FACULTY WORKING PAPERS

Vector Autoregressive Techniques for Structural Analysis

Paul L. Fackler

Faculty Working Paper No. 113 January 1988

DEPARTMENT OF ECONOMICS AND BUSINESS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA

Vector Autoregressive Techniques for Structural Analysis

Paul L. Fackler

Faculty Working Paper No. 113 January 1988

Vector Autoregression (VAR) models have become a widespread tool for forecasting, an application in which their virtues have been well documented (Litterman (1984,1986)). 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 has had the unfortunate consequence of holding out the promise that something could be obtained for nothing. Critics of VAR models (Leamer (1985), 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 that such a simple observation should need to be made at all. The explanation for this seems to lie in the fact that VAR and other time series methods, often are treated as distinct from standard SEMs, even though they are better viewed as special cases of the latter.

In this paper an attempt is made to present the VAR methodology in a way that reflects both the linkages with SEMs and the unique features of VARs that give them a special place in the tool box of applied econometricians. Central to the VAR methodology are the concepts of the Impulse Response Function (IRF) and the Forecast Error Variance Decomposition (FEVD). These rely on the ability to specify a model in terms of primitive, orthogonal shocks. Essentially this means that the covariance matrix of a set of economically interpretable shocks is diagonal. Indeed, it is the central place of these shocks and their interpretation that distinguishes the VAR approach from much of traditional econometric practice, in which the stochastic aspect of a model often is treated as a nuisance rather than as an intrinsic part of the system being examined.

Another feature commonly associated with 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. Actually, the critical feature of the VAR methodology is that the model is identified solely by the restrictions placed

on the contemporaneous interactions among endogenous variables. This feature allows estimation to proceed in a very straightforward manner and enables model specification to be altered at low cost.

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 concerning structural issues that are derived from a VAR or any other SEM will depend on the believability of the identifying assumptions made. The most telling criticisms of the application of VAR methodology is that the identifying restrictions are unbelievable. Most practitioners have identified their models as recursive systems. While there may be situations in which a recursive structure is appropriate they are the exception rather than the rule. Recently, however, several economists have used the features of the VAR methodology in models that are not recursive. Blanchard and Watson, Bernanke, and Sims (1986) all discuss models that have the essential features of a VAR model but without the assumption that the system is recursive in nature.

This paper discusses this generalized approach to VAR models. It is meant to explain the nature of the relationship between VAR models and general dynamic SEMs as well as to serve as a technical reference for those interested in using VAR models. Much of what appears here is implicit in other works but is discussed systematically and in more detail here. The paper also includes explicit expressions, which have not appeared elsewhere, for the Score and Information functions associated with VAR models subject to arbitrary linear parameter restrictions. The first section discusses the general formulation of dynamic SEMs. The particular identifying restrictions of the VAR approach are discussed in the second section. This is followed by a discussion of estimation procedures applicable to VAR models. The construction of the IRF and FEVD is then outlined. The paper concludes with a few comments on the use of VAR models in economics.

DYNAMIC SIMULTANEOUS EQUATION MODELS

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

$$\sum_{s=0}^{\infty} y_{t-s}A(s) = z_tC + \sum_{s=0}^{\infty} v_{t-s}B(s),$$

where y_t and v_t are both (1×k), A(s) and B(s) are (k×k), and z_t is a (1×q) vector of non-stochastic (or strictly exogenous) variables.[2] Furthermore, 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. The term impulses will be applied to these shocks, which represent the independent sources of variation in the system being modeled. The A(s) and B(s) describe the propagation mechanism linking the unobserved impulses to the observed phenomena y.

It will be 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 given by

$$y_t = \sum_{s=1}^{\infty} y_{t-s}G(s) + z_tC^* + u_t,$$

where the u_t are mean zero, serially independent random variables with $Cov(u_t)=\Omega$. The stationarity assumption ensures that the G(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. In the case that B(s)=0, s>0, and that A(s)=0, s>p, this condition will hold exactly, otherwise it represents an approximation. In either case 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.[3] With this assumption the model can be written $y_t = x_t\beta + u_t$, where

$$x_t = [y_{t-1}, y_{t-2}, \ldots, y_{t-p}, z_t],$$

and

$$\beta = \begin{bmatrix} G(1) \\ G(2) \\ \vdots \\ G(p) \\ C^* \end{bmatrix},$$

(β is ((kl+q)×k)).

The lack of serial correlation in both the v_t and the u_t ensure that these two error components are linear combinations of one another, related

according to $u_tA = v_tB$, where, for simplicity of notation, A=A(0) and B=B(0). This, in turn, implies that $\Omega=A^{-T}B'BA^{-1}$.

Using the established terminology, the A(s), B(s), and C are called the structural parameters, while β and Ω are called the reduced form parameters. In general it will not be an easy task to recover all of the former from the latter. For some problems, however, it is enough to recover only the A and B matrices and this may require only knowledge of Ω. In the special case that B(s)=0 for s>0 (i.e. there is no moving average component in the system) it is easy to see that $A(s)=-G(s)A^{-1}$, $C=C^*A^{-1}$. Thus, once A and B are estimated, the other system parameters are obtainable directly from the reduced form parameters. This is an assumption that typically is made in standard treatments of SEMs and suggests that the shocks directly affect the observed variables only as the shocks are realized; all subsequent impacts are through interactions between the variables.[4]

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 log likelihood for this model (for y_t, t=1,...,T) is (see Appendix A for details)

$$\begin{aligned}
\ell &= -\frac{Tk}{2}\ln(2\pi) - \frac{T}{2}\ln\|\Omega\| - \frac{1}{2}\sum_{t=1}^{T}(y_t-x_t\beta)\Omega^{-1}(y_t-x_t\beta)' \\
&= -\frac{Tk}{2}\ln(2\pi) + T(\ln\|AB^{-1}\|) - \frac{1}{2}\operatorname{tr}(B^{-T}A'(Y-X\beta)'(Y-X\beta)AB^{-1}), \\
&= -\frac{Tk}{2}\ln(2\pi) + T(\ln\|AB^{-1}\|) - \frac{1}{2}\operatorname{vec}(AB^{-1})\operatorname{vec}((Y-X\beta)'(Y-X\beta)AB^{-1}),
\end{aligned}$$

where 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.

The number of reduced form parameters in this model equals $k(pk+q)+k(k+1)/2$, corresponding to the β and Ω matrices. If it is assumed that B(s)=0, s>0, then the system has $(p+1)k^2+qk+k^2$ structural parameters, corresponding to the A(s), C, and B matrices and therefore has $(3k^2-k)/2$ more parameters than the reduced form. An order condition for identification thus is that $(3k^2-k)/2$ restrictions need to be imposed on the A(s), C and B matrices. 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 $e_t B$). This reduces the identification problem to one of imposing k^2 restrictions on the A(s) and C matrices.[5] It is clear that identifying restrictions cannot be placed only on A unless it is assumed that A=I, in which case the model is no longer simultaneous.

IDENTIFICATION IN VAR MODELS

In contrast, the VAR approach focuses 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 ability is lost 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 B(s), s>0, and the C matrices.

There are at least two substantive rationales for focusing all identifying 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. On the other hand, it is sometimes difficult for variables to react immediately to new economic developments because of information lags or adjustment costs. Such minimum delay considerations provide one useful source of identifying restrictions. A third reason is that there a significant gain in computational ease when it is possible to separate any restrictions placed on β from those placed on A and B. There is no doubt that 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 $3(k^2-k)/2$ restrictions.[6]

General (linear) restrictions can be represented by

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

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, these being at least $(3k^2-k)/2$ (including normalizing restrictions). A more useful representation of the restrictions can be made, however, in terms of the underlying free parameters of the system, here denoted θ. This general framework is given by

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

where Z, θ, and W are $(2k^2 \times n)$, $(n \times 1)$, and $(2k^2 \times 1)$, respectively, and where $n \leq k(k+1)/2$. While the two representations are equivalent,[7] 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 constraints (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 $\mathrm{vec}(A)=Z_1\theta+W_1$ and $\mathrm{vec}(B)=Z_2\theta+W_2$, this restriction can be imposed by setting $Z_2=0$ $(9 \times n)$ and $W_2=\mathrm{vec}(I_3)$. This imposes $k^2=9$ restrictions and therefore at least $k(k-1)/2=3$ additional restrictions must be imposed. Let these restrictions be $a_{31}=0$, $a_{12}=a_{21}$, and $a_{13}+a_{23}+a_{33}=1$. These restrictions can be imposed directly according to $R_1\mathrm{vec}(A)=r_1$, 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 non-singular 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-multiplied by any non-singular 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, the 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). 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=3, $B=I_k$ and let

$$Z_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad W_1 = 0,$$

and $\theta=(1\ 2\ 1\ 3\ 1\ 1)'$. This model satisfies the order condition for (exact) identification (n=k(k+1)/2=6) with rank(Z)=6 and with

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 2 & 3 & 0 \end{bmatrix}, \quad A^{-1} = \begin{bmatrix} 3 & -3 & -1 \\ -2 & 2 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Note, however, that

$$I(\theta) = \begin{bmatrix} 22 & -8 & -6 & 2 & 0 & 0 \\ -8 & 3 & 3 & -1 & 0 & 0 \\ -6 & 3 & 17 & -7 & 0 & 0 \\ 2 & -1 & -7 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 13 & -13 \\ 0 & 0 & 0 & 0 & -13 & 15 \end{bmatrix},$$

which is rank 5, the first column being equal to the second through fourth columns times (-3 1 2)'.

This example admittedly is somewhat trivial, in that the first two columns of A involve the same variables and hence one would not expect the model to be identified. In cases of higher dimensional k it is quite possible, however, to have other kinds of problems that are not as readily noticeable.

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, while the restrictions on B describe the direct impacts of the shocks on the equations of the system, so that non-diagonal elements of B allow for more than one shock to enter a given equation directly. There often are more than one way to formulate a given model, however. For example, it should be obvious that the model given 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 given 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 the same. 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. Indeed, if general non-linear restrictions were used technically there would be no need to use both matrices explicitly, as either one or the other would suffice.[8] 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 of the main advantages 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 efficiently estimated used OLS. Maximum likelihood estimates of A and B then can be estimated conditional on the estimated values of the reduced form coefficients. This two-stage estimation approach yields FIML coefficient estimates even if the model is over-identified because the identifying restrictions on A and B have no implications for the reduced form coefficients (this differs from 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:

$$\ell = -\frac{Tk}{2} \ln(2\pi) + T \ln\|A\| - \frac{1}{2} \text{vec}(A)'\text{vec}(U'UA).$$

It can be shown (see Appendix A) that

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

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).[9] To derive an estimator for θ note that the FIML estimator for the innovations, U, is $\hat{U}=Y-X\hat{\beta}$, which can be used to construct the estimator for $\text{Cov}(u_t)=\Omega$ given by $\tilde{\Omega}=U'U/T$.

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

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

This strategy yields FIML estimates of θ and β.

In Appendix A it is shown that

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

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

$$\ell(\theta,\hat{\beta}) = -T\left(\frac{k}{2}\ln(2\pi) + \ln(\|A\|) - \frac{1}{2}\ \mathrm{vec}(A)'\mathrm{vec}(\tilde{\Omega}A)\right)$$

and

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

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

It should be noted, however, that $\tilde{\Omega}$ is not necessarily the FIML estimator of Ω, which is instead given by $\hat{\Omega}=\hat{A}^{-T}\hat{A}^{-1}$, where $\mathrm{vec}(\hat{A})=Z1\hat{\theta}+W_1$. $\tilde{\Omega}$ is not FIML because it fails to account for possible over-identifying 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 developed an alternative estimation procedure that exploits this fact.[10] This fact also gave rise to what has become a standard "identification" technique in VAR analysis in which A is set equal to the inverse of the Cholesky decomposition of $\tilde{\Omega}$. This imposes an upper triangular form on A and implies that the system has a recursive structure ala Wold. The "identification" problem thus is reduced to establishing an ordering for the variables in the system.

If the assumption that $B(s)=0$, $s>0$, has been made, the invariance property of ML estimators allows the "structural" coefficient matrices to be derived from the ML estimators of $G(s)$, C^*, and A:

$$\hat{A}(s) = \hat{G}(s)\hat{A}, \qquad s>0,$$

and

$$\hat{C} = \hat{C}^*\hat{A}.$$

For evaluating the quality of the estimators and for hypothesis testing, as well as for checking model identification, it is useful to have an explicit expression for the Information matrix, the inverse of which is equal to the asymptotic $\mathrm{Cov}(\hat{\theta},\hat{\beta})$. The information matrix for this model is

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

the block diagonality of which is demonstrated by noting that the upper right-hand term

$$-E\left[\frac{\partial \mathrm{vec}(X'UAA')}{\partial\theta}\right] = 0,$$

since U is not affected by θ and has expectation zero.[11] This is further manifestation of the separability of the reduced form coefficients and the identifying restrictions on A in this model.

In Appendix A it is shown that

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

where P_k is a permutation matrix constructed by arranging the columns of the k^2 identity matrix in the order:

$$1,k+1,2k+1,..,(k-1)k+1,2,k+2,..,(k-1)k+2,..,(k-1)k,.., \ k^2.$$

It may be noted that P_k is symmetric, that $P_k^{-1} = P_k$, and that, for any (k×k) matrix A, $\mathrm{vec}(A')=P_k\mathrm{vec}(A)$.

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, $T(A^{-T}A^{-1})$:

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

Note that this term is functionally independent of β. Similarly, it can be shown that

$$\frac{\partial^2 \ell}{\partial \mathrm{vec}(\beta)\partial \mathrm{vec}(\beta)'} = -AA'\otimes X'X.$$

x_t contains lagged values of y_t and the expectation of this term involves the autocovariance function of y, which is unknown. Under the stationarity assumption, however, X'X has a well defined asymptotic value and thus AA'⊗X'X provides a consistent estimator of this block of the information matrix.

If the model is generalized to include a non-diagonal B matrix the separation between the reduced form parameters, β, and θ continues to hold. This again allows for a 2-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

with respect to θ can again be evaluated at $\hat{\beta}$, yielding (see Appendix A for details)

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

and

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

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 just identified case. The information matrix again will be block diagonal with the upper left hand block given by:

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

Finally it is noted that there is a special case of the general VAR model 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 quasi-triangular 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 .

IMPULSE ACCOUNTING

The ability to separate the estimation of the reduced form coefficients from that of the contemporaneous coefficients is only one of the features that distinguish VARs from other dynamic SEMs. The other, and more profound, distinction is that the VAR model is identified in such a way that the impact on observable variables of a set orthogonal shocks (impulses) can be given a meaningful interpretation. In the jargon that has grown up around VAR methodology this has been called innovation accounting or, more properly, impulse accounting. Two descriptive measures are used to represent these impacts, the Impulse Response Function (IRF) and the Forecast Error Variance Decomposition (FEVD).

To develop these measures it is necessary to derive the Moving Average (MA) representation for the system. Recall that the u_t are the 1-step ahead forecasting errors, i.e.,

$$u_t = y_t - E_{t-1}[y_t].^{12}$$

By successive substitutions it is straightforward to show that

$$y_t - E_{t-\ell-1}[y_t] = \sum_{s=0}^{\ell} u_{t-s}M(s), \qquad (*)$$

where M(s) can be calculated from the G(s) according to the relationship

$$M(s) = \sum_{i=1}^{s} G(i)M(s-i) = \sum_{i=1}^{\min(s,p)} G(i)M(s-i),$$

(G(i)=0 for i>p). Notice that M(s) involves only the G(i) for $i \leq s$. M(s) also involves M(i), $s-\min(s,p)<i<s$, which gives rise to a straightforward recursive computational algorithm.

From (*) it can be seen that the MA representation[13] describes the (ℓ+1) step ahead forecast error of y_t as a weighted sum of the period t-ℓ through t 1-step ahead forecast errors (innovations). The MA representation, since it involves only the reduced form coefficients, is independent of the identifying restrictions placed on A and B. It is, however, generally not of much interest in itself, since it is not possible to disentangle the separate effects of the typically non-orthogonal forecasting errors. Furthermore, the forecasting errors generally will not have any clear interpretation.

Of more interest is the IRF, which utilizes the relationship

$$y_t - E_{t-\ell-1}[y_t] = \sum_{s=0}^{\ell} v_t BA^{-1}M(s) = \sum_{s=0}^{\ell} v_t R(s),$$

where $R(s)=BA^{-1}M(s)$. The $R(s)$ trace the impact of each of the (orthogonal) system impulses on the observable system variables. For example, $R(s)_{ij}$ represents the impact on variable j of impulse i of one unit (one standard deviation) size s periods previously. The IRF, therefore, measures both the source and the strength of each of the forces affecting a given variable as well as the time path of the response to those forces.

Another mechanism for describing the strength of the forces affectting a given variable is the Forecast Error Variance Decomposition (FEVD). This decomposition measures the proportion of the ℓ-step ahead forecast error variance attributable to each impulse. The ℓ-step ahead forecast error at time t is

$$e_t^{\ell} = \sum_{s=0}^{\ell-1} u_{t+\ell-s} M(s) = \sum_{s=0}^{\ell-1} v_{t+\ell-s} BA^{-1} M(s) = \sum_{s=0}^{\ell-1} v_{t+\ell-s} R(s).$$

The v_{t+s} are serially uncorrelated and mutually orthogonal; hence the covariance of the forecast error is

$$\mathrm{Cov}(e_t^{\ell}) = E\left[\sum_{s=0}^{\ell-1} R(s)'\, v'_{t+\ell-s} v_{t+\ell-s} R(s)\right]$$

$$= \sum_{s=0}^{\ell-1} R(s)'R(s).$$

The variance of the forecast error of the jth variable, the jth diagonal element of $\mathrm{Cov}(e_t^{\ell})$, can be written

$$\sum_{s=0}^{\ell-1} \sum_{i=0}^{k} (R_{ij}(s))^2 = \sum_{i=1}^{k} \left[\sum_{s=0}^{\ell-1} (R_{ij}(s))^2 \right].$$

This suggests that the ℓ-step ahead forecast error variance can be decomposed into components attributable to each of the k impulses in the system. Thus the percentage contribution of the i^{th} impulse to the ℓ-step ahead forecast error variance of the j^{th} variable is

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

Although estimation of the IRF and the FEVD is fairly straightforward, given estimates of A, B, and β, it is difficult but not impossible to construct good measures of their quality. These estimators are based on non-linear functions of A, B, and the G(s). The IRF is a linear function of the MA coefficients, M(s), and their quality is therefore dependent in part on the quality of the estimates of the M(s). These in turn are functions of the G(i), $i \leq \min(s,p)$, which creates two problems. First any sampling errors in the G(i) matrices for low i will be compounded in constructing the M(s). Second, if the order of the autoregression is truncated prematurely (p is too low), the misspecification will result in biased estimates of the G(i) and therefore of the M(s). This is a difficult problem in that such misspecification may be hard to detect. It is possible that the estimated residuals of a low order model are indistinguishable from white noise but that a higher order model is more appropriate for describing the dynamics of the system. Given that most methods of detecting misspecification are based on measures of predictive accuracy, this problem can be extremely hard to uncover. It should be pointed out that this problem is not unique to the VAR approach but can exist in any dynamic SEM model.[14]

Generally meaningful interpretations of VAR models are based on the IRF and FEVD. These measures can be used to trace the partial and the cumulative impact of a given independent source of variation in a system on each of the observable variables. If the model provides a reasonably good representation of the system then this is a powerful tool. On the other hand these measures seem to be fairly sensitive to model specification and therefore may not provide robust results.

SUMMARY

This paper has attempted to point out how VAR models are related to and distinguished from other dynamic SEMs. It is hoped the attempt is helpful in bringing together those economists that attempt to use VAR methods in structural modeling and those that dismiss such attempts. That meaningful interpretation of VAR models depends critically on the identifying restrictions imposed on the model is a fact that many users of these models have neglected.

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 modelling has involved the use of a rather suspect form of identifying restrictions. Furthermore many practitioners seem to be using 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 construct 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 the issue of whether the results are robust to other identification regimes, and, as Bernanke points out, the practice implies a strange prior in which the analyst believes strongly in the recursiveness of the system but is not sure in what order the variables should be arranged.

While it is clear that the recursive model is not generally acceptable, there are at least two reasons for focussing 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 due to 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 whether 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 to economists posed by the VAR methodology.

A final note concerns the place of VAR models in policy analysis. One issue that arises when using statistical models is whether policy is best viewed as the setting of levels of given variables, as a particular setting of the parameters relating variables to one another, or as an external shock to the economy. The latter view is the most consistent with the way VAR

models have been used and implies that policy is endogenous to the system but subject to exogenous shocks. Whatever view is represented in a model it is important to ask whether a given contemplated policy represents an interpolation or an extrapolation of model results. It is probable that no one of these three alternatives is completely satisfactory for all situations. It is also probable that these issues will continue to be the subject of lively debate.

APPENDIX A

Calculation of score functions and Hessians is facilitated by a number of results of matrix algebra and calculus, the more unfamiliar of which are reproduced here, with references to the text by Graham. These results use the convention that the derivative of an n-vector with respect to an m-vector is (m×n), with $[\partial Y/\partial X]_{ij}=\partial Y_j/\partial X_i$.

(1) $\mathrm{tr}(AB) = (\mathrm{vec}(A'))'\mathrm{vec}(B)$ (Table 1, p.121)

(2) $\mathrm{vec}(AYB) = (B'\otimes A)\mathrm{vec}(Y)$ (Eq. 2.13, p.25)

(3) $\frac{\partial \mathrm{tr}(X'AXB)}{\partial X} = AXB + A'XB'$ (Table 6, p.124)

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

(5) $\frac{\partial \mathrm{vec}(AXB)}{\partial \mathrm{vec}(X)} = B\otimes A'$ (Eq. 5.3, p.71)

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

(7) $\frac{\partial Z}{\partial X} = \frac{\partial Y}{\partial X}\frac{\partial Z}{\partial Y}$ where X, Y, and Z are vectors (Table 3, p.122)

The methods discussed by Graham (Second Transformation Principle, p.74) can be used to derive the following results. Noting that, for any (k×k) matrix X, $\mathrm{vec}(X')=P_k\mathrm{vec}(X)$,* these results can also be obtained by the use of the product rule and (5), (6), and (7).

(8) $\frac{\partial \mathrm{vec}(X^{-T})}{\partial \mathrm{vec}(X)} = -P_k(X^{-T}\otimes X^{-1})$

(9) $\frac{\partial \mathrm{vec}(AX^{-1}X^{-T})}{\partial \mathrm{vec}(X)} = -(X^{-1}X^{-T}\otimes X^{-T}A') - P_k(X^{-T}\otimes X^{-1}X^{-T}A')$

(10) $\frac{\partial \mathrm{vec}(X^{-T}AX^{-1}X^{-T})}{\partial \mathrm{vec}(X)} = -(X^{-1}X^{-T}\otimes X^{-T}A'X^{-1}) - P_k(X^{-T}AX^{-1}X^{-T}\otimes X^{-1} + X^{-T}\otimes X^{-1}X^{-T}A'X^{-1})$

---

* P_k is a permutation matrix constructed by arranging the columns of I_{k^2} in the order:
$1,k+1,2k+1,\ldots,k(k-1)+1,2,k+2,\ldots,k(k-1)+2,\ldots,k,2k,\ldots,k(k-1),\ldots,k^2$.
It can be shown, for X and Y both (k×k), that $P_k(X\otimes Y)=(Y\otimes X)P_k$, with the corollary that $P_k(X\otimes X')$ is symmetric.

The log-likelihood for the model discussed in the paper is a function of the vectors, θ and β, corresponding to the contemporaneous and the reduced form parameters. The score function and the diagonal blocks of the Hessian are calculated below, with references to the numbered results given above. These are first given for the special case when B=I and then for the general case.

$$\ell(\theta,\beta) = -\frac{Tk}{2}\ln(2\pi) + T\ln\|A\| - \frac{1}{2}\sum_{t=1}^{T} u_t AA' u_t'$$

$$= -\frac{Tk}{2}\ln(2\pi) + T\ln\|A\| - \frac{1}{2}\sum_{t=1}^{T} \mathrm{tr}(A'u_t'u_tA)$$

$$= -\frac{Tk}{2}\ln(2\pi) + T\ln\|A\| - \frac{1}{2}\mathrm{tr}(A'U'UA)$$

$$= -\frac{Tk}{2}\ln(2\pi) + T\ln\|A\| - \frac{1}{2}\mathrm{vec}(A)'\mathrm{vec}(U'UA) \qquad (1)$$

$$= -\frac{Tk}{2}\ln(2\pi) + T\ln\|A\| - \frac{1}{2}(Z_1\theta+W_1)'(I_k\otimes(Y-X\beta)'(Y-X\beta))(Z_1\theta+W_1)) \qquad (2)$$

$$\frac{\partial \ell}{\partial \mathrm{vec}(\beta)} = -\frac{1}{2}\frac{\partial \mathrm{vec}(U)}{\partial \mathrm{vec}(\beta)}\frac{\partial \mathrm{tr}(AA'U'U)}{\partial \mathrm{vec}(U)} \qquad (7)$$

$$= \frac{1}{2}(I_k\otimes X')\mathrm{vec}(UAA'+UAA') \qquad (5,3)$$

$$= \mathrm{vec}(X'UAA') \qquad (2)$$

$$\frac{\partial \ell}{\partial \theta} = \frac{\partial \mathrm{vec}(A)}{\partial \theta}\left(T\frac{\partial \ln\|A\|}{\partial \mathrm{vec}(A)} - \frac{1}{2}\frac{\partial \mathrm{tr}(AA'U'U)}{\partial \mathrm{vec}(A)}\right) \qquad (7)$$

$$= Z_1'(T\,\mathrm{vec}(A^{-T}) - \mathrm{vec}(U'UA)) \qquad (4,3)$$

$$\frac{\partial^2 \ell}{\partial \mathrm{vec}(\beta)\partial \mathrm{vec}(\beta)'} = \frac{\partial \mathrm{vec}(X'UAA')}{\partial \mathrm{vec}(\beta)}$$

$$= \frac{\partial \mathrm{vec}(U)}{\partial \mathrm{vec}(\beta)}\frac{\partial \mathrm{vec}(X'UAA')}{\partial \mathrm{vec}(U)} \qquad (7)$$

$$= -(I_k\otimes X')(AA'\otimes X) \qquad (5,5)$$

$$= -AA'\otimes X'X$$

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

$$= Z_1'\left(T\frac{\partial \mathrm{vec}(A^{-T})}{\partial \mathrm{vec}(A)} - \frac{\partial \mathrm{vec}(U'UA)}{\partial \mathrm{vec}(A)}\right)Z_1 \qquad (7)$$

$$= -Z_1'(TP_k(A^{-T}\otimes A^{-1}) + (I\otimes U'U))Z_1 \qquad (8,5)$$

$$I(\theta) = TZ_1'(P_k(A^{-T}\otimes A^{-1}) + (I\otimes A^{-T}A^{-1}))Z_1$$

In the general case in which B is not necessarily the identity matrix the gradient and the Hessian with respect to β are unchanged except that AA' is replaced by $AB^{-1}B^{-T}A'$. Other results follow.

$$\ell(\theta,\beta) = -\frac{Tk}{2}\ln(2\pi) + T\ln\|AB^{-1}\| - \frac{1}{2}\sum_{t=1}^{T} u_t AB^{-1}B^{-T}A'u_t'$$

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

$$= -\frac{Tk}{2}\ln(2\pi) + T(\ln\|AB^{-1}\|) - \frac{1}{2}\,\mathrm{vec}(AB^{-1})'\mathrm{vec}((Y-X\beta)'(Y-X\beta)AB^{-1}) \quad (1,2)$$

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

$$+ \frac{\partial \mathrm{vec}(B)}{\partial \theta}\left[-\frac{1}{2}\frac{\partial \mathrm{vec}(B^{-1})}{\partial \mathrm{vec}(B)}\frac{\partial \mathrm{tr}(B^{-T}A'U'UAB^{-1})}{\partial \mathrm{vec}(B^{-1})} - T\frac{\partial \ln\|B\|}{\partial \mathrm{vec}(B)}\right] \quad (7)$$

$$= Z'\begin{bmatrix} T\,\mathrm{vec}(A^{-T}) - \mathrm{vec}(U'UAB^{-1}B^{-T}) \\ \mathrm{vec}(B^{-T}A'U'UAB^{-1}B^{-T}) - T\,\mathrm{vec}(B^{-T}) \end{bmatrix} \quad (3,4,6,2)$$

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

$$Z'\begin{bmatrix} -TP_k(A^{-T}\otimes A^{-1}) - (B^{-1}B^{-T}\otimes U'U) & \cdot \\ -\frac{\partial \mathrm{vec}(U'UAB^{-1}B^{-T})}{\partial \mathrm{vec}(B)} & TP_k(B^{-T}\otimes B^{-1}) + \frac{\partial \mathrm{vec}(B^{-T}A'U'UAB^{-1}B^{-T})}{\partial \mathrm{vec}(B)} \end{bmatrix} Z \quad (8,5)$$

$$\frac{\partial \mathrm{vec}(U'UAB^{-1}B^{-T})}{\partial \mathrm{vec}(B)} = -(B^{-1}B^{-T}\otimes B^{-T}A'U'U) - P_k(B^{-T}\otimes B^{-1}B^{-T}A'U'U) \quad (9)$$

$$\frac{\partial \mathrm{vec}(B^{-T}A'U'UAB^{-1}B^{-T})}{\partial \mathrm{vec}(B)} = -(B^{-1}B^{-T}\otimes Q) - P_k(QB^{-T}\otimes B^{-1} + B^{-T}\otimes B^{-1}Q) \quad (10)$$

where $Q=B^{-T}A'U'UAB^{-1}$ (note that $E[U'U]=TA^{-T}B'BA^{-1}$ and hence $E[Q]=TI_k$).

$$I(\theta) = TZ'\begin{bmatrix} P_k(A^{-T}\otimes A^{-1}) + (B^{-1}B^{-T}\otimes A^{-T}B'BA^{-1}) & \cdot \\ -(B^{-1}B^{-T}\otimes BA^{-1}) - P_k(B^{-T}\otimes A^{-1}) & (B^{-1}B^{-T}\otimes I_k) + P_k(B^{-T}\otimes B^{-1}) \end{bmatrix} Z$$

Appendix B

A quasi-triangular specification is a special case of the general formulation that permits a recursive two stage least squares (2SLS) 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 (k×n) 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 that have not yet been estimated. Note that ι_1 and possibly others may be empty. Let Z_i^* equal the ι_i columns of Z_{1i}, the ith block of Z_1, and initialize θ=0 (n×1) 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 $B_{ii}=(A_{\cdot i}'\tilde{\Omega}A_{\cdot i})^{0.5}$ and set $R_i=A_{\cdot i}'\tilde{\Omega}/B_{ii}$. Notice that the algorithm requires only $\tilde{\Omega}$ 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 $E[V'V]=TI_k$.

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. For strictly exogenous variables to be included will require careful interpretation of model results in some contexts, as in the interpretation of Impulse Response Functions. 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. Maximum likelihood estimation therefore can be performed on either part separately. Note that this implies that lagged y is not useful in predicting current z, suggesting an exogeneity test, which, if rejected, leads to rejection of the exogeneity assumption, but which, if accepted, provides only partial evidence of exogeneity.

3. It should be noted, however, that the 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 will be incorrectly represented.

4. Many theoretical models assume some dynamic structure for model shocks, so that $B(s)\neq 0$, $s>0$. Typically there does not seem to be any compelling rationale, however, for the choice of particular lag structures in such models. The main rationale for working with low order ARMA models seems to be one of parsimony, since it is possible that a given model can be represented using fewer parameters if shocks are modeled as serially correlated. It is also possible in this situation that the effects of omitted variables are less severe. The serially uncorrelated shocks assumption can be rationalized on grounds that the dynamics of the shocks have been solved out of the model by incorporating their effects into the autoregressive parameters of the model. As long as the structural aspect of the model desired is the response of the system to the independent shocks (the v_t), the assumption should not be troublesome.

5. Restrictions could also 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. There are, of course, other identification issues that are common to other SEMS. The first, which thus far has been implicit, is the choice of variables to include in the vectors y_t and z_t. The latter can include variables that are strictly exogenous, which is equivalent to positing a block triangular structural for the A(s) and B(s) in a larger

model that includes both y and the non-deterministic variables of z (Wu). This aspect of the identification issue essentially requires that the current impact of variables omitted from the system, which are manifested in the shocks, must be uncorrelated with both the included exogenous variables and with lagged endogenous variables.

A second issue involves determination of the lag length, p. Typically there are not strong theoretical criteria for making such a determination, except that any relevant technological and seasonal factors should be considered. Beyond this a data-based approach is perhaps the only way to proceed. Such an approach uses some criteria to compare alternate lag lengths. Two common criteria that are simple, indeed mechanical, to apply are the Akaike Information Criteria (AIC) and the Schwartz Criteria (SC). Both of these adjust the maximum of the log-likelihood function for a given lag length by the number of parameters in the model. The maximum of the log-likelihood for a given sample of size T is a function only of $\tilde{\Omega}_p$, the sample error covariance matrix with lag length p, given by

$$\max \ell = -T/2(k(\ln(2\pi)+1) + \ln|\tilde{\Omega}_p|).$$

Clearly maximization of the loglikelihood for a given p is equivalent to minimization of $|\tilde{\Omega}_p|$. However, additional lags will never decrease the value of the log-likelihood. The AIC and SC suggest that the optimal lag length can be determined by finding the p that minimizes $|\tilde{\Omega}_p|+2k(pk+q)/T$ and $|\tilde{\Omega}_p|+\ln(T)k(pk+q)/T$, respectively.

Another remaining issue concerns the identifiability of the reduced form parameters, the necessary and sufficient condition being that X has full column rank. It is assumed that these issues have been addressed and that attention therefore can be directed to the issues specific to VAR models.

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

8. Non-linear restrictions could be written in the form

$$\text{vec}(A) = f(\theta)$$

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

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

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 extension is not pursued here, however.

9. The uniqueness of this estimator is guaranteed when A has full rank, a condition which is also necessary for identification.

10. Bernanke suggests that only in the exactly identified case will the two stage procedure yield FIML estimates. This has been shown to be incorrect.

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

12. Strictly speaking this is only true when z_t contains only deterministic variables. If z_t contains exogenous variables then u_t can be thought of as the forecasting error due to the impulses alone and the analysis remains unchanged. The impacts on the exogenous variables can be analyzed by examination of the regression coefficients alone; there is no need to use a moving average representation to disentangle their effects.

13. Typically the term MA representation is applied to the limit of (*) as $\ell \to \infty$. Furthermore for this actually to be the MA representation of the system it is necessary for the system to be stationary. This can be checked by examining the eigenvalues of

$$\left[\begin{array}{c|c} G(1) & I_{k(p-1)} \\ \vdots & \hline \\ G(p) & 0 \end{array}\right].$$

If all the eigenvalues have modulus less than unity, the system is stationary.

14. For further discussion of the assessment of the IRF and FEVD see Runckle and the accompanying discussion.

REFERENCES

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

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

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

Gordon, R.J. and S.R. King. "The Output Cost of Disinflation in Traditional and Vector Autoregressive Models." Brooking Papers on Economic Activity. 1 (1982): 205-244.

Graham, A. Kroneker Products and Matrix Calculus With Applications. Ellis Horwood Limited. Chichester. 1981.

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

Hausman, J.A., W.K. Newey, and W.E. Taylor. "Efficient Estimation and Identification of Simultaneous Equations Models With Covariance Restrictions." Econometrica. 55 (1987): 849-874.

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

Leamer, E.E. "Let's Take the Con out of Econometrics." American Economic Review. 73 (1983): 31-43.

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

Litterman, R.B. "Specifying Vector Autoregressions for Macroeconomic Forecasting." Federal Reserve Bank of Minneapolis Research Department Staff Report 92. March 1984.

Litterman, R.B. "Forecasting With Bayesian Vector Autoregressions - Five Years of Experience." Journal of the Business and Economic Statistics. 4 (1986): 25-38.

Rothenberg, T.J. "Identification in Parametric Models." Econometrica. 39 (1971): 577-91.

Runckle, D.E. "Vector Autoregressions and Reality." Journal of Business and Economic Statistics. 5 (1987): 437-42. (Also comments, pp. 443-454).

Sargent T.J. "Autoregressions, Expectations, and Advice." American Economic Review. 74 (1984): 408-421.

Sims, C.A. "Macroeconomics and Reality." Econometrica. 48 (1980): 1-48.

Sims, C.A. "Policy Analysis with Econometric Models." Brookings Papers on Economic Activity. 1 (1982): 107-164.

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

Wu, D.M. "Tests of Causality, Predeterminedness and Exogeneity." International Economic Review. 24 (1983): 547-558.