns might be helpful in explaining future bond returns, the effect is not yet statistically significant at  $\alpha = 5\%$ .



## *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 x to y* based on the matrices  $\Sigma_1$  and  $\Sigma_{11}$  as

$$(43) \quad F_{x \rightarrow y} = \ln(|\Sigma_1|/|\Sigma_{11}|),$$

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

$$(44) \quad F_{y \rightarrow x} = \ln(|\Sigma_2|/|\Sigma_{22}|).$$

\*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 that purpose, premultiplying the earlier VAR system of  $\mathbf{y}$  and  $\mathbf{x}$  by

$$\begin{pmatrix} \mathbf{I}_m & -\Sigma_{12}\Sigma_{22}^{-1} \\ \Sigma_{21}\Sigma_{11}^{-1} & \mathbf{I}_k \end{pmatrix}$$

gives a new system of equations, where the first  $m$  equations become (exercise)

$$(45) \quad \mathbf{y}_t = \sum_{i=0}^p \mathbf{C}_{3i}\mathbf{x}_{t-i} + \sum_{i=1}^p \mathbf{D}_{3i}\mathbf{y}_{t-i} + \omega_{1t},$$

with the error  $\omega_{1t} = \nu_{1t} - \Sigma_{12}\Sigma_{22}^{-1}\nu_{2t}$  that is uncorrelated with  $\nu_{2t}$  \* and consequently with  $\mathbf{x}_t$  (important!). That is, we may describe the same structural relationship between  $\mathbf{y}_t$  and  $\mathbf{x}_t$  with contemporaneously uncorrelated error terms for the price of including the current value  $\mathbf{x}_t$  as additional explanatory variable.

$$\begin{aligned} * \text{Cov}(\omega_{1t}, \nu_{2t}) &= \text{Cov}(\nu_{1t} - \Sigma_{12}\Sigma_{22}^{-1}\nu_{2t}, \nu_{2t}) = \\ &= \text{Cov}(\nu_{1t}, \nu_{2t}) - \Sigma_{12}\Sigma_{22}^{-1}\text{Cov}(\nu_{2t}, \nu_{2t}) = \Sigma_{12} - \Sigma_{12} = \mathbf{0} \end{aligned}$$

Similarly, the last  $k$  equations can be written as

$$(46) \quad \mathbf{x}_t = \sum_{i=1}^p \mathbf{E}_{3i} \mathbf{x}_{t-i} + \sum_{i=0}^p \mathbf{F}_{3i} \mathbf{y}_{t-i} + \omega_{2t}.$$

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

$$(47) \quad F_{\mathbf{x} \cdot \mathbf{y}} = \ln(|\Sigma_{11}| / |\Sigma_{\omega 1}|) = \ln(|\Sigma_{22}| / |\Sigma_{\omega 2}|).$$

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

$$(48) \quad F_{\mathbf{x}, \mathbf{y}} = F_{\mathbf{x} \rightarrow \mathbf{y}} + F_{\mathbf{y} \rightarrow \mathbf{x}} + F_{\mathbf{x} \cdot \mathbf{y}}.$$

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

$$(49) \quad \hat{\Sigma}_i = (T - p)^{-1} \sum_{t=p+1}^T \hat{\eta}_{it} \hat{\eta}'_{it},$$

$$(50) \quad \hat{\Sigma}_{ii} = (T - p)^{-1} \sum_{t=p+1}^T \hat{v}_{it} \hat{v}'_{it},$$

$$(51) \quad \hat{\Sigma}_{\omega i} = (T - p)^{-1} \sum_{t=p+1}^T \hat{\omega}_{it} \hat{\omega}'_{it},$$

for  $i = 1, 2$ . For example, returning to the UK stock and bond returns:

$$\begin{aligned} \hat{F}_{\mathbf{x} \rightarrow \mathbf{y}} &= \ln(|\hat{\Sigma}_1| / |\hat{\Sigma}_{11}|) \\ &= \ln(286.4167 / 276.0856) \\ &= 0.036737 \end{aligned}$$

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

$$\text{LR} = (369 - 2) \cdot 0.036737 = 13.48,$$

the same result as we found earlier.

With these estimates one can test the particular dependencies,

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

$$(52) \quad (T - p)\hat{F}_{\mathbf{x} \rightarrow \mathbf{y}} \sim \chi_{mkp}^2.$$

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

$$(53) \quad (T - p)\hat{F}_{\mathbf{y} \rightarrow \mathbf{x}} \sim \chi_{mkp}^2.$$

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

$$(54) \quad (T - p)\hat{F}_{\mathbf{x} \cdot \mathbf{y}} \sim \chi_{mk}^2.$$

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

$$(55) \quad (T - p)\hat{F}_{\mathbf{x}, \mathbf{y}} \sim \chi_{mk(2p+1)}^2.$$

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

Example: (UK returns continued.)

Running the appropriate regressions yields  $|\hat{\Sigma}_{\omega 2}| = 203.9899$ ,  $|\hat{\Sigma}_{22}| = 232.2855$ ,  $|\hat{\Sigma}_2| = 245.4944$ , and  $|\hat{\Sigma}_{\omega 1}| = 242.4533$ , such that

$$\hat{F}_{y \rightarrow x} = 0.05531, \quad \hat{F}_{x \cdot y} = 0.1299,$$

and recalling that  $\hat{F}_{x \rightarrow y} = 0.03674$ :

$$\hat{F}_{x,y} = 0.03674 + 0.05531 + 0.1299 = 0.22195.$$

The LR-statistics of the above measures and the associated  $p$  values for the equity-bond data are reported in the following table:

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

	LR	DF	P-VALUE
x-->y	13.48	8	0.0963
y-->x	20.30	8	0.0093
x.y	47.67	4	0.0000
x,y	81.46	20	0.0000

There is thus linear dependence between bond and stock returns mainly due to instantaneous feedback but also due to Granger-causality from bonds to stocks.

## 4.7 Variance decomposition and innovation accounting

Consider the VAR( $p$ ) model

$$(56) \quad \Phi(L)y_t = \epsilon_t,$$

where

$$(57) \quad \Phi(L) = \mathbf{I}_m - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p$$

is the lag polynomial of order  $p$  with  $m \times m$  coefficient matrices  $\Phi_i$ ,  $i = 1, \dots, p$ .

Provided that the stationarity condition holds we may obtain a vector MA representation of  $y_t$  by left multiplication with  $\Phi^{-1}(L)$  as

$$(58) \quad y_t = \Phi^{-1}(L)\epsilon_t = \Psi(L)\epsilon_t$$

where

$$(59) \quad \Phi^{-1}(L) = \Psi(L) = \mathbf{I}_m + \Psi_1 L + \Psi_2 L^2 + \dots .$$

The  $m \times m$  coefficient matrices  $\Psi_1, \Psi_2, \dots$  may be obtained from the identity

(60)

$$\Phi(L)\Psi(L) = \left(\mathbf{I}_m - \sum_{i=1}^p \Phi_i L^i\right) \left(\mathbf{I}_m + \sum_{i=1}^{\infty} \Psi_i L^i\right) = \mathbf{I}_m$$

as

$$(61) \quad \Psi_j = \sum_{i=1}^j \Psi_{j-i} \Phi_i$$

with  $\Psi_0 = \mathbf{I}_m$  and  $\Phi_i = \mathbf{0}$  when  $i > p$ , by multiplying out and setting the resulting coefficient matrix for each power of  $L$  equal to zero. For example, start with  $L_1^1 = L_1$ :

$$\begin{aligned} -\Phi_1 L_1 + \Psi_1 L_1 &= (\Psi_1 - \Phi_1) L_1 \equiv 0 \\ \Rightarrow \Psi_1 &= \Phi_1 = \Psi_0 \Phi_1 = \sum_{i=1}^1 \Psi_{1-i} \Phi_i \end{aligned}$$

Consider next  $L_1^2$ :

$$\begin{aligned} \Psi_2 L_1^2 - \Psi_1 \Phi_1 L_1^2 - \Phi_2 L_1^2 &\equiv 0 \\ \Rightarrow \Psi_2 &= \Psi_1 \Phi_1 + \Phi_2 = \sum_{i=1}^2 \Psi_{2-i} \Phi_i \end{aligned}$$

The result generalizes to any power  $L_1^j$ , which yields the transformation formula given above.



Now, since

$$(62) \quad y_{t+s} = \Psi(L)\epsilon_{t+s} = \epsilon_{t+s} + \sum_{i=1}^{\infty} \Psi_i \epsilon_{t+s-i}$$

we have that the effect of a unit change in  $\epsilon_t$  on  $y_{t+s}$  is

$$(63) \quad \frac{\partial y_{t+s}}{\partial \epsilon_t} = \Psi_s.$$

Now the  $\epsilon_t$ 's represent shocks in the system. Therefore the  $\Psi_i$  matrices represent the model's response to a unit shock (or innovation) at time point  $t$  in *each* of the variables  $i$  periods ahead. Economists call such parameters dynamic multipliers.

The response of  $y_i$  to a unit shock in  $y_j$  is therefore given by the sequence below, known as the impulse response function,

$$(64) \quad \psi_{ij,1}, \psi_{ij,2}, \psi_{ij,3}, \dots,$$

where  $\psi_{ij,k}$  is the  $ij$ th element of the matrix  $\Psi_k$  ( $i, j = 1, \dots, m$ ).

For example if we were told that the first element in  $\epsilon_t$  changes by  $\delta_1$  at the same time that the second element changed by  $\delta_2, \dots$ , and the  $m$ th element by  $\delta_m$ , then the combined effect of these changes on the value of the vector  $y_{t+s}$  would be given by

(65)

$$\Delta y_{t+s} = \frac{\partial y_{t+s}}{\partial \epsilon_{1t}} \delta_1 + \dots + \frac{\partial y_{t+s}}{\partial \epsilon_{mt}} \delta_m = \Psi_s \delta,$$

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

*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.*

## Example: Exogeneity in MA representation

Suppose we have a bivariate VAR system such that  $x_t$  does not Granger cause  $y_t$ . Then

$$(66) \quad \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \phi_{11}^{(1)} & 0 \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \dots \\ + \begin{pmatrix} \phi_{11}^{(p)} & 0 \\ \phi_{21}^{(p)} & \phi_{22}^{(p)} \end{pmatrix} \begin{pmatrix} y_{t-p} \\ x_{t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}.$$

The coefficient matrices  $\Psi_j = \sum_{i=1}^j \Psi_{j-i} \Phi_i$  in the corresponding MA representation are lower triangular as well (exercise):

$$(67) \quad \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} + \sum_{i=1}^{\infty} \begin{pmatrix} \psi_{11}^{(i)} & 0 \\ \psi_{21}^{(i)} & \psi_{22}^{(i)} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-i} \\ \epsilon_{2,t-i} \end{pmatrix}$$

Hence, we see that variable  $y$  does not react to a shock in  $x$ . Similarly, if there are exogenous variables in a  $m$ -variate VAR, then the implied zero restrictions in the  $\Psi_j$  matrices ensure that the endogenous variables do not react to shocks in the exogenous variables.

## Ambiguity of impulse response functions

Consider a bivariate VAR model in vector MA representation, that is,

$$(68) \quad \mathbf{y}_t = \Psi(L)\epsilon_t \quad \text{with} \quad \mathbb{E}(\epsilon_t \epsilon_t') = \Sigma_\epsilon,$$

where  $\Psi(L)$  gives the response of  $\mathbf{y}_t = (y_{t1}, y_{t2})'$  to both elements of  $\epsilon_t$ , that is,  $\epsilon_{t1}$  and  $\epsilon_{t2}$ .

Just as well we might be interested in evaluating responses of  $\mathbf{y}_t$  to *linear combinations* of  $\epsilon_{t1}$  and  $\epsilon_{t2}$ , for example to unit movements in  $\epsilon_{t1}$  and  $\epsilon_{t2} + 0.5\epsilon_{t1}$ . This may be done by defining new shocks  $\nu_{t1} = \epsilon_{t1}$  and  $\nu_{t2} = \epsilon_{t2} + 0.5\epsilon_{t1}$ , or in matrix notation

$$(69) \quad \nu_t = Q\epsilon_t \quad \text{with} \quad Q = \begin{pmatrix} 1 & 0 \\ 0.5 & 1 \end{pmatrix}.$$

The vector MA representation of our VAR in terms of the new shocks becomes then

(70)

$$\mathbf{y}_t = \Psi(L)\epsilon_t = \Psi(L)Q^{-1}Q\epsilon_t =: \Psi^*(L)\nu_t$$

with

$$(71) \quad \Psi^*(L) = \Psi(L)Q^{-1}.$$

Note that both representations are observationally equivalent (they produce the same  $\mathbf{y}_t$ ), but yield different impulse response functions. In particular,

$$(72) \quad \psi_0^* = \psi_0 \cdot Q^{-1} = \mathbf{I} \cdot Q^{-1} = Q^{-1},$$

which implies that single component shocks may now have contemporaneous effects on more than one component of  $\mathbf{y}_t$ . Also the covariance matrix of residuals will change, since

$$\mathbb{E}(\nu_t\nu_t') = \mathbb{E}(Q\epsilon_t\epsilon_t'Q') \neq \Sigma_\epsilon \quad \text{unless} \quad Q = \mathbf{I}.$$

But the fact that both representations are observationally equivalent implies that we must make a choice which linear combination of the  $\epsilon_{ti}$ 's we find most useful to look at in the response analysis!

## Orthogonalized impulse response functions

Usually the components of  $\epsilon_t$  are contemporaneously correlated, meaning that they have overlapping information to some extent. For example in our VAR(2) model of the equity-bond data the contemporaneous residual correlations are

```
=====
          FTA    DIV    R20    TBILL
-----
FTA      1
DIV      0.910  1
R20     -0.307 -0.211  1
TBILL   -0.150  0.099  0.464  1
=====
```

In impulse response analysis, however, we wish to describe the effects of shocks to a single series only, such that we are able to discriminate between the effects of shocks applied to different series. It is therefore desirable to express the VAR in such a way that the shocks become orthogonal, (that is, the  $\epsilon_{ti}$ 's are uncorrelated). Additionally it is convenient to rescale the shocks such that they have unit variance.

So we want to pick a  $Q$  such that  $\mathbb{E}(\nu_t \nu_t') = \mathbf{I}$ . This may be accomplished by choosing  $Q$  such that

$$(73) \quad Q^{-1} Q^{-1'} = \Sigma_\epsilon$$

since then

$$(74) \quad \mathbb{E}(\nu_t \nu_t') = \mathbb{E}(Q \epsilon_t \epsilon_t' Q') = \mathbb{E}(Q \Sigma_\epsilon Q') = \mathbf{I}.$$

Unfortunately there are many different  $Q$ 's, whose inverse  $\mathbf{S} = Q^{-1}$  act as "square roots" for  $\Sigma_\epsilon$ , that is,  $\mathbf{S} \mathbf{S}' = \Sigma_\epsilon$ .

This may be seen as follows. Choose any orthogonal matrix  $\mathbf{R}$  (that is,  $\mathbf{R} \mathbf{R}' = \mathbf{I}$ ) and set  $\mathbf{S}^* = \mathbf{S} \mathbf{R}$ . We have then

$$(75) \quad \mathbf{S}^* \mathbf{S}^{*'} = \mathbf{S} \mathbf{R} \mathbf{R}' \mathbf{S}' = \mathbf{S} \mathbf{S}' = \Sigma_\epsilon.$$

Which of the many possible  $\mathbf{S}$ 's, respectively  $Q$ 's, should we choose?

Before turning to a clever choice of  $Q$  (resp.  $S$ ), let us briefly restate our results obtained so far in terms of  $S = Q^{-1}$ .

If we find a matrix  $S$  such that  $SS' = \Sigma_\epsilon$ , and transform our VAR residuals such that

$$(76) \quad \nu_t = S^{-1}\epsilon_t,$$

then we obtain an observationally equivalent VAR where the shocks are orthogonal (i.e. uncorrelated with a unit variance), that is,

$$(77) \quad E(\nu_t\nu_t') = S^{-1}E(\epsilon_t\epsilon_t')S'^{-1} = S^{-1}\Sigma_\epsilon S'^{-1} = I.$$

The new vector MA representation becomes

$$(78) \quad y_t = \Psi^*(L)\nu_t = \sum_{i=0}^{\infty} \psi_i^* \nu_{t-i},$$

where  $\psi_i^* = \psi_i S$  ( $m \times m$  matrices) so that  $\psi_0^* = S \neq I_m$ . The impulse response function of  $y_i$  to a unit shock in  $y_j$  is then given by the orthogonalised impulse response function

$$(79) \quad \psi_{ij,0}^*, \psi_{ij,1}^*, \psi_{ij,2}^*, \dots \cdot$$



## Choleski Decomposition & Ordering of Variables

Note that every orthogonalization of correlated shocks in the original VAR leads to contemporaneous effects of single component shocks  $\nu_{ti}$  to more than one component of  $y_t$ , since  $\psi_0 = \mathbf{S}$  will not be diagonal unless  $\Sigma_\epsilon$  was diagonal already.

One generally used method is to choose  $\mathbf{S}$  to be a lower triangular matrix. This is called the Cholesky decomposition which results in a lower triangular matrix with positive main diagonal elements for  $\Psi_0^* = \mathbf{I}_m \mathbf{S} = \mathbf{S}$ , e.g.

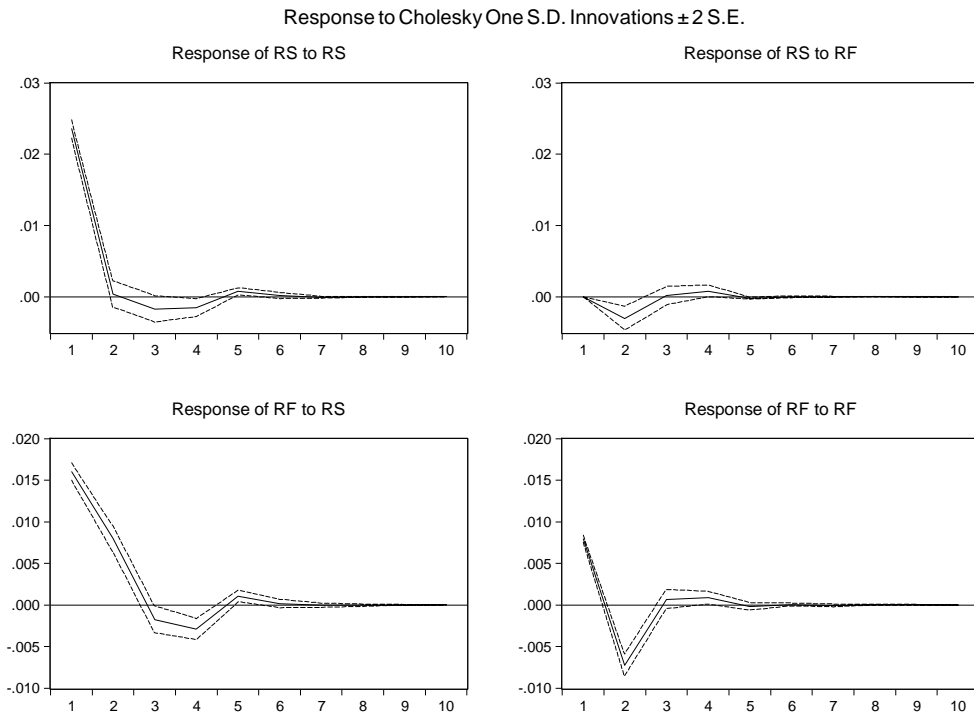
(80)

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \psi_{11}^{*(0)} & 0 \\ \psi_{21}^{*(0)} & \psi_{22}^{*(0)} \end{pmatrix} \begin{pmatrix} \nu_{1,t} \\ \nu_{2,t} \end{pmatrix} + \Psi^*(1)\nu_{t-1} + \dots$$

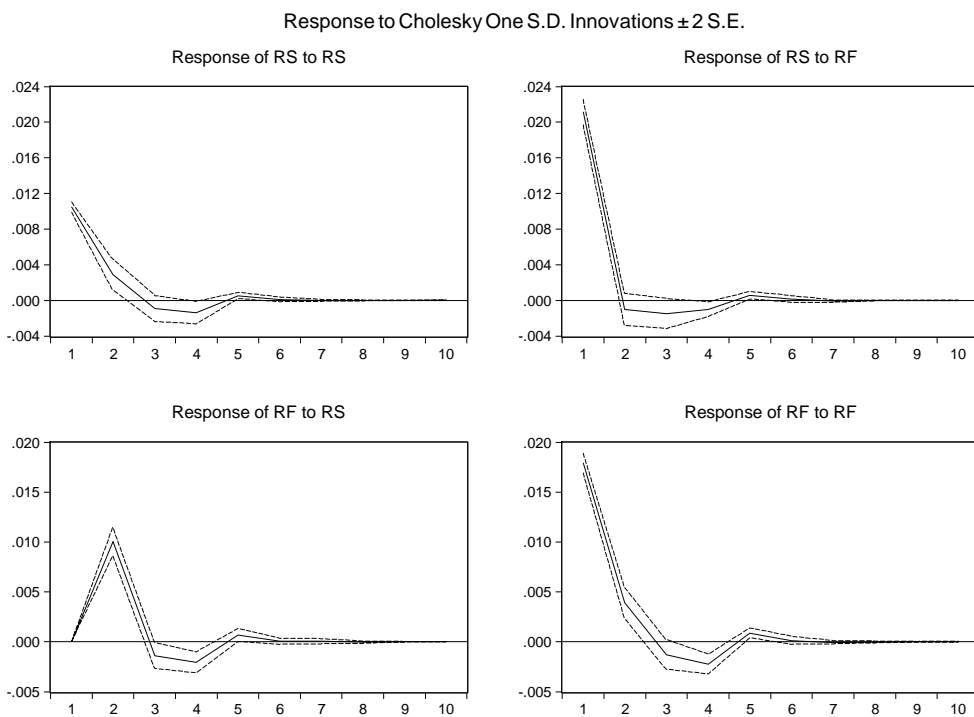
Hence Cholesky decomposition of  $\Sigma_\epsilon$  implies that the second shock  $\nu_{2,t}$  does not affect the first variable  $y_{1,t}$  contemporaneously, but both shocks can have a contemporaneous effect on  $y_{2,t}$  (and all following variables, if we had chosen an example with more than two components). Hence the ordering of variables is important!

Example: (Oil spot and futures continued.)

○. IRF's  $\psi_{ij,k}^*$  when spots are the first series:



○. IRF's  $\psi_{ij,k}^*$  with futures as first series:



## Generalized Impulse Response Function

The preceding approach is somewhat unsatisfactory as its results may be heavily influenced by the ordering of variables, which is a choice made by the researcher rather than a characteristic of the series.

In order to avoid this, the generalized impulse response function at horizon  $s$  to a shock  $\delta_j$  in series  $j$  is defined as<sup>‡‡</sup>

$$(81) \quad GI(s, \delta_j) = \mathbb{E}_t[\mathbf{y}_{t+s} | \epsilon_{jt} = \delta_j] - \mathbb{E}_t[\mathbf{y}_{t+s}].$$

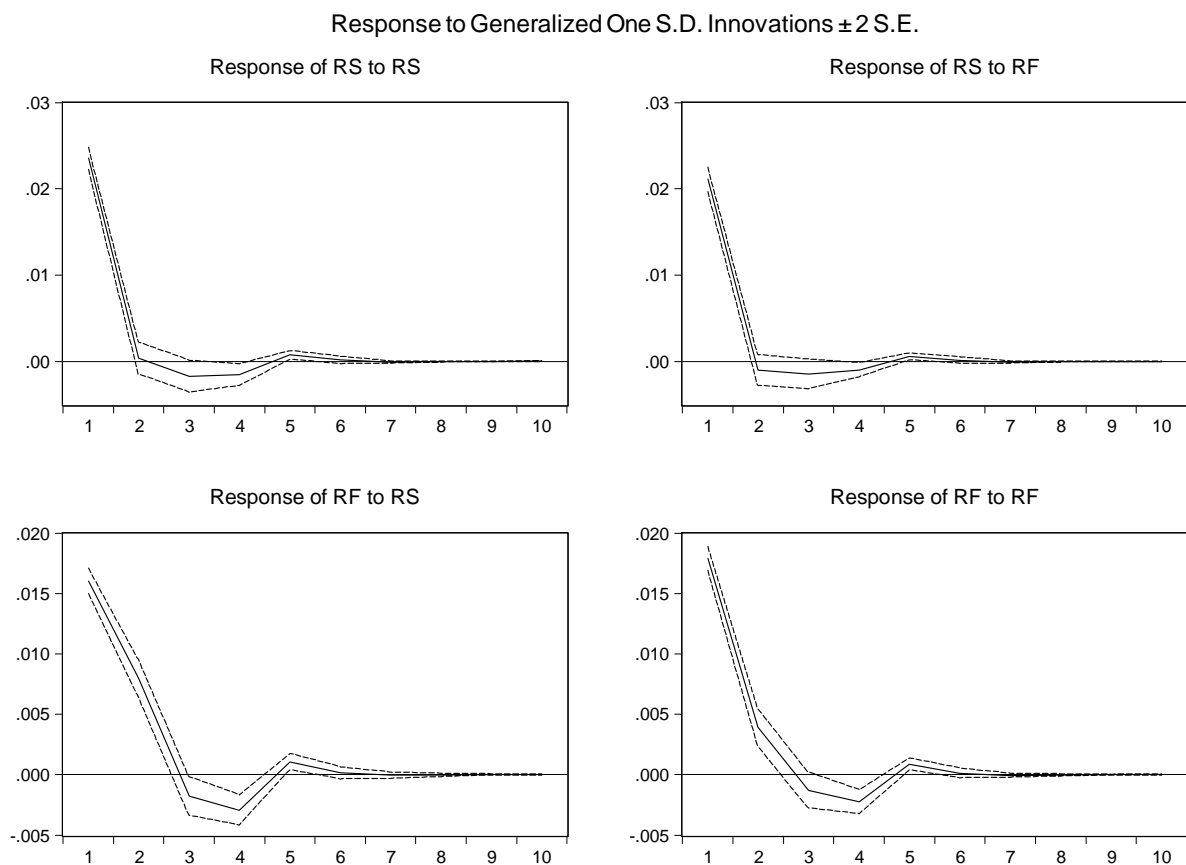
That is the difference in conditional expectations of  $\mathbf{y}_{t+s}$  at time  $t$  whether a shock occurs in series  $j$  or not.

Comparing this with (80) we find that this coincides with the orthogonalized response function using the  $j$ 's series as the first series in the Cholesky decomposition. The other components of the generalized and orthogonalized response functions coincide only if the residual covariance matrix  $\Sigma_\epsilon$  is diagonal.

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

Example: (Oil spot and futures continued.)

EViews output for the generalized impulse response functions is given below:



The generalized impulse responses to the spot returns equal the orthogonalized impulse responses using spot returns as the first Cholesky component, whereas the generalized impulse responses to the futures returns equal the orthogonalized impulse responses using futures returns as the first Cholesky component.

## Variance decomposition

Variance decomposition refers to the breakdown of the forecast error variance into components due to shocks in the series. Basically, variance decomposition can tell a researcher the percentage of the fluctuation in a time series attributable to other variables at selected time horizons.

More precisely, the uncorrelatedness of the orthogonalized shocks  $\nu_t$ 's allows us to decompose the error variance of the  $s$  step-ahead forecast of  $y_{it}$  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  $\sum_{l=0}^{s-1} \psi_{ij}^{*(l)2}$ , as we shall see below.

Consider an orthogonalized VAR with  $m$  components in vector MA representation,

$$(82) \quad \mathbf{y}_t = \sum_{l=0}^{\infty} \psi^*(l) \nu_{t-l}.$$

The  $s$  step-ahead forecast for  $\mathbf{y}_t$  is then

$$(83) \quad \mathbb{E}_t(\mathbf{y}_{t+s}) = \sum_{l=s}^{\infty} \psi^*(l) \nu_{t+s-l}.$$

Defining the  $s$  step-ahead forecast error as

$$(84) \quad \mathbf{e}_{t+s} = \mathbf{y}_{t+s} - \mathbb{E}_t(\mathbf{y}_{t+s})$$

we get

$$(85) \quad \mathbf{e}_{t+s} = \sum_{l=0}^{s-1} \psi^*(l) \nu_{t+s-l}.$$

It's  $i$ 'th component is given by

$$(86) \quad e_{i,t+s} = \sum_{l=0}^{s-1} \sum_{j=1}^m \psi_{ij}^*(l) \nu_{j,t+s-l} = \sum_{j=1}^m \sum_{l=0}^{s-1} \psi_{ij}^*(l) \nu_{j,t+s-l}.$$

Now, because the shocks are both serially and contemporaneously uncorrelated, we get for the error variance

$$\begin{aligned}
 \mathbf{V}(e_{i,t+s}) &= \sum_{j=1}^m \sum_{l=0}^{s-1} \mathbf{V}(\psi_{ij}^{*(l)} \nu_{j,t+s-l}) \\
 (87) \qquad &= \sum_{j=1}^m \sum_{l=0}^{s-1} \psi_{ij}^{*(l)2} \mathbf{V}(\nu_{j,t+s-l}).
 \end{aligned}$$

Now, recalling that all shock components have unit variance, this implies that

$$(88) \qquad \mathbf{V}(e_{i,t+s}) = \sum_{j=1}^m \left( \sum_{l=0}^{s-1} \psi_{ij}^{*(l)2} \right),$$

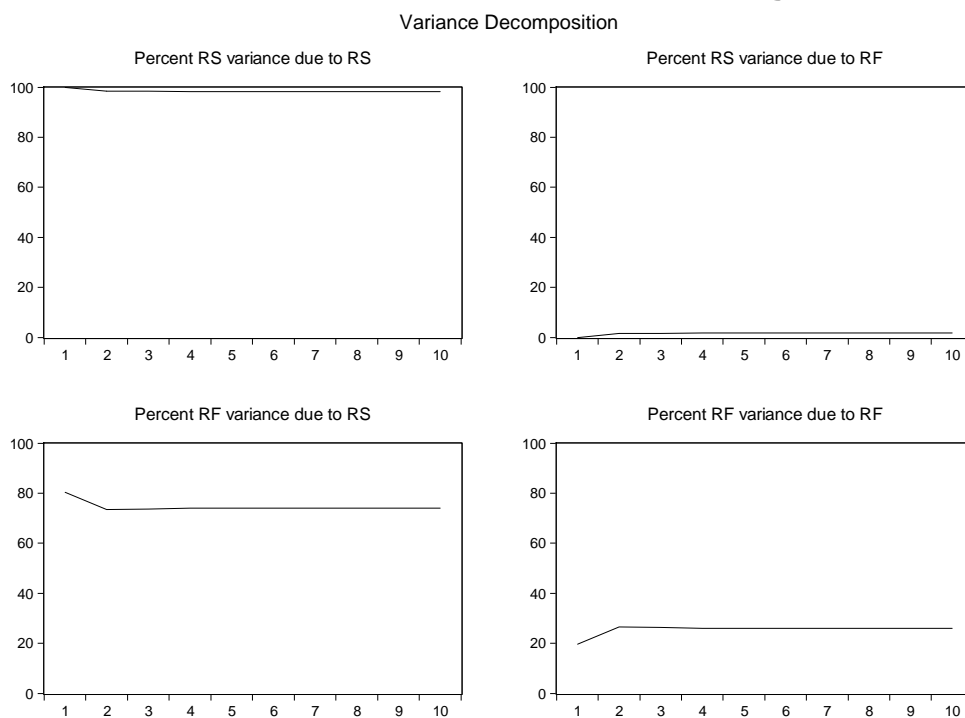
where  $\sum_{l=0}^{s-1} \psi_{ij}^{*(l)2}$  accounts for the error variance generated by innovations to  $y_j$ , as claimed.

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

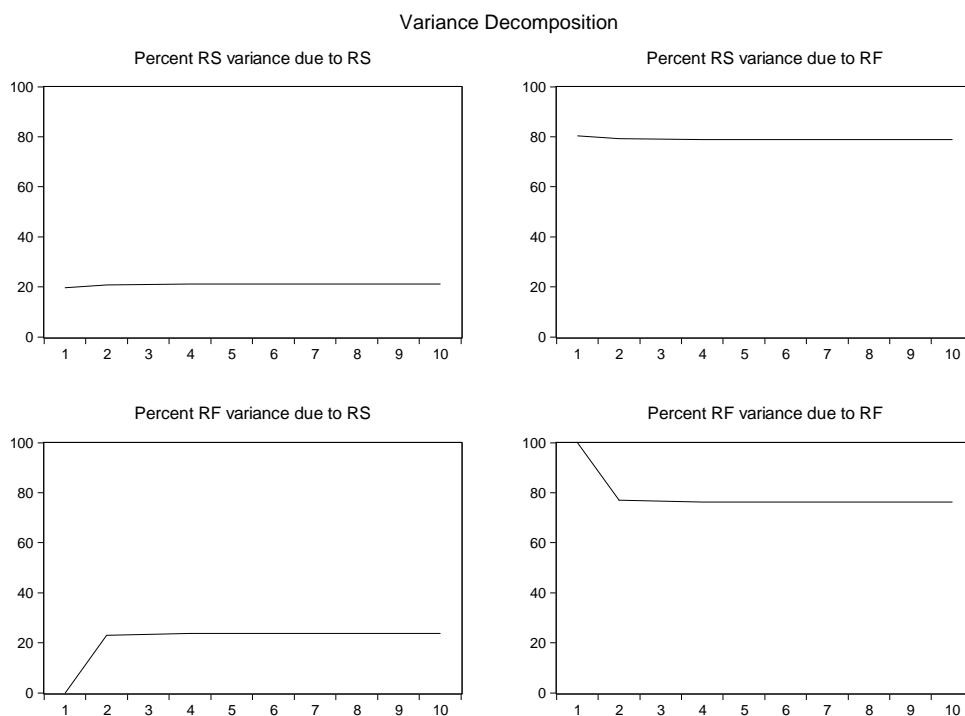
$$(89) \qquad R_{ij,s}^2 = 100 \frac{\sum_{l=0}^{s-1} \psi_{ij}^{*(l)2}}{\sum_{k=1}^m \sum_{l=0}^{s-1} \psi_{ik}^{*(l)2}}.$$

## Example: (Oil spot and futures continued.)

Spot returns as first Cholesky component:



Futures returns as first Cholesky component:





## On the ordering of variables

Here we see very clearly that when the residuals are contemporaneously correlated, i.e.,  $\text{cov}(\epsilon_t) = \Sigma_\epsilon \neq \mathbf{I}$ , the orthogonalized impulse response coefficients and hence the variance decompositions are not unique. There are no statistical methods to define the ordering. It must be done by the analyst!

Various orderings should be tried to check for consistency of the resulting interpretations. 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  $y_t$ , but not on  $y_{1t}$ , the first component, and so on. Of course this is usually a difficult task in practice.

## Variance Decomposition using Generalized Impulse Responses

In an attempt to eliminate the dependence on the ordering of the variables, Pesaran et al suggest to calculate the percentage of forecast error variance in series  $i$  caused by series  $j$  as

$$(90) \quad R_{ij,s}^g{}^2 = 100 \frac{\sum_{l=0}^{s-1} \psi_{ij}^{g(l)2}}{\sum_{k=1}^m \sum_{l=0}^{s-1} \psi_{ik}^{*(l)2}},$$

that is, by replacing the orthogonal impulse response functions  $\psi_{ij}^{*(l)}$  with the corresponding generalized impulse response functions  $\psi_{ij}^{g(l)}$  in the numerator of (89). A problem with this approach is that a proper split-up of the forecast variance really requires orthogonal components, i.e. the percentages above do not sum up to 100%.

Some authors attempt to tackle this problem by renormalizing the  $R_{ij,l}^g$  in (90) as

$$(91) \quad R_{ij,s}^{g'} = 100 \frac{\sum_{l=0}^{s-1} \psi_{ij}^{g(l)2}}{\sum_{k=1}^m \sum_{l=0}^{s-1} \psi_{ik}^{g(l)2}},$$

such that they sum up to 100%, but that only masks the problem, because with non-orthogonal shocks, we always have components of the variance of which we don't know to which series they belong.

For that reason EViews does not have an option to calculate variance decompositions based upon generalized impulse responses.

On the other hand, generalized variance decomposition is still useful for analyzing how important shocks in certain series are relative to shocks in other series, in a way that is not influenced by the subjective ordering of the series by the researcher.

## Calculating $R_{ij,s}^g$ and $R_{ij,s}^{g'}$ with EViews

---

Recall that the generalized impulse response function  $\psi_{ij}^g$  coincides with the orthogonal impulse response function  $\psi_{ij}^*$  using series  $j$  as the first series in the Cholesky decomposition.

This allows us to calculate  $R_{ij,s}^g$  generated by series  $j$  by simply asking EViews to perform a variance decomposition using series  $j$  as the first Cholesky component and only considering that first Cholesky component.

In order to obtain  $R_{ij,s}^{g'}$ , do this for all  $m$  series and divide each  $R_{ij,s}^g$  calculated above by their sum  $\sum_{j=1}^m R_{ij,s}^g$ . Then multiply this with 100, to get a percentage.

Example: (Oil spot and futures continued.)

As an example consider EViews variance decomposition output  $s = 9$  periods after the shock (period 10 in EViews). The variance component due to a shock in the spot series using spots as the first Cholesky component are 98.2644% for the spot series and 74.0526% for the futures series. The variance component due to a shock in the futures series using futures as the first Cholesky component are 78.8621% for the spot series and 76.1674% for the futures series. Hence denoting spots with 1 and futures with 2:

$R_{ij,9}^g$	$j = 1$	$j = 2$
$i = 1$	98.2644	78.8621
$i = 2$	74.0526	76.1674

Example: (continued.)

The renormalized variance components are

$$R_{11,9}^{g'}{}^2 = 100 \cdot \frac{98.2644}{98.2644 + 78.8621} = 55.477,$$

$$R_{12,9}^{g'}{}^2 = 100 \cdot \frac{78.8621}{98.2644 + 78.8621} = 44.523,$$

$$R_{21,9}^{g'}{}^2 = 100 \cdot \frac{74.0526}{74.0526 + 76.1674} = 49.296,$$

$$R_{22,9}^{g'}{}^2 = 100 \cdot \frac{76.1674}{74.0526 + 76.1674} = 50.704.$$

Hence, at this time horizon, shocks to the time series themselves have only a slightly larger impact upon the forecast variance than shocks to the other series, the difference being slightly more pronounced for the spot than for the futures series.

## On estimation of the impulse response coefficients

Consider the VAR( $p$ ) model

$$\Phi(L)y_t = \epsilon_t,$$

with  $\Phi(L) = \mathbf{I}_m - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p$ . Then under stationarity the vector MA representation is

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

When we have estimates of the AR-matrices  $\Phi_i$  denoted by  $\hat{\Phi}_i$ ,  $i = 1, \dots, p$ ; the next problem is to construct estimates  $\hat{\Psi}_j$  for the MA matrices  $\Psi_j$ . Recall that

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

with  $\Psi_0 = \mathbf{I}_m$ , and  $\Phi_j = \mathbf{0}$  when  $i > p$ . The estimates  $\hat{\Psi}_j$  can be obtained by replacing the  $\Phi_i$ 's by their corresponding estimates  $\hat{\Phi}_i$ .

Next we have to obtain the orthogonalized impulse response coefficients. This can be done easily, for letting  $\mathbf{S}$  be the Cholesky decomposition of  $\Sigma_\epsilon$  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  $\Psi_i^*$  are obtained by replacing  $\Psi_t$  with their estimates  $\hat{\Psi}_t$  and using Cholesky decomposition of  $\hat{\Sigma}_\epsilon$ .



## 4.8 Cointegration

### *Motivation and Definition*

Consider the *unbiased forward rate hypothesis*, according to which the futures price of an underlying should equal today's expectation of the underlyings spot price one period ahead, that is,

$$f_t = E_t(s_{t+1}),$$

where  $f_t$  and  $s_t$  denote the logarithm of the futures and the spot prices, respectively. Now, as discussed earlier, rational expectations require that the forecasting errors

$$\epsilon_t := s_{t+1} - E_t(s_{t+1}) = s_{t+1} - f_t$$

are serially uncorrelated with zero mean, in particular  $\epsilon_t$  should be stationary. This appears to be quite a special relationship, because both  $f_t$  and  $s_t$  are  $I(1)$  variables, and for most cases, linear combinations of  $I(1)$  variables are  $I(1)$  variables themselves.

Whenever we can find a linear combination of two  $I(1)$  variables, which is stationary, then we say that the variables are cointegrated.

Formally, consider two  $I(1)$  series  $x_t$  and  $y_t$ .

In general  $u_t = y_t - \beta x_t \sim I(1)$  for any  $\beta$ .

However, if there exist a  $\beta \neq 0$  such that  $y_t - \beta x_t \sim I(0)$ , then  $y_t$  and  $x_t$  are said to be *cointegrated*.

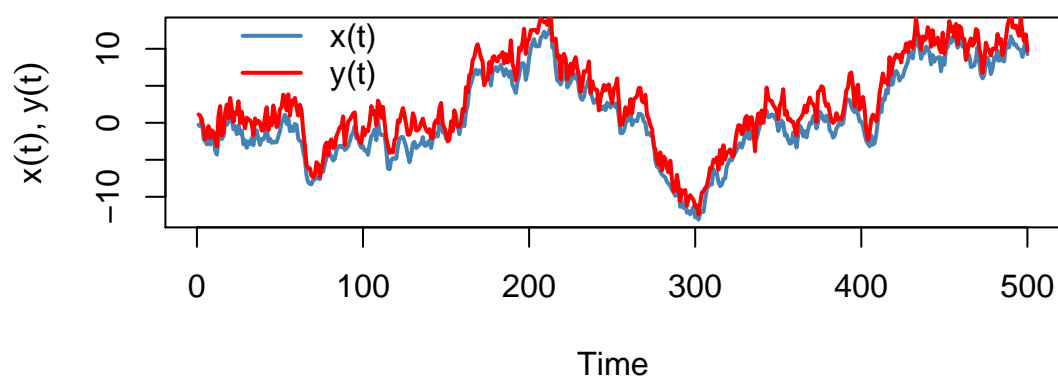
If  $x_t$  and  $y_t$  are cointegrated then  $\beta$  in  $y_t - \beta x_t$  is unique.

$\beta$  is called the cointegration parameter and  $(1, \beta)$  is called the cointegration vector (ci-vector).

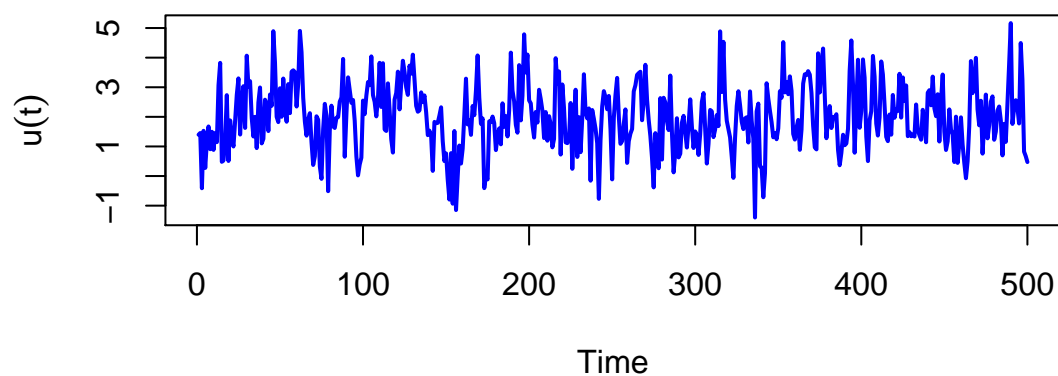
Cointegrated series do not depart "far away" from each other. This will later allow us to set up so called error correction models which can be used to forecast each series based upon the value of the other series, even though none of the series is stationary.

Example: Cointegrated series  $x_t$  and  $y_t$  with ci-vector  $(1, -1)$ , such that  $u_t = y_t - x_t$  is stationary.

### Cointegrated Series



### Stationary $u(t) = y(t) - x(t)$



Remark: If  $y_t - \beta x_t \sim I(0)$ , then  $x_t - \gamma y_t \sim I(0)$ , where  $\gamma = 1/\beta$ .

Note also that for any  $a \neq 0$ ,  $ay_t - a\beta x_t \sim I(0)$ , which implies that the cointegration parameter  $\beta$  is unique when  $a$  is fixed to unity.

Remark: If  $x_t \sim I(0)$  and  $y_t \sim I(0)$  then for any  $a, b \in \mathbb{R}$ ,  $ax_t + by_t \sim I(0)$

If  $x_t \sim I(1)$  and  $y_t \sim I(0)$  then for any  $a, b \in \mathbb{R}$ ,  $a \neq 0$ ,  $ax_t + by_t \sim I(1)$ .

The general definition of cointegration with  $n$  series  $x_{1t}, \dots, x_{nt}$  compiled in a vector  $x_t$  is:

The components of  $x_t = (x_{1t}, \dots, x_{nt})'$  are said to be cointegrated of order  $d, b$ , denoted by  $x_t \sim CI(d, b)$ , if

1. all components of  $x_t$  are integrated of the same order  $d$ ,
2. there exists a vector  $\beta = (\beta_1, \dots, \beta_n) \neq 0$  such that  $\beta x_t \sim I(d - b)$ , where  $b > 0$ .

Note: The most common case is  $x_t \sim CI(1, 1)$ .

Example: *unbiased forward rate hypothesis*

$x_t := (s_{t+1}, f_t)'$  is cointegrated of order 1,1 with cointegrating vector  $\beta = (1, -1)$  since:

1.  $s_{t+1}, f_t \sim I(1)$ , and
2.  $(1, -1)(s_{t+1}, f_t)' = \epsilon_t \sim I(0)$ .

Note: When arguing for cointegration between futures and spot prices we assumed both rational expectations and forward rate unbiasedness. Whether that holds true, is really an empirical matter which needs to be tested.

## Testing for cointegration

### (a) Known ci-relation

If the ci-vector  $(1, -\beta)$  is known (i.e.,  $\beta$  is known, e.g.  $\beta = 1$ ), testing for cointegration means testing for stationarity of

$$(92) \quad u_t = y_t - \beta x_t.$$

Testing can be worked out with the ADF testing.

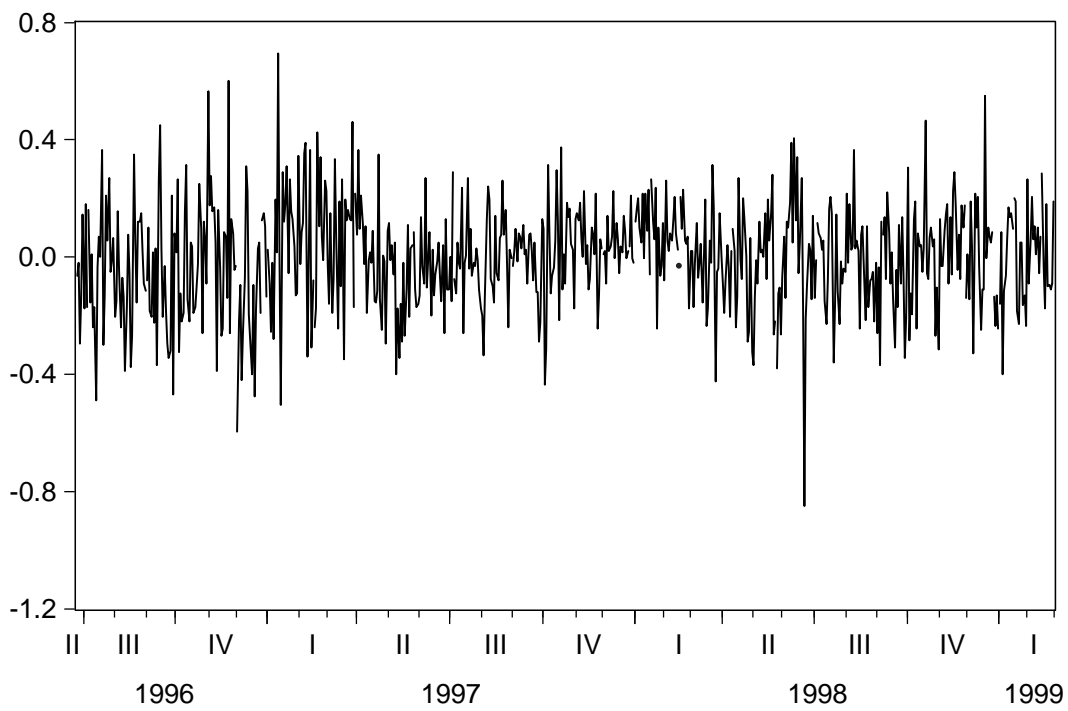
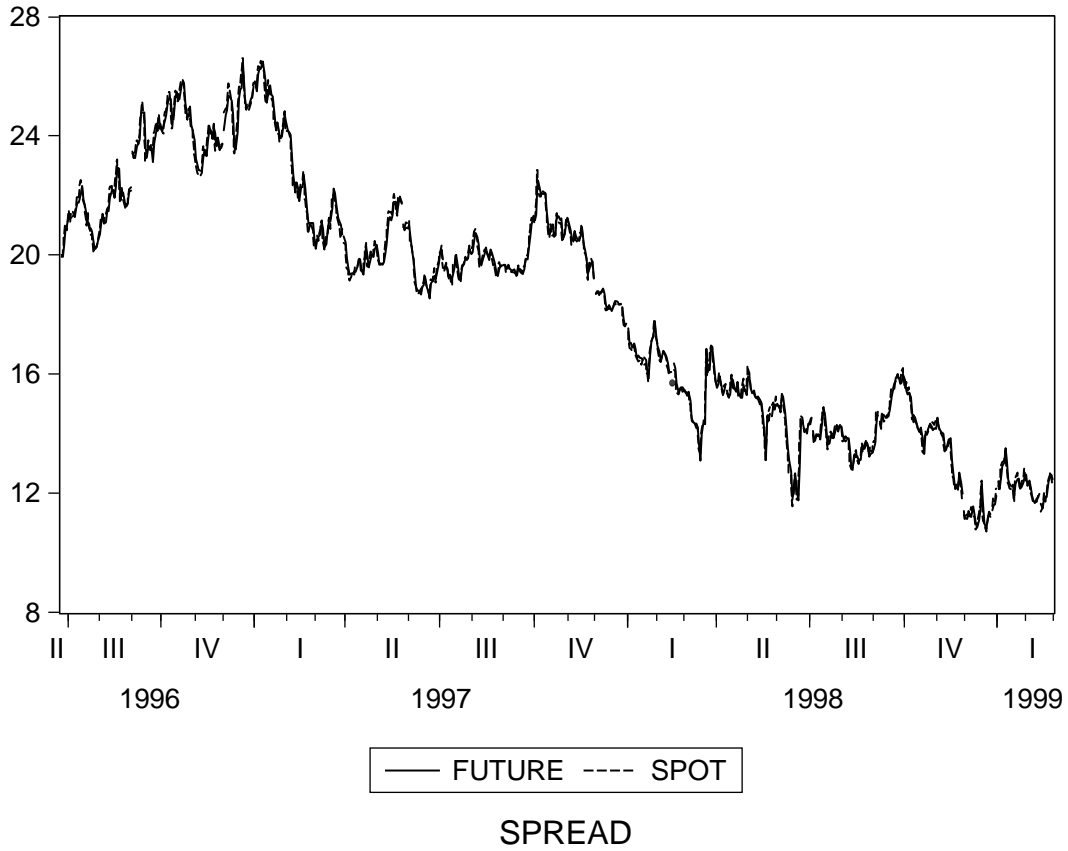
Note that in ADF testing the null hypothesis is that the series is  $I(1)$ .

When applied to ci-testing (with known ci-vector) an ADF test indicates cointegration when the ADF null hypothesis

$$(93) \quad H_0 : u_t \sim I(1)$$

is *rejected*, where  $u_t = y_t - \beta x_t$ .

Example: (Oil spot and futures continued.)



Both the spot and the future prices look integrated, whereas their difference is clearly mean-reverting. Indeed, the ADF-test below confirms that the spot and future prices are cointegrated with cointegrating vector  $\beta = (1, -1)$ , since  $(1, -1)(f_t, s_t)' \sim I(0)$ :

Augmented Dickey-Fuller Unit Root Test on SPREAD

Null Hypothesis: SPREAD has a unit root				
Exogenous: Constant				
Lag Length: 0 (Automatic - based on SIC, maxlag=19)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-23.37333	0.0000
Test critical values:	1% level		-3.439881	
	5% level		-2.865637	
	10% level		-2.569009	
*Mackinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(SPREAD)				
Method: Least Squares				
Date: 03/04/13 Time: 11:37				
Sample (adjusted): 6/25/1996 2/26/1999				
Included observations: 670 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
SPREAD(-1)	-0.900534	0.038528	-23.37333	0.0000
C	-0.002982	0.007077	-0.421379	0.6736
R-squared	0.449895	Mean dependent var		0.000381
Adjusted R-squared	0.449071	S.D. dependent var		0.246763
S.E. of regression	0.183158	Akaike info criterion		-0.553949
Sum squared resid	22.40941	Schwarz criterion		-0.540495
Log likelihood	187.5730	Hannan-Quinn criter.		-0.548738
F-statistic	546.3127	Durbin-Watson stat		2.008427
Prob(F-statistic)	0.000000			



(b) Unknown cointegration relation (Engle-Granger method)

When the cointegration parameter  $\beta$  is unknown, then cointegration testing can be done by running the cointegration regression

$$(94) \quad y_t = \beta_0 + \beta_1 x_t + u_t$$

where  $\beta_0$  and  $\beta_1$  are estimated from the data. If  $u_t$  is stationary, then  $x_t$  and  $y_t$  are cointegrated with cointegration parameter  $\beta_1$ .

Example: (Oil spot and futures continued.)

Dependent Variable: FUTURE Method: Least Squares Date: 03/04/13 Time: 13:10 Sample: 6/24/1996 2/26/1999 Included observations: 671				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.066100	0.032051	2.062342	0.0396
SPOT	0.996215	0.001701	585.5426	0.0000
R-squared	0.998053	Mean dependent var		18.37030
Adjusted R-squared	0.998050	S.D. dependent var		4.149906
S.E. of regression	0.183271	Akaike info criterion		-0.552725
Sum squared resid	22.47055	Schwarz criterion		-0.539286
Log likelihood	187.4393	Hannan-Quinn criter.		-0.547520
F-statistic	342860.2	Durbin-Watson stat		1.801523
Prob(F-statistic)	0.000000			

$\beta_1 \approx 1$  and the residuals are stationary (not shown), implying cointegration with cointegration vector  $(1, -0.9962)$ .

## *Cointegration and Error Correction*

Consider two  $I(1)$  variables  $x_1$  and  $x_2$ , for which the equilibrium relationship  $x_1 = \beta x_2$  holds. Now suppose that the equilibrium is currently disturbed,  $x_{1,t} > \beta x_{2,t}$ , say. In that case there are three possibilities to restore equilibrium:

1. a decrease in  $x_1$  and/or an increase in  $x_2$ ,
2. an increase in  $x_1$  but a larger increase in  $x_2$ ,
3. a decrease in  $x_2$  but a larger decrease in  $x_1$ .

Such a dynamic may be modelled in an error correction model as follows:

$$\begin{aligned}\Delta x_{1,t} &= -\alpha_1(x_{1,t-1} - \beta x_{2,t-1}) + \epsilon_{1,t}, \quad \alpha_1 > 0 \\ \Delta x_{2,t} &= \alpha_2(x_{1,t-1} - \beta x_{2,t-1}) + \epsilon_{2,t}, \quad \alpha_2 > 0\end{aligned}$$

where  $\epsilon_{1,t}$  and  $\epsilon_{2,t}$  are (possibly correlated) white noise processes and  $\alpha_1$  and  $\alpha_2$  may be interpreted as speed of adjustment parameters to the equilibrium. Note that validity of the error correction model above requires  $x_1, x_2 \sim CI(1, 1)$  with cointegrating vector  $(1, -\beta)$ , since both  $\Delta x_{i,t}$  and  $\epsilon_{i,t}$  are assumed to be stationary!

Nothing about this cointegration requirement changes if we introduce lagged changes into the model:

$$\begin{aligned}\Delta x_{1,t} &= a_{10} - \alpha_1(x_{1,t-1} - \beta x_{2,t-1}) \\ &\quad + \sum_{i=1}^p a_{11}(i)\Delta x_{1,t-i} + \sum_{i=1}^p a_{12}(i)\Delta x_{2,t-i} + \epsilon_{1,t}, \\ \Delta x_{2,t} &= a_{20} + \alpha_2(x_{1,t-1} - \beta x_{2,t-1}) \\ &\quad + \sum_{i=1}^p a_{21}(i)\Delta x_{1,t-i} + \sum_{i=1}^p a_{22}(i)\Delta x_{2,t-i} + \epsilon_{2,t}.\end{aligned}$$

This is because  $\epsilon_{i,t}$  and all terms involving  $\Delta x_{1,t}$  and  $\Delta x_{2,t}$  are stationary.

The result, that an error-correction representation implies cointegrated variables, may be generalized to  $n$  variables as follows. Formally the  $I(1)$  vector  $x_t = (x_{1t}, \dots, x_{nt})'$  is said to have an error-correction representation if it may be expressed as

$$\Delta x_t = \pi_0 + \pi x_{t-1} + \sum_{i=1}^p \pi_i \Delta x_{t-i} + \epsilon_t$$

where  $\pi_0$  is a  $(n \times 1)$  vector of intercept terms,  $\pi$  is a  $(n \times n)$  matrix not equal to zero,  $\pi_i$  are  $(n \times n)$  coefficient matrices and  $\epsilon_t$  is a  $(n \times 1)$  vector of possibly correlated white noise.

Then the stationarity of  $\Delta x_{t-i}$ ,  $i = 0, 1, \dots, p$  and  $\epsilon_t$  implies that

$$\pi x_{t-1} = \Delta x_t - \pi_0 - \sum_{i=1}^p \pi_i \Delta x_{t-i} - \epsilon_t$$

is stationary with the rows of  $\pi$  as cointegrating vectors!

It can also be shown that any cointegration relationship implies the existence of an error-correction model. The equivalence of cointegration and error-correction is summarized in Granger's representation theorem:

Let  $x_t$  be a difference stationary vector process. Then  $x_t \sim C(1, 1)$  if and only if there exists an error-correction representation of  $x_t$ :

$$\Delta x_t = \pi_0 + \pi x_{t-1} + \sum_{i=1}^p \pi_i \Delta x_{t-i} + \epsilon_t, \quad \pi \neq 0$$

such that  $\pi x_t \sim I(0)$ .

Note that the  $(n \times n)$  matrix  $\pi$  in the error-correction representation may be decomposed into two  $(n \times r)$  matrices  $\alpha$  and  $\beta$  as  $\pi = \alpha\beta'$ , where  $\beta'$  contains the cointegrating (row) vectors,  $\alpha$  contains the (column) vectors of speed of adjustment parameters to the respective equilibria, and  $r \leq n$  is the rank of  $\pi$ .

Example: For our two-component error correction model we had

$$\Delta x_t = \begin{pmatrix} \Delta x_{1,t} \\ \Delta x_{2,t} \end{pmatrix} = \begin{pmatrix} -\alpha_1 & \alpha_1\beta \\ \alpha_2 & -\alpha_2\beta \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}$$

$$= \pi x_{t-1} + \epsilon_t$$

$$\text{with } \pi = \begin{pmatrix} -\alpha_1 & \alpha_1\beta \\ \alpha_2 & -\alpha_2\beta \end{pmatrix} = \begin{pmatrix} -\alpha_1 \\ \alpha_2 \end{pmatrix} (1 \quad -\beta)$$

and  $\text{rank}(\pi) = 1$ , since the second row is  $-\frac{\alpha_2}{\alpha_1}$  times the first row and the second column is  $-\beta$  times the first column, so there is only 1 linearly independent vector involved.

## Error Correction and VAR

Consider again a multivariate difference stationary series  $y_t = (y_{1t}, \dots, y_{nt})'$ . It has been mentioned earlier, that modelling  $\Delta y_t$  in a vector autoregression model is inappropriate if  $y_t$  is cointegrated. In order to see this important point, assume that  $y_t$  follows a VAR( $p$ ) in levels:

$$y_t = \mu + \sum_{i=1}^p \Phi_i y_{t-i} + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \Sigma).$$

We shall now show that it is always possible to rewrite the VAR in levels as a vector error correction model for the first differences. For that purpose, introduce  $\pi := \sum_{i=1}^p \Phi_i - \mathbf{I}_m$ , such that

$$\Delta y_t = \mu + \pi y_{t-1} + \sum_{i=1}^p \Phi_i (y_{t-i} - y_{t-1}) + \epsilon_t.$$

Now, note that

$$\begin{aligned} y_{t-1} - y_{t-i} &= (y_{t-1} - y_{t-2}) + (y_{t-2} - y_{t-3}) + \dots + (y_{t-i+1} - y_{t-i}) \\ &= \sum_{j=1}^{i-1} \Delta y_{t-j} \end{aligned}$$

such that

$$\begin{aligned} \sum_{i=1}^p \Phi_i (y_{t-i} - y_{t-1}) &= - \sum_{i=1}^p \Phi_i (y_{t-1} - y_{t-i}) = - \sum_{i=1}^p \Phi_i \sum_{j=1}^{i-1} \Delta y_{t-j} \\ &= - \Phi_2 \Delta y_{t-1} - \Phi_3 (\Delta y_{t-1} + \Delta y_{t-2}) - \dots - \Phi_p \sum_{j=1}^{p-1} \Delta y_{t-j} \\ &= \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} \quad \text{where } \Gamma_i = - \sum_{j=i+1}^p \Phi_j. \end{aligned}$$

Therefore,

$$\Delta y_t = \mu + \pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \epsilon_t.$$

Comparing this with an ordinary VAR in differences,

$$\Delta y_t = \mu + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \epsilon_t$$

we notice that such a VAR in differences is misspecified (by leaving out the explanatory variable  $y_{t-1}$ ) whenever  $\pi \neq 0$ , which is exactly what is required for  $y_t$  being cointegrated. Intuitively, for cointegrated series, the term  $\pi y_{t-1}$  is needed in order to model how far the system is out of equilibrium.

## Cointegration and Rank

For notational convenience, consider the simple error correction model

$$\Delta y_t = \pi y_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \Sigma)$$

where  $y_t = (y_{1t}, \dots, y_{nt})'$  as before.

We shall show in the following that we can use the rank of  $\pi$  in order to determine whether  $y_t$  is cointegrated. More precisely, *the number of cointegrating relationships, or cointegrating vectors, is given by the rank of  $\pi$* . There are 3 cases.

1.  $\text{rank}(\pi) = 0$  which implies  $\pi = 0$ .

Therefore the model reduces to  $\Delta y_t = \epsilon_t$ , that is all  $y_{it} \sim I(1)$  since  $\Delta y_t = \epsilon_t \sim I(0)$ , and there is no linear combination of the  $y_{ti}$ 's which is stationary because all vectors  $\beta$  with the property  $\beta y_t \sim I(0)$  have zero entries everywhere. So all components of  $y_t$  are unit root processes and  $y_t$  is not cointegrated.



2.  $\text{rank}(\pi) = r$  with  $1 \leq r < n$ .

Consider first the case  $\text{rank}(\pi) = 1$ , that is, there is only one linearly independent row in  $\pi$ , which implies that all rows of  $\pi$  can be written as scalar multiples of the first. Thus, each of the  $\{\Delta y_{it}\}$  sequences can be written as

$$\Delta y_{it} = \frac{\pi_{ij}}{\pi_{1j}} (\pi_{11}y_{1,t-1} + \pi_{12}y_{2,t-1} + \dots + \pi_{1n}y_{n,t-1}) + \epsilon_{it}.$$

Hence, the linear combination

$$(\pi_{11}y_{1,t-1} + \pi_{12}y_{2,t-1} + \dots + \pi_{1n}y_{n,t-1}) = \frac{\pi_{1j}}{\pi_{ij}} (\Delta y_{it} - \epsilon_{it})$$

is stationary, since both  $\Delta y_{it}$  and  $\epsilon_{it}$  are stationary. So each row of  $\pi$  may be regarded as cointegrating vector of the same cointegrating relationship.

Similarly, if  $\text{rank}(\pi) = r$ , each row may be written as a linear combination of  $r$  linearly independent combinations of the  $\{y_{it}\}$  sequences that are stationary. That is, there are  $r$  cointegrating relationships (cointegrating vectors).

3.  $\text{rank}(\pi) = n \Rightarrow$  the inverse matrix  $\pi^{-1}$  exists. Premultiplying the error correction model with  $\pi^{-1}$  yields then

$$\pi^{-1} \Delta y_t = y_{t-1} + \pi^{-1} \epsilon_t$$

such that all components of  $y_t$  are stationary, since both  $\pi^{-1} \Delta y_t$  and  $\pi^{-1} \epsilon_t$  are stationary. In particular,  $y_t$  is not cointegrated.

## *Johansen's Cointegration tests*

Recall from introductory courses in matrix algebra that the rank of a matrix equals the number of its nonzero eigenvalues, also called characteristic roots. Johansen's (1988) test procedure exploits this relationship for identifying the number of cointegrating relations between non-stationary variables by testing for the number of significantly nonzero eigenvalues of the  $(m \times m)$  matrix  $\pi$  in

$$\Delta x_t = \pi_0 + \pi x_{t-1} + \sum_{i=1}^p \pi_i \Delta x_{t-i} + \epsilon_t.$$

Specifically, the Johansen cointegration test statistics are

1.  $\lambda_{\text{trace}}(r) = -T \sum_{i=1}^m \log(1 - \hat{\lambda}_i)$ , and
2.  $\lambda_{\text{max}}(r, r + 1) = -T \log(1 - \hat{\lambda}_{r+1})$ ,

referred to as *trace statistics* and *maximum eigenvalue statistics*, where  $T$  is the number of usable observations and  $\hat{\lambda}_i$  are the estimated characteristic roots obtained from the estimated  $\pi$  matrix in decreasing order.

The first test statistic

$$\lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^m \log(1 - \hat{\lambda}_i)$$

tests the null hypothesis of less or equal to  $r$  distinct cointegrating vectors against the alternative of  $m$  cointegrating relations, that is a stationary VAR in levels. Note that  $\lambda_{\text{trace}}$  equals zero when all  $\lambda_i = 0$ . The further the estimated characteristic roots are from zero, the more negative is  $\log(1 - \hat{\lambda}_i)$  and the larger is  $\lambda_{\text{trace}}$ .

The second test statistic

$$\begin{aligned} \lambda_{\text{max}}(r, r + 1) &= -T \log(1 - \hat{\lambda}_{r+1}) \\ &= \lambda_{\text{trace}}(r) - \lambda_{\text{trace}}(r + 1) \end{aligned}$$

tests the null of  $r$  cointegrating vectors against the alternative of  $r + 1$  cointegrating vectors. Again  $\lambda_{\text{max}}$  will be small if  $\hat{\lambda}_{r+1}$  is small.

Critical values of both the  $\lambda_{\text{trace}}$  and  $\lambda_{\text{max}}$  statistics are obtained numerically via Monte Carlo simulations.

## Johansens cointegration tests in EViews

Johansens cointegration tests have, contrary to the cointegration tests discussed earlier, the advantage that they are able to identify more than just a single cointegration relationship, which may happen when more than just two series are involved.

In order to perform Johansens cointegration tests in EViews, first set up a VAR in levels for the series you suspect to be cointegrated, and choose then View/Cointegration Test...

You may add exogenous variables which you believe should be subtracted before the linear combination of series becomes stationary.

You should include one lag less than the number of lags you've chosen in setting up your VAR in levels, because the 'Lag Intervals' field in the cointegration test procedure of EViews refers to differences rather than levels.

Given that  $x_t \sim I(1)$  and  $y_t \sim I(1)$ ,

then in the testing procedure there are six different options:

1) No intercept or trend in CE or VAR series:

$$x_t = \text{lags}(x_t, y_t) + e_{xt}, \quad e_{xt} \sim I(0)$$

$$y_t = \text{lags}(x_t, y_t) + e_{yt}, \quad e_{yt} \sim I(0)$$

$$y_t = \beta x_t + u_t$$

Use this only if you are sure that there is no trend and all series have zero mean.

2) Intercept in CE – no intercept in VAR:

$$x_t = \text{lags}(x_t, y_t) + e_{xt}, \quad e_{xt} \sim I(0)$$

$$y_t = \text{lags}(x_t, y_t) + e_{yt}, \quad e_{yt} \sim I(0)$$

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

Use this only if you are sure that there is no trend in any of the series.

3) Intercept in CE and in VAR

$$x_t = \mu_x + \text{lags}(x_t, y_t) + e_{xt}, \quad e_{xt} \sim I(0)$$

$$y_t = \mu_y + \text{lags}(x_t, y_t) + e_{yt}, \quad e_{yt} \sim I(0)$$

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

This is the most common option in empirical work and the default choice in EViews. It allows for both stochastic and deterministic trends in the series.

4) Intercept and trend in CE—only intercept in VAR

$$x_t = \mu_x + \text{lags}(x_t, y_t) + e_{xt}, \quad e_{xt} \sim I(0)$$

$$y_t = \mu_y + \text{lags}(x_t, y_t) + e_{yt}, \quad e_{yt} \sim I(0)$$

$$y_t = \beta_0 + \delta t + \beta x_t + u_t$$

5) Intercept and trend in both CE and VAR

$$x_t = \mu_x + \delta_x t + \text{lags}(x_t, y_t) + e_{xt}, \quad e_{xt} \sim I(0)$$

$$y_t = \mu_y + \delta_y t + \text{lags}(x_t, y_t) + e_{yt}, \quad e_{yt} \sim I(0)$$

$$y_t = \beta_0 + \delta t + \beta x_t + u_t$$

Both options 4 and 5 extend our discussion of cointegration to the situation that a deterministic trend must be subtracted from the linear combination of the  $x_t$  and  $y_t$  series before it becomes stationary. We shall not discuss them further in this course.

EViews has also an option 6, which is just a summary overview of the 5 trend assumptions above, which may be used for an assessment how robust your findings are to different trend assumptions.

Example: (Oil spot and futures continued.)

We consider now logarithmic spot and future prices, such that their differences become log returns, and our results become comparable to those we had earlier when setting up a VAR in logreturns.

We choose a VAR(3) model because 3 lags are suggested by the lag length criteria and its residuals reasonably pass the Portmanteau autocorrelation tests (not shown).

Applying the cointegration tests in EViews using option 3 (both series obviously have a trend) including 2 lags in differences we get the output on the next slide, from which we infer:

1. There is one cointegrating relationship.
2.  $\hat{\beta} = (1, -1.000724)$ .
3.  $\hat{\alpha} = \begin{pmatrix} 0.070491 \\ 0.738774 \end{pmatrix}$ .

Johansen Cointegration Test

Date: 03/05/13 Time: 10:11				
Sample: 6/24/1996 2/26/1999				
Included observations: 671				
Trend assumption: Linear deterministic trend				
Series: LOG(SPOT) LOG(FUTURE)				
Lags interval (in first differences): 1 to 2				
Unrestricted Cointegration Rank Test (Trace)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.199867	150.0553	15.49471	0.0001
At most 1	0.000651	0.437196	3.841466	0.5085
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				
Unrestricted Cointegration Rank Test (Maximum Eigenvalue)				
Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.199867	149.6181	14.26460	0.0001
At most 1	0.000651	0.437196	3.841466	0.5085
Max-eigenvalue 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				
Unrestricted Cointegrating Coefficients (normalized by b*S11*b=I):				
LOG(SPOT)	LOG(FUTURE)			
-195.2418	195.3831			
-4.542973	8.831278			
Unrestricted Adjustment Coefficients (alpha):				
D(LOG(SPOT))	-0.000361	-0.000599		
D(LOG(FUTUR	-0.003784	-0.000401		
1 Cointegrating Equation(s): Log likelihood 3934.610				
Normalized cointegrating coefficients (standard error in parentheses)				
LOG(SPOT)	LOG(FUTURE)			
1.000000	-1.000724			
	(0.00170)			
Adjustment coefficients (standard error in parentheses)				
D(LOG(SPOT))	0.070491			
	(0.17774)			
D(LOG(FUTUR	0.738774			
	(0.13198)			



Estimating the VAR again, changing the VAR type from Unrestricted VAR into Vector Error Correction, yields the output on the following slide, from which we infer the error correction model below:

$$\begin{aligned}
 r_{s,t} = & 0.0705(0.002 + s_{t-1} - 1.0007f_{t-1}) \\
 & + 0.2231r_{s,t-1} + 0.1811r_{s,t-2} \\
 & - 0.3395r_{f,t-1} - 0.2108r_{f,t-2} - 0.00084, \\
 r_{f,t} = & 0.7388(0.002 + s_{t-1} - 1.0007f_{t-1}) \\
 & + 0.4165r_{s,t-1} + 0.2493r_{s,t-2} \\
 & - 0.4683r_{f,t-1} - 0.2552r_{f,t-2} - 0.00077,
 \end{aligned}$$

where  $s$ ,  $f$ ,  $r_s$  and  $r_f$  denote the log prices and log returns in the spot and futures market, respectively.

This implies that our earlier VAR(2) model for spot and future returns was misspecified by omitting the cointegration terms, which also invalidates our earlier analysis on Granger causality and linear dependence on that model. **Hence always test for cointegration between integrated series before trying to fit a VAR in differences!**

Vector Error Correction Estimates

Vector Error Correction Estimates		
Date: 03/05/13 Time: 10:58		
Sample: 6/24/1996 2/26/1999		
Included observations: 671		
Standard errors in ( ) & t-statistics in [ ]		
Cointegrating Eq:	CointEq1	
LOG(SPOT(-1))	1.000000	
LOG(FUTURE(-1))	-1.000724 (0.00170) [-587.831]	
C	0.002015	
Error Correction:	D(LOG(SPOT)	D(LOG(FUTU
CointEq1	0.070491 (0.17774) [ 0.39660]	0.738774 (0.13198) [ 5.59773]
D(LOG(SPOT(-1)))	0.223140 (0.15453) [ 1.44401]	0.416518 (0.11474) [ 3.63003]
D(LOG(SPOT(-2)))	0.181110 (0.10538) [ 1.71869]	0.249335 (0.07825) [ 3.18654]
D(LOG(FUTURE(-1)))	-0.339529 (0.14981) [-2.26641]	-0.468297 (0.11124) [-4.20984]
D(LOG(FUTURE(-2)))	-0.210752 (0.08875) [-2.37453]	-0.255195 (0.06590) [-3.87225]
C	-0.000844 (0.00091) [-0.92553]	-0.000766 (0.00068) [-1.13025]
R-squared	0.028086	0.319991
Adj. R-squared	0.020778	0.314878
Sum sq. resids	0.369797	0.203891
S.E. equation	0.023581	0.017510
F-statistic	3.843330	62.58562
Log likelihood	1565.340	1765.087
Akaike AIC	-4.647808	-5.243179
Schwarz SC	-4.607492	-5.202862
Mean dependent	-0.000725	-0.000708
S.D. dependent	0.023830	0.021155
Determinant resid covariance (dof adj.)	2.82E-08	
Determinant resid covariance	2.77E-08	
Log likelihood	3934.610	
Akaike information criterion	-11.68587	
Schwarz criterion	-11.59180	

## Testing Hypotheses

Recalling the cointegration vector estimate  $\hat{\beta} = (1, -1.000724)$  and the speed of adjustment coefficients  $\hat{\alpha}_1 = 0.070491$  and  $\hat{\alpha}_2 = 0.738774$  it appears that the true cointegration vector might well be  $\beta = (1, -1)$  and the speed of adjustment parameter for the spot return  $\alpha_1 = 0$ .

These hypotheses can be tested. To do so, enter the following VEC Coefficient Restrictions in the 'VEC Restrictions' tab of the Cointegration Test View:

$$A(1,1)=0, \quad B(1,1)=1, \quad B(1,2)=-1.$$

This extends the cointegration test output by the LR-test presented on the next slide, from which we infer, that this set of restrictions cannot be rejected ( $p = 0.85$ ), and a new estimate for the speed of adjustment parameter for the future return of  $\hat{\alpha}_2 = 0.689698$ .

Johansen Cointegration Test

Date: 03/18/13 Time: 08:34 Sample: 6/24/1996 2/26/1999 Included observations: 671 Trend assumption: Linear deterministic trend Series: LOG(SPOT) LOG(FUTURE) Lags interval (in first differences): 1 to 2				
Unrestricted Cointegration Rank Test (Trace)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.199867	150.0553	15.49471	0.0001
At most 1	0.000651	0.437196	3.841466	0.5085
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				
Unrestricted Cointegration Rank Test (Maximum Eigenvalue)				
Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.199867	149.6181	14.26460	0.0001
At most 1	0.000651	0.437196	3.841466	0.5085
Max-eigenvalue 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				
Restrictions:				
A(1,1)=0, B(1,1)=1, B(1,2)=-1				
Tests of cointegration restrictions:				
Hypothesized No. of CE(s)	Restricted Log-likelihood	LR Statistic	Degrees of Freedom	Probability
1	3934.449	0.322756	2	0.850970
1 Cointegrating Equation(s): Convergence achieved after 1 iterations.				
Restricted cointegrating coefficients (standard error in parentheses) LOG(SPOT) LOG(FUTURE) 1.000000 -1.000000 (0.00000) (0.00000)				
Adjustment coefficients (standard error in parentheses) D(LOG(SPOT)) 0.000000 (0.00000) D(LOG(FUTUR 0.689698 (0.05357)				

Estimating the Vector Error Correction model under these restrictions yields the output on the following slide, from which we infer:

$$\begin{aligned}r_{s,t} &= -0.00084 \\ &\quad + 0.2261r_{s,t-1} + 0.1824r_{s,t-2} \\ &\quad - 0.3420r_{f,t-1} - 0.2115r_{f,t-2}, \\ r_{f,t} &= -0.00077 + 0.6897(s_{t-1} - f_{t-1}) \\ &\quad + 0.4197r_{s,t-1} + 0.2509r_{s,t-2} \\ &\quad - 0.4713r_{f,t-1} - 0.2505r_{f,t-2},\end{aligned}$$

where we have rounded the constant  $7 \cdot 10^{-5}$  in the cointegration equation down to zero.

Comparing this with the output for the unrestricted VAR we note that due to the restriction  $\alpha_1 = 0$  the spot return does not react in any way to disturbances in the equilibrium between spot and future prices.

When that happens to a variable, then we say that this variable (here the spot return) is weakly exogeneous with respect to the  $\beta$  parameters.

Vector Error Correction Estimates

Vector Error Correction Estimates		
Date: 03/18/13 Time: 08:39		
Sample: 6/24/1996 2/26/1999		
Included observations: 671		
Standard errors in ( ) & t-statistics in [ ]		
Cointegration Restrictions: A(1,1)=0, B(1,1)=1, B(1,2)=-1 Convergence achieved after 1 iterations. Restrictions identify all cointegrating vectors LR test for binding restrictions (rank = 1): Chi-square(2) 0.322756 Probability 0.850970		
Cointegrating Eq:	CointEq1	
LOG(SPOT(-1))	1.000000	
LOG(FUTURE(-1))	-1.000000	
C	-7.26E-05	
Error Correction:	D(LOG(SPOT) D(LOG(FUTU	
CointEq1	0.000000 (0.00000) [ NA]	0.689698 (0.05357) [ 12.8743]
D(LOG(SPOT(-1)))	0.226088 (0.15433) [ 1.46496]	0.419683 (0.11462) [ 3.66164]
D(LOG(SPOT(-2)))	0.182374 (0.10530) [ 1.73189]	0.250855 (0.07821) [ 3.20766]
D(LOG(FUTURE(-1)))	-0.341978 (0.14960) [-2.28596]	-0.471318 (0.11110) [-4.24219]
D(LOG(FUTURE(-2)))	-0.211485 (0.08869) [-2.38445]	-0.256462 (0.06587) [-3.89347]
C	-0.000843 (0.00091) [-0.92474]	-0.000765 (0.00068) [-1.12973]
R-squared	0.028061	0.319733
Adj. R-squared	0.020753	0.314618
Sum sq. resids	0.369807	0.203968
S.E. equation	0.023582	0.017513
F-statistic	3.839858	62.51153
Log likelihood	1565.331	1764.960
Akaike AIC	-4.647783	-5.242800
Schwarz SC	-4.607466	-5.202484
Mean dependent	-0.000725	-0.000708
S.D. dependent	0.023830	0.021155
Determinant resid covariance (dof adj.)	2.82E-08	
Determinant resid covariance	2.77E-08	
Log likelihood	3934.449	
Akaike information criterion	-11.68539	
Schwarz criterion	-11.59132	

## Exogeneity in possibly cointegrated systems

We mentioned earlier that our original results concerning Granger causality in the oil spot and futures market are void because the series are cointegrated. EViews has an option for testing Granger causality in error correction models, but its results are flawed because it takes only the lagged differences (i.e. returns) into account but not the all important lagged levels (i.e. prices).

In order to test for exogeneity in integrated systems, no matter whether cointegrated or not, Toda and Yamamoto (1995) suggest the following simple procedure:

1. Fit a VAR model in levels of order  $p + 1$ , where  $p$  is the minimum order required to render the residuals white noise.
2. Perform a Wald or LR-test in the usual way, however considering only the first  $p$  lags of the series/block being tested for exogeneity.

Example: (Oil spot and futures continued.)

In estimating a VAR model for the spot and future rates themselves (so not the returns) all information criteria and the LR-test suggest a model of order  $p = 3$ , and the residuals of such a VAR turn out to be white noise.

Hence, following Toda and Yamamoto (1995), we regress spot (futures) rates upon a constant and the first four lags of the the series itself and the other series. In testing for exogeneity of futures (spots), we perform a Wald test, by restricting the first three lags of the future (spot) rates to zero. Both tests reject exogeneity of the other series with a  $p\text{-value} < 1\%$ , so there is strong evidence of Granger causality in both directions.

Note: **In cointegrated systems there must be Granger causality in at least one direction** because otherwise there would be no mechanism to restore equilibrium.