Multivariate Time-Series Analysis

Carlo Favero

2013/2014

Spurious Regressions

To give an intuition of the importance of non-stationarity in time-series and to illustrate the problems related to non-stationarity, consider the results of two regressions reported in the Table , obtained by relating the logarithm of UK stock prices to the log of US dividends and the log of UK dividends.

TABLE 3.1. Regressing UK prices on US dividends and UK dividends

SAMPLE 1973:1 2011:4, Dep. Var LPUK			
Variable	Coefficient	Std. Error	t-Statistic
C	3.38	0.086	38.18
LDUS	1.71	0.036	44.46
R^2 0.9295 S.E. of regression 0.0853 , DW stat 0.13			
C	2.31	0.069	33.04
LDUK	1.22	0.017	72.07
R^2 0.972 S.E. of regression 0.033 , DW stat 0.32			

Understanding spurious regressions

Table 3.2. Univariate models for UK prices and US dividends

SAMPLE 1973:1 2011:4			
Variable	Coefficient	Std. Error	t-Statistic
Dependent variable LPUK			
C	0.111	0.153	0.728
LPUK(-1)	0.986	0.023	25.49
R^2 0.99, S.E. of regr 0.042,			
Dependent variable LDUS			
C	0.0214	0.261	0.082
LDUS(-1)	0.993	0.091	10.90
S.E. of regr 0.011,			

Understanding spurious regressions

LDUS and LPUK can both be approximated by random walk models:

$$
\begin{aligned}
\text{LDUS}_t &= a_0 + \text{LDUS}_{t-1} + \epsilon_{1t}, \\
\text{LPUK}_t &= b_0 + \text{LPUK}_{t-1} + \epsilon_{2t}, \\
\epsilon_{1t} &\sim n.i.d.\left(0, \sigma^2_{\epsilon_1}\right), \\
\epsilon_{2t} &\sim n.i.d.\left(0, \sigma^2_{\epsilon_2}\right).
\end{aligned}
$$

As we already know, recursive substitution yields:

$$
\begin{aligned}
\text{LDUS}_t &= \text{LDUS}_0 + a_0 t + \sum_{i=0}^{t-1} \epsilon_{1t-i}, \\
\text{LPUK}_t &= \text{LPUK}_0 + b_0 t + \sum_{i=0}^{t-1} \epsilon_{2t-i}.
\end{aligned}
$$

Understanding Spurios Regressions

When the following model is estimated:

$$\text{LPUK}_t = \widehat{\alpha} + \widehat{\beta}\text{LDUS}_t + \widehat{\epsilon}_t,$$

the coefficient $\widehat{\beta}$ is significant as both series have a deterministic trend. However, to have a non-spurious relation, we require that the regression also removes the stochastic trend from the dependent variables, leaving stationary residuals.

Clicker

Insert Clicker 1 here

Spurious Regressions

The Durbin−Watson statistic, originally designed to test for the presence of first-order autocorrelation in the residuals, can be re-calibrated to test for stationarity:

$$\mathrm{DW} = \frac{\sum_{i=2}^{T} \left(\widehat{\epsilon}_t - \widehat{\epsilon}_{t-1}\right)^2}{\sum_{i=2}^{T} \widehat{\epsilon}_t^2} \simeq 2\left(1 - \widehat{\rho}\right),$$

where $\widehat{\rho}$ is the OLS coefficient from the regression of $\widehat{\epsilon}_t$ on $\widehat{\epsilon}_{t-1}$. The test was originally tabulated to test the hypothesis H_0: $\rho = 0$; however, critical values for the null of non-stationarity H_0: $\rho = 1$ have been provided by Sargan and Bhargava(1983).

Dynamic Models

TABLE 2.4. Dynamic models for UK stock prices

Dependent variable $LPUK_t$, regression by OLS, 1960:1-1998:1

	Model with US income		Model with UK income	
	Coefficient	S.E.	Coefficient	S.E
c	0.218	0.084	0.416	0.081
$LPUK_{t-1}$	0.9423	0.024	0.857	0.036
$LDUS_{t-1}$	0.124	0.009		
$LDUK_{t-1}$			0.159	0.062
Trend	-0.0007	0.001	0.000004	0.0006
R^2	0.99		0.99	
S.E.	0.0096		0.0092	

Why can a dynamic model solve the spurious regression problem ?

- The log of stock prices and the log of dividends are trending variables, and removing a deterministic trend from them does not deliver stationary time-series.
- However. The dynamic dividend growth model is built on the hypothesis that the log of the dividend price is stationary, this means that, while the log of dividends and the log of prices are non stationary, there exists a linear combination of them that becomes stationary.
- In this case we say that the two series are cointegrated with a cointegrating vector (1,-1)
- In general, we say that two non-stationary variables integrated of order d are cointegrated of order b, if there exists a linear combination of them which is integrated of order $d - b$.

Let us consider the simplest possible multivariate model, i.e. the bivariate model and let us consider the case of the two specific variables to our interest, lp_t, the log of stock prices and ld_t, the log of dividends. We represent the dynamic process as follows:

$$
\begin{aligned}
lp_t &= a_0 + a_1 lp_{t-1} + a_2 ld_{t-1} + \epsilon_{1t}. \\
ld_t &= b_0 + ld_{t-1} + \epsilon_{2t}.
\end{aligned}
$$

Note that system is a multivariate generalization of the univariate autoregressive process than can be re-written as :

$$\begin{aligned}
\mathbf{Y}_t &= \mathbf{A}_0 + \mathbf{A}_1 \mathbf{Y}_{t-1} + \epsilon_t \\
\mathbf{Y}_t &= \begin{bmatrix} lp_t \\ ld_t \end{bmatrix}, \epsilon_t = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \\
\mathbf{A}_0 &= \begin{bmatrix} a_0 \\ b_0 \end{bmatrix}, \mathbf{A}_1 = \begin{bmatrix} a_1 & a_2 \\ 0 & 1 \end{bmatrix}
\end{aligned}$$

The model is therefore naturally called a VAR (Vector Autoregressive Process).

Cointegrated VARs

Cointegration has interesting implications for VAR representation. Consider the following re-parameterization of our VAR :

$$
\begin{aligned}
\Delta lp_t &= a_0 + \alpha \left(lp_{t-1} - \beta_1 ld_{t-1} \right) + \epsilon_{1t}, \\
\Delta ld_t &= b_0 + \epsilon_{2t}. \\
\alpha &= \left(1 - a_1 \right), \quad \beta_1 = \frac{a_2}{1 - a_1}.
\end{aligned}
$$

The estimated dynamic model includes both first differences and levels. The presence of the level variables generates a long-run solution, derived by setting all first differences either to zero (steady state with no deterministic trend) or a constant (steady state). α plays a crucial role and determines the dynamic properties of the system. β_1 defines the long-run relation between *lp* and *ld*.

Cointegrated VARs

- we can interpret $\beta_1 ld$ as the long-run equilibrium level lp^* for the log of prices.
- When $\alpha < 0$, prices increase at time t whenever $lp_{t-1} < lp^*_{t-1}$, and decreases whenever $lp_{t-1} > lp^*_{t-1}$. The system equilibrates in the presence of disequilibrium (i.e. a discrepancy between lp and lp^*). Such error correction features guarantee that $(lp_t - lp^*_t)$ is stationary. This in fact defines cointegration.

Cointegrated VARs

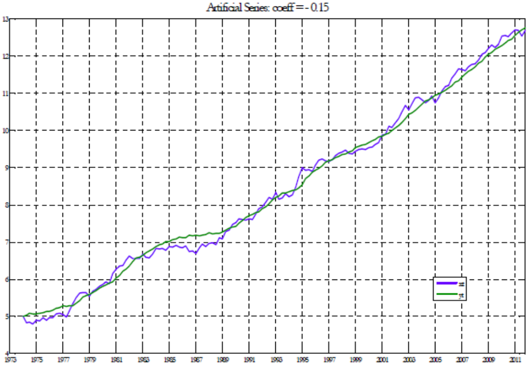
- Cointegration implies an ECM representation, which allows us to re-write a model in levels, involving non-stationary time-series, as a model involving only stationary variables. Such variables are stationary either because they are the first differences of non-stationary variables or because they are stationary linear combinations of non-stationary variables (cointegrating vectors).
- We can reinterpret in terms of cointegration between prices and dividends the results of the predictive regressions of stock market returns on the dividend price ratio.

Clicker

Insert Clicker 2 here

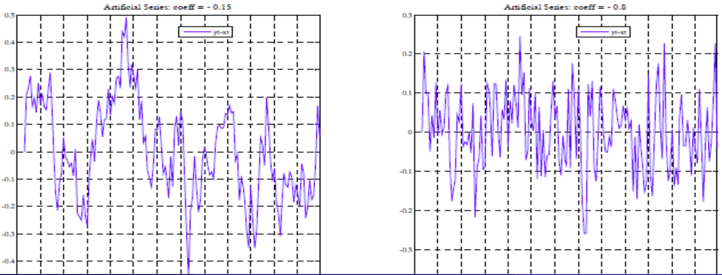
The properties of CVARs

To show the properties of the model, we first generate samples for the two innovation processes; then we generate artificial data for prices and dividends by constructing the above model and solving it dynamically. We do so for a sample of 200 observations. The simulated series in levels ($lp_t$and ld_t) are plotted in the following Figure:



The properties of CVARs

The parameter α in the ECM specification determines the speed of adjustment in the presence of disequilibrium. To illustrate the role of this parameter we report the two series (lp_t-ld_t) generated by taking the same innovations for the sample 1:200. The process (**??**) is used to generate the first time-series of disequilibria (lp_t-ld_t), while the second time-series (lp_t-ld_t^1) is generated by keeping all the parameters unchanged with the exception of α, which is changed from 0.15 to 0.8. The resulting observations for disequilibria are reported in the following Figure



Static Regressions and Dynamic Models

Given the following DGP:

$$\begin{aligned} y_t &= a_1 y_{t-1} + a_2 x_t + a_3 x_{t-1} + u_{1t}, \\ x_t &= b_1 x_{t-1} + u_{2t}, \\ \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} &\sim N.I.D. \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} \right), \end{aligned}$$

a static model is estimated by OLS:

$$\begin{aligned} y_t &= \gamma x_t + \varepsilon_t, \\ \widehat{\gamma} &= \frac{\sum x_t y_t}{\sum x_t^2}. \end{aligned}$$

$$p\lim \widehat{\gamma} = p\lim \left[a_1 \frac{\sum x_t y_{t-1}/T}{\sum x_t^2/T} + a_2 + a_3 \frac{\sum x_t x_{t-1}/T}{\sum x_t^2/T} + \frac{\sum x_t u_{1t}/T}{\sum x_t^2/T} \right].$$

Under the hypothesis $(|b_1| < 1)$, we can substitute for x_t in terms of x_{t-1} and u_{2t} and apply Slutsky's and Cramer's theorems to derive the following result:

$$\begin{aligned} p\lim \widehat{\gamma} &= \frac{a_2 + a_3 b_1}{1 - a_1 b_1}, \\ a_2 &\leq p\lim \widehat{\gamma} \leq \frac{a_2 + a_3}{1 - a_1}. \end{aligned}$$

Cointegration and Forecasting Stock Market returns at different horizons

TABLE 3.4. Forecasting UK Stock-Market Returns at different horizons

Dependent variable $\sum_{j=1}^{k} (h^s_{t+j})$, regression by OLS, 1973:1-2011:4

horizon	β_0^k	β_1^k	R^2	S.E
1-quarter	0.304 (0.093)	0.087 (0.028)	0.0631	0.0069
1-year	1.08 (0.16)	0.31 (0.05)	0.21	0.02
2-year	1.85 (0.198)	0.52 (0.06)	0.34	0.0325
3-year	2.49 (0.2)	0.70 (0.06)	0.46	0.0432

$$\sum_{j=1}^{k} (h^{s,uk}_{t+j}) = \beta_0^k - \beta_1^k (p_t - d_t) + \varepsilon_{t,t+j}$$

$k = 1, 4, 8, 12$

h^s_t are log total annualized real UK stock market returns.

Cointegration and Forecasting Stock Market returns at different horizons

- The pattern of significance in the regressions might depend on the fact that a rolling summation of series integrated of order zero behaves asymptotically as a series integrated of order one and, whenever the regressor is persistent, the well-know occurrence of spurious regression between I(1) variables emerges.
- Having established that estimation and testing using long-horizon variables cannot be carried out using the usual regression methods, Valkanov(2003) proposes propose a rescaled t-statistic, $t/\sqrt{T}$, for testing long-horizon regressions.
- The asymptotic distribution of this statistic, although non-normal, is easy to simulate and the results are applicable to a general class of long-horizon regressions.

Valkanov's test

consider the following DGP:

$$
\begin{aligned}
r^1_{t+1} &= \alpha + \beta lpd_t + \epsilon_{1t}, \\
(1 + \phi L)\, lpd_t &= \mu + \epsilon_{2t}. \\
\phi &= 1 + \frac{c}{T} \\
\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} &\sim N.I.D. \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right)
\end{aligned}
$$

Valkanov's test

The long-horizon variables are

$$Z_t^k = r_{t+1}^k = \sum_{j=1}^{k} r_{t+1}^1$$

The regression at different horizon is run by projecting Z_t^k on lpd_t.The simulation of the relevant distribution requires an estimate of the nuisance parameter c.

Valkanov's test

$$r^1_{t+1} = \rho_0 + \rho lpd_{t+1} + \Delta d_{t+1} - lpd_t$$

assuming that the log-dividends follows an autoregressive process:

$$lpd_{t+1} = \phi lpd_t + u_t.$$

by substituting we have that

$$\begin{aligned} r^1_{t+1} &= \rho_0 - \beta_1 lpd_t + \varepsilon_{t+1} \\ \varepsilon_{t+1} &= \Delta d_{t+1} + u_t. \\ \beta_1 &= (1 - \rho\phi) \end{aligned}$$

Valkanov's test

The k-period horizon return can then be written as follows:

$$
\begin{aligned}
r_{t+1}^{k} &\approx \tilde{k} - \beta_k x_t + \tilde{\epsilon}_{t+1} \\
\beta_k &= \left[(1-\rho\phi) \sum_{i=0:}^{k-1} \phi^i \right]
\end{aligned}
$$

Now, we can write

$$
\beta_k = \left[(1-\rho\phi) \frac{1-\phi^k}{1-\phi} \right]
$$

Valkanov's test

Now, consider the case in which ρ is close to 1, and remember that $\phi = 1 + \frac{c}{T}$ and we can express k in terms of the total length of the available sample as,$k = \lfloor \lambda T \rfloor$, from which $T \approx \frac{k}{\lambda}$. Then:

$$\beta_k = 1 - \left(1 + \frac{c}{T}\right)^k = 1 - \left(1 + \frac{c\lambda}{k}\right)^k$$

$$\lim_{k \longrightarrow \infty} \left(1 + \frac{c\lambda}{k}\right)^k = e^{c\lambda}$$

$$\lim_{k \longrightarrow \infty} \beta_k = \lim_{T \longrightarrow \infty} \beta_{\lfloor \lambda T \rfloor} = 1 - e^{c\lambda}$$

Since we can estimate β_k consistently, we can also find a consistent estimate of c by using the transformation:

$$c^{CONSISTENT} = \frac{1}{\lambda} \log\left(1 - \beta_k\right)$$

Clicker

Insert Clicker 3 here

Cointegration with Multiple Cointegrating Vectors

Consider the case of an econometrician who uses cointegration techniques to investigate simultaneously yields on long-term bonds, short term bonds and the stock market.

- The dynamic dividend growth implies $(lp - ld)$ is stationary
- similarly for bond returns we have that the term spread $(S_t = R_t - r_t)$ is stationary.

Cointegration with Multiple Cointegrating Vectors

$$\begin{aligned} lp_t &= a_0 + a_1 lp_{t-1} + a_2 ld_{t-1} + \\ &+ a_3 R_{t-1,T} + a_4 r_{t-1} + \epsilon_{1t}, \end{aligned}$$

This statistical model fits the data well. As it is found that $a_1 < 1$.The model is reparameterised as follows:

$$\begin{aligned} \Delta lp_t &= a_0 + (a_1 - 1)\left[lp_{t-1} - (lp)^*_{t-1}\right] + u_t \\ (lp)^*_{t-1} &= \frac{a_2}{1-a_1} ld_{t-1} + \frac{a_3}{1-a_1} R_{t-1,T} + \frac{a_4}{1-a_1} r_{t-1}. \end{aligned}$$

Cointegration with Multiple Cointegrating Vectors

As a matter of fact the variables considered might admit two cointegrating relationship, one capturing the stock market dynamics and the other the bond market dynamics:

$$\Delta lp_t = a_0 + (a_1 - 1)\left[lp_{t-1} - ld_{t-1}\right] + a_3\left(R_{t-1,T} - r_{t-1}\right) + u_t$$

we have an **identification** problem

Clicker

Insert Clicker 4 here

The Johansen Procedure

Consider the multivariate generalization of the single-equation dynamic model discussed above, i.e. a vector autoregressive model (VAR) for the vector of, possibly non-stationary, m-variables $\mathbf{y}$:

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + ... + \mathbf{A}_n\mathbf{y}_{t-n} + \mathbf{u}_t.$$

subtract $\mathbf{y}_{t-1}$ from both sides of the VAR to obtain:

$$\Delta\mathbf{y}_t = (\mathbf{A}_1 - \mathbf{I})\,\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + ... + \mathbf{A}_n\mathbf{y}_{t-n} + \mathbf{u}_t.$$

Subtract $(\mathbf{A}_1 - \mathbf{I})\,\mathbf{y}_{t-2}$ from both sides:

$$\Delta\mathbf{y}_t = (\mathbf{A}_1 - \mathbf{I})\,\Delta\mathbf{y}_{t-1} + (\mathbf{A}_1 + \mathbf{A}_2 - \mathbf{I})\,\mathbf{y}_{t-2} + ... + \mathbf{A}_n\mathbf{y}_{t-n} + \mathbf{u}_t.$$

The Johansen Procedure

By repeating this procedure until $n-1$, we end up with the following specification:

$$\begin{aligned}\Delta \mathbf{y}_t &= \Pi_1 \Delta \mathbf{y}_{t-1} + \Pi_1 \Delta \mathbf{y}_{t-2} + \ldots + \Pi \mathbf{y}_{t-n} + \mathbf{u}_t \\ &= \sum_{i=1}^{n-1} \Pi_i \Delta \mathbf{y}_{t-i} + \Pi \mathbf{y}_{t-n} + \mathbf{u}_t,\end{aligned}$$

where:

$$\begin{aligned}\Pi_i &= -\left(I - \sum_{j=1}^{i} \mathbf{A}_j\right), \\ \Pi &= -\left(I - \sum_{i=1}^{n} \mathbf{A}_i\right).\end{aligned}$$

Clearly the long-run properties of the system are described by the properties of the matrix Π.

The Johansen Procedure

There are three cases of interest:

1. rank $(\Pi) = 0$. The system is non-stationary, with no cointegration between the variables considered. This is the only case in which non-stationarity is correctly removed simply by taking the first differences of the variables;
2. rank $(\Pi) = m$, full. The system is stationary;
3. rank $(\Pi) = k < m$. The system is non-stationary but there are k cointegrating relationships among the considered variables. In this case $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where α is an $(m \times k)$ matrix of weights and $\boldsymbol{\beta}$ is an $(m \times k)$ matrix of parameters determining the cointegrating relationships.

The Johansen Procedure

Therefore, the rank of Π is crucial in determining the number of cointegrating vectors.

- The Johansen procedure is based on the fact that the rank of a matrix equals the number of its characteristic roots that differ from zero.
- Having obtained estimates for the parameters in the Π matrix, we associate with them estimates for the m characteristic roots and we order them as follows $\lambda_1 > \lambda_2 > ... > \lambda_m$.
- If the variables are not cointegrated, then the rank of Π is zero and all the characteristic roots equal zero.
  - In this case each of the expression $\ln(1 - \lambda_i)$ equals zero, too.
  - If, instead, the rank of Π is one, and $0 < \lambda_1 < 1$, then $\ln(1 - \lambda_1)$ is negative and $\ln(1 - \lambda_2) = \ln(1 - \lambda_3) = ... = \ln(1 - \lambda_m) = 0$.

The Johansen Procedure

Johansen derives a test on the number of characteristic roots that are different from zero by considering the two following statistics:

$$\lambda_{\text{trace}}(k) = -T \sum_{i=k+1}^{m} \ln\left(1-\widehat{\lambda}_i\right),$$
$$\lambda_{\max}(k,k+1) = -T \ln\left(1-\widehat{\lambda}_{k+1}\right),$$

where T is the number of observations used to estimate the VAR. The first statistic tests the null of at most k cointegrating vectors against a generic alternative. The test should be run in sequence starting from the null of at most zero cointegrating vectors up to the case of at most m cointegrating vectors. The second statistic tests the null of at most k cointegrating vectors against the alternative of at most $k+1$ cointegrating vectors.

Critical values are tabulated by Johansen and they depend on the number of non-stationary components under the null and on the specification of the deterministic component of the VAR.

An example

Consider the VAR representation of our simple dynamic model (**??**) for the two variables, x and y:

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}.$$

This system can be reparameterized as follows in terms of the VECM representation:

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} a_{11} - 1 & a_{12} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix},$$

from which, clearly,

$$\Pi = \begin{pmatrix} a_{11} - 1 & a_{12} \\ 0 & 0 \end{pmatrix}, \quad \alpha = \begin{pmatrix} a_{11} - 1 \\ 0 \end{pmatrix}, \quad \beta' = \begin{pmatrix} 1 & -\frac{a_{12}}{1-a_{11}} \end{pmatrix}.$$

Cointegration in the Bond and Stock Markets

The baseline VAR can be specified as:

$$\begin{bmatrix} lp_t \\ ld_t \\ R_{t,T} \\ r_t \end{bmatrix} = A_0 + A_1 \begin{bmatrix} lp_{t-1} \\ ld_{t-1} \\ R_{t-1,T} \\ r_{t-1} \end{bmatrix} + + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{bmatrix},$$

which could then be reparameterized in VECM form:

$$\begin{bmatrix} \Delta lp_t \\ \Delta ld_t \\ \Delta R_{t,T} \\ \Delta r_t \end{bmatrix} = \Pi_0 + \Pi \begin{bmatrix} lp_{t-1} \\ ld_{t-1} \\ R_{t-1,T} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{bmatrix}.$$

Cointegration in the Bond and Stock Markets

Since we know that there are two cointegrating vectors, we have:

$$\begin{aligned} \Pi &= \alpha\beta', \\ \text{rank}\,\Pi &= 2, \\ \beta' &= \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}. \end{aligned}$$

A possible specification for α is :

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ 0 & 0 \\ 0 & \alpha_{32} \\ 0 & 0 \end{bmatrix}.$$

With the above specification for the loadings, stock market prices adjusts both in presence of disequilibria in the stock and the bond markets, long term bonds react to the spread, while short-term rates and dividends do not respond to disequilibria.

Clicker

Insert Clicker 5 here

Using VAR Models

A Cointegrated VAR, after the identification of the number and shape of cointegrating vector(s), provides a statistical model of the joint distribution of the variables of interests:

$$\begin{aligned} \Delta \mathbf{y}_t &= \boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{y}_{t-1} + \mathbf{u}_t \\ \mathbf{u}_t &\sim N\left(0, \sum\right) \end{aligned} \tag{1}$$

where $\mathbf{y}_t$ is a vector of length N containing the modelled variables.

- The reduced form specification can be adopted directly for forecasting purposes or to describe the dynamic response of the system to innovations to observables, such as the VAR residuals.
- Some further identification choice must be made if the model is to be used for evaluating the response of economic and financial variables to innovations to unobservables, i.e. the "structural" shocks to some of the variables included in the VAR. Impulse response analysis examines the effect of a typical shock, usually one-standard deviation, on the time path of the variables in the

Using VAR Models

- In macroeconomics, the importance of computing impulse responses to structural shocks is related to the fact that the solution of a Dynamic Stochastic General Equilibrium (DSGE) model can be well approximated by a VAR, and VARs have become the natural tool for model evaluation.
  - VAR models are not estimated to yield advice on the best policy but rather to provide empirical evidence on the response of macroeconomic variables to policy impulses in order to discriminate between alternative theoretical models of the economy. It then becomes crucial to identify policy actions using restrictions independent from the theoretical models
- In finance, the use of VAR is more related to forecasting first and second moments of the distributions of returns at different horizons. Macro-finance model concentrate on the different role of permanent versus transitory shocks to understand the comovement between financial and macroeconomic variables.

Identification of VAR

Given the estimation of a VAR the problem of extracting unobservable structural shocks $\boldsymbol{v}_t$ from the observed VAR innovations $\mathbf{u}_t$ is usually addressed by positing the following relations

$$\begin{aligned} \mathbf{A}\mathbf{u}_t &= \mathbf{B}\boldsymbol{v}_t, \\ \boldsymbol{v}_t &\sim N(0, I) \end{aligned}$$

from which we can derive the relation between the variance-covariance matrices of $\mathbf{u}_t$ (observed) and $\boldsymbol{\nu}_t$ (unobserved) as follows:

$$E\left(\mathbf{u}_t\mathbf{u}_t'\right) = \mathbf{A}^{-1}\mathbf{B}E\left(\boldsymbol{v}_t\boldsymbol{v}_t'\right)\mathbf{B}'\mathbf{A}^{-1}.$$

Identification of VAR

Substituting population moments with sample moments we have:

$$\widehat{\sum} = \widehat{\mathbf{A}}^{-1}\mathbf{BI}\widehat{\mathbf{B}}'\widehat{\mathbf{A}}^{-1},$$

$\widehat{\sum}$ contains $n(n+1)/2$ different elements, which is the maximum number of identifiable parameters in matrices **A** and **B.**

- Therefore, a necessary condition for identification is that the maximum number of parameters contained in the two matrices equals $n(n+1)/2$,
- As usual, for such a condition also to be sufficient for identification no equation in should be a linear combination of the other equations in the system .
- As for traditional models, we have the three possible cases of under-identification, just-identification and over-identification. The validity of over-identifying restrictions can be tested via a statistic distributed as a χ^2 with a number of degrees of freedom equal to the number of over-identifying restrictions.

Description of VAR models

After the identification of structural shocks of interest, the properties of VAR models are described using impulse response analysis, variance decomposition and historical decomposition.
Given an identified and estimated estimate structural VAR

$$
\begin{aligned}
\mathbf{y}_t &= \sum_{i=1}^{p} \mathbf{C}_i \mathbf{y}_{t-i} + \mathbf{u}_t, \\
\mathbf{A}\mathbf{u}_t &= \mathbf{B}\boldsymbol{v}_t,
\end{aligned}
$$

we can re-write it as:

$$
\begin{aligned}
\mathbf{A}\mathbf{y}_t &= \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{t-i} + \mathbf{B}\boldsymbol{v}_t, \\
\mathbf{A}^{-1}\mathbf{A}_i &= \mathbf{C}_i
\end{aligned}
$$

which we can express in a compact way as:

$$\begin{aligned} [\mathbf{A} - \mathbf{A}(L)]\,\mathbf{y}_t &= \mathbf{B}\mathbf{v}_t \\ \mathbf{A}(L) &= \sum_{i=1}^{p} \mathbf{A}_i L^i. \end{aligned}$$

By inverting $[\mathbf{A}_0 - \mathbf{A}(L)]$ (under the assumption of invertibility of this polynomial) we obtain the moving average representation for our VAR process:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{C}(L)\,\mathbf{v}_t, \\ \mathbf{y}_t &= \mathbf{C}_0\mathbf{v}_t + \mathbf{C}_1\mathbf{v}_{t-1} + ... + \mathbf{C}_s\mathbf{v}_{t-s}, \\ \mathbf{C}(L) &= [\mathbf{A}_0 - \mathbf{A}(L)]^{-1}, \\ \mathbf{C}_0 &= \mathbf{A}_0^{-1}\mathbf{B}. \end{aligned}$$

To illustrate the concept of an *impulse response function*, we interpret the generic matrix $\mathbf{C}_s$ within the moving average representation as follows:

$$\mathbf{C}_s = \frac{\partial \mathbf{y}_{t+s}}{\partial \mathbf{v}_t}.$$

The generic element $\{i, j\}$ of matrix $\mathbf{C}_s$ represents the impact of a shock hitting the j-th variable of the system at time t on the i-th variable of the system at time $t + s$.

Description of VAR models

Historical decomposition is obtained by using the structural MA representation to separate series in the components (orthogonal to each other) attributable to the different structural shocks.
Finally *forecasting error variance decomposition* (FEVD) is obtained from (**??**) by deriving the error in forecasting $\mathbf{y}_s$ period in the future as:

$$(\mathbf{y}_{t+s} - E_t\mathbf{y}_{t+s}) = \mathbf{C}_0\mathbf{v}_t + \mathbf{C}_1\mathbf{v}_{t-1} + ... + \mathbf{C}_s\mathbf{v}_{t-s}$$

from which we can construct the variance of such forecasting error as:

$$Var\,(\mathbf{y}_{t+s} - E_t\mathbf{y}_{t+s}) = \mathbf{C}_0 I\mathbf{C}_0' + \mathbf{C}_1 I\mathbf{C}_1' + ... + \mathbf{C}_s I\mathbf{C}_s'$$

from which we can compute the share of the total variance attributable to the variance of each structural shock. Note again that such composition makes sense only if shocks are orthogonal to each other.

Identification

In practice, identification requires the imposition of some restrictions on the parameters of **A** and **B.** This step has been historically implemented in a number of different ways.

- Choleski Decomposition
- temporary-persistent decomposition
- sign restrictions
- GIRF

Choleski Decomposition

In the famous article which introduced VAR methodology to the profession, Sims (1980*a*) proposed the following identification strategy, based on the Choleski decomposition of matrices:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ a_{21} & 1 & 0 & 0 \\ . & . & 1 & . \\ a_{n1} & . & a_{nn-1} & 1 \end{pmatrix}, \qquad \mathbf{B} \begin{pmatrix} b_{11} & 0 & 0 & 0 \\ 0 & b_{22} & 0 & 0 \\ . & . & b_{ii}. & . \\ 0 & 0 & 0 & b_{nn} \end{pmatrix}. \tag{2}$$

This is a just-identification scheme. It corresponds to a recursive economic structure, with the most endogenous variable ordered last. A generalization of Choleski is to consider contemporaneous restrictions that do not necessarily lead to a triangular structure of A.

Clicker

Insert Clicker 6 here

CVAR and Identification of shocks

Consider now a CVAR

$$
\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}
$$

$$
\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}.
$$

Model can be re-written as follows :

$$
\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} (1-L) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & 0 \end{pmatrix} \begin{pmatrix} (y_{t-1} - x_{t-1}) \\ \Delta x_{t-1} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}.
$$

CVAR and Identification of shocks

The cointegrating properties of the system suggest the presence of two types of shocks: a permanent one (related to the single common trend shared by the two variables) and a transitory one (related to the cointegrating relation).
To derive long-run responses :

$$\left(\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} (1-L) & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \alpha_{11}L & 0 \\ \alpha_{21}L & 0 \end{pmatrix}\right)\begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22}$$

from which long-run responses are obtained by setting $L = 1$ and by inverting the matrix pre-multiplying variables in the stationary representation of VAR

$$\begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} -\alpha_{11} & 1 \\ -\alpha_{21} & 1 \end{pmatrix}^{-1}\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}$$

CVAR and Identification of shocks

$$\begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \frac{-b_{11}+b_{21}}{\alpha_{11}-\alpha_{21}} & -\frac{b_{12}-b_{22}}{\alpha_{11}-\alpha_{21}} \\ \frac{-\alpha_{21}b_{11}+\alpha_{11}b_{21}}{\alpha_{11}-\alpha_{21}} & \frac{-\alpha_{21}b_{12}+\alpha_{11}b_{22}}{\alpha_{11}-\alpha_{21}} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}.$$

Thus v_{2t} can be identified as the transitory shock by imposing the following restriction:

$$-\alpha_{21}b_{12} + \alpha_{11}b_{22} = 0$$

which, given knowledge of the $\boldsymbol{\alpha}$ parameters from the cointegration analysis, provides the just-identifying restriction for the parameters in **B**. Note that, there is one case in which this identification is equivalent to the Choleski ordering, the case in which $\alpha_{11} = 0$. Note that this is the case in which Δy_t is weakly exogenous for the estimation of b_{21}.

Sign Restrictions

Given the VAR specification:

$$\begin{aligned}\mathbf{y}_t &= \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{t-i} + \mathbf{B}\mathbf{u}_t \\ \Sigma &= \mathbf{B}\mathbf{E}\left(\mathbf{u}_t \mathbf{u}_t'\right)\mathbf{B}' = \mathbf{B}\mathbf{B}'\end{aligned}$$

Consider the Choleski decomposition of Σ, C .
The impulse response function, given the Choleski decomposition could be written as :

$$\mathbf{y}_t = \left[\mathbf{I} - \mathbf{A}\left(L\right)\right]^{-1} \mathbf{C}\mathbf{u}_t$$

Sign Restrictions

All the possible rotation of the Choleski decomposition are obtained as follows:

$$
\begin{aligned}
& [\mathbf{I}-\mathbf{A}(L)]^{-1}\mathbf{CQQ}'\mathbf{u}_t \\
\mathbf{QQ}' &= I
\end{aligned}
$$

The impulse response for $\mathbf{Q}'\mathbf{u}_t$, is then $[\mathbf{I}-\mathbf{A}(L)]^{-1}\mathbf{CQ}$.
The imposition of the sign restrictions then consider Q to generate all possible identification and then select only those that satisfy some sign restriction.

GIRF

If the identification of structural shocks is not an issue of primary interest then Generalized Impulse Response Functions can be used to describe the respoonse of the system to change in observable i.e. the VAR innovations.

Consider again our bivariate CVAR model :

$$\begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = A \begin{pmatrix} (y_{t-1} - x_{t-1}) \\ \Delta x_{t-1} \end{pmatrix} + \mathbf{u}_t$$

$$\mathbf{u}_t \sim N\left(0, \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{pmatrix}\right)$$

GIRF

from the properties of the normal distribution we have that

$$E\left(u_{2t} \mid u_{1t}\right) = \left(\sigma_{11}^2\right)^{-1} \sigma_{12} u_{1t}$$

so the impulse responses can be derived as follows:

$$\begin{aligned} \frac{\partial \left[\begin{array}{c} (y_{t+i} - x_{t+i}) \\ \Delta x_t + i \end{array} \right]}{\partial u_{1t}} &= A^i S \\ S &= \left(\begin{array}{c} 1 \\ \left(\sigma_{11}^2\right)^{-1} \sigma_{12} \end{array} \right) \end{aligned}$$

GIRF seems to be more appropriate when the primary focus of the analysis is the description of the transmission mechanism rather than the structural interpretation of shocks.

Consider a vector $\mathbf{y}_t$ containing two variables x_t and z_t cointegrated with an equilibrium error $S_t = x_t - \beta z_t$.
The Johansen representation for such system will be:

$$
\begin{aligned}
\begin{pmatrix} \Delta x_t \\ \Delta z_t \end{pmatrix} &= \Pi_1 \begin{pmatrix} \Delta x_{t-1} \\ \Delta z_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \end{pmatrix} \begin{pmatrix} 1 & -\beta \end{pmatrix} \begin{pmatrix} x_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}, \\
\begin{pmatrix} \Delta x_t \\ \Delta z_t \end{pmatrix} &= \Pi_1 \begin{pmatrix} \Delta x_{t-1} \\ \Delta z_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \end{pmatrix} S_{t-1} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}
\end{aligned}
$$

Define a matrix M such that

$$
\begin{aligned}
M \begin{pmatrix} \Delta x_t \\ \Delta z_t \end{pmatrix} &= \begin{pmatrix} \Delta x_t \\ \Delta S_t \end{pmatrix} \\
M &= \begin{pmatrix} 1 & 0 \\ 1 & -\beta \end{pmatrix}
\end{aligned}
$$

then we have:

$$
\begin{aligned}
M \begin{pmatrix} \Delta x_t \\ \Delta z_t \end{pmatrix} &= M\Pi_1 \begin{pmatrix} \Delta x_{t-1} \\ \Delta z_{t-1} \end{pmatrix} + M \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \end{pmatrix} S_{t-1} + M \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \\
\begin{pmatrix} \Delta x_t \\ \Delta S_t \end{pmatrix} &= M\Pi_1 M^{-1} \begin{pmatrix} \Delta x_{t-1} \\ \Delta S_{t-1} \end{pmatrix} + M \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \end{pmatrix} S_{t-1} + M \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}
\end{aligned}
$$

Cointegration and PV Models

The system can be rearranged so that it describes levels rather than differences of S_t.
The result is a second order VAR as follows:

$$\begin{pmatrix} \Delta x_t \\ S_t \end{pmatrix} = G_1 \begin{pmatrix} \Delta x_{t-1} \\ S_{t-1} \end{pmatrix} + G_2 \begin{pmatrix} \Delta x_{t-2} \\ S_{t-2} \end{pmatrix} + M \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}$$

Consider the case of the risk free rate and a very long term bond. In such case, under the null of the ET, we have:

$$R_{t,T} = R^*_{t,T} \approx (1-\gamma) \sum_{j=0}^{T-t-1} \gamma^j E[r_{t+j} \mid I_t]$$

Cointegration and PV Models

which could be re-written in terms of spread between long and short-term rates, $S_{t,T} = R_{t,T} - r_t$:

$$S_{t,T} = S^*_{t,T} = \sum_{j=1}^{T-t-1} \gamma^j E[\Delta r_{t+j} \mid I_t]$$

CS construct a bivariate stationary VAR in the first difference of the short-term rate and the spread :

$$\begin{aligned} \Delta r_t &= a(L)\Delta r_{t-1} + b(L)S_{t-1} + u_{1t} \\ S_t &= c(L)\Delta r_{t-1} + d(L)S_{t-1} + u_{2t} \end{aligned}$$

Cointegration and PV Models

Stack the VAR as:

$$\begin{bmatrix} \Delta r_t \\ . \\ . \\ \Delta r_{t-p+1} \\ S_t \\ . \\ . \\ S_{t-p+1} \end{bmatrix} = \begin{bmatrix} a_1 & . & . & a_p & b_1 & . & . & b_p \\ 1 & . & . & 0 & 0 & . & . & 0 \\ 0 & . & . & 0 & 0 & . & . & 0 \\ 0 & . & 1 & 0 & 0 & . & . & 0 \\ c_1 & . & . & c_p & d_1 & . & . & d_p \\ 0 & . & . & 0 & 1 & . & . & 0 \\ 0 & . & . & 0 & 0 & . & . & 0 \\ 0 & . & . & 0 & 0 & . & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta r_{t-1} \\ . \\ . \\ \Delta r_{t-p} \\ S_{t-1} \\ . \\ . \\ S_{t-p} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ . \\ . \\ 0 \\ u_{2t} \\ . \\ . \\ 0 \end{bmatrix}$$

This can be written more succinctly as:

$$z_t = Az_{t-1} + v_t$$

Cointegration and PV Models

The ET null puts a set of restrictions which can be written as :

$$g'z_t = \sum_{j=1}^{T-1} \gamma^j h' A^{j\prime} z_t$$

where $g\prime$ and $h\prime$ are selector vectors for S and Δr correspondingly (i.e. row vectors with 2p elements, all of which are zero except for the p+1st element of $g\prime$ and the first element of $h\prime$ which are unity).

For large T it must be the case that:

$$g' = h'\gamma A (I - \gamma A)^{-1}$$

which implies:

$$g'(I - \gamma A) = h'\gamma A$$

and we have the following constraints on the individual coefficients of VAR:

$$\{c_i = -a_i, \forall i\}, \{d_1 = -b_1 + 1/\gamma\}, \{d_i = -b_i, \forall i \neq 1\}$$

Cointegration and PV Models

1080 JOURNAL OF POLITICAL ECONOMY

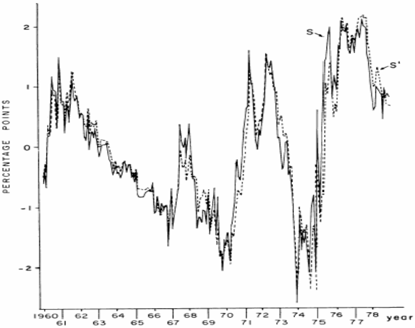


FIG. 1.—Term structure: deviations from means of long-short spread S_t and theoretical spread S'_t.

Cointegration and multivariate trend-shocks decompositions

Having discussed the VECM representation for a vector of m non-stationary variables admitting k cointegrating relationships, let us compare it with the multivariate extension of the Beveridge−Nelson decomposition. Consider the simple case of an I(1) vector $\mathbf{y}_t$ featuring first-order dynamics and no deterministic component:

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \mathbf{u}_t, \tag{3}$$

where α is the $(m \times k)$ matrix of loadings and $\boldsymbol{\beta}$ is the $(m \times k)$ matrix of parameters in the cointegrating relationships. As $\mathbf{y}_t$ is I(1), we can apply the Wold decomposition theorem to $\Delta \mathbf{y}_t$ to obtain the following representation:

$$\Delta \mathbf{y}_t = \mathbf{C}(L)\,\mathbf{u}_t,$$

from which, by applying the algebra illustrated in our discussion of the univariate Beveridge−Nelson decomposition, we can derive the following stochastic trends representation:

Cointegration and multivariate trend-shocks decompositions

The existence of cointegration imposes restrictions on the $\mathbf{C}$ matrices. The stochastic trends must cancel out when the k stationary linear combinations of the variables in $\mathbf{y}_t$ are considered. In other words we must have:

$$\boldsymbol{\beta}'\mathbf{C}(1) = 0.$$

By investigating further the relation between the VECM and the stochastic trend representations, we can give a more precise parameterization of the matrix $\mathbf{C}(1)$.
Note first that VECM is equivalent to:

$$\mathbf{y}_t = \left(I_m + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)\mathbf{y}_{t-1} + \mathbf{u}_t.$$

Pre-multiplying this system by $\boldsymbol{\beta}'$ yields:

$$\begin{aligned}\boldsymbol{\beta}'\mathbf{y}_t &= \boldsymbol{\beta}'\left(I_m + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)\mathbf{y}_{t-1} + \boldsymbol{\beta}'\mathbf{u}_t \\ &= \left(I_k + \boldsymbol{\alpha}\boldsymbol{\beta}'\right)\boldsymbol{\beta}'\mathbf{y}_{t-1} + \boldsymbol{\beta}'\mathbf{u}_t.\end{aligned}$$

Solving this model recursively, we obtain the MA representation for

Cointegration and multivariate trend-shocks decompositions

$$\beta' \mathbf{y}_t = \sum_{i=0}^{\infty} \left(I_k + \alpha \beta' \right)^i \beta' \mathbf{u}_{t-i}.$$

By substituting in the ECM we have the MA representation for $\Delta \mathbf{y}_t$,

$$\Delta \mathbf{y}_t = \sum_{i=1}^{\infty} \alpha \left(I_k + \alpha \beta' \right)^{i-1} \beta' \mathbf{u}_{t-i} + \mathbf{u}_t,$$

from which we have

$$\mathbf{C}(1) = I_n - \alpha \left(\beta' \alpha \right)^{-1} \beta'.$$

Now note the beautiful relation (see Johansen 1995: 40),

$$I_n = \beta_\perp \left(\alpha'_\perp \beta_\perp \right)^{-1} \alpha'_\perp + \alpha \left(\beta' \alpha \right)^{-1} \beta',$$

where $\beta_\perp, \alpha_\perp$ are $((m \times (m-k))$ matrices of rank $m-k$ such that $\alpha'_\perp \alpha = 0$, $\beta'_\perp \beta = 0$.

Cointegration and multivariate trend-shocks decompositions

we have

$$\mathbf{C}(1) = \beta_{\perp} \left(\alpha'_{\perp} \beta_{\perp} \right)^{-1} \alpha'_{\perp},$$

and

$$\mathbf{y}_t = \mathbf{C}^* (L) \, \mathbf{u}_t + \beta_{\perp} \left(\alpha'_{\perp} \beta_{\perp} \right)^{-1} \left(\alpha'_{\perp} \mathbf{z}_t \right),$$

which shows that a system of m variables with k cointegrating relationships features $(m - k)$ linearly independent common trends ($\mathbf{TR}$). The common trends are given by $(\alpha'_{\perp} \mathbf{z}_t)$, while the coefficients on these trends are $\beta_{\perp} \left(\alpha'_{\perp} \beta_{\perp} \right)^{-1}$. Note also that stochastic trends depend on a set of initial conditions and cumulated disturbances,

$$\mathbf{TR}_t = \mathbf{TR}_{t-1} + \mathrm{C}(1) \, \mathbf{u}_t.$$

Our brief discussion should have made clear that the VECM model and the MA model are complementary. As a consequence, the identification problem relevant for the vector of parameters in the cointegrating vectors β is also relevant for the vector of parameters

VECM and common trends representations

.

The joint behaviour of stock prices and dividends under the dynamic dividend growth model is a good empirical example to illustrate VECM and common trend representations. Let decompose (log) stock market prices lp_t in a permanent, information-related component, ld_t, and a temporary cyclical noise component v_t. :

$$
\begin{aligned}
lp_t &= ld_t + v_t, \\
ld_t &= \mu_d + ld_{t-1} + u_t,
\end{aligned}
$$

Dividends are the stochastic trend of stock market prices, which are made of the permanent component and of a transitory component, v_t and u_t are the shocks to the transitory and the permanent component of the system; naturally, they are orthogonal and normally and independently distributed. Dividend and prices are cointegrated, in fact they share the single unobservable common stochastic trend in this system.

VECM and common trends representations

.
We obtain the VAR(1) representation by substituting for ld_t in the first equation from the second equation of :

$$\begin{aligned} \begin{pmatrix} lp_t \\ ld_t \end{pmatrix} &= \begin{pmatrix} \mu_d \\ \mu_d \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} lp_{t-1} \\ ld_{t-1} \end{pmatrix} + \begin{pmatrix} w_t \\ u_t \end{pmatrix}, \\ w_t &= u_t + v_t, \end{aligned}$$

from which we obtain the VECM representation:

$$\begin{pmatrix} \Delta lp_t \\ \Delta ld_t \end{pmatrix} = \begin{pmatrix} \mu_d \\ \mu_d \end{pmatrix} + \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} lp_{t-1} \\ ld_{t-1} \end{pmatrix} + \begin{pmatrix} w_t \\ u_t \end{pmatrix},$$

where

$$\begin{aligned} \Pi &= \alpha\beta'. = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} -1 & 1 \end{pmatrix} \end{aligned}$$

VECM and common trends representations

.

The common trend representation is derived by considering that, as $lp_t - ld_t = v_t$, from which we can write :

$$\begin{pmatrix} \Delta lp_t \\ \Delta ld_t \end{pmatrix} = \begin{pmatrix} \mu_d \\ \mu_d \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} w_t \\ u_t \end{pmatrix} + \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} w_{t-1} \\ u_{t-1} \end{pmatrix},$$

from which:

$$\begin{pmatrix} lp_t \\ ld_t \end{pmatrix} = \begin{pmatrix} \mu_d \\ \mu_d \end{pmatrix} t + \mathbf{C}^* (L) \begin{pmatrix} w_t \\ u_t \end{pmatrix} + C(1)\mathbf{z}_t,$$

where $\mathbf{z}_t$ is a process for which $\Delta \mathbf{z}_t = \begin{pmatrix} w_t \\ u_t \end{pmatrix}$,and

VECM and common trends representations

.

$$\begin{aligned} C(1) &= \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} \alpha'_\perp, \\ \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \left[\begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right]^{-1} \begin{pmatrix} 0 & 1 \end{pmatrix}. \end{aligned}$$

Since in this application $\left(\alpha'_\perp \beta_\perp\right)^{-1} = 1$, dividends and prices have a single common stochastic trend. Such trend can be represented as

$$\alpha'_\perp \left(\begin{pmatrix} \mu_y \\ \mu_y \end{pmatrix} t + \begin{pmatrix} \sum_{i=1}^{t} w_t \\ \sum_{i=1}^{t} u_t \end{pmatrix} \right),$$

and only shocks to the permanent component of prices enter the trend.

Clicker

Insert Clicker 7 here

Risk, Returns and Portfolio Allocation with Cointegrated VARs

.

Consider the continuously compounded stock market return from time t to time $t+1$, $\mathbf{r}_{t+1}$. Define $\boldsymbol{\mu}_t$, the conditional expected log return given information up to time t, as follows:

$$\mathbf{r}_{t+1} = \boldsymbol{\mu}_t + \mathbf{u}_{t+1}$$

where $\mathbf{u}_{t+1}$ is the unexpected log return. Define the k-period cumulative return from period $t+1$ through period $t+k$, as follows:

$$\mathbf{r}_{t,t+k} = \sum_{i=1}^{k} \mathbf{r}_{t+i}$$

The term structure of risk is defined as the conditional variance of cumulative returns, given the investor's information set, scaled by the investment horizon

$$\Sigma_r(k) \equiv \frac{1}{k} Var(\mathbf{r}_{t,t+k} \mid D_t) \quad (4)$$

Inspecting the mechanism: a bivariate case

Consider the continuously compounded stock market return from time t to time $t+1$, $\mathbf{r}_{t+1}$. Define $\boldsymbol{\mu}_t$, the conditional expected log return given information up to time t, as follows:

$$\mathbf{r}_{t+1} = \boldsymbol{\mu}_t + \mathbf{u}_{t+1}$$

where $\mathbf{u}_{t+1}$ is the unexpected log return. Define the k-period cumulative return from period $t+1$ through period $t+k$, as follows:

$$\mathbf{r}_{t,t+k} = \sum_{i=1}^{k} \mathbf{r}_{t+i}$$

The term structure of risk is defined as the conditional variance of cumulative returns, given the investor's information set, scaled by the investment horizon

$$\Sigma_r(k) \equiv \frac{1}{k} Var(\mathbf{r}_{t,t+k} \mid D_t) \tag{5}$$

where $D_t \equiv \sigma\{z_k : k \leq t\}$ consists of the full histories of returns as well as predictors that investors use in forecasting returns.

Inspecting the mechanism: a bivariate case

We illustrate the econometrics of the term structure of stock market risk by considering a simple bi-variate first-order VAR for continuously compounded total stock market returns, r_t^s, and the log dividend price,dp_t:

$$\left(z_t - E_z\right) = \Phi_1 \left(z_{t-1} - E_z\right) + \nu_t$$
$$\nu_t \sim \mathcal{N}(0, \Sigma_\nu)$$

where

$$z_t = \begin{bmatrix} r_t^s \\ dp_t \end{bmatrix}, E_z = \begin{bmatrix} E_{r^s} \\ E_{d-p} \end{bmatrix}$$
$$\Phi_1 = \begin{bmatrix} 0 & \varphi_{1,2} \\ 0 & \varphi_{2,2} \end{bmatrix}$$
$$\begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} \sim \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{matrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{matrix} \right]$$

Inspecting the mechanism: a bivariate case

Given the VAR representation and the assumption of constant Σ_ν

$$
\begin{aligned}
Var_t\left[(z_{t+1}+...+z_{t+k}) \mid D_t\right] \quad = \quad & \Sigma_\nu + (I+\Phi_1)\Sigma_\nu(I+\Phi_1)' + \\
& (I+\Phi_1+\Phi_1^2)\Sigma_\nu(I+\Phi_1+\Phi_1^2)' + ... \\
& +(I+\Phi_1+...+\Phi_1^{k-1})\Sigma_\nu(I+\Phi_1+...+
\end{aligned}
$$

from which we can derive:

$$
\begin{aligned}
\Sigma_r(k) &= \frac{1}{k}\sum_{i=0}^{k-1} D_i \Sigma D_i' \\
D_i &= I + \Phi_1 \Xi_{i-1} \quad i > 0 \\
\Xi_i &= \Xi_{i-1} + \Phi_1^i \quad i > 0 \\
D_0 &\equiv I, \ \ \Xi_0 \equiv I
\end{aligned}
$$

Inspecting the mechanism: a bivariate case

in our simple bivariate example, the term structure of stock market risk takes the form

$$\sigma_r^2(k) = \sigma_1^2 + 2\varphi_{1,2}\sigma_{1,2}\psi_1(k) + \varphi_{1,2}^2\sigma_{2,2}^2\psi_2(k)$$

where

$$\psi_1(k) = \frac{1}{k}\sum_{l=0}^{k-2}\sum_{i=0}^{l}\varphi_{2,2}^i \quad k > 1$$

$$\psi_2(k) = \frac{1}{k}\sum_{l=0}^{k-2}\left(\sum_{i=0}^{l}\varphi_{2,2}^i\right)^2 \quad k > 1$$

$$\psi_1(1) = \psi_2(1) = 0$$

The total stock market risk can be decomposed in three components: i.i.d uncertainty, σ_1^2 , mean reversion, $2\varphi_{1,2}\sigma_{1,2}\psi_1(k)$, and uncertainty about future predictors, $\varphi_{1,2}^2\sigma_{2,2}^2\psi_2(k)$.

Inspecting the mechanism: a bivariate case

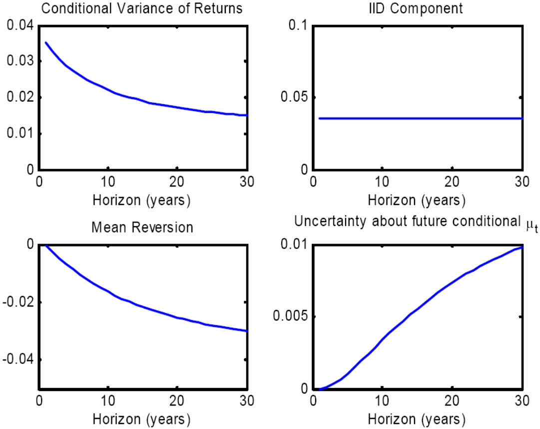
Table 1: A simple bivariate VAR (1910-2008)

$$\left(r^s_{t+1} - E_{r^s}\right) = \varphi_{12}\left(dp_t - E_{dp}\right) + v_{1t+1}$$
$$\left(dp_{t+1} - E_{dp}\right) = \varphi_{22}\left(dp_t - E_{dp}\right) + \nu_{2t+1}$$

$\underset{(t-stat)}{\varphi_{12}}$	$\underset{(t-stat)}{\varphi_{22}}$	$\underset{\varphi_{11}=0,\varphi_{21}=0}{\chi^2_2}$	σ_1	σ_2	$\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}$	$adjR^2_{r^s_{t+1}}$	ac
0.073	0.893	3.128	0.196	0.208	-0.844	0.02	
(1.71)	(19.70)	(0.21)					

Table: The table reports coefficient estimates (with t-statistics in parentheses) and the R^2 statistic for each equation. We also report the standard deviations and correlations of residuals.

Inspecting the mechanism: a bivariate case



Clicker

Insert Clicker 8 here

A VAR with many assets and predictors

$$\mathbf{z}_t = \Phi_0 + \Phi_1 \mathbf{z}_{t-1} + v_t$$

where

$$\mathbf{z}_t = \begin{bmatrix} r_{0t} \\ x_t \\ s_t \end{bmatrix}$$

is a $m \times 1$ vector. with r_{0t} being the log real return on the asset used as a benchmark to compute excess returns on all other asset classes, , x_t being the $n \times 1$ vector of log excess returns on all other asset classes with respect to to the benchmark, and s_t is the $m - n - 1 \times 1$ vector of returns predictors.

v_t is a $m \times 1$ vector of innovations in asset returns and returns' predictors for which standard assumptions apply, i.e.:

$$v_t \sim \mathcal{N}(0, \Sigma_v)$$

where Σ_v is the $m \times m$ variance-covariance matrix.

A VAR with many assets and predictors

Note that

$$\Sigma_v = \begin{bmatrix} \sigma_0^2 & \sigma'_{0x} & \sigma'_{0s} \\ \sigma_{0x} & \Sigma_{xx} & \Sigma'_{xs} \\ \sigma_{0s} & \Sigma_{xs} & \Sigma_{ss} \end{bmatrix}$$

and the unconditional mean and variances-covariance matrix of $\mathbf{z}_t$,assuming that the VAR is stationary and therefore that this moments are well-defined, can be represented as follows:

$$\begin{aligned} \mu_z &= (I_m - \Phi_1)^{-1} \Phi_0 \\ vec(\Sigma_{zz}) &= (I_{m^2} - \Phi_1 \otimes \Phi_1)^{-1} vec(\Sigma_v) \end{aligned}$$

A VAR with many assets and predictors

The conditional mean and variance of the cumulative asset returns at different horizons are instead:

$$E_t(z_{t+1} + ... + z_{t+K}) = \left(\sum_{i=0}^{k-1} (k-i) \, \Phi_1^i \right) \Phi_0 + \left(\sum_{j=0}^{k} \Phi_1^j \right) \mathbf{z}_t$$

$$\begin{aligned} Var_t(z_{t+1} + ... + z_{t+K}) &= \Sigma_v + (I + \Phi_1)\Sigma_v(I + \Phi_1)' + \\ &\quad (I + \Phi_1 + \Phi_1^2)\Sigma_v(I + \Phi_1 + \Phi_1^2)' + ... \\ &\quad +(I + \Phi_1 + ... + \Phi_1^{K-1})\Sigma_v(I + \Phi_1 + ... + \Phi_1^{K-1}) \end{aligned}$$

A VAR with many assets and predictors

Once the conditional moments of excess returns are available the following selector matrix extracts for each period, k-period conditional moments of log real returns:

$$M_r = \begin{bmatrix} 1 & 0_{1xn} & 0_{1x(m-n-1)} \\ \iota_{nx1} & I_{nxn} & 0_{nx(m-n-1)} \end{bmatrix}$$

which implies

$$\frac{1}{k}\begin{bmatrix} E_t\left(r^k_{0,t+1}\right) \\ E_t\left(\mathbf{r}^k_{t+1}\right) \end{bmatrix} = \frac{1}{k}M_r E_t(z_{t+1} + ... + z_{t+K})$$

$$\frac{1}{k}\begin{bmatrix} Var_t\left(r^k_{0,t+1}\right) \\ Var_t\left(\mathbf{r}^k_{t+1}\right) \end{bmatrix} = \frac{1}{k}M_r Var_t(z_{t+1} + ... + z_{t+K})M_r'$$

Therefore after the estimation for the VAR it is possible to derive unconditional and conditional moments for returns and excess returns at all different investment horizons.

Mean-Variance Analysis

The starting point of mean-variance analysis is an expression for the log-returns on the portfolio. The return on the portfolio can be approximated as follows:

$$
\begin{aligned}
r_{p,t+1} &= r_{0,t+1} + \boldsymbol{\alpha}_t' \mathbf{x}_t + \frac{1}{2} \boldsymbol{\alpha}_t' \left(\sigma_x^2 - \Sigma_{xx} \boldsymbol{\alpha}_t \right) \\
\mathbf{x}_t &= \left(\mathbf{r}_{t+1} - r_{0,t+1} \boldsymbol{\iota} \right) \\
\Sigma_{xx} &= Var_t \left(\mathbf{r}_{t+1} - r_{0,t+1} \boldsymbol{\iota} \right) \\
\sigma_x^2 &= diag \left(\Sigma_{xx} \right)
\end{aligned}
$$

given this definition different problems can be addressed:

Mean-Variance Analysis

Campbell-Viceira(2004) show that the optimal weights $\omega_{T,t}$ for the tangency portfolio take the following expression:

$$\begin{aligned}
\omega_{T,t} &= \lambda_f \Sigma_{xx}^{-1} \left[E_t \left(\mathbf{r}_{t+1} - r_{0,t+1} \boldsymbol{\iota} \right) + \frac{1}{2} \boldsymbol{\sigma}_x^2 \right] \\
\lambda_f &= \frac{1}{\left[E_t \left(\mathbf{r}_{t+1} - r_{0,t+1} \boldsymbol{\iota} \right) + \frac{1}{2} \boldsymbol{\sigma}_x^2 \right]' \left(\Sigma_{xx}^{-1} \right)' \boldsymbol{\iota}}
\end{aligned}$$

Consider a k-period horizon we have instead:

$$\begin{aligned}
\omega_{T,t}(k) &= \lambda_f \Sigma_{xx}^{-1}(k) \left[E_t \left(\mathbf{r}_{t+1}^{(k)} - r_{0,t+1}^{(k)} k \boldsymbol{\iota} \right) + \frac{1}{2} \boldsymbol{\sigma}_x^2(k) \right] \\
\lambda_f &= \frac{1}{\left[E_t \left(\mathbf{r}_{t+1}^{(k)} - r_{0,t+1}^{(k)} k \boldsymbol{\iota} \right) + \frac{1}{2} \boldsymbol{\sigma}_x^2(k) \right]' \left(\Sigma_{xx}^{-1}(k) \right)' \boldsymbol{\iota}}
\end{aligned}$$

Mean-Variance Analysis

The typical empirical evidence produced by VAR models is the following term structure of risk: