VAR models in Macro and Finance

Contents

1 VARs and CVARs

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 $\mathbf{y}$:

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

By proceeding in the same way we did for the simple single-equation dynamic model, we can reparameterize the VAR in levels as a model involving levels and the first differences of variables.

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

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

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

$$\Delta\mathbf{y}_t = (\mathbf{A}_1 - \mathbf{I})\,\Delta\mathbf{y}_{t-1} + (\mathbf{A}_1 + \mathbf{A}_2 - \mathbf{I})\,\mathbf{y}_{t-2} + ... + \mathbf{A}_n\mathbf{y}_{t-n} + \mathbf{u}_t. \tag{3}$$

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

$$\begin{aligned}\Delta\mathbf{y}_t &= \mathbf{\Pi}_1\Delta\mathbf{y}_{t-1} + \mathbf{\Pi}_1\Delta\mathbf{y}_{t-2} + ... + \mathbf{\Pi}\mathbf{y}_{t-n} + \mathbf{u}_t \\ &= \sum_{i=1}^{n-1}\mathbf{\Pi}_i\Delta\mathbf{y}_{t-i} + \mathbf{\Pi}\mathbf{y}_{t-n} + \mathbf{u}_t,\end{aligned} \tag{4}$$

where:

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

Clearly the long-run properties of the system are described by the properties of the matrix Π. 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 $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, 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.

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. Here is the intuition on how the tests can be constructed. Having obtained estimates for the parameters in the $\mathbf{\Pi}$ 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 $\ln(1 - \lambda_2) = \ln(1 - \lambda_3) = ... = \ln(1 - \lambda_m) = 0$. 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. Both statistics are small under the null hypothesis. 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. Johansen (1994) has shown in the past some preference for the trace test, based on the argument that the maximum eigenvalue test does not give rise to a coherent testing strategy.

To illustrate briefly the intuition behind the procedure, consider the VAR representation of our simple dynamic model (**??**), introduced in one of the previous sections, for the two variables, x and y:

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}. \tag{5}$$

System (5) can be reparameterized as follows in terms of the VECM representation:

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} a_{11} - 1 & a_{12} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}, \tag{6}$$

from which, clearly,

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

Let us now consider the case when we have more than two variables and work our example on the bond and stock market from the previous section.

1.1 Identification of multiple cointegrating vectors

The Johansen procedure allows us to identify the number of cointegrating vectors. However, in the case of existence of multiple cointegrating vectors, an interesting identification problem arises: α and β are only determined up to the space spanned by them. Thus, for any non-singular matrix ξ conformable by product:

$$\Pi = \alpha\beta' = \alpha\xi^{-1}\xi\beta'.$$

In other words β and $\beta'\xi$ are two observationally equivalent bases of the cointegrating space. The obvious implication is that before solving such an identification problem no meaningful economic interpretation of coefficients in cointegrating vectors can be proposed. The solution is imposing a sufficient number of restrictions on parameters such that the matrix satisfying such constraints in the cointegrating space is unique. Such a criterion is derived by Johansen (1992) and discussed in the works of Johansen and Juselius (1990), Giannini (1992) and Hamilton (1994). Given the matrix of cointegrating vectors $\boldsymbol{\beta}$, we can formulate linear constraints on the different cointegrating vectors using the R_i matrices of dimensions $r_i \times n$. Let us consider the columns of $\boldsymbol{\beta}$, i.e. the parameters in each cointegrating vector, ignoring the normalization constraint to one of one variable in each cointegrating vector. Any structure of linear constraints can be represented as

$$\mathbf{R}_i\boldsymbol{\beta}_i = 0,$$
$$R_i\,(r_i \times n)\,, \quad \boldsymbol{\beta}_i(n \times 1), \quad \text{rank } R_i = r_i.$$

The same constraints can be expressed in explicit forms as

$$\boldsymbol{\beta}_i = \mathbf{S}_i\theta_i,$$

where $S_i\,(n \times (n - r_i))$, $\boldsymbol{\beta}_i(n \times 1)$, $\theta_i\,((n - r_i) \times 1)$, rank $S_i = n - r_i$, $\mathbf{R}_i \mathbf{S}_i = 0$.

A necessary and sufficient condition for identification of parameters in the i-th cointegrating vector is:

$$\text{rank } (\mathbf{R}_i \boldsymbol{\beta}) = r - 1. \tag{7}$$

When (12) is satisfied, it is not possible to replicate the i-th cointegrating vector by taking linear combinations of the parameters in the other cointegrating vectors. In this case, the matrix obtained by applying to the cointegrating space the restrictions of the i-th cointegrating vector has rank $r - 1$.

A necessary condition for identification is immediately derived in that $\mathbf{R}_i \boldsymbol{\beta}$ must have enough rows to satisfy condition (12); therefore, a necessary condition is that each cointegrating vector has at least $r - 1$ restrictions.

A sufficient condition for identification is provided by Johansen by considering the implicit and explicit form of expressing constraints:

Theorem 1 *The i-th cointegrating vector is identified by the constraints* $\mathbf{S}_1$, $\mathbf{S}_2$, ... , $\mathbf{S}_r$, *if for each k=1,...,r-1 and for each set of indices* $1 < j_1 < ... < j_k < r$ *not containing i, we have that rank* $[R_i S_{j_1}, ..., R_i S_{j_k}] > k$

Given identification of the system, we can distinguish the case of just-identification and over-identification. In case of over-identification, the over-identifying restrictions are testable.

To illustrate the procedure let us reconsider our example. Adopting the following vectorial representation of the series: $\left(lp_t\, ld_t\, R_{t,T}\, r_t\right)'$, and leaving aside normalizations, the matrix $\boldsymbol{\beta}$ can be represented as:

$$\begin{pmatrix} \beta_{11} & 0 \\ -\beta_{11} & 0 \\ 0 & \beta_{32} \\ 0 & -\beta_{42} \end{pmatrix}.$$

Given the following general representation of the matrix $\boldsymbol{\beta}$:

$$\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \\ \beta_{41} & \beta_{42} \end{pmatrix},$$

our constraints imply the following specification for the matrices R_i and S_i:

$$R_1 = \begin{pmatrix} 1\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \end{pmatrix}, \quad S_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix},$$

$$R_2 = \begin{pmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,1 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}.$$

The necessary conditions for identification are obviously satisfied, while the sufficient conditions for identification requires that rank $(R_1 S_2) \geq 1$, and rank $(R_2 S_1) \geq 1$. They are also satisfied

$$R_1 S_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \quad R_2 S_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}.$$

1.2 Hypothesis testing with multiple cointegrating vectors

The Johansen procedure allows for testing the validity of restricted forms of cointegrating vectors. More precisely, the validity of restrictions (over-identifying restrictions) in addition to those necessary to identify the long-run equilibria can be tested. The intuition behind the construction of all tests is that when there are r cointegrating vectors, only these r linear combination of variables are stationary; therefore, the test statistics involve comparing the number of cointegrating vectors under the null and the alternative hypotheses. Following this intuition, we understand why only the over-identifying restrictions can be tested. Just-identified models feature the same long-run matrix Π, and therefore, the same eigenvalues of Π. Consider the case of testing restrictions on a set of r identified cointegrating vectors stacked in the matrix $\boldsymbol{\beta}$. The test statistic involves comparing the number of cointegrating vectors under the null and the alternative hypothesis. Let $\widehat{\lambda}_1, \widehat{\lambda}_{2,...}, \widehat{\lambda}_r$ be the ordered eigenvalues of the Π matrix in the unrestricted model, and $\widehat{\lambda}_1^*, \widehat{\lambda}_2^*, ..., \widehat{\lambda}_r^*$ the ordered eigenvalues of the Π matrix in the restricted model. Restrictions on $\boldsymbol{\beta}$ are testable by forming the following test statistic:

$$T\sum_{i=1}^{r}\left[\ln\left(1-\widehat{\lambda}_i^*\right)-\ln\left(1-\widehat{\lambda}_i\right)\right]. \tag{8}$$

Johansen (1992) shows that the statistic (8) has a χ^2-distribution with degrees of freedom equal to the number of over-identifying restrictions. Note that small values of $\widehat{\lambda}_i^*$ with respect to $\widehat{\lambda}_i$ imply a reduction of the rank of Π when the restrictions are imposed and hence the rejection of the null hypothesis. This testing procedure can be extended to tests on restrictions on the matrix of weights $\boldsymbol{\alpha}$ or on the deterministic components (constant and trends) of the cointegrating vectors.

2 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:

$$\begin{aligned} \Delta \mathbf{y}_t &= \boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{y}_{t-1} + \mathbf{u}_t \\ \mathbf{u}_t &\sim N\left(0, \sum\right) \end{aligned} \tag{9}$$

where $\mathbf{y}_t$ is a vector of length N containing the modelled variables. The reduced form specification (9) 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 model.

2.1 An alternative representation of a VECM

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:

$$\begin{pmatrix} \Delta x_t \\ \Delta z_t \end{pmatrix} = \mathbf{\Pi}_1 \begin{pmatrix} \Delta x_{t-1} \\ \Delta z_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \end{pmatrix} \begin{pmatrix} 1 & -\beta \end{pmatrix} \begin{pmatrix} x_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}. \tag{10}$$

$$\begin{pmatrix} \Delta x_t \\ \Delta z_t \end{pmatrix} = \mathbf{\Pi}_1 \begin{pmatrix} \Delta x_{t-1} \\ \Delta z_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \end{pmatrix} S_{t-1} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \tag{11}$$

Define a matrix M such that

$$M \begin{pmatrix} \Delta x_t \\ \Delta z_t \end{pmatrix} = \begin{pmatrix} \Delta x_t \\ \Delta S_t \end{pmatrix}$$

so

$$M = \begin{pmatrix} 1 & 0 \\ 1 & -\beta \end{pmatrix}$$

then we have:

$$
\begin{aligned}
M\begin{pmatrix}\Delta x_t\\ \Delta z_t\end{pmatrix} &= M\mathbf{\Pi}_1\begin{pmatrix}\Delta x_{t-1}\\ \Delta z_{t-1}\end{pmatrix}+M\begin{pmatrix}\alpha_{11}\\ \alpha_{21}\end{pmatrix}S_{t-1}+M\begin{pmatrix}v_{1t}\\ v_{2t}\end{pmatrix}\\
\begin{pmatrix}\Delta x_t\\ \Delta S_t\end{pmatrix} &= M\mathbf{\Pi}_1 M^{-1}\begin{pmatrix}\Delta x_{t-1}\\ \Delta S_{t-1}\end{pmatrix}+M\begin{pmatrix}\alpha_{11}\\ \alpha_{21}\end{pmatrix}S_{t-1}+M\begin{pmatrix}v_{1t}\\ v_{2t}\end{pmatrix}
\end{aligned}
$$

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:

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

3 Identification of VAR

Computing impulse responses to unobservables requires the imposition of some identification assumptions and the orthogonality of structural shocks is a necessary condition to consider the effect of each identified shocks in isolation. The study of the response to the system to an innovation in observables does not require any identification assumption but the contemporaneous linkages between shocks must be modelled.

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. In this context, 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 of the transmission mechanism under empirical investigation, taking into account the potential endogeneity of policy instruments.

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 macroeoconomic variables.Given the estimation of (9) 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:

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

Substituting population moments with sample moments we have:

$$\widehat{\sum} = \widehat{\mathbf{A}}^{-1}\mathbf{BI}\widehat{\mathbf{B}}'\widehat{\mathbf{A}}^{-1}, \tag{12}$$

$\widehat{\sum}$ 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 $n(n+1)/2$, such a condition makes the number of equations equal to the number of unknowns in system (12). As usual, for such a condition also to be sufficient for identification no equation in (12) should be a linear combination of the other equations in the system (see Amisano and Giannini 1996, Hamilton 1994). 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. Once identification has been achieved, the estimation problem is solved by applying generalized method of moments estimation.

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.

4 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. Consider a structural VAR model for a generic vector $\mathbf{y}_t$, containing m variables:

Given an identified and estimated estimate structural VAR

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

we can re-write it as:

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

which we can express in a compact way as:

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

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:

$$\begin{aligned}\mathbf{y}_t&=\mathbf{C}(L)\,\mathbf{v}_t,\\ \mathbf{y}_t&=\mathbf{C}_0\mathbf{v}_t+\mathbf{C}_1\mathbf{v}_{t-1}+...+\mathbf{C}_s\mathbf{v}_{t-s},\\ \mathbf{C}(L)&=[\mathbf{A}_0-\mathbf{A}(L)]^{-1},\\ \mathbf{C}_0&=\mathbf{A}_0^{-1}\mathbf{B}.\end{aligned} \tag{13}$$

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 $t+s$. As s varies we have a function describing the response of variable i to an impulse in variable j. For this function of partial derivative to be meaningful we must allow that a shock to variable j occurs while all other shocks are kept to zero. Of course this is allowed for structural shocks, as they are identified by imposing they are orthogonal to each other. Note, however that the concept of an impulse response function is not applicable to reduced form VAR innovations, which, in general, are correlated to each other.

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 (13) 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:

$$Var\,(\mathbf{y}_{t+s}-E_t\mathbf{y}_{t+s})=\mathbf{C}_0I\mathbf{C}_0'+\mathbf{C}_1I\mathbf{C}_1'+...+\mathbf{C}_sI\mathbf{C}_s'$$

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. Only in this case we can write the variance of the total forecasting error as a sum of variances of the single shocks (as the covariance terms are zero following the orthogonality property of structural shocks).

5 From VAR innovations to structural shocks

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.

5.1 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:

$$\mathbf{A} = \begin{pmatrix} 1 & 0\,0 & 0 \\ a_{21} & 1\,0 & 0 \\ . & .\ 1 & . \\ a_{n1} & .\ a_{nn-1} & 1 \end{pmatrix}, \qquad \mathbf{B} \begin{pmatrix} b_{11} & 0 & 0 & 0 \\ 0 & b_{22} & 0 & 0 \\ . & . & b_{ii} & . \\ 0 & 0 & 0 & b_{nn} \end{pmatrix}. \tag{14}$$

This is obviously a just-identification scheme, where the identification of structural shocks depends on the ordering of variables. It corresponds to a recursive economic structure, with the most endogenous variable ordered last.

Consider for the sake of illustration a bivariate VAR:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}.$$

The MA representation is

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t-1} \\ v_{2t-1} \end{pmatrix}$$
$$+ ... + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^s \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t-s} \\ v_{2t-s} \end{pmatrix},$$

from which impulse response functions, historical decomposition and forecasting error variance decomposition are immediately obtained.

An obvious generalization of Choleski is to consider contemporaneous restrictions that do not necessarily lead to a triangular structure of A. Famous examples of this approach are Blanchard-Perotti, for the analysis of the effects of fiscal policy, and Bernanke and Mihov for the analysis of the monetary policy transmission mechanism in closed economies.

5.2 Structural VARs with long-run restrictions

Often long-run behaviour of shocks provide restrictions acceptable within a wide range of theoretical models. A typical restriction compatible with virtually all macroeconomic models is that in the long-run demand shocks have zero impact on output. Blanchard and Quah (1989) show how these restrictions can be used to identify VARs.

The structural model of interest is specified by posing $\mathbf{A}$ equal to the identity matrix and by imposing no restriction on the $\mathbf{B}$ matrix. We then have the following specification for a generic vector of variables $\mathbf{y}_t$:

$$\mathbf{y}_t = \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{t-i} + \mathbf{B}\mathbf{v}_t$$

from which one can derive the matrix which describes the long-run effect of the structural shocks on the variables of interest as follows:

$$\left(\mathbf{I} - \sum_{i=1}^{p} \mathbf{A}_i\right)^{-1} \mathbf{B}\mathbf{v}_t = -\mathbf{\Pi}^{-1}\mathbf{B}\mathbf{v}_t.$$

Coefficients in $\mathbf{\Pi}$ are obtained from the reduced form, therefore, we are able to impose long-run restrictions given the estimation of the reduced form.

Two points are worth noting:

1. $(I - A_1)$ is -Π, for this matrix to be invertible the VAR must be specified on stationary variables;

2. the long-run restrictions are restrictions on the cumulative impulse response function.

Let us now consider the Blanchard and Quah (1989) dataset. The authors aim at separating demand shocks from supply shocks, they consider a VAR on two variables, the unemployment rate, *UN*, and the quarterly rate of growth of GDP, ΔLY. The VAR is specified with 8 lags, a constant, and a deterministic trend (in the original paper a break in the constant for ΔLY is also allowed but we do not allow it here) as follows:

$$\begin{pmatrix} \Delta LY_t \\ UN_t \end{pmatrix} = A_1 \begin{pmatrix} \Delta LY_{t-1} \\ UN_{t-1} \end{pmatrix} + ... + A_8 \begin{pmatrix} \Delta LY_{t-8} \\ UN_{t-8} \end{pmatrix} + A_9 \begin{pmatrix} 1 \\ t \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}.$$

The structure of interest is the following:

$$\begin{pmatrix} \Delta LY_t \\ UN_t \end{pmatrix} = A_1 \begin{pmatrix} \Delta LY_{t-1} \\ UN_{t-1} \end{pmatrix} + ...A_8 \begin{pmatrix} \Delta LY_{t-8} \\ UN_{t-8} \end{pmatrix} + A_9 \begin{pmatrix} 1 \\ t \end{pmatrix}$$
$$+ \begin{pmatrix} b_{11} \, b_{12} \\ b_{21} \, b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}.$$

To obtain the identifying restrictions consider that

$$\begin{aligned}\begin{pmatrix}\Delta LY_t\\ UN_t\end{pmatrix} &= \left(\mathbf{I}-\sum_{i=1}^{p}\mathbf{A}_i\right)^{-1}\begin{pmatrix}b_{11}\,b_{12}\\ b_{21}\,b_{22}\end{pmatrix}\begin{pmatrix}v_{1t}\\ v_{2t}\end{pmatrix}\\ &= \begin{pmatrix}k_{11}\,k_{12}\\ k_{21}\,k_{22}\end{pmatrix}\begin{pmatrix}b_{11}\,b_{12}\\ b_{21}\,b_{22}\end{pmatrix}\begin{pmatrix}v_{1t}\\ v_{2t}\end{pmatrix}.\end{aligned}$$

Demand shocks are identified by imposing that their long-run impact on the level of output is zero:

$$k_{11}b_{11}+k_{12}b_{21}=0.$$

Note that by imposing the restriction that the cumulative impulse response of the rate of output growth to a demand shock is zero we impose the restriction that the impulse response of the level of output to a demand shock is zero in the long run. As the variables are stationary the long-run response of ΔLY and UN to all shocks is zero by definition.

5.3 CVAR and Identification of shocks

Consider, for simplicity, the case of a bivariate model $\mathbf{y}_t=(y_t,x_t)$, in which variables are non-stationary $I(1)$ but cointegrated with a cointegrating vector $(1,-1)$, so the rank of the $\mathbf{\Pi}$ matrix is 1 and we use the following representation of the stationary reduced form:

$$\begin{pmatrix}\Delta y_t\\ \Delta x_t\end{pmatrix}=\begin{pmatrix}\alpha_{11}\\ \alpha_{21}\end{pmatrix}\begin{pmatrix}1\,-1\end{pmatrix}\begin{pmatrix}y_{t-1}\\ x_{t-1}\end{pmatrix}+\begin{pmatrix}u_{1t}\\ u_{2t}\end{pmatrix} \tag{15}$$

$$\begin{pmatrix}u_{1t}\\ u_{2t}\end{pmatrix}=\begin{pmatrix}b_{11}\,b_{12}\\ b_{21}\,b_{22}\end{pmatrix}\begin{pmatrix}v_{1t}\\ v_{2t}\end{pmatrix}. \tag{16}$$

Model (15) can be re-written as follows :

$$\begin{aligned}\begin{pmatrix}1\,1\\ 0\,1\end{pmatrix}\begin{pmatrix}(1-L)\,0\\ 0\qquad 1\end{pmatrix}\begin{pmatrix}(y_t-x_t)\\ \Delta x_t\end{pmatrix} &= \begin{pmatrix}\alpha_{11}\,0\\ \alpha_{21}\,0\end{pmatrix}\begin{pmatrix}(y_{t-1}-x_{t-1})\\ \Delta x_{t-1}\end{pmatrix}\\ &\quad+\begin{pmatrix}b_{11}\,b_{12}\\ b_{21}\,b_{22}\end{pmatrix}\begin{pmatrix}v_{1t}\\ v_{2t}\end{pmatrix}.\end{aligned} \tag{17}$$

The two representations are completely identical (they feature the same residuals). 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). It seems therefore natural to identify one shock as permanent and the other as transitory. Given that we have a stationary system, the identification of shocks is obtained by deriving long-run responses of the variables of interest to relevant shocks. From (17) we have:

$$\left(\begin{pmatrix} 1\,1 \\ 0\,1 \end{pmatrix}\begin{pmatrix} (1-L)\,0 \\ 0 \quad\quad 1 \end{pmatrix} - \begin{pmatrix} \alpha_{11}L\,0 \\ \alpha_{21}L\,0 \end{pmatrix}\right)\begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} b_{11}\,b_{12} \\ b_{21}\,b_{22} \end{pmatrix}\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix},$$

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

$$\begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} -\alpha_{11}\,1 \\ -\alpha_{21}\,1 \end{pmatrix}^{-1}\begin{pmatrix} b_{11}\,b_{12} \\ b_{21}\,b_{22} \end{pmatrix}\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \tag{18}$$

$$\begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \frac{-b_{11}+b_{21}}{\alpha_{11}-\alpha_{21}} & -\frac{b_{12}-b_{22}}{\alpha_{11}-\alpha_{21}} \\ \frac{-\alpha_{21}b_{11}+\alpha_{11}b_{21}}{\alpha_{11}-\alpha_{21}} & \frac{-\alpha_{21}b_{12}+\alpha_{11}b_{22}}{\alpha_{11}-\alpha_{21}} \end{pmatrix}\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}. \tag{19}$$

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 $\mathbf{B}$. 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}. An application of this identifying scheme is provided in Cochrane(1999) that uses it to identify permanent and transitory components in stock prices.

5.4 Sign Restrictions

Given the VAR specification:

$$\mathbf{y}_t = \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{t-i} + \mathbf{B}\mathbf{u}_t$$
$$\Sigma = \mathbf{BE}\left(\mathbf{u}_t \mathbf{u}_t'\right)\mathbf{B}' = \mathbf{BB}'$$

Consider the Choleski decomposition of Σ, C .

The impulse response function, given the Choleski decomposition could be written as :

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

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

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

The impulse response for $\mathbf{Q}'\mathbf{u}_t$, is then $[\mathbf{I}-\mathbf{A}(L)]^{-1}\mathbf{C}\mathbf{Q}$.

The imposition of the sign restrictions then consider Q to generate all possible identification and then select only those that satisfy some sign restriction.

5.5 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 :

$$
\begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = A \begin{pmatrix} (y_{t-1} - x_{t-1}) \\ \Delta x_{t-1} \end{pmatrix} + \mathbf{u}_t
$$
$$
\mathbf{u}_t \sim N\left(0, \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{pmatrix}\right)
$$

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:

$$
\frac{\partial \begin{bmatrix} (y_{t+i} - x_{t+i}) \\ \Delta x_t + i \end{bmatrix}}{\partial u_{1t}} = A^i S
$$
$$
S = \begin{pmatrix} 1 \\ \left(\sigma_{11}^2\right)^{-1}\sigma_{12} \end{pmatrix}
$$

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. The effect of the shock we are studying with GIRF can be interpreted as the effect on the variables in the model of an intercept adjustment to the particular equation shocked.

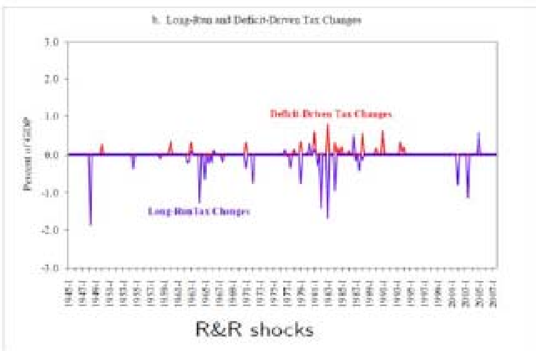
6 Structural Shocks identified independently from VAR

6.1 The R&R narrative approach to fiscal policy shocks

- A time-series of exogenous shifts in taxes is constructed using official documentation, such as congressional reports, etc.to identify the size, timing, and principal motivation for all major postwar tax policy actions
  - legislated tax changes are classified into endogenous (induced by short-run countercyclical concerns) and exogenous, taken to deal with an inherited budget deficit , or driven by concerns about long-run economic growth or politically motivayed: ε_{t-i}^{RR}
- ε_{t-i}^{RR} measure the impact of a tax change at the time it was implemented $(t-i)$ on tax liabilities at time t.
- the effect of ε_{t-i}^{RR} on output is estimated using quarterly data and OLS in a single equation, a truncated (M=12) MA

$$\Delta y_t = a + \sum_{j=0}^{M} b_i e_{t-j}^{RR} + v_t$$

For M=12. Note that this equation is a truncated MA. Impulse responses are read directly off the b_i coefficients.



6.2 The Rudebusch approach to monetary policy shocks

Rudebusch (1998) derives Monetary policy shocks are derived from the thirty-day Federal funds future contracts, which have been quoted on the Chicago Board of Trade since October 1988, and are bets on the average overnight Federal funds rate for the delivery month, the variable included in benchmark VARs. Shocks are constructed as the difference between the Federal funds rate at month t and the thirty-day federal funds future at month $t-1$. Such a choice is based on the evidence that the regression of the Federal funds rate (FF) at t on the thirty-day Federal funds future (FFF) at $t-1$ produces an intercept not significantly different from zero, a slope coefficient not significantly different from one, and serially uncorrelated residuals:

$$\mathrm{FF}_t = -\underset{(0.0436)}{0.037} + \underset{(0.007)}{0.999}\mathrm{FFF}_{t-1} + \widehat{u}_t$$
$$R^2 = 0.99, \qquad \sigma = 0.145, \qquad \mathrm{DW} = 1.86.$$

Note that this procedure produces shocks, labelled FFF, comparable to the reduced form innovations from the VAR and not to the structural monetary policy shocks, because surprises relative to the information available at the end of month $t-1$ may reflect endogenous policy responses to news about the economy that become available in the course of month t. However, if an identification scheme is available, then innovations derived from the future contracts can be transformed in the relevant shocks by applying to them the standard VAR identification procedure. A non-trivial problem with this procedure comes from the fact that Federal funds future are available from 1988 onwards. Future contracts on the one-month Eurodollar are available on a more extended sample. Given that the properties of the series generated by the one-month Eurodollar are very close to the properties of Federal funds future, the direct measurement based on one-month Eurodollar could be used on an extended sample.

7 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:

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{y}_{t-1} + \mathbf{u}_t, \tag{20}$$

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:

$$\mathbf{y}_t = \mathbf{C}^*(L)\,\mathbf{u}_t + \mathbf{C}(1)\,\mathbf{z}_t,$$

where $\mathbf{z}_t$ is a process for which $\Delta\mathbf{z}_t = \mathbf{u}_t$. 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 equation (**??**) is equivalent to:

$$\mathbf{y}_t = (I_m + \boldsymbol{\alpha}\boldsymbol{\beta}')\,\mathbf{y}_{t-1} + \mathbf{u}_t. \tag{21}$$

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

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

Solving this model recursively, we obtain the MA representation for the k cointegrating relationships:

$$\boldsymbol{\beta}'\mathbf{y}_t = \sum_{i=0}^{\infty}(I_k + \boldsymbol{\alpha}\boldsymbol{\beta}')^i\,\boldsymbol{\beta}'\mathbf{u}_{t-i}. \tag{22}$$

By substituting (22) in (**??**) we have the MA representation for $\Delta\mathbf{y}_t$,

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

from which we have

$$\mathbf{C}(1) = I_n - \boldsymbol{\alpha}(\boldsymbol{\beta}'\boldsymbol{\alpha})^{-1}\boldsymbol{\beta}'. \tag{23}$$

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

$$I_n = \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} \alpha'_\perp + \alpha \left(\beta' \alpha\right)^{-1} \beta', \tag{24}$$

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

Using (24) in (23), 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 (**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,

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

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 determining the stochastic trends $\boldsymbol{\alpha}_\perp$. However, there is one aspect in which the two concepts are different. In theory, identified cointegrating relationships on a given set of variables should be robust to augmentation of the information set by adding new variables which should have a zero coefficient in the cointegrating vectors of the VECM representation of the larger information set. This is not true for the stochastic trends. Consider the case of augmenting an information set consisting of m variables admitting k cointegrating vectors to $m+n$ variables. The number of cointegrating vectors is constant while the number of stochastic trends increases by n; moreover, an unanticipated shock in the small system need not be unanticipated in the larger system. Note that we have added 'in theory' to our statement. In practice, given the size of available samples, application of the procedure to analyse cointegration in a larger set of variables might lead the identification of different cointegrating relationships from those obtained on a smaller set of variables.

8 Global VARS

The Global VAR (GVAR) approach advanced in Pesaran, Schuermann and Weiner (2004, PSW) provides a flexible reduced-form framework capable of accommodating a time-varying co-movement across domestic variables and their foreign (in our case euro area) counterpart.

The general specification of a GVAR can be described as follows:

$$\mathbf{x}_{it} = B_{id}\mathbf{d}_t + B_{i1}\mathbf{x}_{it-1} + B_{i0}^{*}\mathbf{x}_{it}^{*} + B_{i1}^{*}\mathbf{x}_{it-1}^{*} + \mathbf{u}_{it}$$

where $\mathbf{x}_{it}$ is a vector of domestic variables, $\mathbf{d}_t$ is a vector of deterministic elements as well as observed common exogenous variables, $\mathbf{x}_{it}^{*}$ is a vector of foreign variables specific to country i. In general $\mathbf{x}_{it}^{*} = \sum_{j\neq i} w_{ji}\mathbf{x}_{jt}$ where w_{ji} is the share of country j in the trade (exports plus imports) of country i. Finally $\mathbf{u}_{it}$ is a vector of country-specific idiosyncratic shocks with $E\left(\mathbf{u}_{it}\mathbf{u}_{jt}'\right) = \Sigma_{ij}, E\left(\mathbf{u}_{it}\mathbf{u}_{jt'}'\right) = 0$, for all i, j and $t \neq t'$.

The construction of the foreign variables allows for the identification of a common component that is different across countries and it is computed as a time-varying linear combination the domestic variables.

Beside being a parsimonious approach to international co-movement the GVAR has also much more flexibility that a VAR in accommodating varying (both in the cross-sectional and in the time-series dimension) co-variation across variables. The GVAR framework can also accommodate long-run solution and the existence of cointegration between the $\mathbf{x}_{it}$ and the $\mathbf{x}_{it}^{*}$. A cointegrating GVAR can be written in VECM format as follows:

$$\Delta\mathbf{x}_{it} = B_{id}\mathbf{d}_t - \Pi_i\mathbf{z}_{it-1} + B_{i0}^{*}\Delta\mathbf{x}_{it}^{*} + \mathbf{u}_{it}$$

where $\mathbf{z}_{it-1} = \left(\mathbf{x}_{it-1}', \mathbf{x}_{it-1}^{*'}\right)'$, $\Pi_i = \left(I - B_{i1}, -B_{i0}^{*} - B_{i1}^{*}\right)$.

9 Finance. Log-linearized Models of Stock and Bond Returns

9.1 Stock Returns and the dynamic dividend growth model

Consider the one-period total holding returns in the stock market, that are defined as follows:[1]

$$H_{t+1}^{s} \equiv \frac{P_{t+1} + D_{t+1}}{P_t} - 1 = \frac{P_{t+1} - P_t + D_{t+1}}{P_t} = \frac{\Delta P_{t+1}}{P_t} + \frac{D_{t+1}}{P_t}, \quad (25)$$

where P_t is the stock price at time t, D_t is the (cash) dividend paid at time t, and the superscript s denotes "stock". The last equality

[1] The use of '$\equiv$' emphasizes that (25) provides a definition. Moreover, ΔX_{t+1} denotes the first difference of a generic variable, or $\Delta X_{t+1} \equiv X_{t+1} - X_t$.

decomposes a discrete holding period return as the sum of the percentage capital gain and of (a definition of) the *dividend yield*, D_{t+1}/P_t. Given that one-period returns are usually small, it is sometimes convenient to approximate them with logarithmic, continuously compounded returns, defined as:

$$r^s_{t+1} \equiv \log\left(1+H^s_{t+1}\right) = \log\left(\frac{P_{t+1}+D_{t+1}}{P_t}\right) = \log\left(P_{t+1}+D_{t+1}\right) - \log\left(P_t\right). \tag{26}$$

Interestingly, while linear returns are additive in the percentage capital gain and the dividend yield components, log returns are not as

$$\log\left(\frac{P_{t+1}+D_{t+1}}{P_t}\right) \neq \log\left(\frac{P_{t+1}}{P_t}\right) + \log\left(\frac{D_{t+1}}{P_t}\right)$$

However, it is still possible to express log returns as a linear function of the log of the price dividend and the (log) dividend growth. Dividing both sides of (25) by $\left(1+H^s_{t+1}\right)$ and multiplying both sides by P_t/D_t we have:

$$\frac{P_t}{D_t} = \frac{1}{\left(1+H^s_{t+1}\right)}\frac{D_{t+1}}{D_t}\left(1+\frac{P_{t+1}}{D_{t+1}}\right).$$

Taking logs (denoted by lower case letters, i.e., $x_t \equiv \log X_t$ for a generic variable X_t), we have:[2]

$$p_t - d_t = -r^s_{t+1} + \Delta d_{t+1} + \ln\left(1+e^{p_{t+1}-d_{t+1}}\right) \tag{27}$$

as $\log(D_{t+1}/D_t) = \log D_{t+1} - \log D_t = \Delta \log D_{t+1} = \Delta d_{t+1}$. Taking a first-order Taylor expansion of the last term about the point $\bar{P}/\bar{D} = e^{\bar{p}-\bar{d}}$ (where the bar denotes a sample average), the logarithm term on the

---

[2] $-r^s_{t+1}$ follows from

$$\begin{aligned} \log\frac{1}{\left(1+H^s_{t+1}\right)} &= \log 1 - \log\left(1+H^s_{t+1}\right) \\ &= -\log\left(1+H^s_{t+1}\right) = -r^s_{t+1} \end{aligned}$$

based on our earlier definitions and the fact that $\log 1 = 0$ for natural logs. Moreover, notice that

$$\frac{P_{t+1}}{D_{t+1}} = e^{\log(P_{t+1}/D_{t+1})} = e^{\log P_{t+1} - \log D_{t+1}} = e^{p_{t+1}-d_{t+1}}$$

right-hand side can be approximated as:

$$\begin{aligned}\ln\left(1+e^{p_{t+1}-d_{t+1}}\right) &\simeq \ln(1+e^{\bar{p}-\bar{d}}) + \frac{e^{\bar{p}-\bar{d}}}{1+e^{\bar{p}-\bar{d}}}[(p_{t+1}-d_{t+1})-(\bar{p}-\bar{d})] \\ &= -\ln(1-\rho)-\rho\ln\left(\frac{1}{1-\rho}-1\right)+\rho(p_{t+1}-d_{t+1}) \\ &= \kappa+\rho(p_{t+1}-d_{t+1})\end{aligned}$$

where

$$\rho \equiv \frac{e^{\bar{p}-\bar{d}}}{1+e^{\bar{p}-\bar{d}}} = \frac{\bar{P}/\bar{D}}{1+(\bar{P}/\bar{D})} < 1 \qquad \kappa \equiv -\ln(1-\rho)-\rho\ln\left(\frac{1}{1-\rho}-1\right).$$

Although $\rho \in (0,1)$ is just a factor that depends on the average price-dividend ratio, in what follows it will be used in a way that resembles a discount factor. At this point, substituting the expression for the approximated term in (27), we obtain that the log price-dividend ratio is defined as:[3]

$$p_t - d_t \simeq \kappa - r^s_{t+1} + \Delta d_{t+1} + \rho(p_{t+1}-d_{t+1}).$$

Re-arranging this expression shows that total stock market returns can be written as:

$$r^s_{t+1} = \kappa + \rho\left(p_{t+1}-d_{t+1}\right) + \Delta d_{t+1} - \left(p_t - d_t\right),$$

or a constant κ, plus the log dividend growth rate (Δd_{t+1}), plus the (discounted, at rate ρ) change in the log price-dividend ratio, $\rho\left(p_{t+1}-d_{t+1}\right) - (p_t-d_t) = \Delta(p_{t+1}-d_{t+1}) - (1-\rho)\left(p_{t+1}-d_{t+1}\right)$. Moreover, by *forward* recursive substitution one obtains:

$$\begin{aligned}(p_t-d_t) &= \kappa - r^s_{t+1} + \Delta d_{t+1} + \rho(p_{t+1}-d_{t+1}) \\ &= \kappa - r^s_{t+1} + \Delta d_{t+1} + \rho\left(\kappa - r^s_{t+2} + \Delta d_{t+2} + \rho(p_{t+2}-d_{t+2})\right) \\ &= (\kappa+\rho\kappa) - (r^s_{t+1}+\rho r^s_{t+2}) + (\Delta d_{t+1}+\rho\Delta d_{t+2}) + \rho^2(p_{t+2}-d_{t+2}) \\ &= (\kappa+\rho\kappa) - (r^s_{t+1}+\rho r^s_{t+2}) + (\Delta d_{t+1}+\rho\Delta d_{t+2}) + \\ &+ \rho^2(\kappa - r^s_{t+3} + \Delta d_{t+3} + \rho(p_{t+3}-d_{t+3})) \\ &= \kappa(1+\rho+\rho^2) - (r^s_{t+1}+\rho r^s_{t+2}+\rho^2 r^s_{t+3}) + (\Delta d_{t+1}+\rho\Delta d_{t+2}+\rho^2\Delta d_{t+3}) + \rho^3(p_{t+3}-d_{t+3}) \\ &= \ldots = \kappa\sum_{j=1}^{m}\rho^{j-1} + \sum_{j=1}^{m}\rho^{j-1}(\Delta d_{t+j} - r^s_{t+j}) + \rho^m\left(p_{t+m}-d_{t+m}\right).\end{aligned}$$

[3] The approximation notation '$\simeq$' appears to emphasize that this expression is derived from an application of a Taylor expansion.

Under the assumption that there can be no rational bubbles, i.e., that[4]

$$\lim_{m \longrightarrow \infty} \rho^m \left(p_{t+m} - d_{t+m}\right) = 0,$$

from

$$\lim_{m \longrightarrow \infty} \sum_{j=1}^{m} \rho^{j-1} = \frac{1}{1-\rho}$$

if $\rho \in (0,1)$, we get

$$(p_t - d_t) = \frac{\kappa}{1-\rho} + \sum_{j=1}^{m} \rho^{j-1} \left(\Delta d_{t+j} - r^s_{t+j}\right).$$

This result shows that the log price-dividend ratio, $(p_t - d_t)$, measures the value of a very long-term investment strategy (buy and hold) which—apart from a constant $\kappa/(1-\rho)$—is equal to the stream of future dividend growth discounted at the appropriate rate, which reflects the risk free rate plus risk premium required to hold risky assets, $r^s_{t+j} \equiv r^f + (r^s_{t+j} - r^f)$.[5] Therefore, for long investment horizons, econometric methods may hope to infer from the data two different types of "information": information concerning the forecasts of future (continuously compounded) dividend growth rates, i.e., Δd_{t+1}, Δd_{t+2}, ..., Δd_{t+m} as $m \longrightarrow \infty$, which are measures of the cash-flows paid out by the risky assets (e.g., how well a company will do); information concerning future discount rates, and in particular future risk premia, i.e., $(r^s_{t+1} - r^f)$, $(r^s_{t+2} - r^f)$, ..., $(r^s_{t+m} - r^f)$ as $m \longrightarrow \infty$. Note that, under the null hypothesis of constancy of returns, the volatility of the price dividend ratio should be completelyt explained by that of the dividend process. The empirical evidence is strongly against this prediction (see the Shiller(1981) and Campbell-Shiller(1987)).

If we decompose future variables into their expected component and the unexpected one (an error term) we can write the relationship between the dividend-yield and the returns one-period ahead and over the long-horizon as follows:

---

[4] This assumption means that as the horizon grows without bounds, the log price-dividend ratio (hence, the underlying price-dividend ratio) may grow without bounds, but this needs to happen at a speed that is inferior to $1/\rho > 1$, so that when $p_{t+m} - d_{t+m}$ is discounted at the rate ρ^m, the limit of the quantity $\rho^m \left(p_{t+m} - d_{t+m}\right)$ is zero.

[5] Here we have assumed that the risk-free interest rate is approximately constant. We shall see that, at least as a first approximation, this is an assumption that holds in practice.

$$r^s_{t+1} = \kappa + \rho E_t\left(p_{t+1} - d_{t+1}\right) + E_t \Delta d_{t+1} - \left(p_t - d_t\right) + \rho u^{pd}_{t+1} + u^{\Delta d}_{t+1}$$

$$\sum_{j=1}^{m} \rho^{j-1} r^s_{t+j} = \frac{\kappa}{1-\rho} + \sum_{j=1}^{m} \rho^{j-1} E_t\left(\Delta d_{t+j}\right) - \left(p_t - d_t\right) + \rho^m E_t\left(p_{t+m} - d_{t+m}\right) + \rho^m u^{pd}_{t+m} + \sum_{j=1}^{m} \rho^{j-1} u^{\Delta d}_{t+j}$$

These two expressions illustrate that when the price dividends ratio is a noisy process, such noise dominates the variance of one-period returnsand the statistical relation between the price dividend ratio and one period returns is weak. However, as the horizon over which returns are defined gets longer, noise tends to be dampened and the predictability of returns given the price dividend ratio increases.

9.2 Bond Returns

The relationship between price and yield to maturity of a constant coupon (C) bond is given by:

$$P^c_{t,T} = \frac{C}{\left(1 + Y^c_{t,T}\right)} + \frac{C}{\left(1 + Y^c_{t,T}\right)^2} + ... + \frac{1 + C}{\left(1 + Y_{t,T}\right)^{T-t}}.$$

When the bond is selling at par, the yield to maturity is equal to the coupon rate. To measure the length of time that a bondholder has invested money for we need to introduce the concept of duration:

$$\begin{aligned}
D^c_{t,T} &= \frac{\frac{C}{\left(1+Y^c_{t,T}\right)} + 2\frac{C}{\left(1+Y^c_{t,T}\right)^2} + ... + (T-t)\frac{1+C}{\left(1+Y_{t,T}\right)^{T-t}}}{P^c_{t,T}} \\
&= \frac{C\sum_{i=1}^{T-t}\frac{i}{\left(1+Y^c_{t,T}\right)^i} + \frac{(T-t)}{\left(1+Y_{t,T}\right)^{T-t}}}{P^c_{t,T}}.
\end{aligned}$$

Note that when a bond is floating at par we have:

$$\begin{aligned}
D^c_{t,T} &= Y^c_{t,T}\sum_{i=1}^{T-t}\frac{i}{\left(1+Y^c_{t,T}\right)^i} + \frac{(T-t)}{\left(1+Y_{t,T}\right)^{T-t}} \\
&= Y^c_{t,T}\frac{\left((T-t)\frac{1}{1+Y^c_{t,T}} - (T-t) - 1\right)\frac{1}{\left(1+Y^c_{t,T}\right)^{T-t+1}} + \frac{1}{1+Y^c_{t,T}}}{\left(1 - \frac{1}{1+Y^c_{t,T}}\right)^2} + \frac{(T-t)}{\left(1+Y_{t,T}\right)^{T-t}} \\
&= \frac{1 - \left(1+Y^c_{t,T}\right)^{-(T-t)}}{1 - \left(1+Y^c_{t,T}\right)^{-1}},
\end{aligned}$$

because when $|x| < 1$,

$$\sum_{k=0}^{n} k x^k = \frac{(nx - n - 1)\, x^{n+1} + x}{(1-x)^2}.$$

Duration can be used to find approximate linear relationships between log-coupon yields and holding period returns. Applying the log-linearization of one-period returns to a coupon bond we have:

$$\begin{aligned} p_{c,t,T} - c &= -r^c_{t+1} + k + \rho\,(p_{c,t+1,T} - c) \\ r^c_{t+1} &= k + \rho p_{c,t+1,T} + (1-\rho)\, c - p_{c,t,T}. \end{aligned}$$

When the bond is selling at par, $\rho = (1+C)^{-1} = \left(1 + Y^c_{t,T}\right)^{-1}$. Solving this expression forward to maturity delivers:

$$r^c_{t+1} = D^c_{t,T} y^c_{t,T} - \left(D^c_{t,T} - 1\right) y^c_{t+1,T},$$

9.3 A simple model of the term structure

Consider the relation between the return on a riskless one period short-term bill, r_t, and a long term bond bearing a coupon C, the one-period return on the long-term bond $H_{t,T}$ is a non-linear function of the log yield to maturity $R_{t,T}$. Shiller (1979) proposes a *linearization* which takes duration as constant and considers the following approximation in the neighborhood $y_{t,T} = y_{t+1,T} = \bar{y} = C$:

$$\begin{aligned} H_{t,T} &\simeq D_T y_{t,T} - (D_T - 1)\, y_{t+1,T} \\ D_T &= \frac{1 - \gamma^{T-t-1}}{1-\gamma} = \frac{1}{1-\gamma_T} \\ \gamma_T &= \left\{1 + \bar{y}\left[1 - 1/(1+\bar{y})^{T-t-1}\right]^{-1}\right\}^{-1} \\ \lim_{T\longrightarrow\infty} \gamma_T &= \gamma = 1/(1+\bar{y}) \end{aligned}$$

solving this expression forward we generate the equivalent of the DDG model in the bond market:

$$y_{t,T} = \sum_{j=0}^{T-t-1} \gamma^j (1-\gamma)\, H_{t+j,T} + \gamma^{T-t} y_{T-1,T}$$

In this case, by equating one-period risk-adjusted returns, we have:

$$E\left[\frac{y_{t,T} - \gamma y_{t+1,T}}{1-\gamma} \mid I_t\right] = r_t + \phi_{t,T} \tag{28}$$

From the above expression, by recursive substitution, under the terminal condition that at maturity the price equals the principal,we obtain:

$$y_{t,T} = y^*_{t,T} + E[\Phi_T \mid I_t] = \frac{1-\gamma}{1-\gamma^{T-t}} \sum_{j=0}^{T-t-1} \gamma^j E[r_{t+j} \mid I_t] + E[\Phi_T \mid I_t] \quad (29)$$

where the constant $\Phi_{t,T}$ is the term premium over the whole life of the bond:

$$\Phi_{t,T} = \frac{1-\gamma}{1-\gamma^{T-t}} \sum_{j=0}^{T-t-1} \gamma^j \phi_{t+j,T}$$

For long-bonds, when $T-t$ is very large, we have :

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

Subtracting the risk-free rate from both sides of this equation we have:

$$\begin{aligned} S_{t,T} = y_{t,T} - r_t &= \sum_{j=1}^{T-1} \gamma^j E[\Delta r_{t+j} \mid I_t] + E[\Phi_T \mid I_t] \\ &= S^*_{t,T} + E[\Phi_T \mid I_t] \end{aligned}$$

10 Linearized Present Value models for Consumption

The accumulation equation for aggregate wealth may be written as:

$$W_{t+1} = (1 + R_{m,t+1})(W_t - C_t) \quad (30)$$

Define $r_{m,t+1} = \log(1 + R_{m,t+1})$, and use lowercase letters to denote log variables throughout. As LL we follow Campbell and Mankiw (1989) and assume that the consumption–aggregate wealth ratio is stationary. In this case the budget constraint may be approximated by taking a first-order Taylor expansion of equation (30) , to obtain

$$\begin{aligned} \Delta w_{t+1} &= r_{m,t+1} + k + \left(1 - \frac{1}{\rho}\right)(c_t - w_t) \quad (31) \\ \rho &= 1 - \exp\left(\overline{c-w}\right) \end{aligned}$$

where k, is a constant of normalization, not relevant for the problem at hand.

By solving (31)forward, we have :

$$c_t - w_t = E_t \left[\sum_{j=1}^{\infty} \rho^j \left(r_{m,t+j} - \Delta c_{t+j} \right) \right] + \frac{\rho k}{1-\rho} \tag{32}$$

LL point out that (32) shows that the consumption–wealth ratio is a function of expected future returns to the market portfolio in a broad range of optimal consumption models, so they concentrate in finding a proxy for $c_t - w_t$ and in assessing its performance for forecasting market returns.

To illustrate how we take equation (**??**) to the data note that, following Campbell(1996), we approximate the log of total wealth as:

$$w_t = v a_t + (1-v) h_t$$

where v is a constant of linearization, equal to the average share of asset holdings in total wealth, a_t is the log of asset holdings and h_t is the log of human capital. While we have available data for financial wealth, the measurement of h_t is not immediate. To find an empirical counterpart of this variable consider that labour income can be interpreted as a dividend on human capital (see Julliard(2004)):

$$1 + R_{h,t+1} = \frac{H_{t+1} + Y_{t+1}}{H_t}$$

Log-linearizing this relation around the steady state human capital-labor income ratio ($\frac{Y}{H} = \frac{1}{\rho_h} - 1$) we have:

$$r_{h,t+1} = (1-\rho_h) k_h + \rho_h (h_{t+1} - y_{t+1}) - (h_t - y_t) + \Delta y_{t+1}$$

By solving this relation forward and by imposing the transversality condition we have:

$$h_t = y_t + \sum_{i=1}^{\infty} \rho_h^{i-1} \left(\Delta y_{t+i} - r_{h,t+i} \right) + k_h$$

so the log of human capital to income ratio is determined by discounted sum of future labour income growth and human capital returns.

Consistently with our linearization for wealth, the total return on wealth can be approximated by:

$$r_{m,t} = v r_{a,t} + (1-v) r_{h,t} + k_r$$

we decompose the unobservable $r_{h,t}$ into a part correlated with $r_{a,t}$ and a part orthogonal to it:

$$r_{h,t} = \beta r_{a,t} + \epsilon_t$$

By substituting all these relationships in the optimality condition we have:

$$\begin{aligned}
c_t - va_t - (1-v)\, y_t = & (1-\psi)\, E_t \left[\sum_{j=1}^{\infty} \rho^j \left(v + (1-v)\, \beta \right) r_{a,t+j} \right] + k + \\
& \sum_{j=1}^{\infty} E_t \rho_h^{j-1} \left(\Delta y_{t+j} - \beta r_{a,t+j} \right) + \eta_t \qquad (33) \\
\eta_t = & \sum_{j=1}^{\infty} \rho_h^{j-1} \epsilon_{t+j}
\end{aligned}$$

where η_t is an unobservable stationary component.

11 Cointegration and Present Value Models

CS tests the ET[6] by considering the case of the risk free rate and a very long term bond.

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

Subtracting the risk-free rate from both sides of this equation we have:

$$\begin{aligned}
S_{t,T} = y_{t,T} - r_t = & \sum_{j=1}^{T-1} \gamma^j E[\Delta r_{t+j} \mid I_t] + E[\Phi_T \mid I_t] \\
= & S^*_{t,T} + E[\Phi_T \mid I_t]
\end{aligned}$$

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

$$S_{t,T} = S^*_{t,T} = \sum_{j=1}^{T-1} \gamma^j E[\Delta r_{t+j} \mid I_t] \qquad (34)$$

[6] In fact CS use de-meaned-variables, that is equivalent to test a weak form of the Expectations Theory, in the sense that de-meaning eliminates a constant risk premium.

(34) shows that a necessary condition for the ET to hold puts constraints on the long-run dynamics of the spread. In fact, the spread should be stationary being a weighted sum of stationary variables. Obviously, stationarity of the spread implies that, if yields are non-stationary, they should be cointegrated with a cointegrating vector (1,-1). However, the necessary and sufficient conditions for the validity of the ET impose restrictions both on the long-run and the short run dynamics.

Assuming[7] that $R_{t,T}$ and r_t are cointegrated with a cointegrating vector (1,-1), CS construct a bivariate stationary VAR in the first difference of the short-term rate and the spread :

$$\begin{aligned} \Delta r_t &= a(L)\Delta r_{t-1} + b(L)S_{t-1} + u_{1t} \\ S_t &= c(L)\Delta r_{t-1} + d(L)S_{t-1} + u_{2t} \end{aligned} \tag{35}$$

Stack the VAR as:

$$\begin{bmatrix} \Delta r_t \\ . \\ . \\ \Delta r_{t-p+1} \\ S_t \\ . \\ . \\ S_{t-p+1} \end{bmatrix} = \begin{bmatrix} a_1 & . & . & a_p & b_1 & . & . & b_p \\ 1 & . & . & 0 & 0 & . & . & 0 \\ 0 & . & . & 0 & 0 & . & . & 0 \\ 0 & . & 1 & 0 & 0 & . & . & 0 \\ c_1 & . & . & c_p & d_1 & . & . & d_p \\ 0 & . & . & 0 & 1 & . & . & 0 \\ 0 & . & . & 0 & 0 & . & . & 0 \\ 0 & . & . & 0 & 0 & . & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta r_{t-1} \\ . \\ . \\ \Delta r_{t-p} \\ S_{t-1} \\ . \\ . \\ S_{t-p} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ . \\ . \\ 0 \\ u_{2t} \\ . \\ . \\ 0 \end{bmatrix} \tag{36}$$

This can be written more succinctly as:

$$z_t = Az_{t-1} + v_t \tag{37}$$

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

$$g' z_t = \sum_{j=1}^{T-1} \gamma^j h' A^{j\prime} z_t \tag{38}$$

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). Since the above expression has to hold for general z_t, and, for large T, the sum converges under the null of the validity of the ET, it must be the case that:

$$g' = h'\gamma A(I - \gamma A)^{-1} \tag{39}$$

[7] In fact, the evidence for the restricted cointegrating vector which constitutes a necessary condition for the ET to hold is not found to be particularly strong in the original CS work.

which implies:

$$g'(I - \gamma A) = h'\gamma A \tag{40}$$

and we have the following constraints on the individual coefficients of VAR(**??**):

$$\{c_i = -a_i, \forall i\}, \{d_1 = -b_1 + 1/\gamma\}, \{d_i = -b_i, \forall i \neq 1\} \tag{41}$$

The above restrictions are testable with a Wald test. By doing so using US data between the fifties and the eighties Campbell and Shiller (1987) rejected the null of the ET. However, when CS construct a theoretical spread $S_{t,T}^*$, by imposing the (rejected) ET restrictions on the VAR they find that, despite the statistical rejection of the ET, $S_{t,T}^*$ and $S_{t,T}$ are strongly correlated.

1080 JOURNAL OF POLITICAL ECONOMY

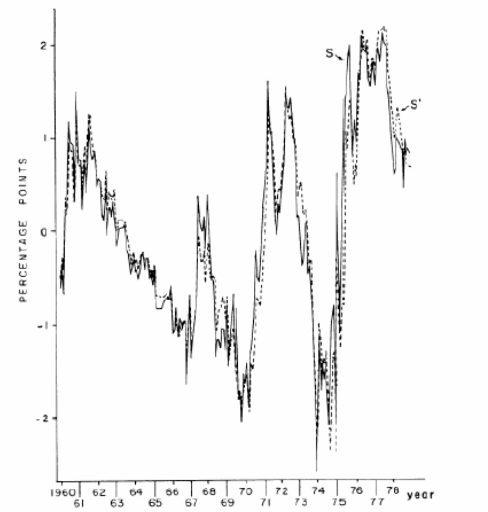


FIG. 1.—Term structure: deviations from means of long-short spread S_t and theoretical spread S_t'.

Things look very different for the stock market when the dynamic dividend growth model with constant rates of return is considered. In this case we have:

$$(p_t - d_t)^* = \sum_{j=1}^{m} \rho^{j-1} (\Delta d_{t+j})$$

and the variable $(p_t - d_t)^*$ can be obtained by imposing the appropriate cross-equation restrictions on a bivariate VAR for the dividend-yield and dividend growth. The relation between the actual and the "theory-consistent" price-dividends looks much different than what it had been obtained for the bond market. This result is consistent from the evidence of predictive regressions relating the dividend price to future returns rather than to future dividend growth.

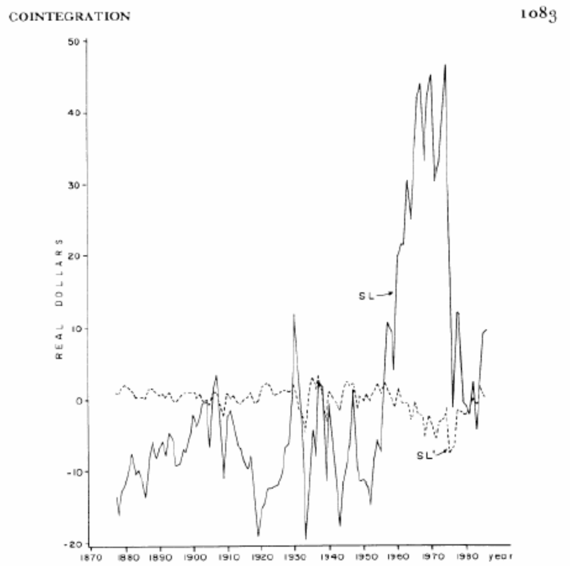


FIG. 2.—Stock market: deviations from means of actual spread (SL_t = Price$_t$ − θ · Dividend$_{t-1}$) and theoretical spread SL'_t, θ = 12.195.

12 Macro. Model Evaluation of a Simple DSGE Model

We consider a small New Keynesian DSGE model of the economy which features a representative household optimizing over consumption, real money holdings and leisure, a continuum of monopolistically competitive firms with price adjustment costs and a monetary policy authority which sets the interest rate. Furthermore, the model is driven by three exogenous processes which determine government spending, g_t, the stationary component of technology, z_t, and the policy shock, $\epsilon_{R,t}$.

A full description of the model can be found in Woodford (2003). Here, we mainly focus on its log-linear representation which takes each variable as deviations from its trend. The model has a deterministic steady state with respect to the de-trended variables: the common component is generated by a stochastic trend in the exogenous process for technology. The model follows Del Negro and Schorfheide (2004) (henceforth, DS) and it reads

$$\tilde{x}_t = E_t\tilde{x}_{t+1} - \frac{1}{\tau}(\tilde{R}_t - E_t\tilde{\pi}_{t+1}) + (1-\rho_G)\tilde{g}_t + \rho_z\frac{1}{\tau}\tilde{z}_t \tag{42}$$

$$\tilde{\pi}_t = \beta E_t\tilde{\pi}_{t+1} + \kappa\left(\tilde{x}_t - \tilde{g}_t\right) \tag{43}$$

$$\tilde{R}_t = \rho_R\tilde{R}_{t-1} + (1-\rho_R)(\psi_1\tilde{\pi}_t + \psi_2\tilde{x}_t) + \epsilon_{R,t} \tag{44}$$

$$\tilde{g}_t = \rho_g\tilde{g}_{t-1} + \epsilon_{g,t} \tag{45}$$

$$\tilde{z}_t = \rho_z\tilde{z}_{t-1} + \epsilon_{z,t} \tag{46}$$

where $\tilde{x}_t$ is the output gap, $\tilde{\pi}_t$ is the inflation rate, $\tilde{R}_t$ is the short-term interest rate and $\tilde{g}_t$ and $\tilde{z}_t$ are two AR(1) stationary processes for government and technology, respectively.

The first equation is an intertemporal Euler equation obtained from the households' optimal choice of consumption and bond holdings. There is no investment in the model and so output is proportional to consumption up to an exogenous process that can be interpreted as time-varying government spending. The net effects of these exogenous shifts on the Euler equation are captured in the process $\tilde{g}_t$. The parameter $0 < \beta < 1$ is the households' discount factor and $\tau > 0$ is the inverse of the elasticity of intertemporal substitution. The second equation is the forward-looking Phillips curve which describes the dynamics of inflation and κ determines the degree of the short-run trade-off between output and inflation.

The third equation describes the behavior of the monetary authority. The central bank follows a nominal interest rate rule by adjusting its instrument to deviations of inflation and output from their respective target levels. The shock $\epsilon_{R,t}$ can be interpreted as unanticipated deviation from the policy rule or as policy implementation error. The set of structural shocks is thus $\epsilon_t = (\epsilon_{R,t}, \epsilon_{g,t}, \epsilon_{z,t})'$ which collects technology, government and monetary shocks.

The model needs to be solved and this can be done by applying the algorithm proposed by Sims (2002). Define the vector of variables as $\tilde{Z}_t = \left(\tilde{x}_t\,\tilde{\pi}_t\,\tilde{R}_t\,\tilde{R}_t^*\,\tilde{g}_t\,\tilde{z}_t\,E_t\tilde{x}_{t+1}\,E_t\tilde{\pi}_{t+1}\right)$ and the vector of shocks as $\epsilon_t = \left(\epsilon_{R,t}\,\epsilon_{g,t}\,\epsilon_{z,t}\right)$. Therefore the previous set of equations, (42) - (46), can be recasted into a set of matrices $(\Gamma_0, \Gamma_1, C, \Psi, \Pi)$ accordingly to the definition of the vectors $\tilde{Z}_t$ and ϵ_t

$$\Gamma_0\tilde{Z}_t = C + \Gamma_1\tilde{Z}_{t-1} + \Psi\epsilon_t + \Pi\eta_t \tag{47}$$

where η_{t+1}, such that $E_t\eta_{t+1} \equiv E_t\left(y_{t+1} - E_t y_{t+1}\right) = 0$, is the expectations error[8].

[8]See Appendix A for a detailed derivation.

As a solution to (47), we obtain the following policy function

$$\tilde{Z}_t = T(\theta)\,\tilde{Z}_{t-1} + R(\theta)\,\epsilon_t \tag{48}$$

and in order to provide the mapping between the observable data and those computed as deviations from the steady state of the model we set the following measurement equations as in DS

$$\Delta \ln x_t = \ln \gamma + \Delta \tilde{x}_t + \tilde{z}_t \tag{49}$$

$$\Delta \ln P_t = \ln \pi^* + \tilde{\pi}_t \tag{50}$$

$$\ln R_t = 4[(\ln R^* + \ln \pi^*) + \tilde{R}_t] \tag{51}$$

which can be also cast into matrices as

$$Y_t = \Lambda_0(\theta) + \Lambda_1(\theta)\,\tilde{Z}_t + v_t \tag{52}$$

where $Y_t = (\Delta \ln x_t, \Delta \ln P_t, \ln R_t)'$, $v_t = 0$ and Λ_0 and Λ_1 are defined accordingly. For completeness, we write the matrices T, R, Λ_0 and Λ_1 as a function of the structural parameters in the model, $\theta = \left(\ln \gamma, \ln \pi^*, \ln r^*, \kappa, \tau, \psi_1, \psi_2, \rho_R, \rho_g, \rho_Z, \sigma_R, \sigma_g, \sigma_Z\right)'$: such a formulation derives from the rational expectations solution.

The evolution of the variables of interest, Y_t, is therefore determined by (48) and (52) which impose a set of restrictions across the parameters on the moving average (MA) representation. Given that the MA representation can be very closely approximated by a finite order VAR representation, DS propose to evaluate the DSGE model by assessing the validity of the restrictions imposed by such a model with respect to an unrestricted VAR representation. The choice of the variables to be included in the VAR is however completely driven by those entering in the DSGE model regardless of the statistical goodness of the unrestricted VAR.

References

[1] Barberis, N., (2000), "Investing for the long run when returns are predictable ", *Journal of Finance*, vol.55, no.1, pages 225-264.

[2] Bec, F., and C. Gollier, (2008), "Asset returns volatility and investment horizon: The French case", *working paper.*

[3] Campbell J., (1999), "Asset prices, consumption, and the business cycle", *Handbook of Macroeconomics*, vol.1, pages 1231-1303.

[4] Campbel J., A. Lo and A. MacKinley, (1997), *The Econometrics of Financial Markets*, Princeton University Press.

[5] Campbell, J., and L. Viceira, (2002), *Strategic Asset Allocation*, Oxford University Press.

[6] Campbell, J., and L. Viceira, (2005), "The term structure of risk-return trade-off", *NBER working paper*, no. 11119.
[7] Del Negro, Marco, and Frank Schorfheide (2004): " Priors from General equilibrium Models for VARs," *International Economic Review*, 45, 643-673.
[8] Del Negro, Marco, and Frank Schorfheide (2006): "How Good is what You've Got? DSGE-VAR as a Toolkit for evaluating DSGE Models," *Federal Reserve Bank of Atlanta Economic Review*
[9] Del Negro, Marco, Frank Schorfheide, Frank Smets and Raf Wouters (2006): "On the Fit of New-Keynesisan Models," *Journal of Business, Economics and Statistics, 25,2, 124-162*.
[10] Merton, R.C., (1969), "Lifetime portfolio selection under uncertainty: the continuous-time case", *Review of Economics and Statistics*, vol. 51, no. 3, pages 247-257.
[11] Merton, R.C., (1973), "An intertemporal capital asset pricing model", *Econometrica*, vol. 41, no. 5, pages 867-887.
[12] Samuelson, P., (1969) "Lifetime portfolio selection by dynamic stochastic programming ", *Review of Economics and Statistics*, vol.51, no.3, pages 239-246.
[13] Stambaugh, R.F., (1999), "Predictive regressions", *Journal of Financial Economics*, vol. 53, no. 3, pages, 375-421.
[14] Ang, Andrew and Geert Bekaert, 2007, Stock Return Predictability: Is It There? The Review of Financial Studies, 20, 651–707.
[15] Ang, A. and A. Maddaloni."Do Demographic Changes Affect Risk Premiums? Evidence from International Data." *Journal of Business*, 78 (2005), 341-380.
[16] Bakshi, Gurdip S., and Zhiwu Chen, 1994, Baby Boom, Population Aging, and Capital Markets, Journal of Business, 67, 2, 165-202.
[17] Van Binsbergen J.H. and R.S.Koijen, 2009, Predictive Regression:a Present Value Approach, mimeo
[18] Boudoukh, Jacob, Richardson, Matthew and Robert F. Whitelaw, 2008, The Myth of Long-Horizon Predictability, The Review of Financial Studies, 21, 4, 1577-1605.
[19] Campbell, J. Y., 1991. A variance decomposition for stock returns. Economic Journal 101, 157–179.
[20] Campbell, J. Y., Lo, A. W., MacKinlay, C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton, NJ.
[21] Campbell, John Y., and Samuel B. Thomson, 2008, Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?, The Review of Financial Studies, 21, 1509-1531.
[22] Campbell, John Y., and Robert Shiller, 1988, Stock Prices,Earnings, and Expected Dividends, Journal of Finance, 43,

661–676.
[23] Campbell, John Y., and Luis M. Viceira, 2002, Strategic Asset Allocation: Portfolio Choice for Long-Term Investors (Oxford University Press).
[24] Campbell, John Y., and Luis M. Viceira, 2005, The Term Structure of the Risk-Return Tradeoff, Financial Analyst Journal, 61, 1, 34-44
[25] Cochrane, J.H., 1994 Permanent and Transitory Components in GNP and Stock Prices, Quarterly Journal of Economics
[26] Cochrane, J. H., 2001. Asset Pricing. Princeton University Press, Princeton, NJ.
[27] Cochrane, John H., 2008, The Dog that Did Not Bark: A Defense of Return Predictability, Review of Financial Studies, 20, 5.
[28] Cochrane J.H., 2008, State-Space versus VAR models of Stock Returns, mimeo
[29] DellaVigna S., and J.Pollet. "Demographics and Industry Returns." *American Economic Review*, 97(2007), 1167-1702.
[30] Erb, C. B.; C. R. Harvey, and T. E. Viskanta. "Demographics and International Investment." *Financial Analysts Journal*, 53 (1996), 4,14-28.
[31] Fama, Eugene and Kenneth R. French, 1988, Dividend Yields and Expected Stock Returns, Journal of Financial Economics, 22, 3-26.
[32] Favero C.A., A.Gozluklu and A.Tamoni (2010) "Demographic Trends, the Dividend-Price Ratio and the Predictability of Long-Run Stock Market Returns" , CEPR working paper 7734, forthcoming in the Journal of Financial and Quantitative Analysis
[33] Geanakoplos, John, Magill, Michael and Martine Quinzii, 2004, Demography and the Long Run Behavior of the Stock Market, Brookings Papers on Economic Activities, 1: 241-325.
[34] Goyal, A., Welch, I., May 2003. Predicting the equity premium with dividend ratios. Management Science 49 (5), 639–654.
[35] Goyal, A. "Demographics, Stock Market Flows, and Stock Returns." *Journal of Financial and Quantitative Analysis*, 39 (2004), 1, 115-142.
[36] Goyal, Amit, and Ivo Welch, 2008, A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. The Review of Financial Studies, 21-4, 1455-1508
[37] Hodrick, Robert, 1992, Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, Review of Financial Studies, 5, 357–386.
[38] Lacerda, F., and Santa-Clara, P. 2010, Forecasting Dividend Growth to Better Predict Returns, Discussion paper.
[39] Lettau, Martin, and Sydney Ludvigson, 2005, Expected Returns

and Expected Dividend Growth, Journal of Financial Economics, 76, 583-626.

[40] Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, Aggregate Wealth and Expected Stock Returns, Journal of Finance, 56, 3, 815-849.

[41] Lettau, Martin, and Stijn Van Nieuwerburgh, 2008, Reconciling the Return Predictability Evidence, Review of Financial Studies, 21, 4, 1607-1652.

[42] Lewellen, Jonathan, 2004, Predicting Returns with Financial Ratios. Journal of Financial Economics 74:209–35.

[43] Nelson, C. C., Kim, M. J., 1993. Predictable stock returns: The role of small sample bias. Journal of Finance 43, 641–661.

[44] Neely C.J., Weller P.A.(2000). Predictability in international asset returns: a reexamination. Journal of Financial and Quantitative Analysis, 35, 601-620

[45] Newey, Whitney K. and Kenneth D. West, 1987, A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica, 55, 3, 703-08.

[46] Newey, Whitney K. and Kenneth D. West, 1994, Automatic Lag Selection in Covariance Matrix Estimation, Review of Economic Studies, 61, 631-653.

[47] Pastor, Lubos, and Robert F. Stambaugh, 2008, Predictive systems: Living with imperfect predictors, Journal of Finance

[48] Pastor, Lubos and Robert F. Stambaugh, 2009, Are Stocks Really Less Volatile in the Long Run?,Working Paper 14757, http://www.nber.org/papers/w14757

[49] Paye, B.S., Timmermann A.(2006), Instability of return prediction models. Journal oF Empirical Finance, 13, 274-315

[50] Pesaran M.H., A.Pick, and A.G.Timmermann(2010), Variable Selection Estimation and Inference for Multi-period forecasting problems, CEPR working paper 7139

[51] Poterba, J. M. "Demographic Structure and Asset Returns." *The Review of Economics and Statistics*, 83 (2001), 4, 565-584.

[52] Rapach D.E., Wohar M.E., 2006, In-sample vs. out-of-sample tests of stock return predictability in the context of data-mining, Journal of Empirical Finance, 13, 231-247

[53] Robertson D. and S.Wright, 2009, The Limits to Stock Return Predictability, mimeo

[54] Oleg Rytchkov, 2008, Filtering Out Expected Dividends and Expected Returns, Working paper.

[55] Rudebusch, G. D. (1998). 'Do Measures of Monetary Policy in a VAR Make Sense?'. *International Economic Review*, 39: 907-931

[56] Siegel Jeremy.J.(1998), Stocks For the Long-Run, 2nd edition, McGraw-Hill
[57] Shiller, Robert J., 2005, Irrational Exuberance, second edition, Princeton University Press.
[58] Schotman, Peter C., Rolf Tschernig, and Jan Budek, Long Memory and the Term Structure of Risk, Journal of Financial Econometrics, Fall 2008, 6 (4), 459495.
[59] Stambaugh, R. F., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.
[60] Taleb, N.N., 2001, Fooled By Randomness. The Hidden Role of Chance in the Markets and in Life, TEXERE XLC
[61] Valkanov, R., 2003. Long-horizon regressions: Theoretical results and applications. Journal of Financial Economics 68, 201–232. 33