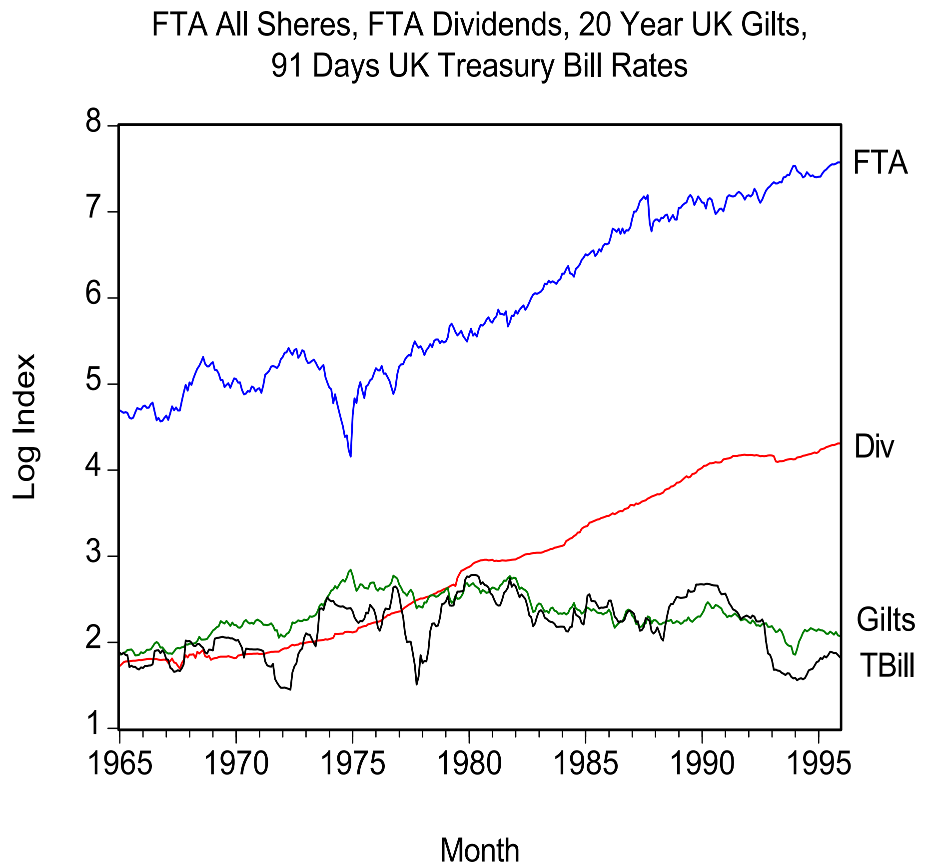
2. Multivariate Time Series

2.1 Background

Example 2.1: Consider the following monthly observations on FTA All Share index, the associated dividend index and the series of 20 year UK gilts and 91 day Treasury bills from January 1965 to December 1995 (372 months)



Potentially interesting questions:

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

2. Are there feedbacks between the markets?

3. How about contemporaneous movements?

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

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

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

Time series models

(i) AR(p)-process

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t \quad (1)$$

or

$$\phi(L)y_t = \mu + \epsilon_t, \quad (2)$$

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

$$\begin{aligned} E(\epsilon_t) &= 0, \\ E(\epsilon_t \epsilon_s) &= \begin{cases} \sigma_\epsilon^2 & \text{if } t = s \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \quad (3)$$

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

$$L^k y_t = y_{t-k} \quad (4)$$

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

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

(5) $$\phi(L) = 0$$

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

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

are less than one in absolute value (are inside the unit circle).

Remark 2.1: Usually the series are centralized such that $\mu = 0$.

(ii) MA(q)-process

(7)
$$\begin{aligned} y_t &= \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \ldots - \theta_q \epsilon_{t-q} \\ &= \mu + \theta(L)\epsilon_t, \end{aligned}$$

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

Remark 2.2: An MA-process is always stationary. But the so called invertibility condition requires that the roots of the characteristic polynomial $\theta(L) = 0$ lie outside the unit circle.

(iii) ARMA(p, q)-process

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

$$\phi(L)y_t = \mu + \theta(L)\epsilon_t. \tag{8}$$

(iv) ARIMA(p, d, q)-process

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

$$(1 - L)^d y_t \sim \text{ARMA}(p, q), \tag{9}$$

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

Remark 2.3: See Definition 1.3 for a precise definition of integrated series.

Example 2.2: Univariate time series models for the above (log) series look as follows. All the series prove to be $I(1)$.

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

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

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

$$\Delta y_t = \rho y_{t-1} + \alpha + \delta t + \sum_{i=1}^{4} \phi_i \Delta y_{t-i} + \epsilon_t, \tag{10}$$

and the hypothesis to be tested is

$$H_0 : \rho = 0 \text{ vs } H_1 : \rho < 0. \tag{11}$$

Test results:

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

2.2 Vector Autoregression (VAR)

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

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

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

In matrix notations

$$(13)\quad \mathbf{y}_t = \mu + \Phi_1 \mathbf{y}_{t-1} + \cdots + \Phi_p \mathbf{y}_{t-p} + \epsilon_t,$$

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

$$(14)\qquad \Phi(L) = \mathbf{I} - \Phi_1 L - \ldots - \Phi_p L^p.$$

Thus finally we get

$$(15)\qquad \Phi(L)\mathbf{y}_t = \epsilon_t.$$

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

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

$$\begin{aligned} \mathbb{E}(\epsilon_t) &= \mathbf{0} \\ \mathbb{E}(\epsilon_t \epsilon_s') &= \begin{cases} \Sigma_\epsilon & \text{if } t = s \\ \mathbf{0} & \text{if } t \neq s \end{cases} \end{aligned} \tag{16}$$

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

$$|\mathbf{I} - \Phi_1 z - \Phi_2 z^2 - \cdots - \Phi_p z^p| = 0 \tag{17}$$

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

Example 2.3: Let us estimate a VAR(1) model for the equity-bond data. First, however, test whether the series are cointegrated. As is seen below, there is no empirical evidence of cointegration (EViews results)

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

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

VAR(2) Estimates:

Sample(adjusted): 1965:04 1995:12
Included observations: 369 after adjusting end points
Standard errors in ( ) & t-statistics in [ ]

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

Continues ...

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

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

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

Defining the order of a VAR-model

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

The underlying assumption is that the residuals follow a multivariate normal distribution, i.e.

$$\epsilon \sim N_m(\mathbf{0}, \Sigma_\epsilon). \tag{18}$$

Akaike's criterion function (AIC) and Schwarz's criterion (BIC) have gained popularity in determination of the number of lags in VAR.

In the original forms AIC and BIC are defined as

$$\text{AIC} = -2 \log L + 2s \tag{19}$$

$$\text{BIC} = -2 \log L + s \log T \tag{20}$$

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

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

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

Under the normality assumption BIC can be simplified to

$$\text{BIC}(j) = \log \left|\widehat{\Sigma}_{\epsilon,j}\right| + \frac{jm^2 \log T}{T}, \tag{21}$$

and AIC to

$$\text{AIC}(j) = \log \left|\widehat{\Sigma}_{\epsilon,j}\right| + \frac{2jm^2}{T}, \tag{22}$$

$j = 0, \ldots, p$, where

$$\widehat{\Sigma}_{\epsilon,j} = \frac{1}{T} \sum_{t=j+1}^{T} \widehat{\epsilon}_{t,j} \widehat{\epsilon}'_{t,j} = \frac{1}{T} \widehat{\mathbf{E}}_j \widehat{\mathbf{E}}'_j \tag{23}$$

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

$$\widehat{\mathbf{E}}_j = \left[\widehat{\epsilon}_{j+1,j}, \widehat{\epsilon}_{j+2,j}, \cdots, \widehat{\epsilon}_{T,j}\right] \tag{24}$$

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

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

$$\text{LR} = T(\log|\hat{\Sigma}_k| - \log|\hat{\Sigma}_p|), \tag{25}$$

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

$$H_0 : \Phi_{k+1} = \cdots = \Phi_p = \mathbf{0} \tag{26}$$

is true, then

$$\text{LR} \sim \chi^2_{df}, \tag{27}$$

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

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

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

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

is used to correct possible small sample bias, where $g = p - k$.

Yet another method is to use sequential LR

$$\text{LR}(k) = (T - m)(\log|\hat{\Sigma}_k| - \log|\hat{\Sigma}_{k+1}|)$$

which tests the null hypothesis

$$H_0 : \Phi_{k+1} = \mathbf{0}$$

given that $\Phi_j = \mathbf{0}$ for $j = k+2, k+3, \ldots, p$.

EViews uses this method in `View -> Lag Structure -> Lag Length Criteria...`.

Example 2.4: Let $p = 12$ then in the equity-bond data different VAR models yield the following results. Below are EViews results.

```
VAR Lag Order Selection Criteria
Endogenous variables: DFTA DDIV DR20 DTBILL
Exogenous variables: C
Sample: 1965:01 1995:12
Included observations: 359
==========================================================
Lag   LogL      LR      FPE       AIC      SC       HQ
----------------------------------------------------------
 0  -3860.59    NA     26324.1   21.530   21.573   21.547
 1  -3810.15  99.473  21728.6*  21.338*  21.554*  21.424*
 2  -3796.62  26.385  22030.2   21.352   21.741   21.506
 3  -3786.22  20.052  22729.9   21.383   21.945   21.606
 4  -3783.57  5.0395  24489.4   21.467   22.193   21.750
 5  -3775.66  14.887  25625.4   21.502   22.411   21.864
 6  -3762.32  24.831  26016.8   21.517   22.598   21.947
 7  -3753.94  15.400  27159.4   21.560   22.814   22.059
 8  -3739.07  27.018* 27348.2   21.566   22.994   22.134
 9  -3731.30  13.933  28656.4   21.612   23.213   22.248
10  -3722.40  15.774  29843.7   21.651   23.425   22.357
11  -3715.54  12.004  31443.1   21.702   23.649   22.476
12  -3707.28  14.257  32880.6   21.745   23.865   22.588
==========================================================
 * indicates lag order selected by the criterion
 LR: sequential modified LR test statistic
     (each test at 5% level)
 FPE: Final prediction error
 AIC: Akaike information criterion
 SC: Schwarz information criterion
 HQ: Hannan-Quinn information criterion
```

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

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

$$H_0 : \Upsilon_1 = \cdots = \Upsilon_h = 0 \tag{28}$$

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

$$Q_h = T \sum_{k=1}^{h} \mathrm{tr}(\hat{\Upsilon}_k' \hat{\Upsilon}_0^{-1} \hat{\Upsilon}_k \hat{\Upsilon}_0^{-1}), \tag{29}$$

where $\hat{\Upsilon}_k = (\hat{\rho}_{ij}(k))$ are the estimated (residual) autocorrelations, and $\hat{\Upsilon}_0$ the contemporaneous correlations of the residuals.

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

Alternatively (especially in small samples) a modified statistic is used

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

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

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

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

```
VAR Residual Portmanteau Tests for Autocorrelations
H0: no residual autocorrelations up to lag h
Sample: 1965:01 1995:12
Included observations: 369
```

Lags	Q-Stat	Prob.	Adj Q-Stat	Prob.	df
1	0.438464	NA*	0.439655	NA*	NA*
2	1.623778	NA*	1.631428	NA*	NA*
3	17.13353	0.3770	17.26832	0.3684	16
4	27.07272	0.7143	27.31642	0.7027	32
5	44.01332	0.6369	44.48973	0.6175	48
6	66.24485	0.3994	67.08872	0.3717	64
7	80.51861	0.4627	81.63849	0.4281	80
8	104.3903	0.2622	106.0392	0.2271	96
9	121.8202	0.2476	123.9049	0.2081	112
10	136.8909	0.2794	139.3953	0.2316	128
11	147.3028	0.4081	150.1271	0.3463	144
12	157.4354	0.5425	160.6003	0.4718	160

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

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

Vector ARMA (VARMA)

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

$$\mathbf{y}_t = \mu + \sum_{i=1}^{p} \Phi_i \mathbf{y}_{t-i} + \epsilon_t - \sum_{j=1}^{q} \Theta_j \epsilon_{t-j} \tag{31}$$

or

$$\Phi(L)\mathbf{y}_t = \mu + \Theta(L)\epsilon_t, \tag{32}$$

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

$$\begin{aligned} \Phi(L) &= \mathbf{I} - \Phi_1 L - \ldots - \Phi_p L^p \\ \Theta(L) &= \mathbf{I} - \Theta_1 L - \ldots - \Theta_q L^q. \end{aligned} \tag{33}$$

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

Autocorrelation and Autocovariance Matrices

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

$$(34) \quad \gamma_{ij}(k) = E(y_{it-k} - \mu_i)(y_{jt} - \mu_j).$$

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

$$(35) \qquad \rho_{i,j}(k) = \frac{\gamma_{ij}(k)}{\sqrt{\gamma_i(0)\gamma_j(0)}}.$$

The kth *autocorrelation matrix* is then

$$(36) \quad \Upsilon_k = \begin{pmatrix} \rho_1(k) & \rho_{1,2}(k) & \dots & \rho_{1,m}(k) \\ \rho_{2,1}(k) & \rho_2(k) & \dots & \rho_{2,m}(k) \\ \vdots & & \ddots & \vdots \\ \rho_{m,1}(k) & \rho_{m,2}(k) & \dots & \rho_m(k) \end{pmatrix}.$$

Remark 2.4: $\Upsilon_k = \Upsilon'_{-k}$.

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

Example 2.5: Cross autocorrelations of the equity-bond data.

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

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

Example 2.6: Consider individual security returns and portfolio returns. For example French and Roll[‡‡] have found that daily returns of individual securities are slightly negatively correlated. The tables below, however suggest that daily returns of portfolios tend to be positively correlated!

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

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

$$r_{pt} = \frac{1}{m}\sum_{i=1}^{m} r_{it} = \frac{1}{m}\iota'\mathbf{r}_t \tag{37}$$

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

$$\text{cov}(r_{pt-1}, r_{pt}) = \text{cov}\left[\frac{\iota'\mathbf{r}_{t-1}}{m}, \frac{\iota'\mathbf{r}_t}{m}\right] = \frac{\iota'\Gamma_1\iota}{m^2}, \tag{38}$$

where Γ_1 is the first order autocovariance matrix.

Therefore

$$m^2 \text{cov}(r_{pt-1}, r_{pt}) = \boldsymbol{\iota}' \Gamma_1 \boldsymbol{\iota} = \left(\boldsymbol{\iota}' \Gamma_1 \boldsymbol{\iota} - \text{tr}(\Gamma_1)\right) + \text{tr}(\Gamma_1), \quad (39)$$

where $\text{tr}(\cdot)$ is the trace operator which sums the diagonal elements of a square matrix.

Consequently, because the right hand side tends to be positive and $\text{tr}(\Gamma_1)$ tends to be negative, $\boldsymbol{\iota}' \Gamma_1 \boldsymbol{\iota} - \text{tr}(\Gamma_1)$, which contains only cross autocovariances, must be positive, and larger than the absolute value of $\text{tr}(\Gamma_1)$, the autocovariances of individual stock returns.

2.3 Exogeneity and Causality

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

(40) $$\mathbf{y}_t = \mathbf{c} + \sum_{i=1}^{p} \mathbf{A}_i' \mathbf{y}_{t-i} + \sum_{i=0}^{p} \mathbf{B}_i' \mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t,$$

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

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

and

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

(42)

where

$$\mathbf{Y}_{t-1} = (\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{y}_1). \tag{43}$$

We can compile this into a matrix form

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{U}, \tag{44}$$

where

$$\begin{aligned} \mathbf{Y} &= (\mathbf{y}_{p+1}, \ldots, \mathbf{y}_T)' \\ \mathbf{X} &= (\mathbf{z}_{p+1} \cdot \ldots, \mathbf{z}_T)' \\ \mathbf{z}_t &= (1, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p}, \mathbf{x}'_t, \ldots, \mathbf{x}'_{t-p})' \\ \mathbf{U} &= (\boldsymbol{\epsilon}_{p+1}, \ldots, \boldsymbol{\epsilon}_T)', \end{aligned} \tag{45}$$

and

$$\mathbf{B} = (\mathbf{c}, \mathbf{A}'_1, \ldots, \mathbf{A}'_p, \mathbf{B}'_0, \ldots, \mathbf{B}'_p)'. \tag{46}$$

The estimation theory for this model is basically the same as for the univariate linear regression.

For example, the LS and (approximate) ML estimator of $\mathbf{B}$ is

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

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

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

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

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

*For an in depth discussion of *Exogeneity*, see Engle, R.F., D.F. Hendry and J.F. Richard (1983). Exogeneity. *Econometrica*, **51:2**, 277–304.

Granger-causality and measures of feedback

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

If the history of x does not help to predict the future values of y, we say that x does not Granger-cause y.*

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

Usually the prediction ability is measured in terms of the MSE (Mean Square Error).

Thus, x fails to Granger-cause y, if for all $s > 0$

$$\begin{aligned} &\text{MSE}(\widehat{y}_{t+s}|y_t, y_{t-1}, \ldots) \\ &\quad = \text{MSE}(\widehat{y}_{t+s}|y_t, y_{t-1}, \ldots, x_t, x_{t-1}, \ldots), \end{aligned} \tag{49}$$

where (e.g.)

$$\begin{aligned} &\text{MSE}(\widehat{y}_{t+s}|y_t, y_{t-1}, \ldots) \\ &\quad = E\left((y_{t+s} - \widehat{y}_{t+s})^2 | y_t, y_{t-1}, \ldots\right). \end{aligned} \tag{50}$$

Remark 2.5: This is essentially the same that y is strongly exogenous to x.

In terms of VAR models this can be expressed as follows:

Consider the $g = m + k$ dimensional vector $\mathbf{z}'_t = (\mathbf{y}'_t, \mathbf{x}'_t)$, which is assumed to follow a VAR(p) model

$$\mathbf{z}_t = \sum_{i=1}^{p} \Pi_i \mathbf{z}_{t-i} + \boldsymbol{\nu}_t, \tag{51}$$

where

$$\begin{aligned} E(\boldsymbol{\nu}_t) &= \mathbf{0} \\ E(\boldsymbol{\nu}_t \boldsymbol{\nu}'_s) &= \begin{cases} \Sigma_\nu, & t = s \\ \mathbf{0}, & t \neq s. \end{cases} \end{aligned} \tag{52}$$

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

$$\begin{aligned} \mathbf{y}_t &= \textstyle\sum_{i=1}^{p} \mathbf{C}_{2i}\mathbf{x}_{t-i} + \sum_{i=1}^{p} \mathbf{D}_{2i}\mathbf{y}_{t-i} + \boldsymbol{\nu}_{1t} \\ \mathbf{x}_t &= \textstyle\sum_{i=1}^{p} \mathbf{E}_{2i}\mathbf{x}_{t-i} + \sum_{i=1}^{p} \mathbf{F}_{2i}\mathbf{y}_{t-i} + \boldsymbol{\nu}_{2t} \end{aligned} \tag{53}$$

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

$$\Sigma_\nu = \begin{pmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{54}$$

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

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 $\Sigma_1 = E(\boldsymbol{\eta}_{1t}\boldsymbol{\eta}'_{1t})$ with $\boldsymbol{\eta}_{1t}$ from the regression

$$\mathbf{y}_t = \sum_{i=1}^{p} \mathbf{C}_{1i}\mathbf{y}_{t-i} + \boldsymbol{\eta}_{1t}. \tag{55}$$

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

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

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

$$H_0 \;:\; \mathbf{C}_{2i} = \mathbf{0}. \tag{56}$$

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

Unrestricted:

$$\hat{\Sigma}_{11} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\nu}_{1t} \hat{\nu}'_{1t}. \tag{57}$$

Restricted:

$$\hat{\Sigma}_{1} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\eta}_{1t} \hat{\eta}'_{1t}. \tag{58}$$

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

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

The LR test is then

$$\text{(59)} \quad \text{LR} = (T-p)\left(\ln|\hat{\Sigma}_1| - \ln|\hat{\Sigma}_{11}|\right) \sim \chi^2_{mkp},$$

if H_0 is true.

Example 2.7: Granger causality between pairwise equity-bond market series

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

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

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

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

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

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

2.4 Geweke's*measures of Linear Dependence

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

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

$$F_{\mathbf{x}\to\mathbf{y}} = \ln(|\Sigma_1|/|\Sigma_{11}|), \tag{60}$$

so that the statement that "$\mathbf{x}$ does not (Granger) cause $\mathbf{y}$" is equivalent to $F_{\mathbf{x}\to\mathbf{y}} = 0$.

Similarly the *measure of linear feedback from* $\mathbf{y}$ *to* $\mathbf{x}$ is defined by

$$F_{\mathbf{y}\to\mathbf{x}} = \ln(|\Sigma_2|/|\Sigma_{22}|). \tag{61}$$

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

It may also be interesting to investigate the *instantaneous causality* between the variables.

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

$$(62) \qquad \begin{pmatrix} \mathbf{I}_m & -\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{12}'\Sigma_{11}^{-1} & \mathbf{I}_k \end{pmatrix}$$

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

$$(63) \mathbf{y}_t = \sum_{i=0}^{p} \mathbf{C}_{3i}\mathbf{x}_{t-i} + \sum_{i=1}^{p} \mathbf{D}_{3i}\mathbf{y}_{t-i} + \omega_{1t},$$

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

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

Similarly, the last k equations can be written as

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

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

Analogously to the linear feedback we can define *instantaneous linear feedback*

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

(65)

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

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

Consequently the linear dependence can be decomposed additively into three forms of feedback.

Absence of a particular causal ordering is then equivalent to one of these feedback measures being zero.

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

$$\hat{\Sigma}_i = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\eta}_{it} \hat{\eta}'_{it}, \tag{67}$$

$$\hat{\Sigma}_{ii} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\nu}_{it} \hat{\nu}'_{it}, \tag{68}$$

$$\hat{\Sigma}_{\omega i} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\omega}_{it} \hat{\omega}'_{it}, \tag{69}$$

for $i = 1, 2$. For example

$$\hat{F}_{\mathbf{x}\to\mathbf{y}} = \ln(|\hat{\Sigma}_1|/|\hat{\Sigma}_{11}|). \tag{70}$$

With these estimates one can test the particular dependencies,

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

$$(T-p)\hat{F}_{\mathbf{x}\to\mathbf{y}} \sim \chi^2_{mkp}. \tag{71}$$

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

$$(T-p)\hat{F}_{\mathbf{y}\to\mathbf{x}} \sim \chi^2_{mkp}. \tag{72}$$

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

$$(T-p)\hat{F}_{\mathbf{x}\cdot\mathbf{y}} \sim \chi^2_{mk}. \tag{73}$$

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

$$(T-p)\hat{F}_{\mathbf{x},\mathbf{y}} \sim \chi^2_{mk(2p+1)}. \tag{74}$$

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

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

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

Example 2.8: The LR-statistics of the above measures and the associated χ^2 values for the equity-bond data are reported in the following table with $p = 2$.

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

	LR	DF	P-VALUE
x-->y	6.41	8	0.60118
y-->x	20.11	8	0.00994
x.y	23.31	4	0.00011
x,y	49.83	20	0.00023

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

2.5 *Variance decomposition and innovation accounting*

Consider the VAR(p) model

(75)
$$\mathbf{\Phi}(L)\mathbf{y}_t = \boldsymbol{\epsilon}_t,$$

where

(76)
$$\mathbf{\Phi}(L) = \mathbf{I} - \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2 - \cdots - \mathbf{\Phi}_p L^p$$

is the (matric) lag polynomial.

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

(77)
$$\mathbf{y}_t = \mathbf{\Phi}^{-1}(L)\boldsymbol{\epsilon}_t = \boldsymbol{\epsilon}_t + \sum_{i=1}^{\infty} \mathbf{\Psi}_i \boldsymbol{\epsilon}_{t-i},$$

where $\mathbf{\Psi}_i$ is an $m \times m$ coefficient matrix.

The error terms $\boldsymbol{\epsilon}_t$ represent shocks in the system.

Suppose we have a unit shock in $\boldsymbol{\epsilon}_t$ then its effect in $\mathbf{y}$, s periods ahead is

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \boldsymbol{\epsilon}_t} = \boldsymbol{\Psi}_s. \tag{78}$$

Accordingly the interpretation of the $\boldsymbol{\Psi}$ matrices is that they represent marginal effects, or the model's response to a unit shock (or innovation) at time point t in *each* of the variables.

Economists call such parameters are as *dynamic multipliers*.

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

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

(79)

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

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

$$\psi_{ij,1}, \psi_{ij,2}, \psi_{ij,3}, \ldots, \qquad (80)$$

where $\psi_{ij,k}$ is the ijth element of the matrix Ψ_k $(i, j = 1, \ldots, m)$.

Generally an impulse response function traces the effect of a one-time shock to one of the innovations on current and future values of the endogenous variables.

What about exogeneity (or Granger-causality)?

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

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \phi_{11}^{(1)} & 0 \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \cdots + \begin{pmatrix} \phi_{11}^{(p)} & 0 \\ \phi_{21}^{(p)} & \phi_{22}^{(p)} \end{pmatrix} \begin{pmatrix} y_{t-p} \\ x_{t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}. \tag{81}$$

Then under the stationarity condition

$$(\mathbf{I} - \mathbf{\Phi}(L))^{-1} = \mathbf{I} + \sum_{i=1}^{\infty} \mathbf{\Psi}_i L^i, \tag{82}$$

where

$$\mathbf{\Psi}_i = \begin{pmatrix} \psi_{11}^{(i)} & 0 \\ \psi_{21}^{(i)} & \psi_{22}^{(i)} \end{pmatrix}. \tag{83}$$

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

Generally, if a variable, or a block of variables, are strictly exogenous, then the implied zero restrictions ensure that these variables do not react to a shock to any of the endogenous variables.

Nevertheless it is advised to be careful when interpreting the possible (Granger) causalities in the philosophical sense of causality between variables.

Remark 2.6: See also the critique of impulse response analysis in the end of this section.

Orthogonalized impulse response function

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

Example 2.9. For example, in the equity-bond data the contemporaneous VAR(2)-residual correlations are

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

Many times, however, it is of interest to know how "new" information on y_{jt} makes us revise our forecasts on $\mathbf{y}_{t+s}$.

In order to single out the individual effects the residuals must be first orthogonalized, such that they become contemporaneously uncorrelated (they are already serially uncorrelated).

Remark 2.7: If the error terms ($\boldsymbol{\epsilon}_t$) are already contemporaneously uncorrelated, naturally no orthogonalization is needed.

Unfortunately orthogonalization, however, is not unique in the sense that changing the order of variables in $\mathbf{y}$ chances the results.

Nevertheless, there usually exist some guidelines (based on the economic theory) how the ordering could be done in a specific situation.

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

$$\boldsymbol{\nu}_t = \mathbf{S}^{-1}\boldsymbol{\epsilon}_t, \tag{84}$$

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

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

The new vector MA representation becomes

$$\mathbf{y}_t = \sum_{i=0}^{\infty} \mathbf{\Psi}_i^* \boldsymbol{\nu}_{t-i}, \tag{85}$$

where $\mathbf{\Psi}_i^* = \mathbf{\Psi}_i \mathbf{S}$ ($m \times m$ matrices) so that $\mathbf{\Psi}_0^* = \mathbf{S}$. The impulse response function of y_i to a unit shock in y_j is then given by

$$\psi_{ij,0}^*, \psi_{ij,1}^*, \psi_{ij,2}^*, \ldots \tag{86}$$

Variance decomposition

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

Because the innovations have unit variances (besides the uncorrelatedness), the components of this error variance accounted for by innovations to y_j is given by

$$\sum_{k=0}^{s} \psi_{ij,k}^{*2}. \tag{87}$$

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

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

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

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

On the ordering of variables

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

It is recommended that various orderings should be tried to see whether the resulting interpretations are consistent.

The principle is that the first variable should be selected such that it is the only one with potential immediate impact on all other variables.

The second variable may have an immediate impact on the last $m-2$ components of $\mathbf{y}_t$, but not on y_{1t}, the first component, and so on. Of course this is usually a difficult task in practice.

Selection of the $\mathbf{S}$ matrix

Selection of the $\mathbf{S}$ matrix, where $\mathbf{SS}' = \Sigma_\epsilon$, actually defines also the ordering of variables.

Selecting it as a lower triangular matrix implies that the first variable is the one affecting (potentially) the all others, the second to the $m-2$ rest (besides itself) and so on.

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

*For example, if the covariance matrix is

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$$

then if $\Sigma = \mathbf{S}\mathbf{S}'$, where

$$\mathbf{S} = \begin{pmatrix} s_{11} & 0 \\ s_{21} & s_{22} \end{pmatrix}$$

We get

$$\begin{aligned} \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} &= \begin{pmatrix} s_{11} & 0 \\ s_{21} & s_{22} \end{pmatrix} \begin{pmatrix} s_{11} & s_{21} \\ 0 & s_{22} \end{pmatrix} \\ &= \begin{pmatrix} s_{11}^2 & s_{11}s_{21} \\ s_{21}s_{11} & s_{21}^2 + s_{22}^2 \end{pmatrix}. \end{aligned}$$

Thus $s_{11}^2 = \sigma_1^2$, $s_{21}s_{11} = \sigma_{21} = \sigma_{12}$, and $s_{21}^2 + s_{22}^2 = \sigma_2^2$. Solving these yields $s_{11} = \sigma_1$, $s_{21} = \sigma_{21}/\sigma_1$, and $s_{22} = \sqrt{\sigma_2^2 - \sigma_{21}^2/\sigma_1^2}$. That is,

$$\mathbf{S} = \begin{pmatrix} \sigma_1 & 0 \\ \sigma_{21}/\sigma_1 & (\sigma_2^2 - \sigma_{21}^2/\sigma_1^2)^{\frac{1}{2}} \end{pmatrix}.$$

Example 2.10: Let us choose in our example two orderings. One with stock market series first followed by bond market series

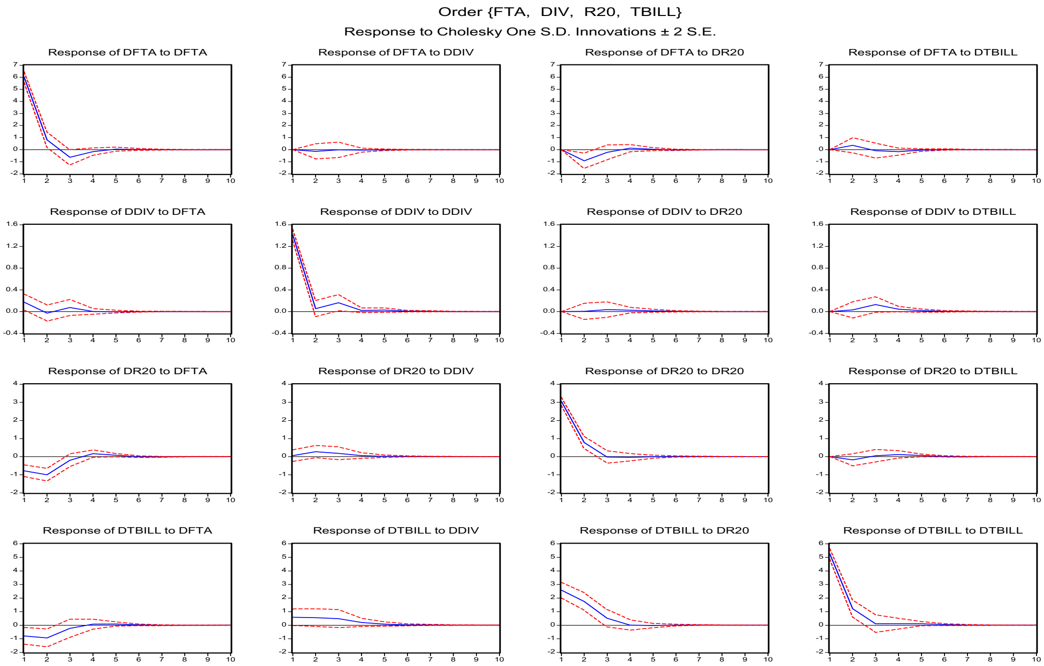
[(I: FTA, DIV, R20, TBILL)],

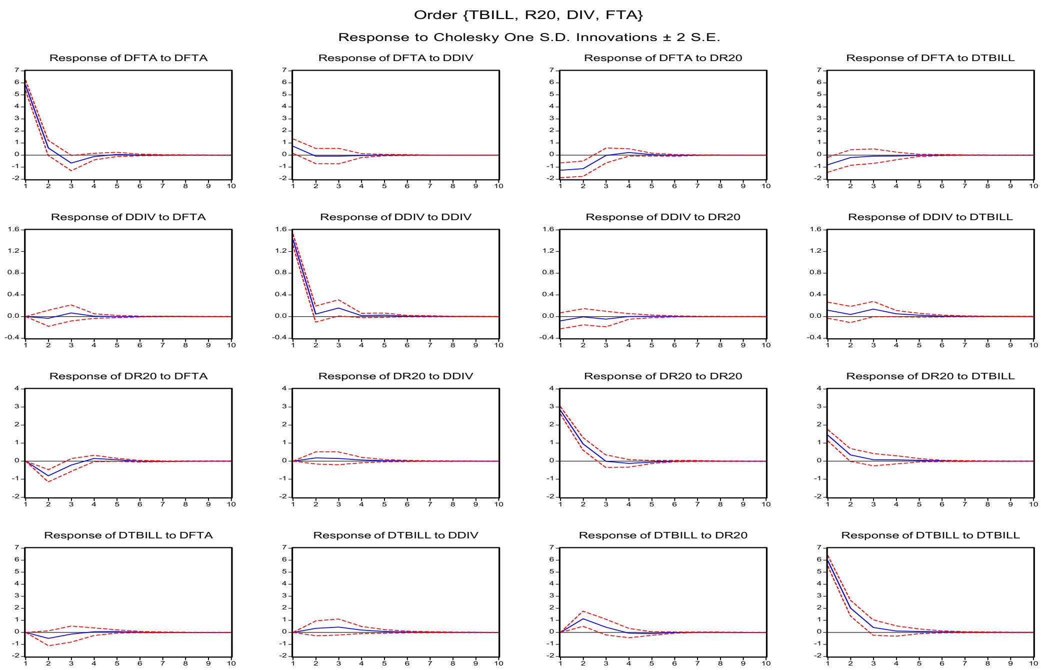
and an ordering with interest rate variables first followed by stock markets series

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

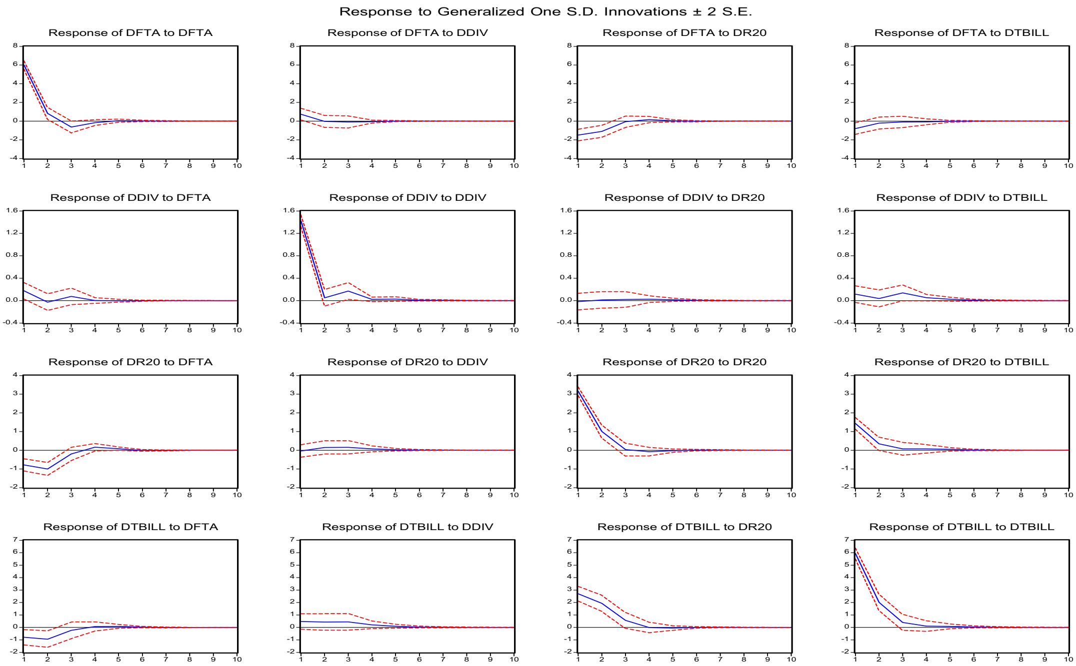
In EViews the order is simply defined in the Cholesky ordering option. Below are results in graphs with I: FTA, DIV, R20, TBILL; II: R20, TBILL DIV, FTA, and III: General impulse response function.

Impulse responses:





Impulse responses continue:



General Impulse Response Function

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

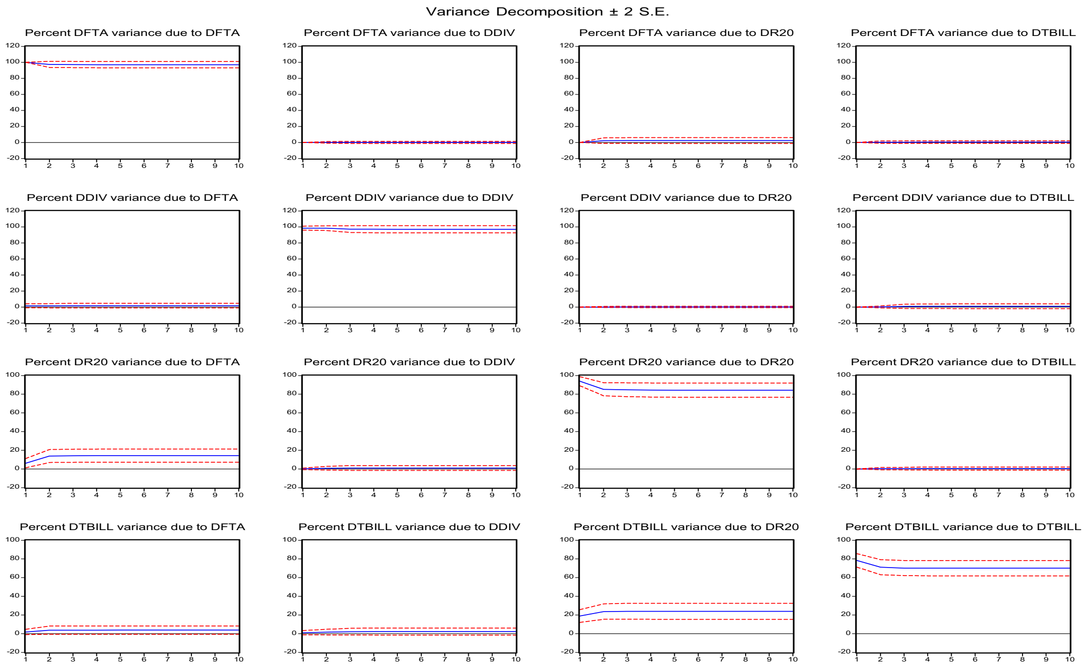
$$GI(j, \delta_i, \mathcal{F}_{t-1}) = \mathbb{E}[\mathbf{y}_{t+j} | \epsilon_{it} = \delta_i, \mathcal{F}_{t-1}] - \mathbb{E}[\mathbf{y}_{t+j} | \mathcal{F}_{t-1}].$$

(89)

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

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

Variance decomposition graphs of the equity-bond data



Accumulated Responses

Accumulated responses for s periods ahead of a unit shock in variable i on variable j may be determined by summing up the corresponding response coefficients. That is,

$$\psi_{ij}^{(s)} = \sum_{k=0}^{s} \psi_{ij,k}. \tag{90}$$

The total accumulated effect is obtained by

$$\psi_{ij}^{(\infty)} = \sum_{k=0}^{\infty} \psi_{ij,k}. \tag{91}$$

In economics this is called the *total multiplier*.

Particularly these may be of interest if the variables are first differences, like the stock returns.

For the stock returns the impulse responses indicate the return effects while the accumulated responses indicate the *price effects*.

All accumulated responses are obtained by summing the MA-matrices

$$\Psi^{(s)} = \sum_{k=0}^{\infty} \Psi_k, \tag{92}$$

with $\Psi^{(\infty)} = \Psi(1)$, where

$$\Psi(L) = 1 + \Psi_1 L + \Psi_2 L^2 + \cdots. \tag{93}$$

is the lag polynomial of the MA-representation

$$\mathbf{y}_t = \Psi(L)\boldsymbol{\epsilon}_t. \tag{94}$$

On estimation of the impulse response coefficients

Consider the VAR(p) model

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \cdots + \Phi_p \mathbf{y}_{t-p} + \boldsymbol{\epsilon}_t \tag{95}$$

or

$$\Phi(L)\mathbf{y}_t = \boldsymbol{\epsilon}_t. \tag{96}$$

Then under stationarity the vector MA representation is

$$\mathbf{y}_t = \boldsymbol{\epsilon}_t + \Psi_1 \boldsymbol{\epsilon}_{t-1} + \Psi_2 \boldsymbol{\epsilon}_{t-2} + \cdots \tag{97}$$

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

$$\Psi_j = \sum_{i=1}^{j} \Psi_{j-i} \Phi_i \tag{98}$$

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

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

$$\Sigma_\epsilon = \mathbf{S}\mathbf{S}', \tag{99}$$

we can write

$$\begin{aligned} \mathbf{y}_t &= \textstyle\sum_{i=0}^{\infty} \Psi_i \epsilon_{t-i} \\ &= \textstyle\sum_{i=0}^{\infty} \Psi_i \mathbf{S}\mathbf{S}^{-1} \epsilon_{t-i} \\ &= \textstyle\sum_{i=0}^{\infty} \Psi_i^* \nu_{t-i}, \end{aligned} \tag{100}$$

where

$$\Psi_i^* = \Psi_i \mathbf{S} \tag{101}$$

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

$$\mathrm{Cov}(\nu_t) = \mathbf{S}^{-1}\Sigma_\epsilon \mathbf{S}'^{-1} = \mathbf{I}. \tag{102}$$

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

Critique of Impulse Response Analysis

Ordering of variables is one problem.

Interpretations related to Granger-causality from the ordered impulse response analysis may not be valid.

If important variables are omitted from the system, their effects go to the residuals and hence may lead to major distortions in the impulse responses and the structural interpretations of the results.