Revista de Análisis Económico, Vol. 3, Nº 2, pp. 119-134 (Noviembre 1988)

VECTOR AUTOREGRESSIVE TECHNIQUES FOR STRUCTURAL ANALYSIS

PAUL L. FACKLER

Department of Economics
North Carolina State University

Abstract:

Vector Autoregressive (VAR) models which do not rely on a recursive model structure are discussed. Linkages to traditional dynamic simultaneous equations models are developed which emphasize the nature of the identifying restrictions that characterize VAR models. Explicit expressions for the Score and Information functions are derived and their role in model identification, estimation and hypothesis testing is discussed.

Introduction

Vector Autoregression (VAR) models have become a widespread tool for forecasting, an application in which their virtues have been well documented (Litterman). As a tool for structural and policy analysis, VAR models are more controversial. The VAR methodology was initially formulated in an attempt to impose minimal restrictions on economic data in the belief that many controversies would never be resolved as long as empirical econometric models were overidentified using what Sims (1980) referred to as incredible restrictions. By imposing minimal restrictions on a model, it was felt that the true structure of the economic system under investigation would emerge.

While this aim was perhaps laudable, it had the unfortunate consequence of holding out the promise that something could be obtained for nothing. Critics of VAR models (Leamer; Cooley and Leroy) point out that in simultaneous equation models (SEMs) it is necessary to make some identifying assumptions to give economically interpretable meaning to model results. It is telling such a simple observation need be made at all. The explanation for this seems to lie in the fact that VAR and other time series methods are often treated as distinct from standard SEMs, even though they are better viewed as special cases of the latter.

There are, of course, special features of VAR models that distinguish them from other SEMs. Central to the VAR methodology are the concepts of the Impulse Response

Function (IRF) and the Forecast Error Variance Decomposition (FEVD), which are both measures of impact of uncertainty in a system caused by the individual shocks that drive the system. For these concepts to make sense, it is necessary to specify a model in terms of a set of primitive, orthogonal shocks that are economically interpretable. Indeed, it is the central place of these shocks and their interpretation that distinguishes the VAR approach from much of traditional econometric practice, which often treats the stochastic aspect of a model as a nuisance rather than as an intrinsic part of the system being examined.

Another distinguishing feature of VAR models is that the associated reduced form model is completely unrestricted. The term VAR itself implies this, and it is common to hear VARs models referred to as unrestricted reduced form models. The critical implication of this feature of the VAR methodology is that these models are identified solely by the restrictions placed on the contemporaneous interactions among endogenous variables.

The important point to be made about these two features is that they are both aspects of the familiar identification problem. Ultimately, the believability of results derived from a VAR or any other SEM will depend on the believability of the identifying assumptions. The most telling criticism of the application of VAR methodology is that the usual practice of imposing a recursive identification on a model is unbelievable. While there may be situations in which a recursive structure is appropriate, they are the exception rather than the rule.

Recently several economists have made use of the features of the VAR methodology in models that are not recursive (Blanchard and Watson, Bernanke, and Sims (1986)). This paper discusses this generalized approach to VAR models. It is meant to clarify the relationship between VAR models and general dynamic SEMs as well as to bring together a number of technical results concerning VAR models. Much of what appears here, while implicit in other works, is discussed systematically an in more detail in this paper. Also included are explicit expressions, which have not appeared elsewhere, for the Score and Information functions associated with VAR models subject to arbitrary linear parameter restrictions. These results facilitate examination of model identification, as well as estimation and hypothesis testing.

The format of the paper is as follows. The first section discusses the general formulation of dynamic SEMs and lays out the notation used. The particular identifying restrictions of the VAR approach are discussed in the second section. This os followed by a discussion of estimation procedures applicable to VAR models. The paper concludes with a few comments on the use of VAR models in economics.

Dynamic simultaneous equations models

A general specification of a dynamic linear SEM can be given by[1]:

$$y_t A = \sum_{s=1}^{\infty} y_{t-s} A_s + z_t C + v_t B,$$

where y_t and v_t are both (1xk) random vectors, A, the A_s and B are (kxk) matrices of coefficients, and z_t is a (1xq) vector of nonstochastic (or strictly exogenous) variables[2]. It is assumed that

$$E[v_t] = 0 \quad \text{and} \quad E[v_t' v_s] = \delta_{st} I_k = \begin{cases} I_k & t=s \\ 0 & t\neq s \end{cases}$$

i.e., the v_t are vectors of serially uncorrelated and mutually orthogonal shocks with unit variance. The term impulses will be applied to these shocks, which represent the independent sources of variation in the system being modeled.

It is also assumed that the system is stationary and, therefore, that both an autoregressive (AR) and a moving average (MA) representation exist and can be obtained from one another by inversion. The AR representation is obtained by postmultiplying the system by A^{-1}:

$$y_t = \sum_{s=1}^{\infty} y_{t-s} A_s^* + z_t C^* + u_t,$$

where $A_s^* = A_s A^{-1}$, $C^* = CA^{-1}$ and $u_t = v_t BA^{-1}$, with $Cov(u_t) = \Omega = A^{-T} B' B A^{-1}$. The u_t are the mean zero, serially independent step-ahead forecast errors (conditional on z_t), also termed the system innovations. The MA representation is given by

$$y_t = \sum_{s=0}^{\infty} u_{t-s} M_s + \sum_{s=0}^{\infty} z_t C^* M_s,$$

where $M_0 = I_k$ and the M_s, s>0, can be calculated from the A_s^* according to the relationship

$$M_s = \sum_{i=1}^{s} A_i^* M_{s-i}.$$

An alternative representation of the system can be written in terms of the orthogonal shocks, v_t,:

$$y_t = \sum_{s=0}^{\infty} v_t R_s + \sum_{s=0}^{\infty} z_t C^* M_s,$$

where $R_s = BA^{-1} M_s$. The R_s describe what is termed the impulse response function (IRF), which traces the impact of each of the (orthogonal) system impulses on the observable system variables. $[R_s]_{ij}$ represents the impact on variable j when impulse i was one unit in size (one standard deviation) s periods previously. The IRF, therefore, measures both the source and the strength of each of the stochastic forces affecting a given variable as well as the time of the response to those forces. The use of the IRF is a hallmark of the VAR methodology. When policy interventions are associated with a particular system impulse, the IRF is the proper tool for analyzing the dynamic impact of that policy[3].

The stationarity assumption ensures that the A_s^* will be close to zero for large enough s. It is therefore convenient and useful to assume that, for s>p, G(s)=0 or, equivalently, $A(s)=0$. The intuition behind the assumption is that the distant past has little or no independent effect on the present; *i.e.*, the effect of the distant past is expressed entirely through the more recent past[4]. With this assumption the model can be written

$$y_t = x_t\beta + u_t, \text{ where } x_t = [y_{t-1}\ y_{t-2} \cdots\ y_{t-p}\ z_t] \text{ and}$$

$$\beta = [A_1^*,\ A_2^*,\ \ldots\ A_p^*,\ C^*]' \ (\beta \text{ is } ((kp+q) \times k)).$$

In standard terminology, A, B, the A_s, and C are called the structural parameters, whereas β and Ω are called the reduced form parameters. Notice that once A and B known, the other structural parameters are obtainable directly from the reduced form parameters.

The stochastic nature of the model can be specified completely by assigning a probability law to the impulses. Here it is assumed that v_t are multinormal; because they are linear combinations of the v_t, the u_t are also multinormal. The loglikelihood for this model (for y_t, t=1,...,T) is

$$\ell = -\frac{1}{2}\left[Tk\ \ln(2\pi) + T\ \ln|\Omega| + \sum_{t=1}^{T} (y_t - x_t\beta)\Omega^{-1}(y_t - x_t\beta)'\right]$$

$$= -\frac{1}{2}[Tk\ \ln(2\pi) + T\ \ln|\Omega| + tr(\Omega^{-1}U'U)]$$

$$= -\frac{1}{2}\ Tk\ \ln(2\pi) + T\ \ln|AB^{-1}| - \frac{1}{2}\ tr(AB^{-1}B^{-T}A'U'U),$$

where $U=Y-X\beta$ and Y and X denote the matrices composed of the T observations on y_t and x_t. It is assumed that X has full column rank, an assumption that ensures the identifiability of the reduced form parameters.

The number of reduced form parameters in this model equals $k(pk+q)+k(k+1)/2$, corresponding to the β and Ω matrices, while there are $(p+1)k^2+qk+k^2$ structural parameters corresponding to A and the A_s, C, and B. There are, therefore, $(3k^2-k)/2$ more structural than the reduced form parameters. An order condition for identification thus is that $(3k^2-k)/2$ restrictions must be imposed on the structural parameters. In the traditional SEM little value is placed on specific knowledge of B, it being considered adequate to estimate B'B, which has only $k(k+1)/2$ free parameters. B'B is the covariance matrix of the (non-orthogonal) structural errors (*i.e.*, the v_tB). This reduces the identification problem to one of imposing k^2 restrictions on A, the A_s and C[5]. Clearly the identifying restrictions cannot be confined to the contemporaneous coefficients matrix A in this case, unless A is completely known *a priori.*

Identification in VAR models

In contrast, the VAR approach concentrates all identifying restrictions on the A and B matrices. The reason for this stems from two features specific to this approach. First, it is considered desirable to be able to trace the impact of each of the impulses on the endogenous variables. This is not possible unless the elements of B can be identified. Second, the modeling philosophy that has developed with the VAR approach deems it

desirable to leave the reduced form parameters associated with the lagged endogenous and exogenous variables (β) relatively unencumbered with model specific restrictions that would be implied by restrictions on the A_s and C.

At least two substantive rationales exist for focusing attention exclusively on restrictions on contemporaneous interactions. Many economic variables are determined in a setting in which the values of past realizations of all variables relevant to a system are known to economic agents and potentially will be used to form expectations about the future state of the economy. These expectations provide a link between past and current realizations of all the variables in a given model. Given the considerable controversy that surrounds expectation formation processes, it is deemed desirable to let the data speak for itself. On the other hand, it is often reasonable to assume that variables do not react immediately to new economic developments because of information lags or adjustment costs. Such minimum delay considerations provide one useful source of identifying restrictions on the contemporaneous interactions among variables. A third reason for using this sort of identification structure is that there is a significant gain in computational ease when it is possible to separate any restrictions placed on β from those placed on A and B. Undoubtedly this has influenced the development of this methodology.

By concentrating on A and B, the contemporaneous coefficients matrices, and leaving the reduced form coefficient matrix β unrestricted, the order condition implies that the number of free parameters in A and B must be less than or equal to $k(k+1)/2$, the number of free parameters in Ω, implying that at least $(3k^2-k)/2$ restrictions must be imposed. Normalization (scaling) will reduce this number to $3(k^2-k)/2$.

General (linear) restrictions can be represented by

$$R\,vec([A\ B]) = r,$$

where R has $2k^2$ columns and the number of rows in both R and r is equal to the number of restrictions imposed on the model. A more useful representation of the restrictions can be made, however, in terms of a vector of underlying free parameters of the system, here denoted θ. This general framework is given by

$$vec([A\ B]) = Z\theta + W,$$

where Z, θ, and W are $(2k^2 \times n)$, $(n \times 1)$, and $(2k^2 \times 1)$, respectively. Viewed in terms of the number of free parameters, the order condition is $n \leqslant k(k+1)/2$. While the two representations are equivalent[6], the parametric representation facilitates estimation, since θ is the vector of underlying parameters to be estimated directly, with Z and W defining the transformation of θ into A and B. This representation allows completely general (linear) constraints to be imposed on A and B, including zero constrinsts (the i^{th} rows of Z and W equal to 0) as well as within –and cross– equation constraints (two or more non-zero elements in the j^{th} column of Z).

A simple example will clarify the relationship between the two methods for representing restrictions. Suppose k=3 and it is assumed that $B=I_k$. Letting $vec(A)=Z_1\theta+W_1$ and $vec(B)=Z_2\theta+W_2$, this restriction can be imposed by setting $Z_2=0$ $(9 \times n)$ and $W_2=vec(I_3)$. This imposes $k^2=9$ restrictions and therefore at least $k(k-1)/2=3$ additional restrictions must be imposed. Let these restrictions be $a_{31}=0$, $a_{21}=a_{12}$, and $a_{13}+a_{23}+a_{33}=1$. These restrictions can be imposed directly according to $R_1\,vec(A)=r_1$, where

$$R_1 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad r_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Note that R and r are not unique and that the same restrictions would be imposed if both were pre-multiplied by any nonsingular (3x3) matrix. The restrictions can also be imposed in parametric fashion by setting

$$Z_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & -1 \end{bmatrix} \quad \text{and} \quad W_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

With Z_1 defined in this way, θ corresponds to $(a_{11}\ a_{21}\ a_{22}\ a_{32}\ a_{13}\ a_{23})'$; but this need not be the case. The same restrictions would be imposed if Z_1 were post-multipled by any nonsingular matrix (with θ appropriately redefined).

The order condition for identification involves simply counting the number of free parameters in the model or, equivalently, the number of restrictions imposed on A and B. As Rothenberg has shown, a necessary and sufficient condition for the local identifiability of any regular point in R^n (*i.e.*, any point, θ, for which the Information matrix $I(\theta)$ has constant rank in a neighborhood of θ) is that $I(\theta)$ be full rank (expressions for $I(\theta)$ are derived in Appendix A and discussed more fully in the next section). In principle this condition should be verifiable by examination of Z and W, which define the restrictions on A and B. Unfortunately no general results appear to be available. As a practical matter the examination of the rank of $I(\theta)$ for a few random values of θ should be sufficient to establish the local identifiability of a given model.

It should be pointed out that neither rank(Z)= n nor rank(A)=rank(B)=k is sufficient to establish the identifiability of a given structure, though these clearly are necessary conditions. An example will suffice to demonstrate this point. Suppose k=4, $B=I_k$ and

$$A = \begin{bmatrix} \theta_1 & \theta_3 & \theta_5 & \theta_8 \\ \theta_2 & 0 & \theta_6 & 0 \\ 0 & \theta_4 & 0 & \theta_9 \\ 0 & 0 & \theta_7 & \theta_{10} \end{bmatrix}$$

(Z_1 is therefore composed of columns, 1, 2, 5, 7, 9, 10, 12, 13, 15, and 16 of I_{16}.) In this case A is invertible (except on a set of measure zero) and satisfies the order condition for (exact) identification (n=k(k+1)/2=10) but $I(\theta)$ has rank 9. This can be verified by choosing at random a value for θ.

It is also important to note that there is an essential redundancy in the A and B matrices. The restrictions imposed on A can be thought of as describing how the variables in the system interact contemporaneously, whereas the restrictions on B describe the direct impacts of the shocks on the equations of the system, so that nondiagonal elements of B allow for more than one shock to enter a given equation directly. Often there is

more than one way to formulate a given model, however. For example, the model defined by

$$A = \begin{bmatrix} a1 & a2 & 0 \\ 0 & a3 & a4 \\ 0 & 0 & a5 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 & b1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and that defined by

$$A = \begin{bmatrix} a1 & a2 & a4 \\ 0 & a3 & a5 \\ 0 & 0 & a6 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

are equivalent. In both the first shock is equated with the innovation to the first variable. Hence it is irrelevant whether its impact on the third variable is said to enter through A or through B. Technically, if general nonlinear restrictions were used, there would be no need to use both matrices explicitly, as either one or the other would suffice[7]. In practice, however, it may be preferable to place restrictions on both matrices if such restrictions can be given a readily interpretable meaning.

Estimation techniques

One main advantage of the VAR model is that the identifying restrictions allow the reduced form parameters to be estimated separately from the contemporaneous coefficients matrices, A and B. The reduced form coefficients can be estimated efficiently using OLS. Maximum likelihood estimates of A and B conditional on the estimated values of the reduced form coefficients then can be estimated. This two-stage estimation approach yields FIML coefficient estimates even if the model is overidentified because the identifying restrictions on A and B are implicitly covariance restrictions and have no implications for the reduced form coeficients β, in contrast to the case of the general SEM.

Details of the estimation strategy proposed here are most easily derived for the case in which $B=I_k$. This restriction implies that each system impulse enters only one equation directly (*i.e.*, B is diagonal), and that the normalization restrictions are applied to B. This results in the log likelihood:

$$\ell = \frac{Tk}{2}\ln(2\pi) + T\ \ln||A|| - \frac{1}{2}\operatorname{vec}(A)'\,(I_k \otimes (Y-X\beta)'\,(Y-X\beta))\operatorname{vec}(A).$$

It can be shown (see Appendix A) that

$$\frac{\partial \ell(\theta,\beta)}{\partial \operatorname{vec}(\beta)} = \operatorname{vec}(X'\,(y-X\beta)AA').$$

Setting this equal to 0 and solving for β yields

$$\hat{\beta} = (X'X)^{-1}\,X'Y,$$

i.e., the OLS estimator (recall that X is assumed to have full column rank)[8].

The fact that the FIML estimator for β is independent of A suggests the two-stage estimation procedure discussed by Sims (1986). In the first stage the OLS estimate of β is calculated. In the second stage numerical optimization methods are used to solve for the FIML estimate of θ:

$$\hat{\theta} = \arg\max_{\theta} \ell(\theta,\hat{\beta}).$$

To implement this strategy first define an estimator of Cov $(u_t)=\Omega$ by $\tilde{\Omega}=\hat{U}'\hat{U}/T$, where $\hat{U}=Y-X\hat{\beta}$ are the least squares residuals. In Appendix A it is shown that

$$\frac{\partial \ell(\theta,\beta)}{\partial \theta} = Z_1' \,(T\,\text{vec}(A^{-T}) - \text{vec}(U'UA)).$$

Evaluating the likelihood and its gradient with respect to θ at $\hat{\beta}$ yields

$$\ell(\theta,\hat{\beta}) = -T[\tfrac{k}{2}\ln(2\pi) + \ln(\|A\|) - \tfrac{1}{2}\text{vec}(A)'\,\text{vec}(\tilde{\Omega}A)]$$

and

$$\frac{\partial \ell(\theta,\hat{\beta})}{\partial \theta} = TZ_1'\,\text{vec}(A^{-T} - \tilde{\Omega}A).$$

Both of these functions involve β and the data only through the estimator $\tilde{\Omega}$, a fact that greatly facilitates estimation of θ.

Note that $\tilde{\Omega}$ is not necessarily the FIML estimator of Ω, which in general is given by $\hat{\Omega}=\hat{A}^{-T}\hat{A}^{-1}$, where vec $(\hat{A})=Z_1\hat{\theta}+W_1$. $\tilde{\Omega}$ is not FIML because it fails to account for possible overidentifying restrictions, though the two estimators should be quite close if the identifying restrictions are good. In the exactly identified case, however, it will always be possible to find an A such that $\tilde{\Omega}=A^{-T}A^{-1}$, which satisfies the first order necessary conditions (FONC) for a maximum. The two estimators therefore will coincide in the exactly identified case. This situation is discussed by Bernanke, who notes that estimates of A can be obtained by solving $A'\tilde{\Omega}A=I_k$ using a nonlinear root finding algorithm[9].

Standard VAR practice implicitly exploits this relationship by setting A equal to the inverse of the Cholesky decomposition of $\tilde{\Omega}$. While this makes estimation easy by eliminating the need for a numerical search, it imposes an upper triangular form on A, implying that the system has a recursive structure. While the "identification" problem is thereby reduced to establishing an ordering for the variables in the system, the believability of such a structure is generally questionable, raising doubts about the validity of model interpretations.

For the general case maximum likelihood methods provide a straightforward estimation framework. The conditional log-likelihood and score functions provided above can be used in conjunction with quasi-Newton nonlinear optimization algorithms. One such algorithm, the method of scoring, also requires the information matrix. This matrix is also useful in checking model identification, as discussed in the previous section, in evaluating the quality of the estimators and in hypothesis testing, its inverse being equal to the asymptotic Cov$(\hat{\theta},\hat{\beta})$. It can be shown that the Information matrix, defined by

$$I(\theta,\beta) = -E\begin{bmatrix} \dfrac{\partial^2 \ell}{\partial\theta\partial\theta'} & \dfrac{\partial^2 \ell}{\partial\theta\partial\,\text{vec}(\beta)'} \\ \dfrac{\partial^2 \ell}{\partial\,\text{vec}(\beta)\partial\theta'} & \dfrac{\partial^2 \ell}{\partial\,\text{vec}(\beta)\partial\,\text{vec}(\beta)'} \end{bmatrix},$$

is block diagonal by noting that the upper right-hand term,

$$-E\left[\frac{\partial\,\text{vec}(X'UAA')}{\partial\theta}\right] = -\frac{\partial\,\text{vec}(AA')}{\partial\theta}\;E[I_k \otimes U'X],$$

is equal to zero, since X'U has expectation zero[10].

In Appendix A it is shown that

$$\frac{\partial^2 \ell}{\partial\theta\partial\theta'} = -Z_1'\,(TP_{k,k}(A^{-T}\otimes A^{-1}) + (I_k \otimes U'U))Z_1,$$

where $P_{m,n}$ is the permutation matrix defined by vec$(A')=P_{m,n}$vec(A), where A is any (mxn) matrix.

The upper left-hand block of the information matrix, which is associated with θ (and here denoted $I(\theta)$), may be obtained by replacing U'U with its expectation, $T(A^{-T}A^{-1})$:

$$I(\theta) = TZ_1'\,(P_{k,k}(A^{-T}\otimes A^{-1}) + (I_k \otimes A^{-T}A^{-1}))Z_1.$$

Note that this term is functionally independent of β.

If the model is generalized to include a nondiagonal B matrix, the separation between β and θ continues to hold. This again allows for a two-step estimation procedure. Indeed the first step is identical and yields the estimator $\tilde{\Omega}$. The likelihood and its gradient with respect to θ can again be evaluated at $\hat{\beta}$, yielding (see Appendix A for details)

$$\ell(\theta,\hat{\beta}) = -T[\tfrac{k}{2}\ln(2\pi) + \ln(\|AB^{-1}\|) - \tfrac{1}{2}\text{vec}(AB^{-1})'\,\text{vec}(\tilde{\Omega}AB^{-1})],$$

and

$$\frac{\partial \ell(\theta,\hat{\beta})}{\partial\theta} = TZ'\begin{bmatrix} \text{vec}(A^{-T} - \tilde{\Omega}AB^{-1}B^{-T}) \\ \text{vec}(B^{-T}A'\tilde{\Omega}AB^{-1}B^{-T} - B^{-T}) \end{bmatrix}$$

In this case any A and B such that $\tilde{\Omega}=A^{-T}B'BA^{-1}$ will satisfy the FONC for a maximum and, again, such a solution will always be possible in the exactly identified case. The information matrix again will be block diagonal with the upper left-hand block given by:

$$I(\theta) = TZ' \begin{bmatrix} P_{k,k}(A^{-T} \otimes A^{-1}) + (B^{-1}B^{-T} \otimes A^{-T}B'BA^{-1}) \\ -(B^{-1}B^{-T} \otimes BA^{-1}) - P_{k,k}(B^{-T} \otimes A^{-1})\ (B^{-1}B^{-T} \otimes I_k) + P_{k,k}(B^{-T} \otimes B^{-1}) \end{bmatrix} Z.$$

Finally, note a special case of the general VAR model that is of interest because it permits a simple recursive two stage least squares (2SLS) algorithm to be used to estimate the coefficients of A and B. The quasitriangular specification is one in which, for some ordering of variables and equations, A has unit diagonal and B is diagonal and in which the ith equation (column of A) involves at most (i-1) elements of θ. This special case is discussed more fully in Appendix B. It is also discussed by Bernanke and used in an empirical application by Blanchard and Watson.

Summary

This paper has discussed the relationship between VAR models and other dynamic SEMs. The distinguishing feature of the VAR methodology is the imposition of identifying restrictions only on the contemporaneous interactions and on the use of orthogonal impulses that can be given economic interpretation. Unfortunately the usual practice of VAR modeling has involved the use of a rather suspect form of identifying restrictions. Furthermore many practitioners seem to impose these restrictions implicitly rather than explicitly, without a clear recognition of the implications. It is not unusual to find discussion of the need to "orthogonalize" the innovations (the u_t) to coinstruct the IRF as if this were a mechanical operation. While the limitations of the usual practice of using a "triangular orthogonalization", with its implication that the system is recursive, seems to be well-recognized, the response by practitioners has been to examine alternative orderings of variables to assess the robustness of the results. This does not address whether the results are robust to other identification regimes, and, as Bernanke points out, the practice implies a strange prior in wich the analyst believes strongly in the recursiveness of the system but is not sure in what order the variables should be arranged.

While clearly the recursive model is not acceptable generally, at least two substantive reasons exist for focusing on the contemporaneous interactions within the system. First, economic theory says very little that is not controversial about the nature of expectations. It is therefore prudent to leave relatively unrestricted the reduced form of the model, which can itself be viewed as a forecasting model. Second, lags in the speed with which variables can respond to shocks because of information lags and adjustment costs lead to a minimum delay rationale for contemporaneous identifying restrictions. Formulating believable identifying restrictions is never a trivial task. Whether VAR models prove to be useful for structural analysis will depend on wheter such considerations will lead to enough restrictions to identify a model. Identifying situations in which this is or is not the case is the challenge VAR methodology poses to economists. By clarifying the unique nature of this methodology and providing technical results useful in its implementation, this paper should aid researchers interested in pursuing this challenge.

Notes

1 Note that this formulation post-multiplies variables by coefficients.

2 The inclusion of deterministic variables in the z_t vector raises no problems. Strictly exogenous variables are those that are uncorrelated with the system impulses and not affected by the endogenous (system) variables. This is essentially equivalent to assuming a block triangular structure for the A(s) and a block diagonal structure for the B(s). This allows the density function for y and z to be partitioned into a part representing y conditioned on z and a part representing the density of z, which is independent of y. Note that this implies that lagged y is not useful in predicting current z.

3 Another measure of the impact of the impulses on the system is given by the forecast error variance decomposition (FEVD), which measures the percentage contribution of the ith impulse to the ℓ-step-ahead forecast error variance of the jth variable:

$$F(\ell)_{ij} = \frac{\sum_{s=0}^{\ell-1} R_{ij}^2(s)}{\sum_{i=1}^{k} \sum_{s=0}^{\ell-1} R_{ij}(s)}$$

4 It should be noted, however, that determination of the lag length is not an easy or trivial matter. It is possible for the system to have a dynamic structure involving relatively high values of p, but that can be approximately represented by a low p system. If standard prediction error methods are used to determine the level of p, the lower value will be chosen and the structural aspects of the system may be incorrectly represented.

5 Restrictions could be imposed on B'B matrix, but this has been done only rarely in practice. However, see Hausman and Taylor and Hausman, Newey, and Taylor.

6 This can be checked by simply setting θ randomly and verifying that $R(Z\theta+W)-r=0$.

7 Non-linear restrictions could be written in the form

$$vec(A) = f(\theta)$$

or, if it is desirable to include B explicitly, in the form

$$vec([A\ B]) = f(\theta).$$

Restrictions of this type have arisen in the context of rational expectations econometric models, where θ is taken to be a vector of "deep" structural parameters representing such things as technology and agent preferences. By defining $Z(\theta)=Df(\theta)$, the results derived in Appendix A and discussed in the section on estimation could be extended in a very straightforward manner. Such an extensión is not pursued here, however.

8 The uniqueness of this estimator is guaranted when A has full runk, a condition also necessary for identification.

9 Bernanke seems to suggest (incorrectly) that only in the exactly identified case will the two-stage procedure yield FIML estimates.

10 The discussion of this point by Bernanke (pp. 13-4) appears to be in error.

Appendix A

Calculation of the score function and Hessian is facilitated by the following eight results of matrix algebra and calculus. Used here is the convention that the derivative of an n-vector with respect to an m-vector is (mxn), with $[\partial Y/\partial X]_{ij}=\partial Y_j/\partial X_i$. With the exception of the product rule (6), which makes use of (2) and (4), references to the text by Graham are provided.

(1) $\mathrm{tr}(XY) = (\mathrm{vec}(X'))'\mathrm{vec}(Y)$ (Table 1, p. 121)

(2) $\mathrm{vec}(XYZ) = (Z'\otimes X)\mathrm{vec}(Y)$ (Et. 2.13, p. 25)

(3) $(X\otimes Y)(Z\otimes W) = XZ\otimes YW$ (Table 2, p. 122)

(4) $\frac{\partial A\mathrm{vec}(X)}{\partial \mathrm{vec}(X)} = A'$ (Table 3, p. 122)

(5) $\frac{\partial \mathrm{vec}(Z)}{\partial \mathrm{vec}(X)} = \frac{\partial \mathrm{vec}(Y)}{\partial \mathrm{vec}(X)} \frac{\partial \mathrm{vec}(Z)}{\partial \mathrm{vec}(Y)}$ (Table 3, p. 122)

(6) $\frac{\partial \mathrm{vec}(YZ)}{\partial \mathrm{vec}(X)} = \frac{\partial \mathrm{vec}(Y)}{\partial \mathrm{vec}(X)} (Z\otimes I) + \frac{\partial \mathrm{vec}(Z)}{\partial \mathrm{vec}(X)} (I\otimes Y')$

(7) $\frac{\partial \mathrm{vec}(AX^{-1}B)}{\partial \mathrm{vec}(X)} = -(X^{-1}B)\otimes(X^{-T}A')$ Table 5, p. 124)

(8) $\frac{\partial \ln|X|}{\partial X} = X^{-T} \qquad |X| > 0$ (Table 6, p. 124)

These numbered results are referred to in the derivations below. Also used is the permutation matrix, $P_{m,n}$, defined by $\mathrm{vec}(X')=P_{m,n}\mathrm{vec}(X)$, where X is (mxn). It can be shown that $P'_{m,n}=P_{n,m}$ and for X (mxn) and Y (pxq), $Y\otimes X=P_{p,m}(X\otimes Y)P_{n,q}$ (Graham, p. 28). Together these results imply that $P_{m,m}(X\otimes X')$ is symmetric.

The loglikelihood for the model discussed in the paper is a function of the vectors, θ and β, coresponding to the contemporaneous and the reduced form parameters. It can be written as

$$M = -\frac{1}{2}Tk\ \ln(2\pi) + T\ \ln|AB^{-1}| - \frac{1}{2}\mathrm{tr}(AB^{-1}B^{-T}A'U'U),$$

where $U=Y-X\beta$, $\mathrm{vec}(A)=Z_1\theta+W_1$, and $\mathrm{vec}(B)=Z_2\theta+W_2$. With $\Omega^{-1}=AB^{-1}B^{-T}A'$ used for notational convenience, the block of the score function associated with the reduced form coefficients, β, can be derived as follows:

$$\frac{\partial \ell}{\partial \mathrm{vec}(\beta)} = -\frac{1}{2}\frac{\partial \mathrm{vec}(U)}{\partial \mathrm{vec}(\beta)} \frac{\partial \mathrm{vec}(U\Omega^{-1})'\mathrm{vec}(U)}{\partial \mathrm{vec}(U)} \qquad (5,1)$$

$$= \frac{1}{2}(I_k\otimes X')(\mathrm{vec}(U\Omega^{-1}) + (\Omega^{-1}\otimes I_k)\mathrm{vec}(U)) \qquad (2,4,6)$$

$$= \frac{1}{2}(I_k\otimes X')\mathrm{vec}(U\Omega^{-1} + U\Omega^{-1}) \qquad (2)$$

$$= \mathrm{vec}(X'U\Omega^{-1}) \qquad (2)$$

This in turn leads to the familiar result for the related block of the Hessian matrix:

$$\frac{\partial^2 \ell}{\partial \mathrm{vec}(\beta)\partial \mathrm{vec}(\beta)'} = \frac{\partial \mathrm{vec}(X'U\Omega^{-1})}{\partial \mathrm{vec}(\beta)}$$

$$= \frac{\partial \mathrm{vec}(U)}{\partial \mathrm{vec}(\beta)} \frac{\partial \mathrm{vec}(X'U\Omega^{-1})}{\partial \mathrm{vec}(U)} \qquad (5)$$

$$= -(I_k\otimes X')(\Omega^{-1}\otimes X) \qquad (2,4)$$

$$= -\Omega^{-1}\otimes X'X \qquad (3)$$

Of more interest in the current contex are the score and Hessian functions related to the "nonlinear" parameter vector, θ. These results are first derived for the simpler case in which $B=I_k$. The relevant blocks for this special case are:

$$\frac{\partial \ell}{\partial \theta} = \frac{\partial \mathrm{vec}(A)}{\partial \theta}\left[T\frac{\partial \ln||A||}{\partial \mathrm{vec}(A)} - \frac{1}{2}\frac{\partial \mathrm{vec}(U'UA)'\mathrm{vec}(A)}{\partial \mathrm{vec}(A)}\right] \qquad (5,1)$$

$$= Z_1'(T\ \mathrm{vec}(A^{-T}) - \mathrm{vec}(U'UA)) \qquad (2,4,6,8)$$

and

$$\frac{\partial^2 \ell}{\partial\theta\partial\theta'} = \frac{\partial Z_1'(T\ \mathrm{vec}(A^{-T}) - \mathrm{vec}(U'UA))}{\partial\theta}$$

$$= Z_1'\left[T\frac{\partial \mathrm{vec}(A^{-T})}{\partial \mathrm{vec}(A)} - \frac{\partial \mathrm{vec}(U'UA)}{\partial \mathrm{vec}(A)}\right]Z_1 \qquad (5,4)$$

$$= -Z_1'(TP_{k,k}(A^{-T}\otimes A^{-1}) + (I\otimes U'U))Z_1 \qquad (7,2)$$

The associated block of the information matrix is found by replacing U'U by its expectation:

$$I(\theta) = TZ_1'(P_{k,k}(A^{-T}\otimes A^{-1}) + (I\otimes A^{-T}A^{-1}))Z_1.$$

In the general case in which B is not necessarily the identity matrix, the analogous results the score function is

$$\frac{\partial \ell}{\partial \theta} = \frac{\partial \mathrm{vec}(A)}{\partial \theta}\left[T\frac{\partial \ln||A||}{\partial \mathrm{vec}(A)} - \frac{1}{2}\frac{\partial \mathrm{vec}(A)'\mathrm{vec}(U'UAB^{-1}B^{-T})}{\partial \mathrm{vec}(A)}\right]$$

$$+ \frac{\partial \mathrm{vec}(B)}{\partial \theta}\left[-\frac{1}{2}\frac{\partial \mathrm{vec}(B^{-1})}{\partial \mathrm{vec}(B)}\frac{\partial \mathrm{vec}(^{-1})'\mathrm{vec}(A'U'UAB^{-1})}{\partial \mathrm{vec}(B^{-1})} - T\frac{\partial \ln||B||}{\partial \mathrm{vec}(B)}\right] \qquad (5,1)$$

$$= Z' \begin{bmatrix} T\,\mathrm{vec}(A^{-T}) - \mathrm{vec}(U'UAB^{-1}B^{-T}) \\ \mathrm{vec}(B^{-T}A'U'UAB^{-1}B^{-T}) - T\,\mathrm{vec}(B^{-T}) \end{bmatrix} \qquad (8,2,3,7)$$

The related block of the Hessian is

$$\frac{\partial^2 \ell}{\partial\theta\partial\theta'} =$$

$$Z' \begin{bmatrix} -TP_{k,k}(A^{-T}\otimes A^{-1}) - (B^{-1}B^{-T}\otimes U'U) \\ -\frac{\partial \mathrm{vec}(U'UAB^{-1}B^{-T})}{\partial \mathrm{vec}(B)} \quad TP_{k,k}(B^{-T}\otimes B^{-1}) + \frac{\partial \mathrm{vec}(B^{-T}A'U'UAB^{-1}B^{-T})}{\partial \mathrm{vec}(B)} \end{bmatrix} Z \qquad (7,2,4,5)$$

where

$$\frac{\partial \mathrm{vec}(U'UAB^{-1}B^{-T})}{\partial \mathrm{vec}(B)} = -(B^{-1}B^{-T}\otimes B^{-T}A'U'U) - P_{k,k}(B^{-T}A'U'U) \qquad (6,7)$$

and

$$\frac{\partial \mathrm{vec}(B^{-T}A'U'UAB^{-1}B^{-T})}{\partial \mathrm{vec}(B)} = -(B^{-1}B^{-T}\otimes Q) - P_{k,k}(QB^{-T}\otimes B^{-1} + B^{-T}\otimes B^{-1}Q) \qquad (6,7)$$

with $Q=B^{-T}A'U'UAB^{-1}$. Note that $E[U'U]=TA^{-T}B'BA^{-1}$ and hence $E[Q]=TI_k$. The associated block of the information matriz, therefore, simplifies to

$$I(\theta) = TZ' \begin{bmatrix} P_{k,k}(A^{-T}\otimes A^{-1}) + (B^{-1}B^{-T}\otimes A^{-T}B'BA^{-1}) \\ -(B^{-1}B^{-T}\otimes BA^{-1}) - P_{k,k}(B^{-T}\otimes A^{-1}) \quad (B^{-1}B^{-T}\otimes I_k) + P_{k,k}(B^{-T}\otimes B^{-1}) \end{bmatrix} Z$$

Finally, the off diagonal block of the Hessian:

$$\frac{\partial^2 \ell}{\partial \mathrm{vec}(\beta)\partial\theta} = \frac{\partial \mathrm{vec}(X'U\Omega^{-1})}{\partial\theta} = \frac{\partial \mathrm{vec}(\Omega^{-1})}{\partial\theta}(I_k\otimes U'X),$$

the expectation of which is clearly zero.

Appendix B

A quasi-triangular specification is a special case of the general formulation that permits a recursive two-stage least squares (2SLS0 estimator to be employed. In the over-identified case this will not result, in general, in the FIML estimator but typically will provide quite good starting values if the FIML estimator is desired. While more general specifications are perhaps possible, it will be assumed here that, for some ordering of variables and equations, A has unit diagonal and B is diagonal. A quasi-triangular system is one in which the ith equation (column of A) involves at most (i-1) elements of θ. This condition is equivalent to the ith (kxn) block of Z_1 having at most (i–1) non-zero columns. If only zero-restrictions are used (in addition to the normalization), a quasi-triangular specification is one in which the ith equation involves at most i variables. It was in this sense that the term was used by Bernanke.

Estimation of such a model with recursive 2SLS involves using the first (i–1) columns of V, the system impulses, to create instruments for the variables included in the ith equation. The procedure can be described as follows. Create a set of index variables ι_i that contain the indexes of the elements of θ that enter the ith equation but have not yet been estimated. Note that ι_1 and possible others may be empty. Let Z_i^* equal the ι_i columns of Z_{1i}, the ith block of Z_1, and initialize θ=) (nx1) and vec(A)=W_1.

On the ith iteration check if ι_i is empty. If not, set

$$\theta_{\iota i} = (Q_i'Q_i)^{-1}Q_i'R_{i-1}A_{\cdot i},$$

where R_{i-1} equals the first (i–1) rows of V'U/T and $Q_i=R_{i-1}Z_i^*$. At this point A will be based only on those elements of θ that have already been estimated (and on W_1). Update A by setting vec(a)=$Z_1\theta+W_1$. On all iterations, set $B_{ii}=(A_{\cdot i}'\hat{\Omega}A_{\cdot i})^{0.5}$ and set $R_i=A_{\cdot i}'\hat{\Omega}/B_{ii}$. Notice that the algorithm requires only $\hat{\Omega}$ and not $\hat{U}$ and that $\hat{V}$ is not directly calculated. Q_i is the projection of the included columns of U in equation i on columns 1 though (i–1) of V, a mapping that is facilitated by the fact that $E[V'V]=TI_k$.

References

BERNANKE, B. S. (1986). "Alternative Explanations of the Money-Income Correlation". NBER Working Paper No. 1842 (February 1986). Also in *Carnegie-Rochester Series on Public Policy*. 25: 49-100.

BLANCHARD, O., and M. W. WATSON (1986). "Are Business Cycles All Alike?" *The American Business Cycle-Continuity and Change*, Ed. Robert J. Gordon, University of Chicago Press. 123-179.

COOLEY, T. F., and S. F. LeROY (1985). "Atheoretical Macroeconometrics: A Critique". *Journal of Monetary Economics*. 16: 283-308.

GRAHAM, A. (1981). *Kroneker Products and Matrix Calculus with Applications*. Ellis Horwood Limited. Chichester.

HAUSMAN, J. A. and W. E. TAYLOR (1983). "Identification in Linear Simultaneous Equations Models with Covariance Restrictions: An Instrumental Variables Interpretation". *Econometrica*. 51: 1527-1549.

HAUSMAN, J.A.; W. K. NEWEY, and W. E. TAYLOR (1987). "Efficient Estimation and Identification of Simultaneous Equations Models with Covariance Restrictions". *Econometrica*. 55: 849-874.

KING, R. G. (1986). "Money and Business Cycles: Comments on Bernanke and Related Literature". *Carnegie-Rochester Series on Public Policy*. 25: 101-116.

LEAMER, E. E. (1985). "Vector Autoregressions for Causal Inference?". *Carnegie-Rochester Conference Series on Public Policy*. 22: 255-304.

LITTERMAN, R. B. (1986). "Forecasting with Bayesian Vector Autoregression - Five Years of Experience". *Journal of the Business and Economic Statistics.* 4: 25-38.
ROTHENBERG, T. J. (1971). "Identification in Parametric Models". *Econometrica.* 39: 577-91.
SARGENT, T. J. (1984). "Autoregression, Expectations, and Advice". *American Economic Review.* 74: 408-421.
SIMS, C. A. (1980). "Macroeconomics and Reality". *Econometrica.* 48: 1-48.
SIMS, C. A. (1982). "Policy Analysis with Econometric Models". *Brookings Papers on Economic Activity.* 1: 107-164.
SIMS, C. A. "Are Forecasting Models Usable for Policy Analysis?". *Federal Reserve Bank of Minneapolis Quarterly Review.* Winter: 2-16.

Revista de Análisis Económico, Vol. 3, Nº 2, pp. 135-158 (Noviembre 1988)

ANALISIS DE LA LEY ANTIMONOPOLIOS EN CHILE*

RICARDO D. PAREDES MOLINA
Departamento de Economía
Universidad de Chile

Abstract:

The purpose of this paper is to provide an explanation to the way in which the Chilean Antitrust Law has been enforced, based on all cases considered by the Resolving Committee, between 1974 and 1987.
Through an analysis based on global indicators such as: number of cases, channels followed and sanctioned, the study concludes that the institutions who enforce the Antitrust Law, were not guided by the principle of social welfare maximization.

1. Introducción

La Ley Antimonopolios en Chile ha tenido una corta tradición especialmente si se juzga su impacto en términos del número de casos tratados por las comisiones encargadas de velar por su cumplimiento.

Sin embargo, resulta altamente inconveniente estimar el impacto de una ley como la que se analizará en el presente estudio, en términos exclusivamente de lo que se aprecia en primera instancia, esto es, los casos efectivamente tratados por la justicia. Ello es así especialmente por el efecto disuasivo que las sanciones tienen sobre la realización de las conductas por eventuales infractores.

* Se agradece el financiamiento de Fondecyt y la disposición de la Fiscalía Nacional Económica, que dio la información para construir la base de datos. También agradezco los comentarios recibidos que mejoraron la presentación de este trabajo; en especial, los de un árbitro anónimo, de A. Iglesias y J. Marshall R., quien, además, me proporcionó la información sobre la Comisión Preventiva. También deseo agradecer a H. Gutiérrez y C. Torres por sus comentarios y críticas en un borrador anterior. Sin embargo, sólo el autor es responsable de las opiniones y errores que persistan.