Papers ESADE núm. 161
noviembre, 1998

VAR MACROECONOMETRICS: FORMULATION, SPECIFICATION, ESTIMATION, IDENTIFICATION AND USES

Fernando C. Ballabriga

Depósito Legal: B-4.761-1992
ISSN: 1132-7278

Index

Abstract

This paper contains a detailed description of the VAR methodology,[1] starting with a brief historical reference, continuing with the topics of formulation, specification, estimation and identification of VAR models, and concluding with a description of their uses.

1 Canova (1995) provides an alternative detailed description of this methodology.

Introduction

Vector Autoregressions (VAR) were introduced by C.A. Sims in two seminal papers. In Sims (1972) a VAR model was used to analyse the long-standing issue of the existence of causality between aggregate money and income. In Sims (1980) VAR models were explicitly proposed as an alternative to standard macroeconometric simultaneous equation models, with the basic argument that these models were specified and identified using incredible economic restrictions. A reasonable alternative, Sims argued, should refrain from using controversial economic restrictions, relying on the data as the most important source of information to identify key macroeconomic interactions. In these terms, VAR models appeared as suitable alternatives.

The proposal soon encountered obstacles that were the source of discussion and research in VAR literature during the 1980s and 1990s. The first was the issue of identification: as raw VAR models are reduced forms they do not go beyond correlations and so cannot be used for macroeconomic analysis. The second obstacle was actually a paradox: VAR methodology aims to avoid prior exclusions, but because of its generous parameterization nature this means that VAR models are not actually operational alternatives to standard macroeconometric models as there are usually too few degrees of freedom if the analyst includes more than a relatively reduced number of variables in the model. In fact, it is uncommon to find VAR models with more than five or six variables.

Both these problems have been reasonally well solved, to the extent that nowadays it is quite widely accepted that VAR methods are among the most successful innovations in empirical macroeconomics of the last two decades.

1. VAR Macroeconometrics

1.1. A Brief History

The Cowles Commission for economic research was headquartered in Chicago from 1939 to 1945. During this time, and particularly during the 1940s, the members established the basis for what we will refer to as "conventional econometrics". More specifically, the Commission made two fundamental contributions to the field of econometrics: it campaigned for the use of statistical inference in economics and developed simultaneous equation models, addressing their identification, estimation and validation.

For three decades, the Commission's economic principles formed the framework of consensus for the profession and monopolised the theory and practice of econometrics. In the specific field of macroeconomics, Klein (1947) marked a turning point in terms of constructing macroeconometric models that could be useful in economic policy-making processes. These models became progressively larger and were systematically used to quantify the macroeconomic impact of different scenarios, which were defined in terms of alternative paths of the model's exogenous variables.

In the second half of the 1970s, two U.S. authors wrote articles that were to become classics: Lucas (1976) and Sims (1980) questioned the uses and principles of the basic construction of conventional macroeconometric models. Their criticism of conventional modelling strategy was so harsh that the authors themselves felt justified in abandoning these models and undertaking a search for alternative strategies which would correct what they considered to be unacceptable features of conventional methodology. Indeed, their articles had a tremendous impact in the United States, triggering the research programmes advocated by the authors. This paper focuses on Sims' criticism of the conventional models.

What did Sims criticise?

He openly criticised the methods used to construct conventional models, arguing as follows:

The validity of the restrictions used to obtain a structural interpretation is crucial if you aim to defend the idea that there is some connection between reality and the model used to represent it. Sims considered that the great majority of the restrictions used to identify conventional macroeconometric models were "incredible" because they had no justification in economic theory. In truth, he argued, there are few powerful sources of restriction that permit identification. Among them is the distinction between technological factors and preferences. In contrast, the number of variables and equations generally included in conventional models is high in comparison. In particular, the much-heralded exogeneity of most of the variables is invention rather than fact.

This is illustrated by looking at the following econometric model:

$$Y_1(t)=F(Y_1(t-s)\,,\,Y_2(t-s),s\geq 0;\delta_F)+u_1(t)$$
$$Y_2(t)=G(Y_2(t-s)\,,\,Y_1(t-s),s\geq 0;\delta_G)+u_2(t) \qquad (1)$$

where t is a time index and, for the sake of convenience, the model's variables have been divided into a Y_1 variable, which represents the private sector, and a Y_2 variable which is the vector of control variables used by economic policy makers. In conventional modelling it has long been common practice to treat the control vector as exogenous, i.e. reducing (1) to the following restricted specification:

$$Y_1(t)=F(Y_1(t-s)\,,\,Y_2(t-s),s\geq 0;\delta_F)+u_1(t)$$
$$Y_2(t)=G(Y_2(t-s)\,,\,s\geq 0;\delta_G)+u_2(t) \qquad (2)$$

on the assumption that the $u_1(t)$ and u_2(t) disturbances are orthogonal. Certainly, the exogeneity of Y_2 guarantees the identification of the F and G equation blocks, but the assumption is very likely unjustified inasmuch as the people in charge of controlling Y_2 obviously respond to the private sector events reflected in the variability of Y_1.

When model identification rests on such a fragile base, its implications in terms of underlying economic interdependencies are hard to take seriously, which disqualifies the model as a tool for empirical analysis.

The alternative proposed by Sims (1980) involved specifying and estimating macroeconometric models that did not include controversial prior restrictions. In fact, he proposed specifying minimally restricted models in which all variables with a clearly economic content would be treated as endogenous. The resulting models are now known as Vector Autoregressions (VAR). Going back to (1), if we assume that s>0, we can write:

$$Y_1(t) = F(Y_1(t-s)\ ,\ Y_2(t-s),\ s>0;\ \beta_F) + \varepsilon_1(t)$$
$$Y_2(t) = G(Y_2(t-s)\ ,\ Y_1(t-s),\ s>0;\ \beta_G) + \varepsilon_2(t) \qquad (3)$$

Further assuming that F (·) and G (·) are linear and that the vector of stochastic disturbance (ε_1, ε_2) is a white noise process, (3) is the VAR representation of endogenous variables (Y_1 , Y_2).

Although frequently poorly interpreted, the philosophy underlying the VAR methodology has always been the same and progress made during the 1980s and 1990s has made it gradually easier to understand. VAR methodology involves accepting the challenge implicit in acknowledging that considerable uncertainty exists as to the true data generating process. The immediate consequence of such an acknowledgement is that an appropriate modelling strategy should enable the analyst to explicitly introduce this uncertainty into the model specification process so that it can be treated systematically and objectively. And this is precisely the idea that justifies the insistence that there be few prior restrictions so that attempts can be made to extract the relevant empirical regularities by giving the economic data the most objective reading possible.

This idea constantly underlies the methods and uses described in the remainder of this paper.

1.2. VAR models

1.2.1. Formulation

In the most common formulation, the autoregressive representation of a stochastic process Y of dimension n is as follows for all t:

$$Y(t) = B(L)\ Y(t) + DZ(t) + \varepsilon(t) \tag{4}$$

where $B(L) = \sum_{s=0}^{m} B_s L^s$ is a polynomial matrix in the lag operator $L(L^s\ Y(t) = Y(t\text{-}s))$, with B_s n x n, $B_0 = 0$, and m being the number of lags in each of the endogenous n components of the Y vector included in the model; Z is a vector with d deterministic components and D, n x d, is its matrix of coefficients and ε is a white noise vector process of dimension n with a zero mean and a covariance matrix Σ. Vector AutoRegression is a natural name for (4) when you note that it relates a vector of variables with its own past.

Alternately, if we go back to the parametrisation implicit in (3), clearly specifying and adding the deterministic component, the autoregressive representation of the stochastic vector Y can be formulated in the following terms:

$$Y(t) = X(t\text{-}1)\beta + \varepsilon(t) \tag{5}$$

where

$$X(t-1)_{n\,x\,nk} = \begin{pmatrix} X_1(t-1)' & 0' & \dots & 0' \\ 0' & X_2(t-1)' & \dots & 0' \\ \dots & \dots & \dots & \dots \\ 0' & 0' & \dots & X_n(t-1)' \end{pmatrix}$$

$$X_i(t-1)_{k\,x\,1} = \begin{pmatrix} Y(t-1) \\ Y(t-2) \\ . \\ . \\ . \\ Y(t-m) \\ Z(t) \end{pmatrix}, i=1,\dots,n \quad ; \quad \beta_{nk\,x\,1} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_n \end{pmatrix}$$

every block of zeros in $X(t\text{-}1)$ is a vector of dimension k, $k=nm+d$, just as are the sub-vectors β_i, i = 1,..., n, which contain stacked i^{th} rows of the matrices of coefficients B_s, s = 1,..., m, and D in (4). The sign "$'$" indicates transposition.

As tools for representing stochastic processes, VAR models provide a very general theoretical framework. Granger and Newbold (1986) state that if the number of lags (m) is not restricted and the possibility that the model's coefficients depend on t is considered, any stochastic process (stationary or otherwise) can be represented as an autoregressive model.[2] In addition to sharing the minimally restrictive spirit of the methodology, this general rule makes VAR models attractive points of departure for econometric modelling and a framework of reference that reveals the restrictions actually included in alternative models, inasmuch as any simultaneous equation models or time series econometric models can be expressed in the reduced form (4)-(5).

1.2.2. The Unrestricted VAR model

The Unrestricted Vector Autoregression model (UVAR) is obtained when the number of lags (m) included in the autoregressive representation (4)-(5) is sufficient to guarantee that the structure of the residuals is statistically compatible with the white noise hypothesis. The adjective "unrestricted" reflects the fact that the UVAR model includes a minimum of the prior restrictions needed to become operative: the selection of a set of variables (n), specification of the algebraic relation (linear) that connects them

---

2 Although (4) and (5) do not include the possibility of time varying, this will be explicitly discussed somewhat further on.

and a para-metrisation (k) which gives the analyst a sufficient degree of freedom to produce statistically acceptable estimates.

The UVAR model has largely been used in applications of VAR methodology. There are probably two reasons for this: the first is that it is sufficiently general to be widely used as a tool for representing stochastic processes and the other is that it is easy to estimate, using the ordinary least squares (OLS) method, which we will now discuss.

Estimation

Let us take as our reference the linear regression framework used in (5). As is well known, the OLS method has desirable properties in this framework. Specifically, it is the most efficient unbiased linear estimator for small samples, as well as being consistent and asymptotically normal.

Do these properties survive in the UVAR framework?

In order to answer this question, note that the difference between the UVAR and linear regression frameworks is that the UVAR's explanatory variables are stochastic while in linear regression they are deterministic. The deterministic character of the explanatory variables in the linear regression model ensures that the OLS estimator is unbiased in small samples, giving rise to the Gauss-Markov theorem, and also facilitates the derivation of its asymptotic properties.

When explanatory variables are stochastic, the analyst must distinguish between their stationarity or non-stationarity and consider their relation to the model error term. In particular, if the model considered is stationary and the explanatory variables are statistically independent of the error term, the OLS estimator maintains all the properties it had in the deterministic framework of regression.

Certainly the UVAR model can be stationary but its explanatory variables are not independent of the error term. More concretely, if we have T observations generated by model (5), we have:

$$E\left[\varepsilon(t-s)'\, X(t-1)\right] \neq 0, \quad s \geq 2 \tag{6}$$

Although this characteristic is an added complication, it is not insurmountable. Indeed, and lagged correlations in (6) aside, it is true that *a)* the UVAR model's vector of disturbances by definition constitutes a sequence of random independent vectors and, in consequence, *b)* there is no correlation between the current vector of the disturbance and the model's explanatory variables; that is to say:

$$E\left[\varepsilon(t)'\, X(t-1)\right] = 0 \tag{7}$$

Conditions *a)* and *b)* are sufficient to be able to apply the Mann-Wald and Cramér theorems (see, for example, Harvey [1981]). When combined, these theorems can be used to demonstrate that in a stationary model with stochastic explanatory variables, the OLS estimator asymptotically maintains the same properties as in the linear regression framework. A stationary UVAR model can therefore be estimated by using the OLS method.

In the non-stationary framework, the existence of unit roots and possible cointegration relations among the components of vector Y does not affect the asymptotic properties of the OLS estimator. Sims, Stock and Watson (1990) demonstrate that the estimator is consistent and Park and Phillips (1989) and Ahn and Reinsel (1990) demonstrate that it has the same asymptotic properties as the maximum likelihood estimator when cointegration restrictions are taken into account.

Yet another question that arises when estimating a UVAR model is whether applying the OLS method to each of the system's n equations is not less efficient than estimating them globally. More specifically, let us assume that we have a T-size sample and, in accordance with the notation in (5), we represent the T observations of our model in the following form:

$$\underset{Tn\,x\,1}{Y} = X\beta + \underset{Tn\,x\,1}{\varepsilon} \tag{8}$$

where

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{pmatrix} = \begin{pmatrix} Y_1(1) \\ \cdot \\ Y_1(T) \\ Y_2(1) \\ \cdot \\ Y_2(T) \\ \cdot \\ \cdot \\ Y_n(1) \\ \cdot \\ Y_n(T) \end{pmatrix} ; \quad \varepsilon \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} \varepsilon_1(1) \\ \cdot \\ \varepsilon_1(T) \\ \varepsilon_2(1) \\ \cdot \\ \varepsilon_2(T) \\ \cdot \\ \cdot \\ \varepsilon_n(1) \\ \cdot \\ \varepsilon_n(T) \end{pmatrix}$$

$$\underset{Tn\,x\,kn}{X} = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & X_n \end{pmatrix}$$

$$\underset{Tn\,x\,kn}{X_i} = \begin{pmatrix} X_i(1-1)' \\ X_i(2-1)' \\ \cdot \\ \cdot \\ \cdot \\ X_i(T-1)' \end{pmatrix} = \begin{pmatrix} Y(1-1)' & \ldots & Y(1-m)' & Z(1)' \\ Y(2-1)' & \ldots & Y(2-m)' & Z(2)' \\ \ldots & \ldots & \ldots & \ldots \\ Y(T-1)' & \ldots & Y(T-m)' & Z(T)' \end{pmatrix} \quad ; i=1,\ldots,n$$

Alternately, we can separate the n regression models included in (8) and write:

$$\underset{T\,x\,1}{Y_i} = X_i \beta_i + \underset{T\,x\,1}{\varepsilon_i} \quad ; i=1,\ldots,n \tag{9}$$

where

$$Y_i = \begin{pmatrix} Y_i(1) \\ \cdot \\ \cdot \\ \cdot \\ Y_i(T) \end{pmatrix}; \varepsilon_i = \begin{pmatrix} \varepsilon_i(1) \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_i(T) \end{pmatrix}$$

The question which then arises is whether applying the OLS method to each of the n models in (9) is or is not less efficient than applying the Generalised Least Squares (GLS) method in (8). In order to formally express this, we will write the covariance structure of the disturbances in (9) as follows:

$$Cov(\varepsilon_i, \varepsilon_j) = E\varepsilon_i \varepsilon_j' = \underset{T\,x\,T}{\Gamma_{ij}} \quad ; i, j=1,\ldots,n \tag{10}$$

where

$$\Gamma_{ij} = \sigma_{ij} I_T$$
$$[\sigma_{ij}] = \Sigma$$

Likewise, the covariance structure of the disturbances in (8) can be written as:

$$Cov(\varepsilon) = E\varepsilon\varepsilon' = \underset{Tn\,x\,Tn}{\Gamma} \tag{11}$$

where

$$\Gamma = [\Gamma_{ij}] = \Sigma \otimes I_T$$

and $\otimes$ represents the Kronecker product.

The GLS estimator for β (β_{GLS}) is then:

$$\beta_{GLS} = (X' \Gamma^{-1} X)^{-1} X' \Gamma^{-1} Y \tag{12}$$

While the OLS estimator is:

$$\beta_{OLS} = \begin{pmatrix} (X_1' X_1)^{-1} X_1' Y_1 \\ \vdots \\ (X_n' X_n)^{-1} X_n' Y_{n1} \end{pmatrix} \tag{13}$$

and our question is whether β_{GLS} is more efficient than β_{OLS}. Once again, the regression theory provides a suitable frame of reference for answering this question. In this case, the most useful frame of reference is SURE (Seemingly Unrelated Regressions).

A standard result in the SURE framework is that the GS and OLS estimates coincide when there is no contemporaneous correlation between the components of the error term (Σ is diagonal in our notation) and/or when the explanatory variables are the same in each of the n equations. There is no increase in efficiency in either of these cases (see Harvey [1981], for example).

In a UVAR model Σ is rarely diagonal, but in contrast all the equations share the same explanatory variables. That is to say, $X_i = \overline{X}$, i=1,...,n. We therefore write:

$$\beta_{GLS} = \begin{pmatrix} (\overline{X}' \overline{X})^{-1} \overline{X}' Y_1 \\ \vdots \\ (\overline{X}' \overline{X})^{-1} \overline{X}' Y_n \end{pmatrix} = \beta_{OLS} \tag{14}$$

The result (14) justifies the common practice of using a single equation to estimate UVAR systems.

1.2.3. The Bayesian VAR model

As can be seen from (4) and (5), the generality of the autoregressive representation is based on broad parametrisation: the number of coefficients increases exponentially with the number of variables included and proportionate to the number of lags of each variable.

The UVAR model described in the preceding section aims to directly profit from this generality without any additional restriction being imposed on the lag structure once *m* has been selected. As a result, the model tends to be highly parametrised. Note, for example, that a UVAR model with five endogenous variables, only four lags and a constant term per equation will have a total of 105 coefficients to be estimated.

Highly parametrised models are not, however, the most suitable for empirical analysis of macroeconomic data because macroeconomic information tends to be scarce and, furthermore, has a high content of random variability. A highly parametrised model in conjunction with scarce and extremely random sample information causes overfit, i.e. the resulting model is overfitted in that it basically reflects random empirical variability (noise) rather than systematic variability (signal).

In this context, analysts who want to use the UVAR framework to analyse macroeconomic series are forced to specify models of a limited size. Indeed, it is not customary to encounter UVAR applications that contain more than five or six variables. This is a truly paradoxical obstacle: as we have already mentioned, VAR methodology aims to provide an alternative to conventional econometric modelling by avoiding controversial prior exclusions, but the UVAR framework is not really an operational alternative because its generous parametrisation quickly uses up the available degrees of freedom, even in models of a limited size, and the resulting models are over-parametrised.

Litterman (1980) and Doan, Litterman and Sims (1984) proposed using the Bayesian dimension of VAR methodology in an attempt to solve the problem of over-fit in UVAR models other than by using the standard solution that resorts to economic theory and statistical contrasts as sources of exclusion restrictions and is commonly used in conventional simultaneous equation econometric models. In consonance with its spirit of minimum restriction, BVAR methodology aims to avoid the influence of random variability in estimation without forcing the analyst to choose whether to include or exclude lags of the different variables. Consequently, the resulting model maintains the prior generality of the autoregressive representation.

The Bayesian solution appears natural when one realises how unsatisfactory it is to have to decide whether to include/exclude lags in situations where the analyst never knows in advance and for certain if the value of a particular coefficient is zero or when he is not absolutely ignorant of the value of the coefficients of the

model, situations which are common in econometric analysis. With the Bayesian approach such exclusions can be avoided and the available prior information can be more realistically expressed by assigning probability distributions to the coefficients of the model.

More concretely, the aforementioned authors proposed to complement the autoregressive representation with a prior distribution of its coefficients. Without being either diffuse or placing the entire weight on a single value, this alternative offers a reasonable range of uncertainty and therefore can be altered by the sample information when there are substantial differences in both sources of information. So long as the prior information is not excessively diffuse, it will probably only be altered by systematic, not random, variability, thus reducing the risk of over-fitting.

Putting this idea into practice involves formally specifying a probability distribution for the vector of coefficients β and combining it with (4)-(5). The result of this combination is a BVAR (Bayesian Vector Autoregression) model.

Estimation

With the Bayesian approach, β is a random vector rather than a vector of parameters. Therefore, characterising the stochastic behaviour of $Y(t)$ conditional on $X(t\text{-}1)$ requires explicit assumptions as regards both β and $\varepsilon(t)$. In the BVAR approach, the following hypotheses are common:

$$\begin{gathered} \beta \mid X(t-1) \sim N(\bar{\beta}(t-1), \Omega(t-1)) \\ \varepsilon(t) \mid X(t-1) \sim N(0, \Sigma) \\ \beta \text{ and } \varepsilon(t) \text{ are independent} \end{gathered} \tag{15}$$

The hypothesis of normality is not inevitable, but it is certainly convenient. In other words, what we are really searching for is a flexible model that will allow us to include prior information in our analysis. The hypothesis of normality enables us to exploit the convenient properties of the Gaussian framework in order to obtain this model.

Let us begin by noting that, from a Bayesian standpoint, the problem of estimating our econometric model boils down to applying the Bayes theorem in order to obtain in every t the posterior distribution of $[\beta \mid X(t\text{-}1), Y(t)]$ by combining the prior distribution of $[\beta \mid X(t\text{-}1)]$ in (15) and the sample information. We will concentrate first on obtaining the posterior distribution and then proceed to discuss the choice of prior information.[3]

---

3 From the strictly Bayesian viewpoint Σ would also be part of the problem of estimation. In other words, the problem would involve obtaining a posterior distribution from a prior distribution for $[\beta, \Sigma \mid X(t\text{-}1)]$. However, literature on BVAR models customarily proceeds by conditioning in Σ and focusing attention on the vector of coefficients β. This paper will work within this framework.

Theil's mixed estimation technique (Theil [1971]) provides a suitable framework for obtaining the posterior distribution of our vector of coefficients. On the one hand, it permits us to combine different pieces of available information (in this case, prior and sample information) and, on the other, it can be interpreted in Bayesian terms.[4] In order to apply the mixed estimation technique, we first need to express our prior information in the form of dummy observations. Note specifically that the distribution in the first line of equation (15) can be expressed as follows:

$$r(t\text{-}1) = R(t\text{-}1)\,\beta + \eta(t\text{-}1) \tag{16}$$

where

$$\underset{nk\,x\,1}{r(t-1)} = W(t-1)^{-1}\ \bar{\beta}(t-1)$$
$$\underset{nk\,x\,nk}{R(t-1)} = W(t-1)^{-1}$$
$$W(t-1)\ W(t-1)' = \Omega(t-1)$$
$$\eta(t-1) \sim N(0,I)$$

As mentioned earlier, (16) is our prior information about the vector of coefficients β. A second piece of information stems from (5), which defines the connection between the vector of observables $Y(t)$ and β, and which we repeat below for the sake of convenience:

$$Y(t) = X(t\text{-}1)\,\beta + \varepsilon(t) \tag{17}$$

The vector of disturbances $\varepsilon(t)$ is characterised by the second line of (15). According to the third line, it is independent of the vector of disturbances $\eta(t\text{-}1)$ in (16).

The linear stochastic restrictions (16) and (17) contain the information on β which is available in t, and can be combined as follows:

$$\begin{bmatrix} r(t-1) \\ Y(t) \end{bmatrix} = \begin{bmatrix} R(t-1) \\ X(t-1) \end{bmatrix} \beta + \begin{bmatrix} \eta(t-1) \\ \varepsilon(t) \end{bmatrix} \tag{18}$$

---

4 Posterior distribution can be obtained in several different ways. For example, Ballabriga (1991,1997) uses the updating scheme provided by the Kalman filter.

where

$$\begin{bmatrix} \eta(t-1) \\ \varepsilon(t) \end{bmatrix} \sim N\left(0, \begin{bmatrix} I & 0 \\ 0 & \Sigma \end{bmatrix}\right)$$

Theil's mixed estimator, β, β_{MIX}, is then obtained by applying the GLS method to (18). The result is as follows:

$$\beta_{MIX}(t) = \left[\Omega(t-1)^{-1} + X(t-1)'\Sigma^{-1}X(t-1)\right]^{-1}\left[\Omega(t-1)^{-1}\bar{\beta}(t-1) + X(t-1)'\Sigma^{-1}Y(t)\right] \tag{19}$$

$$Cov\left(\beta_{MIX}(t)\right) = \left[\Omega(t-1)^{-1} + X(t-1)'\Sigma^{-1}X(t-1)\right]^{-1}$$

The question now is what connection there is between the estimators in (19) and the posterior distribution of $[\beta \mid X(t\text{-}1), Y(t)]$. And the answer is obtained by giving Theil's mixed estimation technique a Bayesian interpretation, considering (16) as a second sample independent of the observables $[Y(t), X(t\text{-}1)]$ in (17). Using the prior information included in the form of a dummy sample in (18), we shall proceed as though our prior information about β was diffuse (uniform). By then combining the likelihood of (18) with β's diffuse prior information we obtain its posterior distribution, which turns out to be normal, with the mean and variance expressed in (19) (see Theil [1971]). In other words, we write:

$$\left[\beta \mid X(t-1), Y(t)\right] \sim N\left(\bar{\beta}(t),\ \Omega(t)\right) \tag{20}$$

where

$$\bar{\beta}(t) = \beta_{MIX}(t)$$
$$\Omega(t) = Cov\left(\beta_{MIX}(t)\right)$$

We can therefore conclude that conditioning on Σ, (19) actually provides a way to update prior information on $[\beta \mid X(0)]$. This enables us to process the sample T observations in order to obtain the posterior distribution of $[\beta \mid X(T)]$, thus completing the Bayesian estimation process.

Time variation of the coefficients

The description of the BVAR methodology thus far has been based on the assumption that the vector of coefficients β has a distribution that does not vary over time and which successive sample observations can estimate with progressively greater precision. However, analysts commonly believe that their sample data might contain structural changes. This belief can be explicitly

included in the model by accepting as part of the prior information the possibility that the distribution of the vector of coefficients β might change over time.

Time variation is a standard characteristic of BVAR models. It makes the specification process more flexible and provides a useful way to detect possible structural changes in the sample without having to explicitly model the source of the change (for example, a change in economic policy regime).

The most common way to incorporate time variation into the BVAR framework is by specifying the law of motion of β as a first order autoregression process. This law of motion is considered sufficient to detect possible shifts in the model's linear structure and also permits the analysis to be kept within the Gaussian framework. In fact, the framework described in the foregoing section can be easily generalised in order to incorporate this type of time variation. Specifically, the model's vector of coefficients now takes the following form:

$$\beta(t) = \begin{bmatrix} \beta_1(t) \\ \beta_2(t) \\ \cdot \\ \cdot \\ \cdot \\ \beta_n(t) \end{bmatrix} \tag{21}$$

where the added time index indicates that the stochastic properties of the vector depend on time. As a consequence, the characterisation of the stochastic behaviour of $Y(t)$ conditioned in $X(t\text{-}1)$ requires that the hypotheses contained in (15) be extended in order to account for this time dependence. The extension used is the following:

$$\begin{gathered} \beta(t) = S\beta(t\text{-}1) + u(t) \\ \beta(t-1) \mid X(t-1) \sim N(\bar{\beta}(t-1), \Omega(t-1)) \\ u(t) \mid X(t-1) \sim N(0, \varphi) \\ \varepsilon(t) \mid X(t-1) \sim N(0, \Sigma) \\ \beta(t\text{-}1),\ u(t) \text{ and } \varepsilon(t) \text{ are independent} \end{gathered} \tag{22}$$

where S and φ are square matrices of order nk.

The prior distribution of $[\beta(t\text{-}1) \mid X(t\text{-}1)]$ (i.e., previously the first line of (15)) is now obtained by combining the first three lines of (22), which gives us:

$$\left[\beta(t-1) \mid X(t-1)\right] \sim N(\beta^*(t-1), \Omega^*(t-1)) \qquad (23)$$

where

$$\beta^*(t-1) = S\bar{\beta}(t-1)$$
$$\Omega^*(t-1) = S\Omega(t-1)\,S' + \varphi$$

The analysis contained in the foregoing section is thus valid if in all t, $\bar{\beta}(t-1)$ and $\Omega(t-1)$ are replaced by $\beta^*(t-1)$ and $\Omega^*(t-1)$, respectively, resulting in the following updating scheme:

$$\beta_{MIX}(t) = \left[\Omega^*(t-1)^{-1} + X(t-1)'\Sigma^{-1}X(t-1)\right]^{-1}\left[\Omega^*(t-1)^{-1}\beta^*(t-1) + X(t-1)'\Sigma^{-1}Y(t)\right] \qquad (24)$$

$$Cov\left[\beta_{MIX}(t)\right] = \left[\Omega^*(t-1)^{-1} + X(t-1)'\Sigma^{-1}X(t-1)\right]^{-1}$$

Lastly, note that the time variation framework generates, as a particular case, the framework without time variation when S is the identity and φ is the null matrix, in which case the hypotheses in (22) are identical to the hypotheses in (15) and the updating schemes (19) and (24).

Prior information

As mentioned in reference to (19), if the updating scheme in (24) is to be operational the first sample period (t=1) must necessarily include an initial specification for the matrix Σ and the prior distribution in (23), which in itself requires that the matrixes S, φ and $\Omega(0)$, and the vector $\beta(0)$, be specified. This initial specification is what defines the model's prior information.

The choice of prior information is, without a doubt, the most distinctive feature of the BVAR model specification process. This information can, in principle, take a variety of forms and come from a variety of sources and it is precisely this that makes the method attractive. However, and as mentioned earlier, within the framework of BVAR methodology its principal aim is to reduce the risk of overfit without affecting the generality of the model. In this sense, it is purely instrumental information which, as such, does not pretend to be true on average, but does aim to provide a realistic range of data-generating mechanisms, from among which analysts can choose the most appropriate for explaining the variability of their sample data.

Moreover, and in consonance with their instrumental nature, the source of prior information that is common to the BVAR framework is statistical-empirical and lacks economic content. This economic "neutrality" aims to make the resulting specification acceptable to a wide spectrum of analysts, whose visions about the true structure of the economy analysed might differ significantly.

Specifically, the backbone of prior information consists of three empirical regularities which are characteristic of statistical time series analysis.

1) The current value of a series is the best way to predict its future value (random walk). This hypothesis reflects the behaviour of many economic time series.

2) Recent lagged values of a time series usually contain more information about their current value than lagged values dating from further back in time.

3) The lagged values of a time series usually contain more information about their current value than the lagged values of other variables.

As indicated by the formal description of the model contained in the foregoing paragraphs of this section, and in particular (23), prior information has been included by means of specifying a normal multivariate distribution. As previously mentioned, the assumption of normality is not inevitable. We would like to stress this once again: the real aim is to ensure that the prior information contains regularities 1) to 3) above. The assumption of normality allows us to include them and furthermore facilitates analysis by enabling us to use the Gaussian framework. In fact, the most direct procedure is by defining (23) in t=1 as a set of normal independent nk distributions, one for each of the model's coefficients, parametrised individually in accordance with regularities 1) to 3). However, this strategy of individual parametrisation leads to overfit, which is precisely what we are striving to avoid.

An alternative is to assume prior independence among nk distributions, but introducing a functional dependence among all of them and a reduced set of parameters which makes it possible to control their basic dimensions so that they reflect regularities 1) to 3). In BVAR jargon these parameters are called hyperparameters in order to distinguish them from the term parameter as used in standard econometrics.

Figure 1 illustrates the prior density function for a representative equation of system (5) and shows how to include regularities 1) to 3):

1) is included by specifying an average equal (or close to) one for the distribution of the coefficient of the first own lag and equal to zero for the remaining coefficients.

2) is included by reducing distribution variance as the lag increases so that the greater the lag, the greater the likelihood that the distribution of its coefficient will be around zero.

Lastly, 3) is introduced by assigning a higher value to own lags (row 1) than to the lags of other variables (row 2), increasing the likelihood of these latter being zero.

Figure 1 also gives an idea of the nature of the set of control hyperparameters. Thus, one of the hyperparameters usually controls the average coefficient of the first own lag. A second hyperparameter controls the distribution variance of the coefficients of the own lags and a third controls the coefficients of other variables' lags. A fourth hyperparameter controls the speed at which the variance of the coefficients diminishes as the lag under consideration increases. Moreover, it is usually assumed that the analyst does not have specific information about the deterministic component and therefore the prior distribution of the coefficient is diffuse (row 3).

An additional hyperparameter is usually specified in order to control the overall extent of uncertainty with which the model coefficients are introduced. This is crucial in determining the relative weight assigned to the prior and sample information, respectively. In terms of Figure 1, an increase in this hyperparameter would provoke a generalised increase in the variance of the distributions so that the relative weight of the prior information would be reduced.

Certainly, in specific applications the analyst may wish to control other relevant dimensions of the prior information (for example, seasonal or long-term dimensions), but the dimensions described here are common to all BVAR applications.

Returning now to a formal description of the model, we will transcribe these ideas in terms of their defining elements (23).

Starting with the vector $\overline{\beta}$, the specification is as follows:

$$\overline{\beta}_i(0)=\begin{bmatrix}0\\ \cdot\\ \cdot\\ \cdot\\ \tau_1\\ \cdot\\ \cdot\\ \cdot\\ 0\end{bmatrix} \quad ; i=1,2,\ldots,n \tag{25}$$

where the hyperparameter τ_1 occupies the i^{th} position and represents the prior average of the coefficient of the first own lag of the dependent variable in equation i. The prior average coefficients for the remaining retards, own or otherwise, are equal to zero.

As mentioned earlier, prior information usually assumes independence between the components of $\beta(0)$, i.e. it starts from a diagonal matrix $\Omega(0)$, whose principal diagonal elements are specified as follows:

$$\sigma_{ij}^2(l)=\left(\frac{\tau_2}{l^{\tau_4}}\right)\sigma_{\varepsilon_i}^2 \qquad ; i=1, \ldots, n \qquad ; i=j \qquad ; l=1, \ldots, m$$

$$\sigma_{ij}^2(l)=\left(\frac{\tau_2\,\tau_3}{l^{\tau_4}}\right)\left(\frac{\sigma_{\varepsilon_i}^2}{\sigma_{\varepsilon_j}^2}\right) \qquad ; i=1, \ldots, n \qquad ; i\neq j \qquad ; l=1, \ldots, m \quad (26)$$

$$\sigma_{ih}^2=\tau_2\,\tau_5\,\tau_{\varepsilon_i}^2 \qquad ; i=1, \ldots, n \qquad ; h=1, \ldots, d$$

where i represents equation, j endogenous variable, l lag and h deterministic component. τ_2 controls the overall degree of uncertainty with which the prior information is introduced to the model estimation process; as τ_2 increases the prior distribution loses its precision, becoming diffuse in the limit. τ_3 controls the degree of uncertainty of the lags of other variables relative to own lag uncertainty. In the limit, when τ_3 is equal to zero the prior information defines a model composed of n univariate AR(m) processes. τ_4 controls the speed at which the variance decreases with the lag and τ_5 controls the relative uncertainty of the deterministic component. Lastly, $\sigma_{\varepsilon_i}^2$ and $\sigma_{\varepsilon_j}^2$ represent the elements of the principal diagonal of Σ and are a measure of the size of the fluctuations in variables i and j. Their role in the prior information is twofold: on the one hand, they enable you to compare the degree of uncertainty with the range of fluctuations. On the other, they introduce a correction to possible differences in the units for measuring the variables included in the model.

Although it is possible to hyperparametrise Σ, usual practice has been to condition in Σ, estimating this from the resulting residues in AR(m) univariate models estimated by the OLS method.

It now remains to characterise the time variation of the model, which depends on the S and φ matrices and is specified as follows:

$$\begin{aligned} S &= diag\,(S_1, \ldots, S_n) \\ \underset{k\times k}{S_i} &= diag\,(\tau_6) \quad ; i=1,2,\ldots,n \\ \varphi &= diag\,(J_1, \ldots, J_n)\,\Omega(0) \\ \underset{k\times k}{J_i} &= diag\,(\tau_7) \quad ; i=1,2,\ldots,n \end{aligned} \qquad (27)$$

where τ_6 controls the coefficients of the first order autoregression process that characterises the law of motion of the vector of coefficients β, and τ_7 controls the degree of time variation actually

introduced in the model. Note in particular that with τ_6=1 and τ_7=0, you obtain a version of the model which does not vary over time. Note too that time variation is proportionate to the prior variance matrix of vector $\beta(0)$, which permits a relative evaluation of the degree of time variation.

At this point, you will have observed that the specification of the prior information shown above is incomplete inasmuch as it depends on τ, which is an unknown vector of hyperparameters. From a strictly Bayesian viewpoint, prior information should not contain unknown (hyper)parameters. Indeed, a strictly Bayesian implementation would require that distributions be specified for the hyperparameters and then integrated in the relevant range in order to obtain the posterior distribution. However, BVAR applications usually opt for two alternative approximations to this tedious procedure.

The first involves using the posterior distribution associated with a particular numerical value of the vector τ which directly reflects the empirical regularities 1) to 3) described above. For example:

$$\tau = \begin{bmatrix} 1 \\ 0{,}2 \\ 0{,}5 \\ 1 \\ 10^6 \\ 1 \\ 0{,}001 \end{bmatrix} \tag{28}$$

This was standard procedure in the first uses of the methodology and is the same as assuming that the τ vector is a degenerate random vector with probability one in the specific choice (28).

The second approximation procedure involves using the posterior distribution associated with a particular numerical value selected in accordance with some goodness of fit criterion. Two criteria which are commonly used are the minimisation of a loss function defined in statistical terms of predictive capacity and the maximisation of the model's likelihood function.

Focusing on this latter criterion, we note that, assuming normality, the likelihood of our model is as follows:

$$\prod_{t=1}^{T} L\left(Y(t) \mid X(t-1), \Sigma, \tau\right) = (2\pi)^{-T/2} \tag{29}$$

$$\prod_{t=1}^{T} \left|\sigma(t-1)\right|^{-1/2} \exp\left[-1/2\,(Y(t)-X(t-1)\,\beta^*(t-1))'\,\sigma(t-1)^{-1}\,(Y(t)-X(t-1)\beta^*(t-1))\right]$$

where

$$\sigma(t-1) = X(t-1)\ \Omega^{*}(t-1)\ X(t-1)' + \Sigma$$

The approximation criterion based on the likelihood function then involves maximising (29) with respect to τ and obtaining the posterior distribution associated with this optimum vector. The Bayesian justification of this procedure is that it can reasonably approximate the complete integration process. Specifically, if a diffuse prior distribution is assigned to τ, the posterior distribution of the vector of coefficients β will be a weighted average of the posterior distributions associated with each specific value of τ, with the weights being the value of the likelihood at that specific value. So, by choosing the posterior distribution associated with the value of τ maximised in (29) we are in fact using the most heavily weighted posterior distribution in the integration process. When high likelihood values of τ give rise to similar associated posterior distributions, the procedure provides a reasonably approximate approach to the true posterior distribution.

We shall conclude this section by stressing that the Bayesian VAR methodology is very flexible in that it enables us to use our sample information to analyse a wide range of parameters. This in turn gives a wide generality of statistical representations that embraces everything from the univariate AR model to the UVAR model, which is a particular case of the BVAR framework obtained when the prior information selected is diffuse, i.e. when τ_2 tends to infinity, in which case $\Omega^{*}(0)^{-1}$ tends to zero and, as can clearly be seen in (24), the updating scheme of the prior information generates the OLS estimation of the model.

The efficiency of joint estimation

In the case of the UVAR model we concluded that single equation estimates are efficient because all the equations have the same explanatory variables. But does the same hold true of the BVAR model?

Using Theil's mixed technique to make our estimates enables us to again use the SURE framework to demonstrate that the answer to the above question is negative. Specifically, we attempt to verify whether the set of explanatory variables continues to be the same for all equations in the system when prior information is introduced. This can be done by going back to equation (18) which combines prior and sample information and in which X_{t-1} and R_{t-1} are the set of explanatory variables. Generally speaking $X_{t-1} \neq R_{t-1}$, which means that the BVAR model contains two blocks of equations whose explanatory variables differ and the results of the UVAR framework are therefore not applicable.

In fact, the condition that makes single equation estimates efficient in the BVAR framework is that the a priori variance of the

coefficients is a multiple of the residual variance in each one of the equations.[5] However, the type of prior information generally used in the BVAR applications described in the foregoing section does not meet this condition because each equation stresses the own lags, which means that the variance of own lag coefficients is greater than the variance of the other coefficients in the equation. Therefore, joint estimation of all the model equations is a requisite for efficiency in the BVAR framework.

Cointegration

Readers will have observed that at no point in our description of the BVAR model have we referred to whether the modelled stochastic process is stationary or non-stationary. In fact, we have deliberately avoided doing so in order to demonstrate that the Bayesian approach to estimation is equally suitable in both cases. As mentioned earlier, the important thing is that the prior information presents the sample information with a wide range of choice regardless of whether or not the process is a stationary one. Moreover, likelihood —the other feature of the estimation process— is also unaffected by whether or not the process is stationary inasmuch as the assumption that the joint density of the sample is normal is not contingent on whether the process analysed is stationary or not. Thus, in principle, there is no real reason for adopting different approaches to analysing stationary and non-stationary series.

Nevertheless, it is true that this stance has been criticised in cases of non-stationary processes with unit roots and potential cointegration relations. Lütkepohl (1991), Clements and Mizon (1991) and Phillips (1991) have all suggested that by starting with prior information that treats all coefficients as independent both within and across equations and assigns a mean equal, or close to, one to the first own lag and zero to the others, the Bayesian estimate of VAR models tends to be biased towards systems made up of univariate AR models, and is incapable of detecting the possible shared stochastic tendencies that characterise cointegrated processes. Sims (1991) has countered by suggesting that such criticism is ill-founded. He argues that, due to the property of superconvergence which characterises unit roots and cointegration relations in the data, these aspects of estimation tend to be clearly revealed regardless of the type of prior information used.

Álvarez and Ballabriga (1994) contributed to the discussion by altering the customary prior information of BVAR models so that it explicitly acknowledges the possible existence of cointegration in the process analysed. The authors carried out a brief Monte Carlo experiment, using a cointegrated process that makes it possible to weigh the capacity of different estimation methods to detect the long-term relationship. When the prior is fine-tuned, the results obtained support Sims' thesis rather than that of his critics.

---

5 See Doan, Litterman and Sims (1984).

1.3. Identifying VAR models

The foregoing methodological description contains no economic argument other than the minimum implied by the selection of the economic variables to be analysed. Readers might be bewildered by this, but it is one of the things that helps establish a sort of brand image for VAR methodology: the clear differentiation between the statistical and economic aspects of the analysis which respectively define the VAR model specification and identification processes.

The foregoing discussion focuses on model specification. Standard UVAR and Bayesian (BVAR) specification methods have been proposed but in both cases the objective was to exploit the statistical generality of the autoregressive representation (4)-(5) without contaminating it with arguments of an economic nature. The result is therefore a purely statistical model. Or, more precisely, a reduced-form model which, as we all know, is the econometric term for statistical representation models which have no economic content.

Certainly, obtaining a reduced form can in itself constitute an objective if your aim is simply to predict and/or analyse a set of correlations. However, when the objectives of your analysis include other issues, such as evaluating the effectiveness of monetary policy or the relative importance of supply and demand factors in explaining macroeconomic variability, the reduced form is insufficient and should be no more than a step midway along the path to a statistical model in a structural form that contains the economic information necessary to be able to reply to the questions of interest. As we noted earlier, VAR methodology aims to be an operational alternative to conventional macroeconometric simultaneous equation models, whose principal aim is precisely to reply to questions of interest such as those mentioned above. Therefore, the additional effort involved in obtaining an interpretable model from the reduced form is generally unavoidable when applying VAR (UVAR, BVAR) methodology. This is the model identification stage.

Sims' seminal work (1980) and its succeeding applications soon revealed that VAR model identification was one of the weakest points in the proposed methodology. Indeed, the opinion that VAR models were simply reduced forms and, as such, not valid in terms of quantifying economic relations quickly became widespread.

However, this criticism was not strictly justified inasmuch, as we will see in the formal description set out below, initial applications of the methodology used a contemporaneous causal chain which is the equivalent of a recursive simultaneous equation structural model. Nevertheless, it is true that a recursive structure is rarely suitable for describing economic reality. VAR models could therefore certainly be criticised in terms of their identification, not because it was lacking but because its credibility was questionable. This is a new paradox if we recall that, when originally proposed, the incredible identification of simultaneous equation models was the motivating criticism.

Because it was one of the most controversial features of the methodology, the identification stage accounted, and continues to account, for a good part of the academic discussion on the subject of VAR models. A satisfactory solution to the problem was found and present-day identification methods are considerably better than those used in the initial applications and simultaneously respect the largely unrestricted spirit of the methodology.

Formal description

As is a well-known fact, identifying an econometric model is a generic problem that refers to the model per se, not to the methodology of modelisation. A common way to address the identification problem is by presenting it as the obtention of a structural model from a reduced form. The structural model can be interpreted in economic terms and will be identifiable if it consists of statistically distinguishable equations which, as such, can be retrieved from the statistical variability summarised in the reduced form.

As mentioned earlier, the problem is not solved by opting for one methodology or another: all methodologies must address it. What can differ from one methodology to another is the way the problem is approached. Thus, conventional simultaneous equation models manage to make their equations statistically distinguishable by including or excluding variables which are treated as exogenous. Taken to unwarranted extremes, this strategy provides an identification that is simply an illusion or, as Sims (1980) puts it, incredible.

In contrast, VAR methodology rejects the assumption of exogeneity and uses an identification strategy that combines a minimum of exclusion restrictions with conditions on the probabilistic structure of the model's error term. More concretely, a VAR model is referred to as structural when the statistical distinction in its equations is obtained by imposing a set of restrictions (not necessarily of exclusion) on the contemporaneous matrix of coefficients, such that the components of the model's error term are orthogonal which, in turn, allows us to interpret them as primitive sources of economic variability.

Orthogonal error terms are not a general requisite for conventional structural models and reveal a profound conceptual difference in terms of whether, economically speaking, the relevant variability is the "total" or the "unexpected". Conventional models proceed as though the total was the relevant variability and therefore do not insist that disturbances be orthogonal. However, orthogonal disturbances are essential when you want to analyse the dynamic implications of the model in the conviction that relevant variability is the unexpected.

The orthogonality requirement also explains why "model identification" and "orthogonalisation of the error component" are commonly used as interchangeable terms in VAR model literature.

To illustrate this more clearly, let us repeat the reduced form used in (4) , i.e.[6]

$$Y(t) = B(L)\ Y(t) + DZ(t) + \varepsilon(t) \qquad (30)$$

where the components of $\varepsilon(t)$ are generally correlated, with a covariance matrix equal to Σ in all t. Identifying a VAR model can then be seen as obtaining a linear combination of $\varepsilon(t)$, which is a new vector of disturbances whose components are orthogonal and interpretable in economic terms. Or, to put it more formally, obtaining an A, $n \times n$ matrix so that, in all t:

$$A\varepsilon(t) = \nu(t) \qquad (31)$$

where the components of $\nu(t)$ represent isolated sources of economic variability (fiscal or monetary, private or public, supply or demand, etc.), which means that their variance and covariance matrices are diagonal and can be normalised to the identity without loss of generality. Note that the A matrix is the link between the reduced and structural versions of the VAR model. In fact, premultiplying the reduced format VAR model (30) by A gives you the structural VAR model:

$$AY(t) = AB(L)\ Y(t) + ADZ(t) + \nu(t) \qquad (32)$$

Or its equivalent:

$$C(L)\ Y(t) = GZ(t) + \nu(t) \qquad (33)$$

where

$$C(L) = A(I - B(L))$$
$$G = AD$$

Note that model (32)-(33) is actually a conventional structural model with the peculiarity that all its predetermined variables, except the deterministic component, are lagged endogenous variables and that the error component is orthogonal: peculiarities which are precisely what define it as a structural VAR model.

Equations (32)-(33) also enable us to more clearly see how VAR models can be identified by using a combination of restrictions in

---

6 For the sake of simplicity we will continue using a model whose coefficients are not time dependent.

the contemporaneous coefficient matrix and condition on the probabilistic structure of the error term: matrix *A* contains the coefficients of impact and should be chosen in such a way as to guarantee both the fulfilment of (31) and make a statistical distinction between the system's equations so that the structure is effectively identified. A choice of *A* which guarantees both conditions involves specifying *A* as a lower triangular matrix. This was the usual choice in the first applications of the methodology and, as can now be clearly seen, is the equivalent of a contemporaneous causal chain and makes model (32)-(33) a recursive structural model.

The recursive structure appears to be technically correct in that it produces an orthogonal vector and a structure composed of distinguishable equations. However, as has already been mentioned, recursive structures are not generally appropriate for describing economic reality because they do not include the simultaneous relations which normally characterise this reality. They can therefore be criticised for using contemporaneous time restrictions that are difficult to believe and, as a consequence, they fail in their attempt to isolate sources which could be believably interpreted in economic terms, which is a fundamental feature of the identification process.

The literature on VAR model identification has gone beyond recursiveness to consider more general specifications of matrix *A* which produce more credible structural models. We used two types of identification restrictions to obtain these specifications: short- and long-term.

Short-term restrictions are zero in certain positions in Matrix *A* and are normally justified by the fact that some economic agents receive information flows with a certain delay. For example, the delay with which monetary authorities receive information about macroeconomic evolution could justify the assumption that interest rates do not respond to disturbances in production and price levels at the exact moment they occur: two zeros in matrix *A* which enable us to identify the supply and demand of liquid balances in the economy. The use of the adjective "short-term" is clear: the restriction is exclusively limited to the contemporaneous effect of certain disturbances.

Long-term restrictions are usually based on economic theory and, as their name indicates, restrict the long-term effect of certain disturbances on certain variables. For example, the model can introduce the restriction that monetary disturbances do not have real long-term effects, a restriction which is based on the widely accepted principle of long-term monetary neutrality.

The use of long-term restrictions requires a stationary model so that the long-term effects are well-defined, i.e. they are not explosive. In formal terms, imposing these restrictions is tantamount to restricting certain linear combinations in the matrix of long-term effects associated with the moving average (MA) representation of the structural model. Note specifically that, with a

stationary model, the polynomial $C(L)$ in (33) can be inverted, giving rise to the following MA representation for the non-deterministic component of $Y(t)$:

$$Y(t) - M(L)DZ(t) = M(L)A^{-1}\nu(t) \tag{34}$$

where

$$M(L) = (I - B(L))^{-1}$$

The matrix of the long-term effects of the various structural disturbances is the sum of the matrices which define the $M(L)\ A^{-1}$ polynomial, each of which determines the effect of the disturbances in a different time horizon. Thus, the long-term effects matrix is written as:

$$M(L=1)\ A^{-1} = \sum_{i=0}^{\infty} M_i\ A^{-1} \tag{35}$$

And, as mentioned earlier, these long-term restrictions are tantamount to restricting certain linear combinations of the elements of the matrix shown in (35), which could be expressed as follows:

$$\underset{p\times n^2}{C}\ vec\left[\underset{n^2\times 1}{M(1)}\ A^{-1}\right] = \underset{p\times 1}{c} \tag{36}$$

where we have used the vec(·) operator and p represents the number of restrictions.[7] In the particular case which eliminates the long-term effect of the disturbance i in the variable j, the value of c will be nil and the C matrix will have the dimension $1 \times n^2$, with a one in the entry $[(j-1)*n] + i$, $i, j = 1,...,n$, and zeros in the others.

It is important to note that this set of short and long-term restrictions creates a very frugal framework for identification from the standpoint of restrictions: when exclusion is used, it is used only with contemporaneous impacts, without excluding possible lagged effects. When the lagged effects are restricted the restrictions are loose, conditioning only the long-term effect. This is in line with our earlier observation that the VAR framework of identification respects the little-restrictive spirit of the methodology.

---

7 The vec(·) operator turns matrixes $m \times n$ into vectors $mn \times 1$ stacking their n columns.

Estimating the structural model

The estimation of structural VAR models still remains to be discussed, i.e. estimating the G matrix and the $C(L)$ matrix polynomial, in accordance with equation (35). Let us therefore go back to equations (31) and (32)-(33).

Let us start by observing that in accordance with (31), the compatibility between the variance and covariance matrices of disturbances in reduced and structural models means that the relation between Σ and the matrix of contemporary effects A is as follows (remember that the covariance matrix of the structural error term was normalised to the identity):

$$A\Sigma A' = I \tag{37}$$

or its equivalent:

$$\Sigma = A^{-1} A^{-1'} \tag{38}$$

Note that (32)-(33) suggest that the structural model could be estimated in two steps:

Step 1 —> Estimate the matrixes of coefficients D and $B(L)$ of the VAR reduced form. Then take the resulting sequence of errors $\hat{\varepsilon}(t)$ and use them to estimate Σ, $\hat{\Sigma}$, starting with the resulting sequence of errors.

Step 2 —> Use $\hat{\varepsilon}(t)$ and $\hat{\Sigma}$ from Step 1 together with (36) and (38) to obtain the maximum likelihood estimator for matrix A.

By combining the estimators D and $B(L)$ of the first step and the estimator A of the second you will then be able to obtain the estimators G and $C(L)$ of the structural model.

Step 1 does not introduce any new elements. It simply estimates reduced VAR by using the methods described in sections 1.2.2 or 1.2.3 of this paper, depending on whether the classical or Bayesian version of the model is used.

Step 2 aims to maximise, with respect to the A coefficients matrix, the sample likelihood of the series of reduced disturbances obtained in Step 1, bearing in mind possible long-term restrictions (36) and the compatibility requirement (38). To specify, note that when normality is assumed the likelihood of the series $\hat{\varepsilon}(t)$, $t = 1,2,...,T$ is as follows, when logs are used and constants ignored:

$$-\frac{T}{2} \ln |\Sigma| - \frac{1}{2} \sum_{t=1}^{T} \hat{\varepsilon}(t)' \, \Sigma^{-1} \, \hat{\varepsilon}(t) \tag{39}$$

In Step 2 there is the problem of obtaining the A matrix which maximises (39) subject to (36) and (38). Note that the number of different conditions in (38) is not n^2 but $(n^2 + n)/2$, inasmuch as Σ is symmetrical. The maximum number of contemporaneous coefficients different than zero that can be determined when using the conditions in (36) and (38) is $[(n^2 + n)/2] + p$. The remaining coefficients are equal to zero, constituting short-term identification restrictions. Thus, when long-term restrictions (p=0) are not used, the number of short-term restrictions (zero in matrix A) should be at least equal to $(n^2 - n)/2$.

The 2-step estimation procedure described above is, in fact, the procedure used in VAR methodology applications. It is attractive on two grounds. First, it is in line with the idea of clearly separating the statistical restriction specifications from economic identification which, as we have mentioned earlier, is a distinctive characteristic of VAR methodology. Secondly, the method generates efficient estimators of structural coefficients inasmuch as they are equivalent to those obtained by directly estimating G and $C(L)$ using the maximum likelihood method. The reason is that, when normality is assumed, the information matrix of the likelihood of model (33) is diagonal with respect to $[D, B(L)]$ and A, a condition established in Durbin (1970) in order to justify the efficiency of the 2-step procedure.

Pioneering articles on the identification framework described in this section are: Bernanke (1986), Blanchard and Watson (1986), Sims (1986) and Blanchard and Quah (1989).

1.4. Uses of VAR models

In the section on identification we noted that VAR differs from conventional methodology in that requires structural disturbances to be orthogonal. The objective of this is to isolate primitive sources of economic variability which stem from the behaviour of either supply or demand in the public or private sector or in the export or domestic sector of the economy. It should therefore come as no surprise that the analyst who uses VAR modelling methods tends to be genuinely interested in the dynamic effects of these primitive disturbances on the evolution of the observable variables that characterise the economic framework to be studied, i.e. in the effects of ν on Y, to use the terms of our notation.

For the same reason, neither should it come as a surprise that the analyst is more interested in recovering the structural MA representation in (34) than the structural autoregression representation in (32)-(33) because it is the MA representation that directly shows the effect of ν on Y. In fact, as we will see below, typical uses of VAR models almost all deal with obtaining and analysing the MA representation of the model.

These uses include calculating the model's impulse response function; the forecast error, variance decomposition and obtaining

future projections. Generally speaking, VAR, like all econometric models, are used for hypothesis testing and future projections. However, VAR models emphasise the sources of primitive variability, which means that hypotheses are not tested simply on the basis of the statistical significance of certain structural coefficients, but on the global pattern of interrelations embedded in the model and shown through the impulse response function and the variance decomposition. Similarly, the exogenous variability which can condition future projections does not stem from the variability of observables determined outside the model but from the very sources of primitive disturbance contained in the model.

Before continuing, we should briefly comment on the existence of the MA representation of the model. As is well known, this representation exists in the stationary case because the autoregressive model can be inverted. However, the same is not true of the non-stationery case associated with the existence of unit roots, which is so common in empirical analysis in economics. In this case, the auto-regressive model cannot be inverted and therefore has no MA representation because the sequence $M(L)\ A^{-1}$ in (34) is explosive.

Does this mean that analysis should be limited to the stationary framework? In order to verify that the answer is no, go back to equation (33) and try replacing the left-hand side of the formula with $Y(t\text{-}s)$, $s = 1,....,H$, using the inherent probabilistic mechanism (33). This tedious but simple process will yield the first H terms of the MA form of the model, expressing $Y(t)$ as the sum of two components:

$$Y(t) = \sum_{s=0}^{H-1} M_s\ A^{-1}\ v(t-s) + E_{t-H}\ Y(t) \qquad (40)$$

Or its equivalent:

$$Y(t) = \sum_{s=0}^{H-1} \overline{M}_s\ v(t-s) + E_{t-H}\ Y(t) \qquad (41)$$

where

$$\overline{M} = M_s\ A^{-1} \quad ; \quad s = 0,...,H-1$$

$E_{t-H} \equiv$ the expected value with the information available in $t\text{-}H$

The first component in (41) represents the contribution to the $Y(t)$ value of the innovations that occurred between the periods $t\text{-}H\text{+}1$ and t, both inclusive. This contribution is determined by the sum of the first H terms of the model's MA form. The second component is the average projection of $Y(t)$ based on available information in the

t-H period, which depends on the vector of $\tilde{Y}$ observables between the periods *t-H* and *t-H-m*+1 (remember that *m* stands for the number of lags) and the deterministic component between the periods *t-H*+1 and *t*.

The advantage of decomposition (41) is that it exists for any finite *H* and regardless of whether or not the process analysed has unit roots. The description of uses set out below is therefore based on this decomposition.

The impulse response function

As its very name implies, the impulse response function quantifies the effect of isolated impulses equal to one in each of the model's *n* disturbances on the system's *n* variables over a time horizon of *H* periods. In other words, the function quantifies the effect in Y_i (t), i=1,...., *n* of the disturbance ν_j $(t\text{-}s)$=1, j=1,...., *n*, which occurred *s* periods earlier, *s*=0,...,*H*-1. Calculating the function for the entire system thus gives you *n* x *n* series of *H* length.

You will immediately see that these *n* x *n* series are the same as those which make up the sequence of matrices $\overline{M}_s$, *s*=0,..., *H*-1 of the first component of (41). In order to verify this, observe the following sequence of disturbances:

$$\begin{aligned} \nu(t-\bar{s})' &= (0,...,1_j,0...,0) \ , \ 0 \le \bar{s} \le H-1 \\ \nu(t-s)' &= 0 \qquad\qquad\quad , \ s \neq \bar{s} \end{aligned} \tag{42}$$

In other words, the j^{th} component of ν is disturbed in one unit in the period $t-\bar{s}$. Note that when you use sequence (42) to calculate the first component of (41) the result obtained is equal to the j^{th} column of the $\overline{M}_{\bar{s}}$ matrix, which proves that the response of $Y_i(t)$ to an impulse of a unit which occurred *s* periods earlier in the ν_j disturbances is given by the element $\overline{M}_s(i,j)$. Note too that, according to decomposition (41), this response should be interpreted as the deviation from the average projection $E_{t-H}Y(t)$ induced by the specific impulse.

Thus we conclude that the impulse response function constitutes a tool for assessing the explanatory power of the different sources of variability (disturbance) included in the model, and that calculating them for an *H* time series is the same as calculating the first *H* term of the structural model's MA form.

Forecast error variance decomposition

An alternate way of assessing the relative importance of the various sources of disturbance is by analysing how they contribute to the model's forecasting error. The motivation for this analysis can be clearly seen in (41), when one notes that its first component represents both the contribution to the value of $Y(t)$ of the

disturbances that occurred between t-H+1 and t and the error of predicting $Y(t)$ with the information available in t-H. Analysis of contributions to forecasting error thus provide information about the relevant sources of variability in an H time horizon.

The specific way to analyse these contributions involves calculating the forecast error variance for a particular horizon and isolating the percentages of this variance attributable to each of the disturbances contained in the model. This is the origin of the expression "variance decomposition" which gives this procedure its name. More specifically yet, the error variance involved in forecasting Y with an H time horizon is the variance of the first component of (41), which is as follows (remember that the variance of ν was normalised to the identity):

$$var\left[\sum_{s=0}^{H-1} \overline{M}_s\ \nu(t-s)\right] = \sum_{s=0}^{H-1} \overline{M}_s\ \overline{M}_s' \qquad (43)$$

The formal exercise involves breaking down (43) into components that represent the percentage of the forecast error variance associated with Y_i which is explained by the contribution of the component ν_j , i, j=1,...,n in the H time horizon. This decomposition would not be feasible if the elements of the ν vector were lagged or contemporaneously correlated, inasmuch as the covariances cannot be clearly attributed. Such correlations are, however, null, which makes decomposition possible.

Indeed, given the orthogonality of the ν elements, the variance of any of their linear combinations would always be the sum of the variances of each of the elements involved. Thus, isolating their contribution from the total variance is simply a matter of isolating the terms associated with each disturbance and adding up their variances. In this particular case, the linear combination analysed is the first component of (41), which we repeat below:

$$\sum_{s=0}^{H-1} \overline{M}_s\ \nu(t-s) \qquad (44)$$

Note then that the terms of (44) which correspond to the ν_j between the t-H+1 and t periods are those associated with the j^{th} columns of the matrices $\overline{M}_s$, s=0,...,t-H+1. These terms can be algebraically isolated by postmultiplying the $\overline{M}_s$ matrixes by the R_j instrumental matrix, defined with zeros in all their entries except in (j,j) which has a one. More concretely, if the sum of all these terms is known as P_j we find that:

$$P_j = \sum_{s=0}^{H-1} \overline{M}_s\ R_j\ \nu(t-s) \quad ;\ j = 1, \ldots, n \qquad (45)$$

where

$$\overline{M}_s R_j = \begin{bmatrix} 0 & \dots & \overline{M}_s(1,j) & \dots & 0 \\ 0 & \dots & \overline{M}_s(2,j) & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \overline{M}_s(n,j) & \dots & 0 \end{bmatrix}$$

It is clear then that forecast error (44) can be expressed as the sum of the n components in (45). In other words, we find that:

$$\begin{aligned} \sum_{s=0}^{H-1} \overline{M}_s \, v(t-s) &= \sum_{j=1}^{n} \left[\sum_{s=0}^{H-1} \overline{M}_s \, R_j \, v(t-s) \right] \\ &= \sum_{j=1}^{n} P_j \end{aligned} \tag{46}$$

This formula isolates the contribution of each of the n disturbance components to the forecast error and provides the basis for calculating their contribution to the variance of this error. Specifically, given the orthogonality of the components P_j, j=1,...,n the variance of the forecast error in (43) is immediately expressed as follows:

$$\begin{aligned} \mathcal{P} \equiv var \left[\sum_{s=0}^{H-1} \overline{M}_s \, v(t-s) \right] &= \sum_{s=0}^{H-1} \overline{M}_s \, \overline{M}_s' \\ &= var \left[\sum_{j=1}^{n} P_j \right] \\ &= \sum_{j=1}^{n} var \, P_j \\ &= \sum_{j=1}^{n} \mathcal{P}_j \end{aligned} \tag{47}$$

where

$$var \, P_j = \mathcal{P}_j \quad ; \quad j = 1, \dots, n$$

The variance in the forecast error of Y_i with an H horizon is then the $\mathcal{P}(i,i)$ element, and the proportion of this variance explained by the v_j disturbance is written as $\mathcal{P}_j(i, i)/ \mathcal{P}(i, i)$, i, j=1, ..., n, a proportion which enables us to judge the relative importance of the different sources of variability included in the model.

Future projections

Unlike calculations of the impulse response function and variance decomposition, both of which are specific to VAR methodology, future projections are common in econometric models The term "projection" should be understood here in a broad sense, including both forecasting and simulation. Choosing the name "projection" rather than the more common terms "forecasting" and "simulation" is justified by an aim to interpret these exercises in the general sense of projecting any aspect of the future distribution of the variables included in the model and not just their average dimensions. The emphasis on this theme can be said to be a distinguishing feature of the VAR framework and we will discuss this further in the following section. For the time being, however, we will focus on obtaining average projections.

VAR methodology jargon distinguishes between unconditional and conditional projections. The former refer to projections generated by the model with the information available in the period which defines the origin of the projection and without imposing any condition on the future path of the model's variables. Conditional projections, on the other hand, place certain restrictions on the future evolution of some of the model's variables; for example, on the future path of interest rates or wages. It can certainly be argued that the distinction is not very clear inasmuch as the unconditional projections are conditioned by the information available at the time of prediction, but nevertheless this is the terminology used.

In any case, let's go back now to (41) in order to clarify the distinction. Formulated for the period $t = T + h$, $h \geq 1$, it is expressed as follows:

$$Y(T+h) = \sum_{s=0}^{h-1} \overline{M}_s \; v(T+h-s) + E_T \; Y(T+h) \qquad (48)$$

Now let's assume that sample information is available up until period T and that the forecasting horizon is $h = H$. Using the information available in T, the average unconditional projection of $Y(T+H)$ is thus $E_T\, Y(T+H)$ in accordance with (48): in other words, the result of making the forecasting error for horizon H equal to zero.

By using (48) and the reduced form of the model (30) we can explicitly calculate $E_T \; Y(T+H)$. Specifically, we can calculate the average unconditional projection of $Y(T+h)$ for the horizon $h = 1,\ldots, H$. According to (48), such a projection for $h = 1$ is $E_T\, Y(T+1)$, and by referring to (30) we can immediately verify that:

$$\begin{aligned} E_T Y(T+1) &= B(L)Y(T+1) + DZ(T+1) \\ &= B_1 Y(T) + B_2 Y(T-1) + \ldots + B_m Y(T-m+1) + DZ(T+1) \end{aligned} \qquad (49)$$

Likewise, and in accordance with (48), we obtain an average projection equal to $E_T\,Y(T+2)$ for $h = 2$. According to (30) and (49) this is written as:

$$E_T Y(T+2) = B_1 E_T Y(T+1) + B_2 Y(T) + \ldots + B_m Y(T-m+2) + DZ(T+2) \tag{50}$$

Continuing in the same vein for horizons $h = 3,\ldots,H$, the formula is as follows:

$$E_T Y(T+H) = B_1 E_T Y(T+H-1) + \ldots + B_m E_T Y(T-m+H) + DZ(T+H) \tag{51}$$

That is to say, the average unconditional projection using information available in T and horizon H is obtained by replacing the lagged variables in the reduced form of the model with their own unconditional projections made with the information available in T.

Conditional projections use the same information as their unconditional counterparts as well as information about certain restrictions on the evolution of some of the model's variables between the origin and the final period of the projection horizon. In other words, there are restrictions on certain components of the vectors $Y(T+1)$, $Y(T+2)$,..., $Y(T+H)$. Generally speaking, any linear combination of these components can be restricted, but the most common type of restriction involves setting future values for some of them (for example, by charting the path wages are likely to take in the future) so that we can predict the consequences which, according to the model, such evolution would have on the rest of the economy.

Charting the future path of an exogenous variable is a trivial matter. Inasmuch as the variable is determined outside the model, all you have to do is set its value at the desired level. However, when the variable is endogenous, which is always the case of VAR models, things become more complicated. The variable must then be determined in the model, which means that any restriction must necessarily affect the sources of variability included in the model, i.e. in terms of model disturbances. This is immediately apparent in (48) where you can clearly see that, given the information available in T and therefore $E_T Y(T+h)$, restricting $Y(T+h)$ is tantamount to restricting the forecast error with horizon h. In other words, it is the equivalent of saying that:

$$\sum_{s=0}^{h-1} \overline{M}_s \; v(T+h-s) = \tilde{Y}(T+h) - E_T\, Y(T+h) \tag{52}$$

where $\tilde{Y}(T+h)$ represents the restricted value of $Y(T+h)$. More particularly, the components $Y_i(T+h)$ in the horizons $h=1,\ldots,\bar{h}$ with $1 \leq \bar{h} \leq H$, can be restricted, thus imposing a future path on the i^{th} component of the Y vector.

Formula (54) reveals that placing r restrictions on the future evolution of a VAR model's variables is the same as placing linear r restrictions on the model's future disturbance vectors, and can generally be written as:

$$\underset{r \times nH}{Q} \quad \underset{nH \times 1}{V} = \underset{r \times 1}{q} \tag{53}$$

where V contains stacked disturbance vectors $\nu(T+1)$, $\nu(T+2)$, ..., $\nu(T+H)$. Q is appropriately defined in terms of the $\bar{M}_s$ matrixes in order to incorporate type (52) restrictions and q contains the constants that define the linear r restrictions imposed.

The average conditional projection of the disturbances between $T+1$ and $T+H$ is thus given by the average of the vector V conditioned to (53), $E[V \mid QV = q]$, and the average conditioned projection of $Y(T+H)$ is immediately obtained from (48), with $h=H$ and taking on both sides of the equation expectations conditional on (53) and on the information available in T, which results in:

$$E_T\left[Y(T+H) \mid QV=q\right] = E_T\, Y(T+H) + \sum_{s=0}^{H-1} \bar{M}_s\, E\left[\nu(T+H-s) \mid QV = q\right] \tag{54}$$

In other words, the average conditional projection is the unconditional projection adjusted by the conditional contribution of the disturbances in the forecast time horizon.

To conclude, it should be stressed that, unlike the impulse response function and variance decomposition, future projections do not necessarily depend on model identification. Indeed, in the case of unconditional projections, they are clearly independent because, as we have already seen, they only use the statistical variability summarised in the reduced form of the model. Nor are conditional projections dependent when the conditions in themselves restrict forecast error because, according to the following expression, forecasting error is independent of identification:

$$\sum_{s=0}^{H-1} \bar{M}_s\, \nu(T+H-s) = \sum_{s=0}^{H-1} M_s\, \varepsilon(T+H-s) \tag{55}$$

where the relation between ν and ε given in (31) and the definition of $\overline{M}_s$ given in (41) are used. Identification is only necessary when the restriction in the forecast error involves imposing specific paths on some component of the vector ν, inasmuch as in this case the aim is to restrict the behaviour of a specific economic agent or sector, which means that it is necessary to first identify the source of the economic variability of the model associated thereto.

Measuring Uncertainty

Although not explicitly mentioned, the three uses described above involve calculations that are uncertain because they are based on a stochastic model with estimated coefficients which in turn are random variables. More specifically, the impulse response function and the variance decomposition directly depend on the coefficient matrices $\overline{M}_s$ of the MA form of the model, as is clearly shown by (42) and (47), respectively. Moreover, future projections depend directly on the coefficients of the autoregressive form (D, B_s) which determine $E_T Y(T+H)$ in (48), as well as on the first component in (48), which in turn depends on the $\overline{M}_S$ coefficients and the error term ν.

The immediate conclusion is that the impulse response function, variance decomposition and future projections are all stochastic magnitudes which, as such, can be characterised by using their corresponding distribution. The description contained above was limited to obtaining point estimates, average or otherwise, of these magnitudes. However, as is well known, point estimates provide very little information. Ideally, the analyst should attempt to characterise aspects of the distribution which give the most precise possible idea of the uncertainty surrounding his calculations. Such a characterisation is always possible, to a greater or lesser extent, when the model utilised includes a complete stochastic distribution of all its variables. This is always the case with VAR models and is discussed in the following paragraphs.

Let us begin with the impulse response function and the variance decomposition which are a direct function of the $\overline{M}_S$ coefficient matrices. As you will recall, according to (41) these matrices are written as:

$$\overline{M}_S = M_s A^{-1} \tag{56}$$

According to (34), the M_s matrices which make up the matrix polynomial $M(L)$ are defined on the basis of the matrices of the reduced form polynomial $B(L)$:

$$M(L) = (I - B(L))^{-1} \tag{57}$$

Combining (56) and (57) therefore gives us the following:

$$\overline{M}(L) = (I - B(L))^{-1} A^{-1} \tag{58}$$

In other words, the matrix polynomial with $\overline{M}_s$ coefficient matrices depends directly, not linearly, on the matrix polynomial of the reduced form B_s coefficient matrices and on matrix A with the contemporaneous coefficients that determine the model's identification scheme.

Formula (58) explicitly reveals the stochastic nature of the impulse response function and variance decomposition, showing that its distribution depends on the distributions of the coefficients in $B(L)$ and A. However, it also reveals that this dependency is highly non-linear, which means that obtaining average responses and decompositions is not the same as using the average of $B(L)$ and A on the right-hand side of (58). Moreover, and more generally speaking, it is difficult to analytically obtain their distributions. This forces the analyst to use Monte Carlo methods which enable him to empirically characterise the distributions by drawing on the $B(L)$ and A distributions, both of which we assume to be normal. This is in fact customary practice for obtaining the confidence intervals which are typically reported in the applications of the methodology. Note though that the procedure used is generally approximate because it draws on $B(L)$ and calculates $\overline{M}(L)$ conditioning on a particular identification scheme, i.e. treating the A matrix as a constant.

Now let us discuss future projections. As mentioned earlier, this involves projecting any aspect of the future distribution of the variables contained in the model and not simply its average dimensions. A good way to proceed then is by immediately considering the mechanism that generates future values. We will therefore use the VAR model, taking as our reference the T period and the error term expressed, according to (31), as a function of the structural disturbance vector ν:

$$Y(T+s) = B_1 Y(T+s-1) + B_2 Y(T+s-2) + \ldots + B_m Y(T+s-m) + \\ + DZ(T+s) + A^{-1}\, \nu(T+s) \tag{59}$$

$s \geq 1$

Now observe that, given the path $s = 1,2,\ldots,H$ it is possible to obtain realisations of the observable vectors $Y(T+1),\ldots, Y(T+H)$ drawing on the $B(L)$, D, and A distributions and on the disturbance vectors $\nu(T+1),\ldots, \nu(T+H)$ and then successively substituting the Y values in (61). This once again leads us to use Monte Carlo methods as a way to empirically characterise the joint distribution of the future path of the model's variables.

Just as in the case of responses and decompositions, it is customary to use an approximate procedure. In this case, it tends to ignore the sampling error associated with the coefficients, which are treated as constants. Unconditional projections can then be made by successively drawing on the distribution of the vector V defined in (53) which is, according to our assumptions, a multivariate normal with a zero mean and a covariance matrix equal to the identity. Likewise, conditional projections are made by drawing on the distribution of $[V \mid QV = q]$ which, under the assumption of normality we have used throughout this paper, is normal and written as:

$$\left[V \mid QV = q\right] \sim N\left[Q'\left(QQ'\right)^{-1} q \ , \ I - Q'\left(QQ'\right)^{-1} Q\right] \quad (60)$$

In the case of either conditional or unconditional projections, the analyst can not only empirically characterise the average dimensions of future distribution, but can also obtain confidence intervals and, more generally, calculate the probability of any event associated with the future evolution of the variables included in the model, which is essential given the considerable degree of uncertainty surrounding the behaviour of the economy.

Bibliography

AHN, S.R.; REINSEL, G.C. "Estimation for Partially Onstacionary Multivariate Autoregressive Models", *Journal of the American Statistical Association*, 1990, no. 85, pp. 813-823.

ÁLVAREZ, L.J.; BALLABRIGA, F.C. *BVAR Models in the Context of Cointegration: A Monte Carlo Experiment.* Madrid: Banco de España, 1994. [Documento de Trabajo, no. 9405].

BALLABRIGA, F.C. "Instrumentación de la metodología VAR". *Cuadernos Económicos de ICE*, 1991, no. 48, pp. 85-104.

BALLABRIGA, F.C. *Bayesian Vector Autoregressions.* Barcelona: ESADE, 1997. [Papers ESADE, no. 155].

BERNANKE, B. "Alternative Explanations of the Money Income Correlation", *Carnegie-Rochester Conferences Series in Public Policy*, 1986, no. 25, pp. 49-100.

BLANCHARD, O.; WATSON, M. "Are Business Cycles All Alike?". In: GORDON, R., ed. *The American Business Cycle.* Chicago: University of Chicago Press, 1986.

BLANCHARD, O.; QUAH, D. "The Dynamic Effects of Aggregate Demand and Supply Disturbances", *American Economic Review*, 1989, no. 79, pp. 655-673.

CANOVA, F. "Vector Autoregressive Models: Specification, Estimation, Inference and Forecasting". In: PESARAN; HASHEM and WICKENS, eds. *Handbook of Applied Econometrics.* Blackwell Publishers Ltd., 1995.

CLEMENTS, M.P.; MIZON, G.E. "Empirical Analysis of Macroeconomic Time Series: VAR and Structural Models", *European Economic Review*, 1991, no. 35, pp. 887-917.

DOAN, T; LITTERMAN, R; SIMS, C. "Forecasting and Conditional Projections Using Realistic Prior Distributions", *Econometric Review*, 1984, no. 3(1), pp. 1-100.

DURBIN, S. "Testing for Serial Correlation in Least Square Regression When Some of the Regressors Are Lagged Dependent Variables", *Econometrica*, 1970, no. 38, pp. 410-421.

GRANGER, C.W.J.; NEWBOLD, P. *Forecasting Economic Time Series.* Academic Press, 1986.

HARVEY, A.C. *The Econometric Analysis of Time Series.* New York: John Wiley & Sons, 1981.

KLEIN, L. R. "The Use of Econometric Models as a Guide to Economic Policy", *Econometrica*, 1947, no. 15(2), pp. 111-151.

LITTERMAN, R. "Techniques for Forecasting with Vector Autoregressions". Minneapolis: University of Minnesota. 1980. Ph. D. Thesis.

LUCAS, R.E. "Econometric Policy Evaluation: A Critique", *Carnegie-Rochester Conferences on Public Policy*, 1976, no. 1, pp. 19-46.

LÜTKEPOHL, H. *Introduction to Multiple Time Series Analysis*. New York: Springer Verlag, 1991.

PARK, J.Y.; PHILLIPS, P.C.B. "Statistical Inference in Regressions with Integrated Processes. Part 2", *Econometric Theory*, 1989, no. 5, pp. 95-131.

PHILLIPS, P.C.B. "Bayesian Routes and Units Roots: De Rebus Prioribus Semper Est Disputandum", *Journal of Applied Econometrics*, 1991, no. 6, pp. 435-473.

SIMS, C.A. "Macroeconomics and Reality", *Econometrica*, 1980, no. 48(1), pp. 1-48.

SIMS, C.A. "Are Forecasting Models Usable for Policy Analysis?", *Quarterly Review*, Winter 1986, pp. 2-16. Minneapolis: Federal Reserve Bank of Minneapolis.

SIMS, C.A. "Comment on 'Empirical Analysis of Macroeconomic Time Series: VAR and Structural Models' by Clements and Mizon". *European Economic Review*, 1991, no. 35, pp. 922-932.

SIMS, C.; STOCK, J.; WATSON, M. "Inference in Linear Time Series Models with Some Unit Roots", *Econometrica*, 1990, no. 58, pp. 113-144.

THEIL, H. *Principles of Econometrics*. New York: John Wiley, 1971.

FIGURE 1
BASIC TRAITS OF THE PRIOR INFORMATION